

Західноукраїнський національний університет  
Міністерство освіти і науки України

Хмельницький національний університет  
Міністерство освіти і науки України

Кваліфікаційна наукова праця  
на правах рукопису

ЛІП'ЯНИНА-ГОНЧАРЕНКО ХРИСТИНА ВОЛОДИМИРІВНА

УДК 004.6+004.9

ДИСЕРТАЦІЯ

ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ЗАСАДИ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ  
ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ДАНИХ  
ТЕРИТОРІАЛЬНИХ ГРОМАД

05.13.06 – інформаційні технології  
(12 – інформаційні технології)

Подається на здобуття наукового ступеня доктора технічних наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і  
текстів інших авторів мають посилання на відповідне джерело  
\_\_\_\_\_ Х.В. Лип'яніна-Гончаренко

Ідентифікаційні номери дисертації  
засвідчують



секретар СВР

Хмельницький - 2025

Андрій Меленюк

## АНОТАЦІЯ

*Лип'яніна–Гончаренко Х. В.* Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу соціально–економічних даних територіальних громад. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.06 «Інформаційні технології» (12 – Інформаційні технології). – Хмельницький національний університет, Хмельницький національний університет, Хмельницький, 2025.

Відсутність відомих інтегрованих моделей, методів та засобів інтелектуального аналізу соціально–економічних даних територіальних громад (ТГ), які одночасно задовольняли б вимоги до адаптивності в умовах динамічних і нестационарних процесів, інтеграції різнорідних джерел інформації, забезпечення високої точності прогнозування та прозорості результатів, створює актуальну науково–прикладну проблему. Одним із шляхів її вирішення є розроблення теоретичних та прикладних засад інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ, яка дозволить формувати обґрунтовані управлінські рішення, підвищувати соціальну стійкість та забезпечувати сталий розвиток громад завдяки адаптивному врахуванню специфіки соціально–економічних процесів.

*Наукова новизна результатів дисертаційної роботи.* Отримано такі нові наукові результати:

1. Вперше запропоновані узагальнені принципи синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ, особливістю якої є здатність інтегрувати різнорідні дані, враховувати динамічність соціально–економічних процесів, що забезпечило підвищення соціальної стійкості інфраструктури громад за рахунок інтеграції гетерогенних даних, адаптивного вибору методів аналізу та формування управлінських рішень.

2. Вперше розроблено методологію аналізу соціально–економічних даних ТГ як сукупність методів класифікаційного, кластерного та гібридного аналізу, а також прогнозування соціально–економічних даних, особливістю якої є адаптивність для роботи з нестационарними даними, реальним часом для оперативного прийняття

рішень та інтеграцією різномірних джерел даних, що дозволило підвищити точність прогнозування, класифікації та прозорість управлінських рішень.

3. Розроблено новий метод формування навчальної вибірки для нестационарних процесів на основі RFM та кластерного аналізу, що, на відміну від існуючих підходів, забезпечує сегментацію та кластеризацію, який дозволив підвищити точність опрацювання нових вхідних наборів даних у задачах прогнозування, класифікації та аналізу нестационарних соціально–економічних даних.

4. Вперше розроблено метод ансамблевих адаптивних прогнозних моделей для багатовимірного аналізу, особливістю якого є оцінювання на «ковзному вікні» та метамоделі першого і другого рівнів на основі оптимізації функції Лагранжа, що забезпечило точність прогнозів багатовимірних нестационарних процесів із врахуванням їхньої динамічності.

5. Удосконалено метод класифікаційного аналізу кількісних соціально–економічних даних, який, на відміну від відомих підходів, забезпечує інтеграцію структурованих, неструктурованих і напівструктурованих даних, що дозволило підвищити точність класифікації кількісних показників для реалізації управлінських рішень.

6. Удосконалено метод класифікаційного аналізу текстових соціально–економічних даних, який, на відміну від відомих підходів, забезпечує інтеграцію текстових джерел різної структури та їхній аналіз за допомогою технологій обробки природної мови, що дозволило підвищити ефективність виявлення закономірностей для реалізації управлінських рішень.

7. Удосконалено метод кластерного аналізу соціально–економічних даних, який, на відміну від відомих підходів, забезпечує ідентифікацію груп об'єктів з подібними характеристиками, що дозволило підвищити точність розподілу ресурсів для реалізації управлінських рішень.

8. Удосконалено метод прогнозування соціально–економічних даних, який, на відміну від відомих підходів, забезпечує високоточне прогнозування динамічних процесів шляхом інтеграції структурованих, неструктурованих і напівструктурованих даних із застосуванням адаптивних методів інтелектуального аналізу, що дозволило підвищити точність прогнозування для реалізації управлінських рішень.

9. Удосконалено метод гібридного аналізу соціально–економічних даних, який, на відміну від відомих підходів, забезпечує врахування кількісних та якісних показників завдяки інтеграції різнорідних даних і багаторівневого підходу до аналізу, що дозволило підвищити гнучкість і адаптивність процесу прийняття рішень для реалізації управлінських рішень.

*Практична цінність отриманих результатів дисертаційної роботи* полягає в розробці інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ, що дозволяє подолати існуючі обмеження традиційних методів обробки даних. Запропонована технологія забезпечує адаптивність до динамічних змін соціально–економічних процесів завдяки використанню методів машинного навчання. Це дозволяє оперативно реагувати на зміни в соціально–економічних показниках та враховувати специфіку різнорідних джерел даних, таких як статистичні показники, відкриті дані та результати соціологічних опитувань. Практичне застосування такої технології сприятиме точному прогнозуванню ризиків, що є критично важливим для управління в умовах сучасних викликів.

Інтеграція різнорідних джерел даних у межах запропонованої технології дозволяє вирішувати завдання комплексного аналізу та моделювання стану ТГ. Це забезпечує можливість виявлення тенденцій, прогнозування змін і оцінювання впливу прийнятих рішень на соціально–економічний стан громад. Створення модулів для обробки як структурованих, так і неструктурованих даних спрощує процес інтеграції інформації з різних джерел, усуваючи технічні бар'єри.

Прозорість і зрозумілість результатів аналізу, забезпечена розробленою інформаційною системою з пояснюваними результатами, дозволяє управлінцям ТГ ухвалювати обґрунтовані рішення. Розроблені інструменти сприяють не лише підвищенню точності при аналізі даних, але й покращенню комунікації між зацікавленими сторонами завдяки доступності інформації для кінцевих користувачів. У підсумку, впровадження технології інтелектуального аналізу даних дозволяє ТГ досягати сталого розвитку, підвищувати соціальну стійкість та адаптуватися до кризових умов.

Результати дослідження активно використовуються в освітньому процесі Західноукраїнського національного університету. Вони інтегровані в навчальні курси для студентів і аспірантів за спеціальністю «Комп'ютерні науки», такі як:

«Інтелектуальний аналіз даних», «Машинне навчання», «Методи та засоби штучного інтелекту», «Інтелектуальна обробка даних у складних системах». Це дозволяє вдосконалити підготовку спеціалістів у галузі обробки та аналізу великих даних.

Наукові результати дисертації впроваджено в рамках міжнародних і державних проєктів, таких як, зокрема: Erasmus+: KA2 CBHE «Міждоменні компетенції для забезпечення здорової та безпечної роботи у 21 столітті» (WORK4CE) 619034–EPP–1–2020–1–UA–EPPKA2–CBHE–JP; Creative Europe: AURA – Auralisation of Acoustic Heritage Sites using Augmented and Virtual Reality (nr. 101008547); ERASMUS–EDU–2023–PI–ALL–INNO під назвою "My Farm" (Proposal number: 101140288); держбюджетної наукової роботи молодих учених (номер держреєстрації 0117U003871) «Методи та засоби структурно–статистичної ідентифікації ієрархічних об'єктів за характерними точками їх контурів»; держбюджетної наукової роботи молодих учених (номер держреєстрації 0117U003871) «TruScanAI: інструмент виявлення фейкової інформації на основі технологій ШІ для боротьби з дезінформацією»; науково–дослідної роботи ІОСУ2023 «К» «Інтелектуальні методи, моделі та технології соціально–економічного розвитку ТГ в умовах сьогодення» (протокол № 4 від 09.11.2022 р.); перспективного плану розвитку наукового напрямку «Технічні науки» Західноукраїнського національного університету (номер державної реєстрації 0121U114705).

Крім того, результати дослідження реалізовано у діяльності ІТ–компанії «МагнетікВан», Великоберезовицької територіальної громади, громадська організація «Прогресивні люди» та Департаменту цифрової трансформації Тернопільської ОДА.

**Ключові слова:** соціально–економічні дані, методологія, інформаційна технологія, інтелектуальний аналіз даних, класифікація, кластеризація, прогнозування, адаптивні методи, сталий розвиток, територіальні громади.

## ANNOTATION

*Lypianina–Honcharenko Kh. V.* Theoretical and Applied Foundations of Information Technology for the Intelligent Analysis of Socio–Economic Data of Territorial Communities. – A qualifying scientific work presented in manuscript form.

The thesis (dissertation) for the degree of Doctor of technical sciences for specialty 05.13.06 «Information technology» (12 – Information technology). – Khmelnytskyi National University, Khmelnytskyi National University, Khmelnytskyi, 2025.

The absence of known integrated models, methods, and tools for the intelligent analysis of socio–economic data of territorial communities (TCs) that simultaneously meet the requirements for adaptability under dynamic and non–stationary processes, the integration of heterogeneous information sources, as well as ensuring high forecasting accuracy and result transparency, constitutes *a pressing scientific and applied problem*. One approach to addressing this challenge is the development of theoretical and applied foundations of information technology for the intelligent analysis of socio–economic data of TCs. This technology will enable the formulation of well–grounded managerial decisions, enhance social resilience, and ensure the sustainable development of communities by adaptively accounting for the specific nature of socio–economic processes.

*Scientific novelty of the dissertation results.* The following new scientific results have been obtained:

1. For the first time, generalized principles have been proposed for the synthesis of information technology for the intelligent analysis of TC socio–economic data. A key feature is its ability to integrate heterogeneous data and consider the dynamic nature of socio–economic processes. This integration has increased the social resilience of community infrastructures through the adaptive selection of analytical methods and the formulation of managerial decisions.

2. For the first time, a methodology for the analysis of socio–economic TC data has been developed, comprising classification, clustering, hybrid analysis methods, and socio–economic data forecasting. The methodology’s key feature is its adaptability for working with non–stationary data and real–time decision–making, as well as the integration of

heterogeneous data sources, which has improved forecasting accuracy, classification, and the transparency of managerial decisions.

3. A new method for forming a training dataset for non-stationary processes based on RFM analysis and cluster analysis has been developed. Unlike existing approaches, it ensures segmentation and clustering, improving the accuracy of processing new input data sets for forecasting, classification, and the analysis of non-stationary socio-economic data.

4. For the first time, a method of ensemble adaptive forecasting models for multidimensional analysis has been developed. A distinctive feature is the evaluation on a "sliding window" and first- and second-level metamodels based on Lagrange function optimization. This ensures accurate forecasting of multidimensional non-stationary processes, taking into account their dynamism.

5. The method of classification analysis of quantitative socio-economic data has been improved. Unlike known approaches, it ensures the integration of structured, unstructured, and semi-structured data, thereby increasing the accuracy of quantitative indicator classification for implementing managerial decisions.

6. The method of classification analysis of textual socio-economic data has been improved. Unlike known approaches, it integrates textual sources of varying structures and applies natural language processing technologies. This enhances the efficiency of identifying patterns for managerial decision-making.

7. The method of cluster analysis of socio-economic data has been improved. Unlike known approaches, it identifies groups of objects with similar characteristics, thereby increasing the accuracy of resource allocation for managerial decision-making.

8. The method of forecasting socio-economic data has been improved. Unlike known approaches, it provides highly accurate forecasting of dynamic processes by integrating structured, unstructured, and semi-structured data using adaptive intelligent analysis methods, thereby improving forecasting accuracy for implementing managerial decisions.

9. The method of hybrid analysis of socio-economic data has been improved. Unlike known approaches, it accounts for both quantitative and qualitative indicators by integrating heterogeneous data and employing a multi-level analytical approach. This enhances the flexibility and adaptability of the decision-making process for managerial purposes.

*Practical value of the dissertation results.* The development of an information technology for the intelligent analysis of socio–economic TC data overcomes the limitations of traditional data processing methods. The proposed technology ensures adaptability to dynamic changes in socio–economic processes by employing machine learning methods. This enables timely responses to changes in socio–economic indicators and the consideration of the specific nature of heterogeneous data sources, such as statistical indicators, open data, and survey results. The practical application of this technology will facilitate accurate risk forecasting, which is critically important for management under contemporary challenges.

Integrating heterogeneous data sources within the proposed technology enables comprehensive analysis and modeling of the TC's state. This provides the ability to identify trends, forecast changes, and evaluate the impact of decisions on the community's socio–economic status. The development of modules for processing both structured and unstructured data streamlines the integration of information from various sources, removing technical barriers.

The transparency and clarity of analytical results, ensured by the developed information system with explainable outputs, allow TC managers to make well–founded decisions. The proposed tools not only enhance data analysis accuracy but also improve communication among stakeholders, thanks to the accessibility of information to end users. In summary, implementing the intelligent data analysis technology enables TCs to achieve sustainable development, strengthen social resilience, and adapt to crisis conditions.

The research results are actively utilized in the educational process of the West Ukrainian National University. They have been integrated into courses for undergraduate and graduate students in Computer Science, such as "Intelligent Data Analysis," "Machine Learning," "Methods and Tools of Artificial Intelligence," and "Intelligent Data Processing in Complex Systems." This improves the training of specialists in the field of big data processing and analysis.

The scientific results of the dissertation have been implemented within international and national projects, including Erasmus+: KA2 CBHE "Cross–domain competences for health and safe work in the 21st century" (WORK4CE) 619034–EPP–1–2020–1–UA–



EPPKA2–CBHE–JP; Creative Europe: AURA – Auralisation of Acoustic Heritage Sites using Augmented and Virtual Reality (no. 101008547); ERASMUS–EDU–2023–PI–ALL–INNO "My Farm" (Proposal no. 101140288); the state–funded research project for young scientists (state registration no. 0117U003871), "Methods and Tools of Structural and Statistical Identification of Hierarchical Objects by Characteristic Points of Their Contours"; the state–funded research project for young scientists (state registration no. 0117U003871) "TruScanAI: A Tool for Detecting Fake Information Using AI Technologies for Combating Disinformation"; the IOSU2023 “K” research project "Intelligent Methods, Models, and Technologies for the Socio–Economic Development of TCs under Current Conditions" (Protocol no. 4 dated 09.11.2022); and the perspective plan for developing the "Technical Sciences" research area at the West Ukrainian National University (state registration no. 0121U114705).

In addition, the research results have been implemented in the activities of the IT company "MagneticOne", the Velykoberezovytska Territorial Community, the public organization "Progressive People," and the Department of Digital Transformation of the Ternopil Regional State Administration.

**Keywords:** socio–economic data, methodology, information technology, intelligent data analysis, classification, clustering, forecasting, adaptive methods, sustainable development, territorial communities.

## СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

Статті у наукових виданнях, включених до Переліку наукових фахових видань України:

1. Лип'яніна–Гончаренко, Х. (2024). Узагальнений принцип синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ. *Measuring and Computing Devices in Technological Processes*, (1), 359–367. <https://doi.org/10.31891/2219-9365-2024-77-48>
2. Лип'яніна–Гончаренко, Х. (2024). Методологія аналізу соціально–економічних даних територіальних громад. *Herald of Khmelnytskyi National University. Technical Sciences*, 337(3(2)), 446–450. <https://doi.org/10.31891/2307-5732-2024-337-3-67>
3. Лип'яніна–Гончаренко, Х. (2024). Методи аналізу соціально–економічних даних територіальних громад для адаптивного управління ресурсами. *Computer Systems and Information Technologies*, (3), 92–97. <https://doi.org/10.31891/csit-2024-3-12>
4. Лип'яніна–Гончаренко, Х. В. (2023). Метод формування навчальної вибірки для масивів даних на основі машинного навчання. *Наукові записки НаУКМА. Комп'ютерні науки*, (6), 30–35. <https://doi.org/10.18523/2617-3808.2023.6.30-35>
5. Лип'яніна–Гончаренко, Х., & Юрків, Х. (2023). Методи бустингового машинного навчання для нестационарних часових рядів. *Measuring and Computing Devices in Technological Processes*, (3), 19–30. <https://doi.org/10.31891/2219-9365-2023-75-2>
6. Лип'яніна–Гончаренко, Х., Комар, М., Юрків, Х., & Лук'янчук, В. (2023). Концептуальна модель інтелектуальної оцінки наслідків техногенних катастроф. *Вісник Хмельницького національного університету. Серія: Технічні науки*, 329(6), 230–237. <https://doi.org/10.31891/2307-5732-2023-329-6-230-237>
7. Комар, М., Лип'яніна–Гончаренко, Х., Кіт, І., Мадараш, Р., & Юрків, Х. (2023). Інтелектуальний метод виявлення джерел мультілінгвальної дезінформації. *Measuring and Computing Devices in Technological Processes*, (2), 221–230. <https://doi.org/10.31891/2219-9365-2023-74-31>
8. Лип'яніна–Гончаренко, Х. В. (2022). Інтелектуальний метод формування людських ресурсів на короткостроковий проєкт. *Вісник Черкаського державного*

технологічного університету, (3), 49–58. <https://doi.org/10.24025/2306-4412.3.2022.259775>

9. Ліп'яніна–Гончаренко, Х., Комар, М., Саченко, А., & Лендюк, Т. (2022). Оцінка інвестиційних ризиків віртуальної ІТ–компанії на основі машинного навчання. *Measuring and Computing Devices in Technological Processes*, (3), 45–60. <https://doi.org/10.31891/2219-9365-2022-71-3-6>

10. Ліп'яніна–Гончаренко, Х. В., Комар, М. П., Саченко, А. О., & Лендюк, Т. В. (2022). Метод виявлення фіктивних підприємств на підставі Гаусового наївного класифікатора Байєса. *Scientific Bulletin of UNFU*, 32(5), 92–96. <https://doi.org/10.36930/40320513>

11. Ліп'яніна–Гончаренко, Х. (2023). Інтелектуальний метод вибору локації для старту бізнесу в розумному місті. *Системні технології*, 4(147), 132–140. <https://doi.org/10.34185/1562-9945-4-147-2023-12>

12. Ліп'яніна–Гончаренко, Х., Комар, М., Саченко, А., & Лендюк, Т. (2022). Метод формування контексту реклами та цільової аудиторії на основі навчання асоціативних правил. *Вісник Хмельницького національного університету. Серія: Технічні науки*, 313(5), 279–287. <https://doi.org/10.31891/2307-5732-2022-313-5-279-287>

13. Ліп'яніна–Гончаренко, Х. В. (2023). Метод генерування рекламного зображення на основі ключових слів. *Телекомунікаційні та інформаційні технології*, 81(4). <https://doi.org/10.31673/2412-4338.2023.043745>

14. Ліп'яніна–Гончаренко, Х., & Кіт, І. (2023). Метод удосконалення рекламних текстів на основі генеративних моделей. *Вісник Приазовського державного технічного університету. Серія: Технічні науки*, (46), 6–13. <https://doi.org/10.31498/2225-6733.46.2023.288087>

15. Ліп'яніна–Гончаренко, Х. В. (2023). Метод генерування рекламного зображення на основі відеопотоку. *Науковий вісник Ужгородського університету. Серія: Математика і інформатика*, 43(2), 130–135. [https://doi.org/10.24144/2616-7700.2023.43\(2\).130-135](https://doi.org/10.24144/2616-7700.2023.43(2).130-135)

Статті у періодичних виданнях, включених до категорії «А» Переліку наукових фахових видань України, або у закордонних виданнях, проіндексованих у базах даних *Web of Science Core Collection* та/або *Scopus*:

16. Lipianina–Honcharenko, K., Wolff, C., Sachenko, A., Kit, I., & Zahorodnia, D. (2023). Intelligent method for classifying the level of anthropogenic disasters. *Big Data and Cognitive Computing*, 7(3), 157. <https://doi.org/10.3390/bdcc7030157> (Індексована в наукометричній базі *Web of Science* та *Scopus*, що відноситься до Q2 відповідно до *SCImago Journal & Country Rank*)

17. Lipianina–Honcharenko, K., Wolff, C., Sachenko, A., Desyatnyuk, O., Sachenko, S., & Kit, I. (2023). Intelligent information system for product promotion in internet market. *Applied Sciences*, 13(17), 9585. <https://doi.org/10.3390/app13179585> (Індексована в наукометричній базі *Web of Science* та *Scopus*, що відноситься до Q2 відповідно до *SCImago Journal & Country Rank*)

18. Lipianina–Honcharenko, K., Komar, M., Osolinskyi, O., Shymanskyi, V., Havryliuk, M., & Semaniuk, V. (2024). Intelligent waste–volume management method in the smart city concept. *Smart Cities*, 7(1), 78–98. <https://doi.org/10.3390/smartcities7010004> (Індексована в наукометричній базі *Web of Science* та *Scopus*, що відноситься до Q1 відповідно до *SCImago Journal & Country Rank*)

19. Lipianina–Honcharenko, K., Komar, M., Melnyk, N., & Komarnytsky, R. (2024). Sustainable information system for enhancing virtual company resilience through machine learning in smart city socio–economic scenarios. *Economics*, 12(2). <https://doi.org/10.2478/eoik-2024-0022> (Індексована в наукометричній базі *Web of Science* та *Scopus*, що відноситься до Q2 відповідно до *SCImago Journal & Country Rank*)

20. Lipianina–Honcharenko, K., Bodyanskiy, Y., Kustra, N., & Ivasechko, A. (2024). OLTW–TEC: Online learning with sliding windows for text classifier ensembles. *Frontiers in Artificial Intelligence*, 7, 1401126. <https://doi.org/10.3389/frai.2024.1401126> (Індексована в наукометричній базі *Scopus* та відноситься до Q2 відповідно до *SCImago Journal & Country Rank*)

21. Bodyanskiy, Y. V., Lipianina–Honcharenko, K. V., & Sachenko, A. O. (2022). Ensemble of adaptive predictors for multivariate nonstationary sequences and its online learning. *Radio Electronics, Computer Science, Control*, (4(67)), 91–97, <https://doi.org/10.15588/1607-3274-2023-4-9> (Індексована в наукометричній базі Web of Science)

22. Lipianina–Honcharenko, K., Bodyanskiy, Y., & Sachenko, A. (2023). Intelligent information system of the city's socio–economic infrastructure. *System Research and Information Technologies*, (3), 108–120. <https://doi.org/10.20535/srit.2308-8893.2023.3.08> (Індексована в наукометричній базі Scopus та відноситься до Q4 відповідно до SCImago Journal & Country Rank)

*Монографії (розділи у колективних монографіях):*

23. Lipyanina, H., Sachenko, O., Lendyuk, T., Sachenko, A., & Vasylykiv, N. (2020). Intelligent method of forming the HR management short–term project. In *Advances in Intelligent Systems and Computing* (pp. 1045–1055). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-63270-0\\_71](https://doi.org/10.1007/978-3-030-63270-0_71)

24. Lipianina–Honcharenko, K., Wolff, C., Chyzhovska, Z., Sachenko, A., Lendiuk, T., & Grodskyi, S. (2022). Intelligent method for forming the consumer basket. In *Communications in Computer and Information Science* (pp. 221–231). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-16302-9\\_17](https://doi.org/10.1007/978-3-031-16302-9_17) (Індексована в наукометричній базі Scopus)

25. Lipyanina, H., Maksymovych, V., Sachenko, A., Lendyuk, T., Fomenko, A., & Kit, I. (2020). Assessing the investment risk of virtual IT company based on machine learning. In *Communications in Computer and Information Science* (pp. 167–187). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-61656-4\\_11](https://doi.org/10.1007/978-3-030-61656-4_11) (Індексована в наукометричній базі Scopus)

26. Lipianina–Honcharenko, K., Lendiuk, T., Sachenko, A., Osolinskyi, O., Zahorodnia, D., & Komar, M. (2022). An intelligent method for forming the advertising content of higher education institutions based on semantic analysis. In *Communications in Computer and Information Science* (pp. 169–182). Cham: Springer International Publishing.

[https://doi.org/10.1007/978-3-031-14841-5\\_11](https://doi.org/10.1007/978-3-031-14841-5_11) (Індексована в наукометричній базі Scopus)

Статті у матеріалах міжнародних конференцій, які індексуються у наукометричних базах Scopus та Web of Science:

27. Lipyanina–Goncharenko, H., Brych, V., Sachenko, S., Lendyuk, T., Bykovyy, P., & Zahorodnia, D. (2021). Method of forming a training sample for segmentation of tender organizers on a machine learning basis. In *COLINS* (pp. 1843–1852). <https://ceur-ws.org/Vol-2870/paper134.pdf> (Індексована в наукометричній базі Scopus та Web of Science)

28. Lipianina–Honcharenko, K., Kit, I., Zahorodnia, D., & Osolinskyi, O. (2023). Optimization method of advertising texts based on generative models. In *Proceedings of the 3rd International Workshop on Information Technologies: Theoretical and Applied Problems 2023* (Vol. 3628, pp. 126–135). <https://ceur-ws.org/Vol-3628/paper10.pdf> (Індексована в наукометричній базі Scopus)

29. Komar, M., Savchyshyn, R., Lipianina–Honcharenko, K., & Osolinskyi, O. (2023). Intelligent method for counting cars from satellite images. In *IntSol* (pp. 295–303). [https://ceur-ws.org/Vol-3538/Short\\_1.pdf](https://ceur-ws.org/Vol-3538/Short_1.pdf) (Індексована в наукометричній базі Scopus)

30. Lipianina–Honcharenko, K., Bodyanskiy, Y., Sachenko, A., Kit, I., Podchasova, T., & Lendiuk, T. (2023). An intelligent method of predicting the demand for goods/services in crisis conditions. In *Proceedings of the 3rd International Workshop of IT–Professionals on Artificial Intelligence (ProfIT AI 2023)* (Vol. 3641, pp. 224–232). <https://ceur-ws.org/Vol-3641/short2.pdf> (Індексована в наукометричній базі Scopus)

31. Lipyanina, H., Sachenko, S., Lendyuk, T., Brych, V., Yatskiv, V., & Osolinskyi, O. (2021). Method of detecting a fictitious company on a machine learning basis. In *Advances in Computer Science for Engineering and Education IV* (pp. 138–146). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-80472-5\\_12](https://doi.org/10.1007/978-3-030-80472-5_12) (Індексована в наукометричній базі Scopus)

32. Krysovaty, A., Lipianina–Honcharenko, H., Sachenko, S., Desyatnyuk, O., Banasik, A., & Lukasevych–Krutnyk, I. (2022). Recognizing the fictitious business entity on a logistic regression basis. In *Proceedings of the 3rd International Workshop on*

*Intelligent Information Technologies & Systems of Information Security* (Vol. 3156, pp. 315–327). <https://ceur-ws.org/Vol-3156/paper15.pdf> (Індексована в наукометричній базі Scopus )

33. Krysovatyu, A., Lipyanina–Goncharenko, H., Desyatnyuk, O., & Sachenko, S. (2021). Classification method of fictitious enterprises based on Gaussian Naive Bayes. In *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)* (pp. 210–216). <https://doi.org/10.1109/CSIT52700.2021.9648584> (Індексована в наукометричній базі Scopus )

34. Krysovatyu, A., Lipyanina–Goncharenko, H., Sachenko, S., & Desyatnyuk, O. (2021). Economic crime detection using support vector machine classification. In *Modern Machine Learning Technologies and Data Science Workshop (MoMLeT&DS 2021)* (Vol. 2917, pp. 830–840). <https://ceur-ws.org/Vol-2917/paper46.pdf> (Індексована в наукометричній базі Scopus )

35. Lipianina–Honcharenko, K., Sachenko, A., Semaniuk, V., Badasian, A., & Kopania, Ł. (2023). Intelligent method for selecting a business location in a smart city. In *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. (pp. 1184–1188). <https://doi.org/10.1109/IDAACS58523.2023.10348823> (Індексована в наукометричній базі Scopus)

36. Gramyak, R., Lipyanina–Goncharenko, H., Sachenko, A., Lendyuk, T., & Zahorodnia, D. (2022). Intelligent method for choosing a competitive product based on emotional feedback coloring. In *Proceedings of the 2nd International Workshop on Intelligent Information Technologies & Systems of Information Security* (Vol. 2853, pp. 346–357). <https://ceur-ws.org/Vol-2853/paper31.pdf> (Індексована в наукометричній базі Scopus)

37. Lipianina–Honcharenko, K., Lendiuk, T., Sachenko, A., & Wołoszyn, J. (2022). Method of forming the context of advertising and target audience based on associative rule learning. In *Proceedings of the Fifth International Workshop on Computer Modeling and Intelligent Systems (CMIS–2022)* (Vol. 3137, pp. 98–107). <https://ceur-ws.org/Vol-3137/paper9.pdf> (Індексована в наукометричній базі Scopus)

38. Lipyanina, H., Sachenko, S., Lendyuk, T., & Sachenko, A. (n.d.). Targeting model of HEI video marketing based on a classification tree. In *16th International Conference on ICT in Education, Research, and Industrial Applications. Integration, Harmonization, and Knowledge Transfer (ICTERI 2020)* (Vol. 2732, pp. 487–498). <https://ceur-ws.org/Vol-2732/20200487.pdf> (Індексована в наукометричній базі Scopus)

39. Lipianina–Honcharenko, K., Lukasevych–Krutnyk, I., Butryn–Boka, N., & Sachenko, A. (2022). Intelligent method for identifying fraudulent online stores. In *2021 IEEE 8th International Conference on Problems of Infocommunications, Science, and Technology (PIC S&T)*. <https://doi.org/10.1109/PICST54195.2021.9772195> (Індексована в наукометричній базі Scopus)

40. Koziuk, V., & Lipyanina–Goncharenko, H. (2021). Intelligent method of predicting the discount rate trend. In *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. IEEE. <https://doi.org/10.1109/IDAACS53288.2021.9660835> (Індексована в наукометричній базі Scopus)

41. Lipianina–Honcharenko, K., Sachenko, A., Wolff, C., & Bodyanskiy, Y. (2023). Simulation model for determining the quality of life in Ukrainian cities during the war. In *2023 IEEE European Technology and Engineering Management Summit (E-TEMS)* (pp. 97–101). <https://doi.org/10.1109/E-TEMS57541.2023.10424587> (Індексована в наукометричній базі Scopus)

42. Pariy, V., Lipianina–Honcharenko, K., Brukhanskyi, R., Sachenko, A., Tkachyk, F., & Lendiuk, D. (2023, November). Intelligent verbal interaction methods with non–player characters in metaverse applications. In *2023 IEEE 5th International Conference on Advanced Information and Communication Technologies (AICT)* (pp. 67–71). <https://doi.org/10.1109/AICT61584.2023.10452688> (Індексована в наукометричній базі Scopus)

43. Lipianina–Honcharenko, Kh., Soia, M., Yurkiv, Kh., & Ivasechko, A. (2024). Evaluation of the effectiveness of machine learning methods for detecting disinformation in Ukrainian text data. In *Proceedings of the Seventh International Workshop on Computer*



*Modeling and Intelligent Systems (CMIS-2024)* (pp. 97–109). <https://ceur-ws.org/Vol-3702/paper9.pdf> (Індексована в наукометричній базі Scopus)

44. Sachenko, A., Lendiuk, T., Lipianina–Honcharenko, K., Dobrowolski, M., Boguta, G., & Bytsyura, L. (2024). Method of determining the text sentiment by thematic rubrics. In *COLINS* (pp. 404–414). <https://ceur-ws.org/Vol-3688/paper26.pdf> (Індексована в наукометричній базі Scopus)

45. Lipianina–Honcharenko, K., Melnychuk, A., Yurkiv, K., Hladiy, G., & Telka, M. (2024). Integrated approach to the international aspects of online dispute resolution formation. In *Proceedings of the First International Workshop of Young Scientists on Artificial Intelligence for Sustainable Development* (pp. 88–98). <https://ceur-ws.org/Vol-3716/paper7.pdf> (Індексована в наукометричній базі Scopus)

46. Komar, M., Lipianina–Honcharenko, K., Domanskyi, V., & Melnyk, N. (2024). Regression–based method for real–time solar power plant efficiency forecasting. In *Proceedings of the Modern Machine Learning Technologies Workshop (MoMLeT 2024)* (pp. 235–245). <https://ceur-ws.org/Vol-3711/paper14.pdf> (Індексована в наукометричній базі Scopus)

47. Sachenko, A., Lendiuk, T., Lipianina–Honcharenko, K., Koval, V., Hladiy, G., & Halias, Y. (2024). Evaluation of ensemble machine learning models for movie recommendation systems. In *Proceedings of the Modern Machine Learning Technologies Workshop (MoMLeT 2024)* (pp. 273–286). <https://ceur-ws.org/Vol-3711/paper17.pdf> (Індексована в наукометричній базі Scopus)

48. Lipianina–Honcharenko, K., Maika, N., Sachenko, S., Kopania, L., & Soia, M. (2024). A cyclical approach to legal document analysis: Leveraging AI for strategic policy evaluation. In *CEUR Workshop Proceedings* (Vol. 3736, pp. 201–211). <https://ceur-ws.org/Vol-3736/paper15.pdf> (Індексована в наукометричній базі Scopus)

*Публікації, які додатково відображають наукові результати дисертації:*

49. Крисоватий, А. І., Лип'яніна–Гончаренко, Х. В., та ін. (2022). *Метод виявлення фіктивних підприємств на основі машинного навчання*. Авторське право на твір № 111509, Україна. Опубл. 31.03.2022, Бюл. № 69.

## ЗМІСТ

Перелік умовних скорочень .....	21
Вступ .....	23
Розділ 1. Стан і перспективи розвитку аналізу соціально–економічних даних в інфраструктурі територіальної громади .....	40
1.1 Особливості соціально–економічних даних для інфраструктури територіальної громади .....	40
1.2 Огляд існуючих методів аналізу соціально–економічних даних для територіальних громад.....	48
1.3 Огляд методів аналізу даних, які можуть бути застосовані для інфраструктури ТГ .....	69
1.4 Висновки, постановка наукової проблеми та задач дослідження.....	85
Розділ 2. Принципи синтезу інформаційної технології та методологія інтелектуального аналізу соціально–економічних даних територіальних громад ....	88
2.1 Узагальнений принципи синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних територіальних громад	88
2.2 Методологія інтелектуального аналізу соціально–економічних даних територіальних громад.....	94
2.3 Дослідження даних в управлінні територіальними громадами на стаціонарність .....	99
2.4 Метод формування навчальної вибірки для нестаціонарних даних на основі RFM та кластерного аналізу.....	104
2.5 Метод ансамблю адаптивних прогнозних моделей для багатовимірних нестаціонарних послідовностей в реальному часі .....	112
2.6 Структура потоків даних інформаційної технології інтелектуального аналізу .....	121
2.7 Висновки .....	127
Розділ 3. Методи та засоби класифікаційного аналізу кількісних соціально–економічних даних територіальних громад.....	129

3.1	Метод класифікаційного аналізу кількісних соціально–економічних даних територіальних громад .....	129
3.2	Метод оцінки інвестиційного ризику віртуальної ІТ–компанії на основі машинного навчання.....	131
3.3	Метод виявлення фіктивного підприємства на основі машинного навчання .....	134
3.4	Експериментальні результати .....	137
3.5	Висновки .....	155
Розділ 4. Методи та засоби класифікаційного аналізу текстових соціально–економічних даних територіальних громад.....		157
4.1	Метод класифікаційного аналізу текстових соціально–економічних даних територіальних громад .....	157
4.2	Інтелектуальний метод виявлення дезінформації.....	159
4.3	Інтелектуальний метод формування короткострокового проєкту управління персоналом.....	166
4.4	Експериментальні результати .....	169
4.5	Висновки .....	180
Розділ 5. Методи та засоби кластерного аналізу соціально–економічних даних територіальних громад.....		182
5.1	Метод кластерного аналізу соціально–економічних даних територіальних громад.....	182
5.2	Інтелектуальний метод формування споживчого кошика .....	185
5.3	Інтелектуальний метод пошуку альтернативних постачальників .....	188
5.4	Експериментальні результати .....	192
5.5	Висновки .....	205
Розділ 6. Методи та засоби прогнозування соціально–економічних даних територіальних громад.....		207
6.1	Метод прогнозування соціально–економічних даних територіальних громад .....	207
6.2	Інтелектуальний метод прогнозування попиту на продукцію .....	210

6.3	Інтелектуальний метод вибору локації для старту бізнесу в територіальних громадах.....	213
6.4	Експериментальні результати.....	216
6.5	Висновки.....	228
	Розділ 7. Методи та засоби гібридного аналізу соціально–економічних даних територіальних громад.....	230
7.1	Метод гібридного аналізу соціально–економічних даних територіальних громад.....	230
7.2	Інтелектуальний метод класифікації рівня техногенних катастроф..	233
7.3	Інтелектуальний метод прогнозування обсягів відходів громади.....	237
7.4	Експериментальні результати.....	241
7.5	Висновки.....	258
	Висновки.....	260
	Список використаних джерел.....	263
	Додаток А. Структурні схеми методів.....	318
	Додаток Б. Опис вхідних параметрів для методів.....	332
	Додаток В. Експертний висновок.....	336
	Додаток Г. Список публікацій здобувача та відомості про апробацію результатів дисертації.....	339
	Додаток Д. Акти впровадження результатів дисертаційної роботи.....	347

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

- 5G – п’яте покоління мобільного зв’язку (5th Generation Mobile Networks)
- ANN – штучна нейронна мережа (Artificial Neural Network)
- API – інтерфейс програмування додатків (Application Programming Interface)
- AR – доповнена реальність
- AUC – площа під кривою (Area Under the Curve)
- AWS – хмарні сервіси Amazon Web Services
- Big Data – великі дані
- CNN – згортокова нейронна мережа (Convolutional Neural Network)
- CSV – значення, розділені комами (Comma-Separated Values)
- DBMS – система управління базами даних (Database Management System)
- GAN – генеративно-змагальна мережа (Generative Adversarial Network)
- GPS – глобальна система позиціонування (Global Positioning System)
- HR – людські ресурси
- IoT – інтернет речей (Internet of Things)
- JSON – JavaScript Object Notation
- KNN – метод найближчих сусідів (K-Nearest Neighbors)
- LDA – латентний Діріхле-розподіл (Latent Dirichlet Allocation)
- LSTM – довготривала короткострокова пам’ять (Long Short-Term Memory)
- MAE – середня абсолютна похибка (Mean Absolute Error)
- MSE – середньоквадратична похибка (Mean Squared Error)
- NLP – обробка природної мови (Natural Language Processing)
- PCA – метод головних компонент (Principal Component Analysis)
- ReLU – випрямляюча лінійна функція активації (Rectified Linear Unit)
- RFM – аналіз частоти покупок, останніх покупок та грошової вартості (Recency, Frequency, Monetary)
- RMSE – квадратичне середнє відхилення (Root Mean Squared Error)
- RNN – рекурентна нейронна мережа (Recurrent Neural Network)

ROC – крива робочих характеристик приймача (Receiver Operating Characteristic)

SQL – мова структурованих запитів (Structured Query Language)

SVM – метод опорних векторів (Support Vector Machine)

TF-IDF – частота термінів, обернена до частоти в документах (Term Frequency–Inverse Document Frequency)

XGBoost – екстремальне градієнтне підсилення (Extreme Gradient Boosting)

МН – машинне навчання

СУБД – система управління базами даних

ТГ – територіальна громада

ШІ – штучний інтелект

## ВСТУП

**Актуальність дослідження.** Сучасний розвиток інформаційних технологій та стрімке зростання обсягів соціально–економічних даних, що генеруються територіальними громадами (ТГ), створюють нові можливості та виклики для управління і прийняття рішень. Ці дані є цінним ресурсом для органів місцевого самоврядування, підприємств та інших зацікавлених сторін, оскільки містять інформацію про демографічні зміни, економічну активність, екологічні умови, соціальні настрої тощо. Однак, для того щоби повністю використати потенціал цих даних, необхідно розробити методи їх аналізу, які враховують специфіку нестационарності та високу динамічність соціально–економічних процесів.

Вивченню методів, моделей та засобів інтелектуального аналізу даних присвячено ряд робіт українських та іноземних учених: Саченка А. [347, 346], Бодяньського Є. [57, 58], Бармака О. [372, 344], Говорущенко Т. [169], Гнатчук Є. [395], Комара М. [208, 207], Єгорова Є. [106], Федорової Ю. [475], Чепурненка А. [76], Аленді А.А. (Alhendi, A. A.) [14], Алама С. (Alam, S.) [11], Андерссона Дж. (Andersson, J.) [26], Андеобу Л. (Andeobu, L.) [25], Арасу Б. С. (Arasu, B. S.) [30], Бенассі М. (Behnassi, M.) [46], Камеро А. (Camero, A.) [69], Чіанга Л.–Л. (Chiang, L.–L.) [77], Чіу Ч.–Ц. (Chiou, C.–C.) [79], Чивери М. (Civera, M.) [82], Комендадора Б.Е.В. (Comendador, B. E. V.) [98], Кости Н. Л. (Costa, N. L.) [88], Донга К. (Dong, Q.) [103], Ейхера А. (Eicher, A.) [107], Ел Морра К. (El Morr, C.) [108], Ел Уаді Дж. (El Ouadi, J.) [109], Ерруссо Х. (Errouso, H.) [111], Еспозіто Е. (Esposito, E.) [112], Етіаті Л. (Etaati, L.) [113], Фаніні Б. (Fanini, B.) [114], Френті П. (Fränti, P.) [119], Фройнда Й. (Freund, Y.) [120], Фу Д. (Fu, D.) [121], Фустоса Дж. (Fustos, J.) [122], Гедіклі А. (Gedikli, A.) [130], Гіти Т.В. (Geetha, T. V.) [131], Герке С. (Gerke, S.) [132], Гоше А. (Ghose, A.) [133], Гіллеспі А.А. (Gillespie, A. A.) [134], Голдані М. Х. (Goldani, M. H.) [135], Гомеса М. (Gomes, M.) [136], Гопнараяна А. (Gopnarayan, A.) [137], Говіндасам Р. (Govindasamy, R.) [138], Гриви А. (Griva, A.) [141], Гю І.Г.В. (Gue, I. H. V.) [143], Герцоні М. (Guerzoni, M.) [144], Гуїча Дж. А. (Guich, J. A.) [145],

Гуо С. (Guo, S.) [146, 147], Гупти І. П. (Gupta, Y. P.) [148], Гуркана Ф. (Gurcan, F.) [149], Гу С. (Gu, X.) [142] та багатьох інших. У цих роботах висвітлено різні аспекти інтелектуального аналізу даних, зокрема методи МН, адаптивні алгоритми та їх застосування в різних сферах.

Незважаючи на значний внесок цих досліджень у розвиток методів інтелектуального аналізу даних, існує низка невирішених проблем, пов'язаних зі специфікою соціально–економічних даних ТГ. Зокрема, існуючі методи не завжди враховують нестационарність часових рядів, високу динамічність змін соціально–економічних показників. Це призводить до недостатньої точності прогнозування для прийняття управлінських рішень.

Традиційні підходи до аналізу соціально–економічних даних часто не здатні адаптивно реагувати на швидкі зміни в них, що є критично важливим в умовах сучасних викликів, таких як: економічні кризи, екологічні проблеми, техногенні катастрофи та інформаційні загрози тощо. Відсутність інструментів для аналізу та прогнозування ускладнює процес прийняття рішень, планування розвитку інфраструктури та оптимізації використання ресурсів у ТГ.

Крім того, зростає потреба у впровадженні інтелектуальних систем підтримки прийняття рішень, які б інтегрували передові методи машинного навчання (МН), здатні обробляти масиви даних у режимі реального часу. Це особливо актуально для забезпечення соціальної стійкості та розвитку ТГ, управління екологічними ситуаціями, розвитку підприємницької діяльності та просування товарів на інтернет–ринку. Тому існує нагальна потреба у розробці нових інтелектуальних моделей, методів та інформаційних технологій, які б враховували специфіку соціально–економічних даних ТГ, забезпечували високу точність прогнозування, адаптивність до змінних умов та у прийнятті рішень.

Аналіз сучасного стану досліджень та практики обробки соціально–економічних даних ТГ виявив низку суттєвих протиріч, що обмежують можливості використання цих даних для управління та прийняття рішень. Зокрема, існують такі протиріччя:



1) між швидким зростанням обсягів соціально–економічних даних, які є гетерогенними та динамічними, і обмеженими можливостями сучасних методів їх інтеграції, аналізу та прогнозування. Традиційні інструменти не здатні адаптуватися до постійних змін соціально–економічних процесів і не враховують всю різноманітність джерел даних;

2) між необхідністю швидкого реагування на сучасні виклики, такі як пандемії, економічна нестабільність чи військова агресія, і відсутністю інтегрованих інформаційних систем, що забезпечують точний аналіз і оперативне прогнозування. Наявні технології часто мають низьку адаптивність до кризових умов, що знижує їхню точність у реальному часі;

3) між вимогою до інтеграції різнорідних джерел даних для забезпечення комплексного аналізу (наприклад, статистичних показників, відкритих даних, соціальних опитувань) і технічними бар'єрами їхньої обробки. Такі бар'єри включають проблеми сумісності, нерівномірності даних і високі обчислювальні витрати;

4) між потребою в прозорості та зрозумілості результатів аналізу для прийняття обґрунтованих рішень і складністю існуючих методів, які часто незрозумілі для кінцевих користувачів, включаючи управлінців ТГ.

Для розв'язання зазначених протиріч у дисертаційній роботі сформульовано актуальну *науково–прикладну проблему* щодо розроблення методологічних основ інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ. Ця технологія забезпечить адаптивність до змінних умов за рахунок використання методів машинного навчання для моделювання нестационарних і динамічних процесів; інтеграцію різнорідних джерел даних через створення інструментів для збирання, обробки та аналізу структурованих і неструктурованих даних; прозорість і зрозумілість завдяки моделям з пояснюваними результатами, що сприятимуть прийняттю управлінських рішень; а також оперативність і точність прогнозування через впровадження адаптивних прогнозних моделей з ковзним вікном для роботи в режимі реального часу.

### **Зв'язок роботи з науковими програмами, планами, темами, грантами.**

Наведені в дисертації дослідження проводились в рамках міжнародних та державних науково–дослідних проєктів, зокрема: Erasmus+: KA2 CBHE «Міждоменні компетенції для забезпечення здорової та безпечної роботи у 21 столітті» (WORK4CE) 619034–EPP–1–2020–1–UA–EPPKA2–CBHE–JP; Creative Europe: AURA – Auralisation of Acoustic Heritage Sites using Augmented and Virtual Reality (nr. 101008547); ERASMUS–EDU–2023–PI–ALL–INNO під назвою "My Farm" (Proposal number: 101140288); держбюджетної наукової роботи молодих учених (номер держреєстрації 0117U003871) «Методи та засоби структурно–статистичної ідентифікації ієрархічних об'єктів за характерними точками їх контурів» (термін виконання: 01.01.2019 р. – 31.12.2020 р.); держбюджетної наукової роботи молодих учених (номер держреєстрації 0117U003871) «TruScanAI: інструмент виявлення фейкової інформації на основі технологій ШІ для боротьби з дезінформацією» (термін виконання: 01.01.2024 р. – 31.12.2025 р.); науково–дослідної роботи ІОСУ2023 «К» «Інтелектуальні методи, моделі та технології соціально–економічного розвитку ТГ в умовах сьогодення» (протокол № 4 від 09.11.2022 р., термін виконання: 01.01.2023–31.12.2027); перспективного плану розвитку наукового напрямку «Технічні науки» Західноукраїнського національного університету (номер державної реєстрації 0121U114705, термін виконання: 01.04.2023 р. – 31.12.2023 р.).

Отже, **метою дисертаційної роботи** є забезпечення соціальної стійкості соціально–економічної інфраструктури ТГ шляхом реалізації управлінських рішень на основі інформаційної технології інтелектуального аналізу даних.

Досягнення цієї мети зумовило потребу теоретичних розробок, визначення та послідовного вирішення таких **задач**:

- 1) провести дослідження сучасного стану та визначити перспективи розвитку аналізу соціально–економічних даних у межах інфраструктури ТГ;
- 2) розробити узагальнений принцип синтезу інформаційної технології для інтелектуального аналізу соціально–економічних даних ТГ;
- 3) розробити методологію аналізу соціально–економічних даних громад, що включає адаптивні методи класифікації, кластеризації та прогнозування;

- 4) розробити новий метод формування навчальної вибірки для нестационарних даних на основі RFM та кластерного аналізу;
- 5) розробити метод ансамблевих адаптивних прогнозних моделей для багатовимірного аналізу;
- 6) розробити метод класифікаційного аналізу кількісних соціально–економічних даних;
- 7) розробити метод класифікаційного аналізу текстових соціально–економічних даних;
- 8) розробити метод кластерного аналізу соціально–економічних даних;
- 9) розробити метод прогнозування соціально–економічних даних;
- 10) розробити метод гібридного аналізу соціально–економічних даних;
- 11) оцінити ефективність запропонованих методів та інформаційної технології.

*Об'єкт дослідження* – процеси обробки інформації ТГ у динамічному середовищі, які впливають на їх розвиток при стійкому управлінні ресурсами.

*Предмет дослідження* – інтелектуальні методи, моделі та інформаційні технології аналізу соціально–економічних даних ТГ для підвищення точності прогнозування, оптимізації управління ресурсами та забезпечення стійкого розвитку громад.

**Методи дослідження.** Для досягнення поставленої мети в дисертаційній роботі використовуються комплекс теоретичних і практичних методів, спрямованих на вирішення визначених завдань. Теоретичний аналіз застосовано для вивчення існуючих підходів до обробки та аналізу соціально–економічних даних у контексті управління ТГ. Математичне моделювання використовувалося для розробки нових інтелектуальних моделей і методів, зокрема ансамблевих адаптивних предикторів та методів формування навчальної вибірки. Алгоритми МН були залучені для реалізації інтелектуальних методів прогнозування, класифікації та оптимізації в управлінні соціальними, екологічними та бізнес–процесами ТГ. Експериментальне моделювання та комп'ютерні симуляції проводилися для оцінки запропонованих методів, з використанням реальних та синтетичних даних. Для верифікації і підтвердження

практичної значущості результатів застосовувалися статистичні методи аналізу та порівняння з існуючими підходами.

**Наукова новизна результатів дисертаційної роботи.** Отримано такі нові наукові результати:

1) вперше запропоновані узагальнені принципи синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ, особливістю якої є здатність інтегрувати різнорідні дані, враховувати динамічність соціально–економічних процесів, що забезпечило підвищення соціальної стійкості інфраструктури громад за рахунок інтеграції гетерогенних даних, адаптивного вибору методів аналізу та формування управлінських рішень;

2) вперше розроблено методологію аналізу соціально–економічних даних ТГ як сукупність методів класифікаційного, кластерного та гібридного аналізу, а також прогнозування соціально–економічних даних, особливістю якої є адаптивність для роботи з нестационарними даними, реальним часом для оперативного прийняття рішень та інтеграцією різнорідних джерел даних, що дозволило підвищити точність прогнозування, класифікації та прозорість управлінських рішень;

3) розроблено новий метод формування навчальної вибірки для нестационарних процесів на основі RFM–аналізу та кластерного аналізу, що, на відміну від існуючих підходів, забезпечує сегментацію та кластеризацію, для підвищення точності опрацювання нових вхідних наборів даних у задачах прогнозування, класифікації та аналізу нестационарних соціально–економічних даних;

4) вперше розроблено метод ансамблевих адаптивних прогнозних моделей для багатовимірного аналізу, особливістю якого є оцінювання на «ковзному вікні» та метамоделі першого і другого рівнів на основі оптимізації функції Лагранжа, що забезпечило точність прогнозів багатовимірних нестационарних процесів з врахуванням їхньої динамічності;

5) удосконалено метод класифікаційного аналізу кількісних соціально–економічних даних, який, на відміну від відомих підходів, забезпечує інтеграцію структурованих, неструктурованих і напівструктурованих даних, що дозволило

підвищити точність класифікації кількісних показників для реалізації управлінських рішень;

6) удосконалено метод класифікаційного аналізу текстових соціально–економічних даних, який, на відміну від відомих підходів, забезпечує інтеграцію текстових джерел різної структури та їхній аналіз за допомогою технологій обробки природної мови, що дозволило підвищити ефективність виявлення закономірностей для реалізації управлінських рішень;

7) удосконалено метод кластерного аналізу соціально–економічних даних, який, на відміну від відомих підходів, забезпечує ідентифікацію груп об'єктів з подібними характеристиками, що дозволило підвищити точність розподілу ресурсів для реалізації управлінських рішень;

8) удосконалено метод прогнозування соціально–економічних даних, який, на відміну від відомих підходів, забезпечує високоточне прогнозування динамічних процесів шляхом інтеграції структурованих, неструктурованих і напівструктурованих даних із застосуванням адаптивних методів інтелектуального аналізу, що дозволило підвищити точність прогнозування для реалізації управлінських рішень;

9) удосконалено метод гібридного аналізу соціально–економічних даних, який, на відміну від відомих підходів, забезпечує врахування кількісних та якісних показників завдяки інтеграції різнорідних даних і багаторівневого підходу до аналізу, що дозволило підвищити гнучкість і адаптивність процесу прийняття рішень для реалізації управлінських рішень.

**Практичне значення отриманих результатів.** Розробка інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ дозволяє подолати існуючі обмеження традиційних методів обробки даних. Запропонована технологія забезпечує адаптивність до динамічних змін соціально–економічних процесів завдяки використанню методів машинного навчання. Це дозволяє оперативно реагувати на зміни в соціально–економічних показниках та враховувати специфіку різнорідних джерел даних, таких як статистичні показники, відкриті дані та результати соціологічних опитувань. Практичне застосування такої технології

сприятиме точному прогнозуванню ризиків, що є критично важливим для управління в умовах сучасних викликів.

Інтеграція різнорідних джерел даних у межах запропонованої технології дозволяє вирішувати завдання комплексного аналізу та моделювання стану ТГ. Це забезпечує можливість виявлення тенденцій, прогнозування змін і оцінювання впливу прийнятих рішень на соціально–економічний стан громад. Створення модулів для обробки як структурованих, так і неструктурованих даних спрощує процес інтеграції інформації з різних джерел, усуваючи технічні бар'єри.

Прозорість і зрозумілість результатів аналізу, забезпечено розробленою інформаційною системою з пояснюваними результатами, дозволяє управлінцям ТГ приймати обґрунтовані рішення. Розроблені інструменти сприяють не лише підвищенню точності при аналізі даних, але й покращенню комунікації між зацікавленими сторонами завдяки доступності інформації для кінцевих користувачів. У підсумку, впровадження технології інтелектуального аналізу даних дозволяє ТГ досягати сталого розвитку, підвищувати соціальну стійкість та адаптуватися до кризових умов.

**Реалізація результатів та впровадження.** Результати дисертаційної роботи впроваджено в освітні процеси Західноукраїнського національного університету для підготовки здобувачів вищої освіти на першому та другому (освітніх) рівнях за спеціальністю «Комп'ютерні науки» через вдосконалення існуючих навчальних курсів: «Інтелектуальний аналіз даних», «Машинне навчання», «Методи та засоби штучного інтелекту», «Інтелектуальна обробка тексту та природної мови», «Методи та засоби обробки зображень», а також для підготовки здобувачів вищої освіти на третьому (освітньо–науковому) рівні за спеціальністю «Комп'ютерні науки» під час викладання дисциплін «Інтелектуальна обробка даних у складних системах» та «Методи обробки й аналізу експериментальних даних» (акт впровадження від 30 жовтня).

Результати дисертаційної роботи впроваджено при виконанні науково–дослідних робіт та проектів:

– Erasmus+:KA2 CBHE «Міждоменні компетенції для забезпечення здорової та безпечної роботи у 21 столітті / «Cross-domain competences for health and safe work in the 21st century» (WORK4CE) 619034-EPP-1-2020-1-UA-EPPKA2-CBHE-JP (довідка від 27 жовтня 2024р.);

– Creative Europe: AURA – Auralisation of acoustic heritage sites using Augmented and Virtual Reality (nr. 101008547) (акт впровадження №20 від 29 листопада 2023р.);

– ERASMUS-EDU-2023-PI-ALL-INNO, «My Farm» (nr. 101140288) (довідка впровадження від 27 жовтня 2024р.);

– договір БФ/4-2021 від 01.06.2021 р. «Виконання завдань перспективного плану розвитку наукового напрямку «Технічні науки» Західноукраїнського національного університету» (номер державної реєстрації 0121U114705) (акт впровадження від 29 жовтня 2024р.);

– держбюджетна наукова робота молодих учених (номер держреєстрації 0117U003871) «TruScanAI: інструмент виявлення фейкової інформації на основі технологій ШІ для боротьби з дезінформацією» (акт впровадження від 29 жовтня 2024р.);

– науково-дослідна робота ІОСУ2023 «К» «Інтелектуальні методи, моделі та технології соціально-економічного розвитку ТГ в умовах сьогодення». Протокол № 4 від 09.11.2022 р. (акт впровадження від 29 жовтня 2024р.).

Переважна більшість цих науково-дослідних робіт і проектів виконувалась під науковим керівництвом д. т. н. проф. Анатолія Саченка, неоцінімі ідеї, консультації і поради якого зробили суттєвий вклад на дану дисертаційну роботу, і якому авторка висловлює глибоку подяку.

Велику подяку авторка висловлює також проф. Карстену Вольфу (Carsten Wolff) за періодичні консультації і дискусії з основних аспектів дисертаційної роботи і положень наукової новизни, що сприяло покращенню якості дисертації та підготовці ключових статей для періодичних міжнародних видань, індексованих науко метричними базами Web of Science і Scopus.

Результати дисертаційної роботи апробовано та впроваджено у ряді підприємств та установ:

–ІТ–компанії «МагнетікВан», м. Тернопіль (акт впровадження №15 від 26 вересня 2024);

–Департамент цифрової трансформації Тернопільської обласної державної адміністрації (акт впровадження №569 від 20 жовтня 2024 року);

–Великобerezовицька ТГ (акт впровадження №336 від 28 жовтня 2024 року);

–Громадська організація «Прогресивні люди» (акт впровадження №23 від 18 жовтня 2024).

**Особистий внесок здобувача.** Всі положення, які виносяться на захист, належать автору особисто. Роботи [456, 460, 461, 466, 467] виконувалися без співавторів.

У роботах, опублікованих у співавторстві, здобувач особисто виконував ролі головного дослідника, автора ідеї, аналітика даних, автора методології, експериментатора та програміста. Відповідно до пунктів наукової новизни дисертаційної роботи зроблено такі внески:

– розроблено узагальнений принцип синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ [241, 245, 256, 459];

– розроблено методологію аналізу соціально–економічних даних громад з адаптивними методами класифікації, кластеризації та прогнозування [208, 239, 242, 245, 246, 250, 251, 255–257, 346, 455, 458];

– метод формування навчальної вибірки для нестационарних даних на основі RFM–аналізу та кластерного аналізу [264, 458, 466];

– метод ансамблю адаптивних прогнозних моделей для багатовимірного аналізу [56, 242, 458].

– метод класифікаційного аналізу кількісних соціально–економічних даних [208, 215–217, 239, 245, 249, 251, 256, 259, 262, 263, 459, 465, 468].

– метод класифікаційного аналізу текстових соціально–економічних даних з використанням технологій обробки природної мови [139, 242, 243, 245, 247, 248, 250, 254, 256, 261, 315, 346, 347, 455, 459, 460, 462, 464].

– метод кластерного аналізу соціально–економічних даних для ідентифікації груп об’єктів з подібними характеристиками [245, 255, 309, 457].



– метод прогнозування соціально–економічних даних з інтеграцією адаптивних методів інтелектуального аналізу [207, 211, 214, 240, 253, 456, 457].

– метод гібридного аналізу соціально–економічних даних з врахуванням кількісних та якісних показників [246, 245, 256, 257, 457, 461, 467, 469].

Особистий внесок інших співавторів у спільних публікаціях: у статті [56] Саченко А. (Sachenko A.) виконував керівництво, концептуалізацію та методологію дослідження; Бодянський Є. (Bodianskiy Y.) здійснював математичне редагування, наукове консультування, обговорення результатів дослідження, рецензування та коригування рукопису. У статті [139] Загородня Д. (Zahorodnia D.) виконувала обов'язки автора–кореспондента, адміністрування дослідження; Лендюк Т. (Lendiuk T.) виконував адміністрування дослідження; Грам'як Р. (Gramyak R.) займався опрацюванням результатів експериментів; Саченко А. (Sachenko A.) здійснював наукове консультування, керівництво проектом, концептуалізацію дослідження. У статті [207] Комар М. (Komar M.) виконував рецензування та коригування рукопису; Доманський В. (Domanskyi V.) працював над візуалізацією результатів дослідження; Мельник Н. (Melnyk N.) займалися опрацюванням результатів експериментів. У статті [208] Комар М. (Komar M.) та Савчишин Р. (Savchyshyn R.) виконували рецензування та коригування рукопису; Осолінський О. (Osolinskyi O.) спільно із здобувачем працював над оглядом відомих методів та рішень. У статтях [215, 216, 217] Десятнюк О. (Desyatnyuk O.), Крисоватий А. (Krytsovatiy A.), Саченко С. (Sachenko S.), Лукасевич–Крутник І. (Lukasevych–Krutnyk I.) виконували рецензування та коригування рукопису; Бадасян А. (Badasian A.) займався опрацюванням результатів експериментів. У статті [239] Саченко А. (Sachenko A.) виконував керівництво проектом, концептуалізацію та методологію дослідження; Бодянський Є. (Bodianskiy Y.) наукове консультування, рецензування та коригування рукопису; Волф К. (Wolff C.) брав участь у методології дослідження, рецензуванні та коригуванні рукопису. У статті [240] Саченко А. (Sachenko A.) виконував керівництво проектом, концептуалізацію та методологію дослідження; Бодянський Є. (Bodianskiy Y.) здійснював математичне редагування, наукове консультування, рецензування та коригування рукопису; Лендюк Т. (Lendiuk T.) виконав підготовку чернетки

рукопису; Подчасова Т. (Podchasova T.) виконувала рецензування та коригування рукопису; Кіт І. (Kit I.) займався візуалізацією результатів дослідження. У статті [241] Саченко А. (Sachenko A.) виконував керівництво проектом, концептуалізацію та методологію дослідження; Бодянський Є. (Bodyanskiy Y.) здійснював наукове консультування, рецензування та коригування рукопису. У статті [242] Бодянський Є. (Bodyanskiy Y.) здійснював математичне редагування, наукове консультування, рецензування та коригування рукопису; Івасечко А. (Ivasechko A.) виконувала обов'язки автора–кореспондента, адміністрування дослідження, візуалізацію результатів дослідження; Кустра Н. (Kustra N.) виступала як рецензент. У статті [243] Загородня Д. (Zahorodnia D.) виконувала обов'язки автора–кореспондента, адміністрування дослідження; Кіт І. (Kit I.) працював над візуалізацією результатів дослідження; Осолінський О. (Osolinskyi O.) спільно зі здобувачем працював над оглядом відомих методів та рішень. У статті [245] Комар М. (Komar M.) виконував рецензування та коригування рукопису; Комарницький Р. (Komarnytsky R.) та Мельник Н. (Melnyk N.) займалися опрацюванням результатів експериментів. У статті [246] Комар М. (Komar M.) виконував рецензування та коригування рукопису; Гаврилюк М. (Havryliuk M.) та Шиманський В. (Shymaskyi V.) виступали як рецензенти; Осолінський О. (Osolinskyi O.) спільно зі здобувачем працював над оглядом відомих методів та рішень та виконував коригування рукопису. У статті [247] Саченко А. (Sachenko A.) виконував керівництво проектом, концептуалізацію та методологію дослідження; Лендюк Т. (Lendiuk T.) виконав адміністрування дослідження, підготовку чернетки рукопису; Воложин Ю. (Woloszyn J.) виконував рецензування та коригування рукопису. У статті [248] Саченко А. (Sachenko A.) виконував керівництво проектом, концептуалізацію та методологію дослідження; Загородня Д. (Zahorodnia D.) займалась адміністрування дослідження; Лендюк Т. (Lendiuk T.) виконував підготовку чернетки рукопису; Комар М. (Komar M.) виконував рецензування та коригування рукопису; Осолінський О. (Osolinskyi O.) спільно зі здобувачем працював над оглядом відомих методів та рішень. У статті [249] Саченко А. (Sachenko A.) виконував керівництво проектом, концептуалізацію та методологію дослідження; Лукасевич–Крутник І. (Lukasevych–Krutnyk I.) виступала

як рецензент; Бутрин–Бока Н. (Butryn–Boka N.) займалася опрацюванням результатів експериментів. У статті [250] Коранія Ї. та Майка Н. (Maika N.) виступали як рецензенти; Саченко С. також виконувала (Sachenko S.) рецензування та коригування рукопису; Соїя М. (Soia M.) займалася опрацюванням результатів експериментів. У статті [251] Гладій Г. (Hladiy G.) виконував коригування рукопису; Мельничук А. (Melnychuk A.) виконувала рецензування; Юрків Х. (Yurkiv K.) займалась опрацюванням результатів експериментів; Телька М. (Telka M.) виконував підготовку чернетки рукопису. У статті [253] Саченко А. (Sachenko A.) виконував керівництво проектом, концептуалізацію та методологію дослідження; Коранія Ї. та Семанюк В. (Semaniuk V.) виступали як рецензенти; Badasian A. займалась опрацюванням результатів експериментів. У статті [254] Івасечко А. (Ivasechko A.) виконував адміністрування дослідження та візуалізацію результатів дослідження; Соїя М. (Soia M.) та Юрків Х. (Yurkiv K.) займалися опрацюванням результатів експериментів. У статті [255] Саченко А. (Sachenko A.) виконував керівництво проектом, концептуалізацію та методологію дослідження; Волф К. (Wolff C.) брав участь у методології дослідження, рецензуванні та коригуванні рукопису; Лендюк Т. (Lendiuk T.) виконував адміністрування дослідження, підготовку чернетки рукопису; Гродський С. (Grodskiy S.), Чижовська З. (Chyzhovska Z.) займалися опрацюванням результатів експериментів. У статті [256] Саченко А. (Sachenko A.) виконував керівництво проектом, концептуалізацію дослідження, методологію дослідження; Wolff C. також брав участь у методології дослідження, рецензуванні та коригуванні рукопису; Десятнюк О. (Desyatnyuk O.), Саченко С. (Sachenko S.) виконували рецензування та коригування рукопису; Кіт І. (Kit I.) працював над візуалізацією результатів дослідження. У статті [257] Саченко А. (Sachenko A.) виконував керівництво проектом, концептуалізацію та методологію дослідження; Волф К. (Wolff C.) брав участь у методології дослідження, рецензуванні та коригуванні рукопису; Загородня Д. (Zahorodnia D.) виконувала адміністрування дослідження, підготовку чернетки рукопису; Кіт І. (Kit I.) працював над візуалізацією результатів дослідження. У статті [261] Максимович В. (Maksymovych V.) виконувала рецензування та коригування рукопису; Саченко А. (Sachenko A.) виконував

керівництво проєктом, концептуалізацію та методологію дослідження; Лендюк Т. (Lendiuk T.) виконував адміністрування дослідження, підготовку чернетки рукопису; Кіт І. (Kit I.) працював над візуалізацією результатів дослідження. У статті [261] Саченко А. (Sachenko A.) виконував керівництво проєктом, концептуалізацію та методологію дослідження; Лендюк Т. (Lendiuk T.) виконував адміністрування дослідження, підготовку чернетки рукопису; Саченко О. (Sachenko O.), Васильків Н. (Vasylykiv N.) виконували рецензування та коригування рукопису. У статті [262] Саченко А. (Sachenko A.) виконував керівництво проєктом, концептуалізацію та методологію дослідження; Лендюк Т. (Lendiuk T.) виконував адміністрування дослідження, підготовку чернетки рукопису; Саченко С. (Sachenko S.) виконував рецензування та коригування рукопису. У статті [263] Лендюк Т. (Lendiuk T.) виконував адміністрування дослідження, підготовку чернетки рукопису; Саченко С. (Sachenko S.) та Яцків В. (Yatskiv V.) виконували рецензування та коригування рукопису; Брич В. (Brych V.) займався опрацюванням результатів експериментів. У статті [264] Загородня Д. (Zahorodnia D.) виконувала підготовку чернетки рукопису; Лендюк Т. (Lendiuk T.) виконував адміністрування дослідження; Саченко С. (Sachenko S.) виконувала рецензування та коригування рукопису; Брич В. (Brych V.) займався опрацюванням результатів експериментів; Биковий П. (Bykovyy P.) працював над візуалізацією результатів дослідження. У статті [315] Саченко А. (Sachenko A.) виконував керівництво проєктом, концептуалізацію та методологію дослідження; Ткачик Ф. (Tkachuk F.), Парій В. (Pariy V.) спільно з автором працювали над оглядом відомих методів та рішень; Брюханський Р. (Brukhanyskiy R.), Лендюк Д. (Lendyuk D.) займалися опрацюванням результатів експериментів. У статті [346] Саченко А. (Sachenko A.) виконував керівництво проєктом, концептуалізацію та методологію дослідження; Лендюк Т. (Lendiuk T.) виконував адміністрування дослідження, підготовку чернетки рукопису; Добровольський М. (Dobrowolski M.) виконував рецензування та коригування рукопису; Богута Г. (Boguta G.), Бицюра Л. (Bytsyura L.) займалися опрацюванням результатів експериментів. У статті [347] Саченко А. (Sachenko A.) виконував керівництво проєктом, концептуалізацію та методологію дослідження; Лендюк Т. (Lendiuk T.) виконував адміністрування

дослідження, підготовку чернетки рукопису; Коваль В. (Koval V.), Гладій Г. (Hladiy G.) виконували рецензування та коригування рукопису; Галіас Ю. (Halias Y.) займався опрацюванням результатів експериментів. У статті [455] Комар М. виконував рецензування та коригування рукопису, методологію дослідження; Кіт І. працював над візуалізацією результатів дослідження; Мадараш Р., Юрків Х. займалися опрацюванням результатів експериментів. У статті [462] Кіт І. працював над візуалізацією результатів дослідження. У статті [463] Юрків Х. займалася опрацюванням результатів експериментів. У статті [464] Саченко А. виконував керівництво проектом, концептуалізацію та методологію дослідження; Лендюк Т. виконував адміністрування дослідження, підготовку чернетки рукопису; Комар М. виконував рецензування та коригування рукопису. У статті [465] Саченко А. виконував керівництво проектом, концептуалізацію та методологію дослідження; Лендюк Т. виконував адміністрування дослідження, підготовку чернетки рукопису; Комар М. виконував рецензування та коригування рукопису. У статті [468] Саченко А. виконував керівництво проектом, концептуалізацію та методологію дослідження; Комар М. виконував рецензування та коригування рукопису; Лендюк Т. виконував адміністрування дослідження, підготовку чернетки рукопису. У статті [469] Комар М. виконував рецензування та коригування рукопису; Юрків Х. займалася опрацюванням результатів експериментів; Лукянчук В. спільно з автором працював над оглядом відомих методів та рішень.

**Апробація результатів дисертації.** Основні результати дисертаційного дослідження багаторазово доповідалися та обговорювалися на 19 міжнародних і національних конференціях, а саме: 10th, 11th, 12th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS) (18–21 вересня 2019 року, Мец, Франція; 22–25 вересня 2021 року, Краків, Польща; 7–9 вересня 2023 року, Дортмунд, Німеччина); 4th, 5th IEEE International Conference on Advanced Information and Communication Technologies (AICT) (20–24 вересня 2021 року, Львів, Україна; листопад 2023 року, Київ, Україна); 5th, 7th International Workshop on Computer Modeling and Intelligent Systems (CMIS) (12 травня 2022 року, Запоріжжя, Україна; 3 травня 2024 року,

Запоріжжя, Україна); International Workshop on Intelligent Information Technologies & Systems of Information Security (ИТ&SIS) (23–25 вересня 2021 року, Хмельницький, Україна; 29–30 вересня 2022 року, Хмельницький, Україна); International Workshop on Modern Machine Learning Technologies (MoMLeT) (20–22 вересня 2021 року, Львів, Україна; 31 травня – 1 червня 2024 року, Львів, Україна); International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI) (20–22 листопада 2023 року, Ватерлоо, Канада); 16th IEEE International Conference on Computer Sciences and Information Technologies (CSIT) (22–25 вересня 2021 року, Львів, Україна); IEEE European Technology and Engineering Management Summit (E-TEMS) (9–11 березня 2023 року, Каунас, Литва); First International Workshop of Young Scientists on Artificial Intelligence for Sustainable Development (YS-AISD) (10–11 травня 2024 року, Тернопіль, Україна); International Conference on Computational Linguistics and Intelligent Systems (COLINS) (22–23 квітня 2021 року; 18–19 квітня 2024 року); III International Scientific Symposium «Intelligent Solutions» (INTSOL–2023) (26–28 квітня 2023 року); IEEE International Conference on Problems of Infocommunications, Science and Technology (PIC S&T) (5–7 жовтня 2021 року); 16th International Conference on ICT in Education, Research and Industrial Applications (ICTERI 2020) (6–8 жовтня 2020 року, Харків, Україна); Advances in Computer Science for Engineering and Education IV, частина Міжнародної конференції з комп'ютерних наук, інженерії та освітніх застосувань (ICCSEEA) (15–17 січня 2021 року); International Workshop on IT Project Management (ITPM) (14–17 вересня 2021 року).

**Публікації.** За темою дисертації опубліковано 49 наукових праць, зокрема:

–15 статей у наукових виданнях, включених до Переліку наукових фахових видань України [455–469];

–7 статей у періодичних виданнях, включених до категорії «А» Переліку наукових фахових видань України, або у закордонних виданнях, проіндексованих у базах даних Web of Science Core Collection та/або Scopus [56, 241, 242, 245, 246, 256, 257]. Відповідно до класифікації SCImago Journal and Country Rank, 1 статтю віднесено до квартилю Q1 [246], 4 статті – до квартилю Q2 [242, 245, 256, 257], 1 статтю – до квартилю Q4 [241];

–4 монографій (розділи у колективних монографіях) [248, 255, 259, 261];

–22 статті у матеріалах міжнародних конференцій, які індексуються у наукометричних базах Scopus та Web of Science [139, 207, 208, 215–217, 239, 240, 243, 247, 249–251, 253, 254, 262–264, 315, 346, 347];

–1 публікація, яка додатково відображає наукові результати дисертації [474].

**Структура та обсяг дисертації.** Дисертаційна робота складається з анотацій, вступу, 7 розділів, висновків, 5 додатків на 45 сторінках та списку використаних джерел з 475 найменувань на 52 сторінках. Загальний обсяг дисертації становить 361 сторінок, з них 299 сторінок основного тексту, 76 рисунків та 15 таблиць.

## РОЗДІЛ 1.

### СТАН І ПЕРСПЕКТИВИ РОЗВИТКУ АНАЛІЗУ СОЦІАЛЬНО– ЕКОНОМІЧНИХ ДАНИХ В ІНФРАСТРУКТУРІ ТЕРИТОРІАЛЬНОЇ ГРОМАДИ

#### **1.1 Особливості соціально–економічних даних для інфраструктури територіальної громади**

Соціальна стійкість є ключовим чинником сталого розвитку ТГ, який визначає здатність спільнот не лише адаптуватися до викликів та дестабілізуючих подій, але й трансформувати отриманий досвід у можливості для зростання. Для забезпечення управління ТГ необхідно враховувати динамічну природу соціально–економічних процесів, що відображають різноманітність потреб, ресурсів та можливостей громад. Одним з дієвих інструментів для досягнення соціальної стійкості є використання інтелектуального аналізу даних, що дозволяє моделювати соціально–економічну інфраструктуру ТГ, виявляти слабкі місця та розробляти оптимальні управлінські рішення.

Соціально–економічні дані для ТГ охоплюють широкий спектр інформації, яка характеризує життєдіяльність населення та функціонування інфраструктури. Ці дані включають демографічні показники, рівень доходів, зайнятість, стан освітніх та медичних закладів, транспортної системи, рівень доступу до цифрових технологій, а також екологічні аспекти. Особливістю таких даних є їх гетерогенність, неструктурованість і нерівномірність, що ускладнює їх обробку та аналіз традиційними методами.

Територіальні громади [454] – жителі, об'єднані постійним проживанням у межах села, селища, міста, що є самостійними адміністративно–територіальними одиницями, або добровільне об'єднання жителів кількох сіл, селищ, міст, що мають єдиний адміністративний центр.



ТГ створюються для управління ресурсами, організації громадського життя, розвитку інфраструктури та забезпечення потреб своїх членів. Управління такою громадою передбачає вибірні органи, такі як місцева рада та виконавчі комітети, які відповідають за прийняття рішень на місцевому рівні. Ці громади відіграють ключову роль у децентралізації влади, надаючи більше повноважень та ресурсів на місцевому рівні, що дозволяє громадянам більш активно брати участь в управлінні та розвитку своїх територій. Це також сприяє підвищенню прозорості та відповідальності місцевої влади перед громадою.

Інфраструктура [451] – це комплекс специфічних засобів, методів, процесів, а також різноманітних споруд і систем комунікацій, що створюють умови для нормального функціонування економічної, соціальної, екологічної та інших сфер суспільного життя. Ця сукупність елементів забезпечує підтримку, розвиток і відтворення суспільства, надаючи необхідні ресурси для його повноцінної діяльності. До складу інфраструктури входять окремі галузі економіки та система технічних, технологічних, організаційних, соціальних і комунікаційних зв'язків між усіма її складовими елементами.

Залежно від ролі, яку інфраструктура відіграє в економіці та суспільстві, її умовно поділяють на виробничу та соціальну. Окрім цього, інфраструктура може бути класифікована за іншими напрямками, наприклад, як регіональна, інноваційна тощо. Такий розподіл обґрунтований тим, що окремі галузі й сфери діяльності водночас обслуговують потреби підприємств (організацій, установ, компаній) та населення [480].

До виробничої інфраструктури входять підрозділи підприємств, які не беруть безпосередньої участі у створенні основної продукції, але забезпечують всі необхідні умови для функціонування основних виробничих процесів. Сюди належать допоміжні та обслуговуючі підрозділи (цехи, дільниці, господарства), комунікаційні мережі, природоохоронні споруди, засоби збору та обробки інформації, складські приміщення. Також до цієї категорії відносять діяльність, пов'язану з інжинірингом, матеріально–технічним забезпеченням, оптовою торгівлею тощо [451].

Соціальна інфраструктура, у свою чергу, зосереджена на задоволенні соціально–побутових і культурно–освітніх потреб працівників. До неї належать заклади громадського харчування, спортивні, освітні та культурні установи, заклади для відпочинку, медичні установи, житлово–комунальні об’єкти [451].

Окрему вагому складову інфраструктури становить інформаційна інфраструктура, яка відповідає за накопичення, зберігання та передачу виробничої, наукової, соціальної, екологічної та іншої інформації. Вона забезпечує обмін знаннями та даними, необхідними для ефективного функціонування всіх сфер життєдіяльності [451].

Отже, інфраструктура ТГ є комплексною системою [393] (Рис. 1.1), яка включає в себе різноманітні компоненти, кожен з яких відіграє важливу роль у забезпеченні сталого та функціонування середовища громади. Основні елементи інфраструктури охоплюють транспортну систему, енергетику, водопостачання, управління відходами, освіту, охорону здоров'я, зелені простори, культурні та розважальні заклади, а також бізнес–сектор. Кожен з цих аспектів взаємодіє та впливає на інші, формуючи інтегровану мережу, яка сприяє підвищенню якості життя мешканців і стійкості середовища громади.



Рис. 1.1. Інфраструктура територіальної громади [393]

Соціально–економічні аспекти ТГ мають значний вплив на сучасне життя громади, пропонуючи нові можливості для покращення його якості, економічного розвитку та задоволення потреб громадян. Реалізація потенціалу цих технологій вимагає інтегрованого підходу та активної участі усіх зацікавлених сторін.

Концепція використання технологій в управлінні інфраструктурою ТГ охоплює різноманітні форми технологій та інновацій для покращення умов життя громади. Ця ідея ґрунтується на впровадженні технологічних рішень у різні аспекти управління громадою та щоденного життя її мешканців. Використання передових технологій, як–от IoT [292], ШІ [335], великі дані та інші цифрові інструменти, дозволяють оптимізувати функції ТГ.

У світі даних розуміння та ефективне використання різних типів масивів даних є ключовим для вирішення сучасних аналітичних завдань. Структуровані, неструктуровані та напівструктуровані масиви даних вносять свій унікальний вклад у спектр доступних інформаційних ресурсів (Рис. 1.2). Розуміння їх особливостей та потенціалу для конкретних застосувань дозволяє ефективно структурувати та аналізувати величезні обсяги даних, які щодня генеруються у цифровому світі.

Структуровані масиви даних відзначаються визначеною і консистентною організацією, що робить їх легко читаними та оброблюваними як машинами, так і людьми. Вони зазвичай використовують добре визначену схему або модель даних, яка їх структурує. Наприклад, для випадку масивів у NumPy, структуровані масиви можуть мати компонування з простіших типів даних, організованих як послідовність іменованих полів. Це дозволяє чітко визначити поля та їх типи в масиві, забезпечуючи легку маніпуляцію та відображення структурованих даних [383].

Неструктуровані масиви даних, навпаки, не мають попередньо визначеної організації або схеми. Це можуть бути тексти, відео, зображення та інші типи даних, які не вписуються в традиційні структури баз даних. Неструктуровані дані часто вимагають складніших методів для їх обробки та аналізу, таких як ШІ або обробка природної мови, для перетворення їх у структуровані дані.



Рис. 1.2. Класифікація масивів даних [383]

Напівструктуровані масиви даних займають проміжне місце між структурованими та неструктурованими даними. Вони можуть включати елементи обох типів, такі як документи XML [314, 452] або JSON [270], які містять як структуровані, так і неструктуровані дані.

Порівняльний аналіз цих трьох типів масивів даних підкреслює їх відмінності та специфічні застосування. Структуровані дані зазвичай використовуються для точних і швидких запитів, а також легко інтегруються між різними системами. Неструктуровані дані, хоча і складніші для обробки, надають багатший контекст і можуть бути корисними для більш глибокого аналізу. Напівструктуровані дані ж поєднують елементи обох попередніх типів, пропонуючи гнучкість у використанні та можливість адаптації під різні потреби [384].

Сучасні дослідження в області економічної інфраструктури та інтелектуального управління соціально-економічними для ТГ відкривають нові перспективи для розробки комплексних підходів та інноваційних рішень, спрямованих на підвищення ефективності, безпеки та стійкості систем громад.

Статистичний аналіз та МН є двома основними стовпами сучасного аналізу даних (Рис. 1.3), кожен з яких вносить свій унікальний вклад у розуміння та

інтерпретацію великих обсягів інформації. У той час, як статистичний аналіз забезпечує фундаментальні інструменти для опису, вивчення та висновків з даних, МН розширює ці можливості, використовуючи алгоритми для виявлення складних шаблонів та автоматичного вивчення з досвідом. Обидва підходи взаємодоповнюють один одного, пропонуючи потужні інструменти для аналітиків даних, які прагнуть вирішувати складні завдання, від прогнозування тенденцій до виявлення прихованих зв'язків у даних.



Рис. 1.3. Основні методи інтелектуального аналізу даних [93]

У наукових джерелах розглядаються різні аспекти інтелектуальних систем управління соціально–економічними процесами: зокрема, у статті [307] детально досліджено проблеми організації таких систем, визначено їхні рівні інтелектуального розвитку, запропоновано конкретні кроки для інтелектуалізації та продемонстровано ефективність застосування цих підходів на практиці. У публікації [414] фонд Hans–Böckler–Stiftung та його дослідницький підрозділ «Робота майбутнього» представляють оновлену модель системи підтримки прийняття рішень (DSS) Енцо Вебера, яка спрямована на вирішення суттєвих прогалин у соціальному забезпеченні

працівників платформної економіки. Крім того, у статті [400] аналізуються принципи розробки інтелектуальних ситуаційних систем підтримки прийняття рішень для інноваційного розвитку інфраструктури мегаполісів, підкреслюючи їхню важливість у сучасному міському управлінні. Зокрема, у дослідженні [430] пропонується нова концептуальна основа, яка розглядає системи ШІ місцевого самоврядування через призму відповідальних міських інновацій, визначаючи перспективні дослідницькі питання в цій сфері. У роботі [285] здійснений аналіз впливу управління на стійкість міст, підкреслюючи роль ІКТ у підвищенні ефективності політики, підзвітності та прозорості в міських системах.

Загалом, ці дослідження роблять значний вклад у розробку та впровадження передових інтелектуальних систем управління, що дозволяють громадам більш ефективно адаптуватися до швидких змін, оптимізувати управління ресурсами та покращувати якість життя громадян.

На основі викладеного, випливає що розробка інтелектуальної інформаційної технології аналізу соціально–економічних даних в контексті ТГ буде відігравати важливу роль у забезпеченні стабільності соціально–економічної інфраструктури. Через аналіз масивів даних, зібраних з різних джерел, включаючи опитування мешканців і статистичні показники, така система дозволяє глибоко зрозуміти потреби громадян, виявити тенденції та відслідковувати зміни у соціальній, екологічній, економічній і суспільній сферах. Використання інтелектуальних методів обробки даних сприяє ефективному управлінню інфраструктурою громади, підвищенню рівня комфорту життя, зменшенню екологічного впливу та підтримці соціально–політичної стабільності, особливо в умовах сучасних викликів, таких як військова агресія.

У сучасних реаліях, коли ТГ стикаються з безпрецедентними викликами, такими як пандемія, економічна нестабільність та військова агресія, виникає нагальна потреба в ефективних інструментах для підтримки стійкості та розвитку. Розробка інтелектуальної інформаційної технології аналізу соціально–економічних даних ТГ є актуальною, оскільки вона надає можливість оперативно реагувати на зміни, прогнозувати потенційні ризики та приймати обґрунтовані управлінські рішення. Це сприятиме підвищенню соціальної згуртованості, оптимізації використання ресурсів

та забезпеченню стійкого розвитку громад в умовах динамічних змін і невизначеності. Таким чином, дослідження відповідає нагальним потребам суспільства та має значний практичний потенціал для покращення якості життя населення.

Основною задачею розробленої інтелектуальної інформаційної системи забезпечення соціально–економічної інфраструктури ТГ є інтеграція кількісних та якісних показників в управління розвитком ТГ, що дозволяє ефективно розподіляти повноваження між державними і регіональними органами влади у процесі планування соціальних програм та генеральних планів розвитку. Цей комплекс задач (Рис. 1.4), має на меті створення умов для підвищення рівня якості життя населення та може слугувати основою для розробки подібних систем у майбутньому.

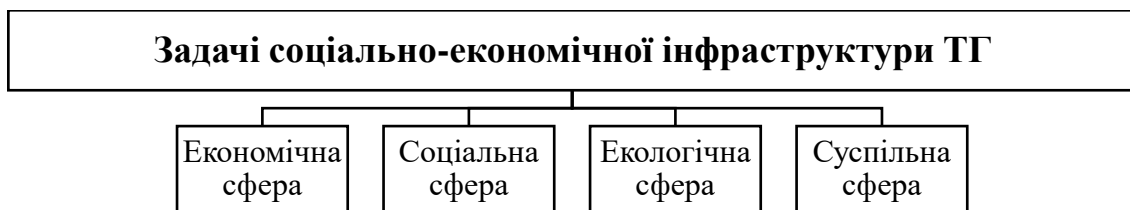


Рис. 1.4. Комплекс задач інтелектуального аналізу даних у соціально–економічній інфраструктурі ТГ [393]

Комплекс задач охоплює чотири основних рівні соціально–економічної інфраструктури міста: економічний, соціальний, екологічний та суспільний [393]. Кожен рівень визначається відповідними показниками та оцінками:

- 1) соціальний рівень включає ставлення населення до культурних аспектів ТГ, умов проживання та іншого;
- 2) екологічний рівень оцінюється через поточну екологічну ситуацію в ТГ;
- 3) економічний рівень зосереджується на стабільності економіки ТГ, матеріальному становищі мешканців та ринку праці;
- 4) суспільний рівень враховує оцінку мешканцями інформаційної безпеки та особистого соціального самопочуття.

З наведеного аналізу випливає, що існує протиріччя між наявністю великого обсягу соціально–економічних даних, які характеризують різні аспекти життєдіяльності ТГ, та обмеженими можливостями їх інтеграції, обробки та використання для прийняття управлінських рішень. Це протиріччя ускладнюється гетерогенністю, нерівномірністю і динамічністю таких даних, що вимагає застосування сучасних інструментів аналізу та моделювання.

Водночас, актуальність вирішення цього протиріччя зростає в умовах сучасних викликів, таких як пандемія, економічна нестабільність та військова агресія, які потребують оперативних, обґрунтованих та адаптивних рішень для забезпечення соціально–економічної стабільності та розвитку громад.

Подолання цього протиріччя можливе через розробку інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ, яка дозволить:

- інтегрувати кількісні та якісні показники для отримання цілісного уявлення про стан соціально–економічної інфраструктури;
- виявляти тенденції та прогнозувати ризики на основі даних, отриманих з різних джерел, таких як статистичні показники, опитування мешканців і відкриті дані;
- розробляти управлінські рішення, що сприяють підвищенню соціальної стійкості та якості життя населення.

Таким чином, обґрунтування необхідності створення такої технології, її структура та підходи до реалізації мають стати основою для вирішення виявленої проблеми в наступних підрозділах дослідження.

## **1.2 Огляд існуючих методів аналізу соціально–економічних даних для територіальних громад**

### **1.2.1 Аналіз інтелектуальних методів забезпечення соціальної стійкості**

В сучасних умовах динамічних соціально–економічних змін забезпечення соціальної стійкості ТГ стає пріоритетним завданням для органів місцевого



самоврядування та державної політики в цілому. Інтеграція інтелектуальних методів аналізу соціально–економічних даних відкриває нові можливості для глибокого розуміння потреб громад, прогнозування соціальних тенденцій та розробки ефективних управлінських стратегій.

У зв'язку із зазначеним, подальше дослідження буде присвячене детальному аналізу різних інтелектуальних методів, спрямованих на забезпечення соціальної стійкості. Зокрема, розглядатимуться інтелектуальні методи формування споживчого кошика та аналізу споживчої поведінки, інтелектуальні технології у туризмі та культурній спадщині, інтелектуальні методи управління екологічними ситуаціями, інтелектуальні методи інтерпретації юридичних текстів та виявлення дезінформації. Ці аспекти дозволять комплексно оцінити потенціал інтелектуальних технологій у різних сферах суспільного життя та їхню роль у зміцненні соціальної стійкості ТГ.

Дослідження [193] визначило інструменти, що впливають на формування споживчого кошика на різних рівнях і в різних сферах, включаючи національне законодавство, політику уряду та ініціативи місцевої влади. Досліджено їх взаємодію з різними типами коротких ланцюгів поставок у Францію, які відіграють важливу роль у місцевих харчових системах, таких як прямий маркетинг, магазини виробників, урбаністичне фермерство та постачання продуктів для громадського харчування. У дослідженні [388] вивчалася еволюція взаємозв'язків коливань і передачі між категоріями індексу споживчих цін, а також були створені зважені мережі причинно–наслідкових зв'язків Грейнджера (WGCN) для кожної країни групи G7. У роботі [358] індекс споживчих цін у Великій Британії аналізується як складна мережа, де застосовані методи кластеризації та оптимізації для дослідження динаміки мережі протягом часу. В статті [336] розглядається порівняння моделей споживання в десяти азійських країнах за допомогою трьох відомих системних моделей попиту: Роттердамської, CBS та III.

У дослідженні [138] було зроблено спробу прогнозувати сегмент середньоатлантичного ринку вина, враховуючи купівельну поведінку, установки споживачів та соціально–демографічні характеристики, для чого застосовано кластерний аналіз. В роботі [330] за допомогою апріорного алгоритму проаналізовано

ринковий кошик, щоб виявити закономірності в покупках на основі транзакційних даних. В дослідженні [77], використовуючи аналітику великих даних, було проаналізовано близько 44 тис. записів транзакцій в магазинах для 26 тис. клієнтів тайванської роздрібною мережі, щоби виявити зв'язок між особистісними рисами споживачів, країною походження та характеристиками брендів пива, а також прогнозування життєвої вартості клієнтів.

У дослідженні [32] аналізується ефективність різних методів кластеризації даних для ідентифікації прибуткових сегментів споживачів в секторі гостинності Великої Британії. В роботі [27] використано модель RFM (Recency, Frequency and Monetary) із застосуванням алгоритму K-Means для сегментації даних, причому результати такого аналізу порівнюються на основі параметрів, таких як частота, обсяг та свіжість продажів. У статті [188] представлено модель прогнозування, що базується на індивідуальній поведінці споживачів і методах аналізу даних. Модель використовує дані з бази даних супермаркету та додаткові дані з Amazon, які містять інформацію про покупки клієнтів.

У статті [61] проаналізовано вплив великих даних на роздрібну торгівлю, з розглядом п'яти ключових аспектів даних, таких як: клієнти, продукція, час, геопросторове розташування та канали продажу. У дослідженні [313] розглядається використання передових технологій у маркетингу та роздрібній торгівлі, пропонуючи нові підходи, які розвиваються завдяки технологічному прогресу, з особливою увагою до впливу інтелектуальних технологій на поведінку споживачів і інтеграцію маркетингу та роздрібною торгівлі з такими дисциплінами, як психологія, медіадослідження та соціологія, для більш складного та сучасного аналізу. У дослідженні [342] фокус робиться на споживачах, які беруть активну участь у створенні цінності, щоби концептуалізувати спільне створення інтелектуального досвіду та систему обслуговування.

У дослідженні [265] використаний інтуїтивний алгоритм нечіткої кластеризації для аналізу даних покупців супермаркету з огляду на витрати на певні категорії товарів. У роботі [295] розроблено модифікацію в алгоритм CDFCM для усунення виявлених недоліків, що привело до створення модифікованого динамічного

нечіткого с–середнього алгоритму. Дослідження [294] зосереджується на сегментації клієнтів за допомогою методу К–середніх в рамках неконтрольованого МН. Клієнти класифікуються з урахуванням демографічних, географічних, психографічних та поведінкових даних.

У контексті дослідження інтелектуальної інформаційної технології аналізу соціально–економічних даних ТГ, значна увага приділяється використанню технологій доповненої та віртуальної реальності (AR та VR) у сфері туризму та культурної спадщини, що відображено в ряді сучасних наукових робіт.

У дослідженні [153] аналізуються фактори, які впливають на туристичний досвід у контексті культурного туризму, з акцентом на вплив технологій VR та AR на навчальний досвід відвідувачів. У дослідженні [102] використовується інтегрована модель прийняття технологій, структура стимул–реакції та теорія потоку для дослідження впливу мобільних додатків AR на поведінку споживачів у туристичній сфері.

Стаття [354] зосереджена на аналізі, виявленні та обговоренні переваг і ключової ролі використання великих даних та ШІ у туристичній галузі. У роботі [234] проведено порівняльний аналіз традиційних туристичних інформаційних послуг і послуг, що використовуються в контексті інтелектуального туризму. У статті [420] розглядається підтримка прийняття рішень щодо планування туристичного маршруту за допомогою моделі нейронних мереж і буферного аналізу.

У статті [311] описано простий підхід до створення розмовного агента за допомогою передових інструментів, який передбачає перетворення типів та характеристик, пов'язаних з туризмом, на розмовні компоненти в прямому та узагальненому форматах. Дослідження SpaceBook [43], досліджує потенціал віртуального туристичного гіда, який працює без використання рук та очей, відповідаючи на запитання через інтерфейс розмовного діалогу та інформуючи користувача про цікаві особливості під час навігації туриста по різних напрямках. У роботі [165] розглядається розробка методів, які інтегрують різні аспекти відтворення аудіо, геолокації та інших сенсорних функцій смартфонів для створення захоплюючих геолокаційних історій у мобільному додатку.

Існує чимало досліджень, які зосереджуються на двох основних аспектах: застосуванні технологій AR та VR у сфері архітектури та культурної спадщини. Багато робіт досліджують можливості цих технологій, зокрема їхній вплив на сприйняття архітектурного простору [343], а також різноманітність застосувань у контексті культурної спадщини [16, 47]. Зокрема, в області AR проводяться дослідження, що включають розробку таксономії AR для мистецтва та культурної спадщини [87], створення систем візуалізації об'єктів культурної спадщини з використанням AR [154], використання фотограмметрії та AR для відтворення культурних об'єктів [324], розробку AR–систем для поглиблення знань про культурну спадщину [339], а також методики мультиспектрального 3D–запису та документування для мобільних додатків, пов'язаних з культурною спадщиною [67].

В роботі [222] акцентується увага на покращенні функцій AR через інтеграцію глибокого навчання, семантичного аналізу та графів знань. У роботах [125, 380] розглядається використання AR у поєднанні з МН для підтримки досвіду користувача, зокрема через інтелектуальні мобільні додатки. Робота [21] зосереджується на розпізнаванні об'єктів для культурної спадщини, а [148, 232] досліджують оптимізацію AR–додатків з використанням глибокого навчання.

У роботі [333] представлено прототип системи AR для культурних об'єктів, яка сприяє культурному туризму. Стаття [71] оглядає різні аспекти мобільної AR, включаючи інтерфейси користувача та техніки візуалізації. У роботах [99] та [197] обговорюється використання AR у промисловості та музейному секторі, відповідно. Дослідження [79, 114, 312] фокусуються на використанні AR для розширення сприйняття культурної спадщини, освіти та розваг, а [341] досліджує прийняття технології AR серед відвідувачів історичних місць.

Робота [15] присвячена розробці мультимодального фреймворка для інтелектуальних віртуальних агентів у змішаному середовищі дійсності. Цей фреймворк інтегрує різні технології, включаючи просторове відображення та розпізнавання мови. Стаття [148] звертає увагу на параметри продуктивності та оптимізацію AR–додатків з використанням технологій глибокого навчання.

У сфері управління туристичною областю активно використовуються системи підтримки прийняття рішень, засновані на інформаційних технологіях моделювання та аналізу туристичного попиту. Одні дослідники застосували метод, що базується на складеному пошуковому індексі та узагальнених динамічних факторах моделі для точного прогнозування туристичного попиту, але не врахували якісні фактори [233]. Інші дослідження включали використання трьох моделей сімейства ARMA, заснованих лише на аналізі туристичного потоку, що обмежує врахування інших важливих факторів [310]. Додаткові методи оцінювання туристичного попиту зосереджуються на використанні пошукових запитів у Google і Baidu, які дозволяють визначати тенденції уподобань користувачів туристичних продуктів [428]. Також існує підхід, що включає аналіз даних за допомогою теорії грубого набору, який дозволяє легше керувати категоричними змінними і виявляти зв'язки між поведінкою туристів та туристичними продуктами [74]. В цілому, для оцінки стану туристичної галузі використовуються різноманітні методи, як-то регресійний аналіз, гравітаційні моделі, SSA-аналіз, аналіз кривих попиту та когнітивне моделювання [472]. Окремі дослідження також зосереджені на аналізі прогресу в управлінні інформаційними технологіями у сфері туризму, зокрема такі, що оцінюють стратегічні напрями та еволюцію eTourism за останні десятиліття [296].

У сучасному світі, де швидкість та точність інформації є критичними, особливо під час надзвичайних екологічних ситуацій, використання інтелектуального аналізу даних стає ключовим фактором в ефективному управлінні такими подіями.

Ряд досліджень уособлюють собою розвиток методів керування надзвичайними ситуаціями за допомогою аналізу соціальних мереж, особливо Twitter. Ці дослідження фокусуються на різних аспектах: від стадій катастроф, що відображаються в соціальних мережах [297], до розробки систем для ефективного використання твітів у мобілізації ресурсів [51]. Також вони включають застосування мультиагентних систем для реального часу управління та аналізу екстрених ситуацій [352], порівняння алгоритмів МН для класифікації твітів [137] та розробку моделей для оптимізації розподілу ресурсів на основі інформації з мікроблогів [44, 282]. Ці дослідження підкреслюють значення соціальних медіа як інструменту для

оперативного реагування на надзвичайні ситуації. Дослідження [183] пропонує нову концептуальну основу IDSS для боротьби з катастрофами, з особливою увагою до лісових пожеж та хвиль холоду/спеки. IDSS використовує великі дані, зібрані з API і ШІ, щоб допомогти особам, які ухвалюють рішення, приймати їх більш швидко та точно.

Інші дослідження розширюють підхід до управління надзвичайними ситуаціями, з використанням глибоких технологій МН та аналізом даних. ChangeOS [447] пропонує структуру для виявлення семантичних змін, тоді як інші дослідження зосереджені на визначенні часу реагування екстрених служб [40] та розробці систем виявлення аварійних ситуацій [34]. Дослідження [192] демонструють використання гібридних підходів, що комбінують аналіз тексту з визначенням місцезнаходження і розширені моделі МН для більш точного виявлення постраждалих у надзвичайних ситуаціях. Вони підкреслюють потенціал інтеграції різних технологічних рішень для поліпшення управління надзвичайними ситуаціями.

Зростання кількості автомобілів у міських територіях приводить до підвищення рівня викидів шкідливих газів, що негативно впливає на екологічну ситуацію та якість повітря. Значна частина цих викидів пов'язана з пошуком вільних місць для паркування, що спричиняє додатковий трафік і затори. Тому дослідження, спрямовані на оптимізацію використання місць для паркувань та управління транспортними потоками, стають критично важливими для зменшення екологічного навантаження на міста. У роботі [237] наголошується на цій проблемі, а саме: багато досліджень у цій галузі включають розробку методів прогнозування кількості вільних місць для паркування. У роботі [204] також обговорюється ця проблема, а у [62] пропонується модель, що передбачає наявність місця для паркування в реальному часі на основі опитувань дорожнього руху. Роботи [17, 39] вносять вклад, пропонуючи альтернативні техніки виявлення місць для паркування: перша використовує дані про сусідні транспортні засоби та відомі фіксовані інфраструктури, а друга базується на техніці обробки зображень. У роботах [10, 69, 111, 181, 365] розширена ця тематика, де пропонуються різноманітні підходи та моделі, включаючи глибоке навчання, періодичні нейронні мережі, регресійне моделювання та гібридні згорнуті моделі для

прогнозування заповнення місць для паркувань. Водночас, дослідження [293, 301, 381, 382] зосереджуються на аналізі даних міського вантажного транспорту, попиту на паркування, та різних стратегіях прогнозування, що має велике значення для міського планування. Крім цього, інтелектуальне виявлення та підрахунок автомобілів набуває все більшої важливості у контексті управління дорожнім рухом. У роботі [378] представлена система виявлення та підрахунку транспортних засобів, що базується на комп'ютерному зорі. У роботах [390, 422, 427, 449] розширено цей напрям, використані різні підходи, такі як інтеграція комп'ютерного бачення та моделювання дорожнього руху, застосування глибоких нейронних мереж та ефективні методи підрахунку транспортних засобів. Окрім того, сучасні компанії, які мають сервери, активно розробляють платформи для автоматизації процесу МН, такі як: AWS SageMaker [182], Azure МН Studio [113], IBM Watson Studio [73], Google Cloud AutoML [53], Oracle AutoML Pipeline [426], Dataiku [93], DataRobot [94].

Сучасні дослідження в області ШІ активно розвиваються, зокрема у сфері юридичної інформатики, де ШІ використовується для аналізу, пошуку та інтерпретації юридичних текстів. У ряді досліджень [22, 210, 355] особлива увага приділяється розробці та використанню інтелектуальних систем для аналізу та інтерпретації юридичних документів. Ці дослідження зосереджені на технологіях обробки природної мови, МН та екстракції знань, що дозволяє підвищити точність та ефективність юридичних досліджень. У статті [224] підкреслюється значення інтелектуальних систем пошуку в праві для спрощення юридичних досліджень, а у дослідженні [303] обговорено використання інформаційних систем на базі ШІ у судових процесах.

У дослідженні [136] основна увага приділяється систематизації знань про новітні практики інтелектуального пошуку інформації в контекстах, аналогічних до сфери податкового права. Дослідження [126] висвітлює, як ШІ та МН змінюють роботу юридичних професіоналів та юридичних фірм, а у дослідженні [366] розглядається трансформація традиційного правознавства в розумну юридичну систему за допомогою автоматизації. В статті [123] розглядається застосування ШІ урядами та комерційними організаціями для автоматизованого прийняття рішень,

тоді як дослідження [132, 159] акцентують на етичних та юридичних викликах використання ШІ. Ці роботи підкреслюють важливість комплексного підходу до інтерпретації юридичних документів, який би включав глибокий аналіз семантичних та контекстуальних зв'язків, а також адаптацію до специфічних потреб користувача, що є критичним для розуміння складної юридичної інформації та ефективного вирішення юридичних задач.

ШІ та глибоке навчання стали ключовими у виявленні дезінформації. Стаття [356] висвітлює тематичний аналіз застосування ШІ для автоматичного виявлення дезінформації, а роботи [441] демонструють використання МН та глибокого навчання для створення моделей виявлення фейкових новин, зокрема BERT-Based Domain Adaptation Neural Network. У статтях [24, 36] представлено систему EXMULF та використання двонаправленої LSTM-рекурентної нейронної мережі для цих цілей. Статті [9, 135] описують різноманітні підходи та методи, від класифікаторів мовних ознак до автоматизованого виявлення дезінформації на різних платформах, підкреслюючи значення ШІ та МН у цьому напрямку. У статті [318] розглянуто стилістичні, семантичні, орфографічні та морфологічні особливості для побудови класифікаторів фейкових новин, однак виявлено, що для точного виявлення дезінформації необхідна перевірка фактів. Стаття [121] описує систему DISCO, яка використовує графові методи МН для обробки користувацьких запитів і виявлення дезінформації. У дослідженні [18] фокус робиться на виявленні дезінформації, пов'язаної з COVID-19, використовуючи МН, де моделі BERT показали кращі результати.

У підсумку аналізу сучасних досліджень, присвячених інтелектуальним методам забезпечення соціальної стійкості ТГ, можна відзначити значний прогрес у розробці та впровадженні різноманітних технологічних рішень. Застосування методів МН (Таблиця 1.1), глибокого навчання, аналізу великих даних та інтелектуальних систем підтримки прийняття рішень дозволяє більш точно прогнозувати соціальні та економічні тенденції, ефективно сегментувати споживчі ринки, оптимізувати управління транспортними потоками та ресурсами, а також забезпечувати високий рівень безпеки та екологічної стабільності. Ці технології не лише підвищують



ефективність управлінських процесів, але й сприяють підвищенню якості життя мешканців ТГ через більш гнучке та адаптивне реагування на виклики сучасності.

Таблиця 1.1

Ключові результати аналізу інтелектуальних методів  
забезпечення соціальної стійкості для ТГ

Дослідження	Методи та технології	Основні напрями	Ключові результати
[27, 32, 77, 138, 188, 193, 330, 336, 358, 388]	Кластеризація (K–Means, RFM), регресійне моделювання, глибоке навчання	Формування споживчого кошика, прогнозування споживчої поведінки	Ідентифікація прибуткових сегментів споживачів, прогнозування життєвої вартості клієнтів, оптимізація продажів
[102, 153, 183, 354, 378, 390, 422, 427, 449]	AR, VR, великі дані, ШІ, комп'ютерний зір	Туризм, управління екстремими ситуаціями, транспорт, паркування	Покращення туристичного досвіду, швидке реагування на катастрофи, зниження рівня викидів, оптимізація транспортних потоків
[9, 18, 24, 25, 36, 109, 121, 135, 236, 318, 356, 438, 441]	Глибоке навчання, комп'ютерний зір, МН, графові методи	Управління відходами, виявлення дезінформації	Підвищення точності сортування відходів, ефективне виявлення фейкових новин, підтримка прозорості у громадських процесах

Однак, попри досягнутий прогрес, залишається необхідність у подальшій розробці та адаптації інтелектуальних методів для вирішення специфічних задач, що виникають у різних аспектах соціально–економічного управління. Наприклад, у сфері управління транспортом та паркуванням важливо враховувати динамічні зміни у потоках автомобілів та їхній вплив на екологічну ситуацію. У сфері туризму та культурної спадщини необхідно розробляти системи, які б ефективно інтегрували технології AR та VR для покращення досвіду відвідувачів та підтримки культурних об'єктів. Аналогічно, у сфері юридичної інформатики та виявлення дезінформації важливо створювати методи, що забезпечуватимуть високу точність та надійність

аналізу складних текстових даних. Таким чином, розвиток інтелектуальних методів повинен бути орієнтований на конкретні потреби та особливості кожної сфери, що буде забезпечувати їхню максимальну ефективність і практичну значимість.

Перспективи подальших досліджень у цій галузі полягають у глибшій інтеграції різних технологічних підходів та створенні гібридних моделей, які б комбінували найкращі характеристики існуючих методів. Важливим напрямом є також розробка адаптивних систем, здатних швидко реагувати на зміни у соціально–економічному середовищі та забезпечувати стійкість ТГ перед новими викликами. Загалом, подальший розвиток інтелектуальних методів забезпечення соціальної стійкості ТГ є критично важливим для досягнення сталого розвитку та покращення якості життя в умовах швидкозмінного світу, особливо у період війни.

### **1.2.2 Аналіз інтелектуальних методів управління підприємницькою діяльністю**

У сучасних умовах динамічних ринкових змін та технологічного прогресу інтелектуальні методи управління стають ключовими інструментами для підприємств, що прагнуть забезпечити свою конкурентоспроможність та стійкість. Цей підрозділ присвячений аналізу новітніх підходів у розробці та впровадженні бізнес–моделей, що інтегрують передові технології, такі як МН, ШІ та доповнена реальність. Особлива увага приділяється застосуванню чат–ботів у різних сферах діяльності, оптимізації ланцюгів постачань, управлінню ризиками, а також використанню методів МН для аналізу та прогнозування економічних процесів. Окрім цього, розглядаються інноваційні підходи до забезпечення безпеки підприємств та боротьби з економічною злочинністю за допомогою сучасних інформаційних технологій. Таким чином, дослідження інтелектуальних методів управління підприємницькою діяльністю демонструє широкий спектр можливостей для підвищення ефективності та адаптивності бізнесу в умовах швидких змін.

Хоча бізнес–моделі принципово пов'язані з технологічними інноваціями, сама конструкція бізнес–моделі по суті відокремлена від технологій. Бізнес–модель можна

визначити як систему, яка вирішує проблему ідентифікації клієнтів, в процесі взаємодії з їхніми потребами та задоволеннями з урахуванням монетизації цінності [35]. Тому ряд дослідників, зокрема автори [97, 321], запропонували концепцію бізнес–моделі у рамках традиційної стратегії конкурентної переваги. Разом з тим, інноваційні технології самі по собі впливатимуть на можливості бізнес–моделей [42]. Класифікація бізнес–моделей із застосуванням теоретико–шаблонного підходу розглянута в роботах [55, 277, 337]. У [227] розроблено ресурсно–функціональну модель безпеки (яка складається з часткових показників та компонентів економічної безпеки бізнесу) та також застосовано ресурсно–функціональний підхід до розрахунків. У роботі [164] проведена кількісна оцінка рівня фінансової платоспроможності країн на основі використання багатовимірного методичного інструментарію оцінки фінансових показників розвитку країни, що приводить до побудови відповідних інтегральних індексів безпеки.

Згідно з останніми дослідженнями, сучасні технології продовжують змінювати сферу управління підприємством. Зокрема, у [287] розроблено система, яка виявляє та відстежує обличчя на мобільних пристроях, інтегрується з хмарними сервісами для розпізнавання осіб з особистого контактного списку користувача. Дослідження [285] показало, що користувачі легко сприймають AR, яка має позитивний когнітивний ефект. Інші дослідження [286] розглядають використання адаптивних систем та інструментів AR для підтримки навчання вчителів, а [360] досліджують вплив AR на командну роботу, особливо щодо мотивації та емоційної активації.

У контексті міського планування, запропоновано рамкову структуру для підтримки розумних маневрів міста [320], а також таксономічне рішення для оцінки готовності технологій до Індустрії 4.0 [334]. В індустріальному секторі розроблений додаток AR для монтажу котушок для труб [220], а також запропонована методологія застосування AR у реальних промислових дослідженнях [274]. У соціальному підприємстві, розглянуто міжнародне соціальне підприємництво з використанням інноваційних технологій [178], а у роботі [359] описано використання екологічних віртуальних підприємств. AnyLogic [28, 252] дозволяє аналізувати трафік продукції,

а також розробляти гібридні моделі для оптимізації основних параметрів, включаючи агентне та багатоагентне моделювання [437].

У сфері віртуальних підприємств та інформаційних систем роботи [112, 117, 348] представляють концептуальні моделі віртуальних підприємств та інноваційні інфраструктури інформаційних систем. Особливу увагу приділено вибору партнерів у роботі [103], де цей процес сформульовано як гібридна багатокритеріальна проблема прийняття рішень з використанням нечітких чисел. Дослідження [288] описує алгоритм групового прийняття рішень для вирішення неповних і непослідовних даних.

Окремий напрям досліджень присвячений використанню чат-ботів у HR та рекрутингу. У роботі [394] розглядаються технології у HR, [54] – обґрунтовується використання ШІ в рекрутингу, [106] – розглядаються можливості чат-ботів у HR. Автоматизація вимог і класифікації з використанням інтелектуальних розмовних чатів згадується у роботі [332]. Глибоке навчання й адаптація самонавчання для відбору HR-кандидатів описано в [370], а в [448] – розробка Microsoft XiaoIce. У роботі [89] представлено чат-сервіс SuperAgent; [306] – наведено стилі відповідей існуючих чат-систем. У [305] зроблено аналіз публікацій, пов'язаних з чат-ботами за останні 5 років. У [424] будується нова модель для оптимізації розподілу ресурсів, а у [329] пропонується метод відбору функцій для вибору HR-ресурсів. У роботах [273, 276, 299] автори зосереджені на аналітичних параметрах та моделях розвитку людських ресурсів. У [84, 287, 298] розглядаються чат-боти як консультанти у педіатрії, навчанні та вивченні мов. У [155] пропонується чат-бот, що навчається на основі діалогу. Аналіз взаємодії людина-людина для розробки людино-чат-бот взаємодій досліджується у [81, 369]. В роботах [49, 86, 377] описується застосування чат-ботів у різних сферах, включаючи освіту, туризм, продаж цифрових послуг. У [314] зроблений аналіз прогресу в галузі ШІ, зосередившись на чат-ботах у роздрібній торгівлі. У [86] наведено методи тестування чат-ботів.

У роботах [66, 177, 405] акцент робиться на стратегіях управління ризиками, зокрема, в контексті IT-проектів та інноваційних і фінансових ризиків віртуальних підприємств. У статтях [279] розглядається управління ризиками в рамках проектного

управління. Значна частина досліджень [158, 213, 283, 392] присвячена використанню алгоритмів МН, включаючи стохастичний градієнтний спуск, методи ансамблю, векторні машини підтримки, нейронні мережі та інші, для різноманітних застосувань, від штучного відбору облігацій до класифікації фінансових великих даних.

У дослідженні [280] розглядається використання LSTM–мереж для аналізу попиту в управлінні ланцюгами поставок, що вимагає значних обчислювальних ресурсів. Також дослідження [271] зосереджуються на прогнозуванні попиту у конкретних сферах, як–от таксі та сільське господарство, але їхні методи обмежені відсутністю широкого застосування. Дослідження [7, 14], які фокусуються на спеціалізованих моделях для прогнозування попиту на харчові добавки під час пандемії та в сфері водно–енергетичного зв'язку, також мають обмеження через свою специфічність. Дослідження [434] розглядає застосування МН для прогнозування кількості пацієнтів в екстрених медичних установах, але його вузька спеціалізація обмежує застосування.

Дослідження [371] розробляє інтелектуальний підхід для вибору фінансування ланцюга постачань, зосереджуючись на сегментації постачальників за ризиком платежів, водночас маючи обмежений фокус на кризові ситуації. Також, дослідження [23] пропонують стійкі та сталі моделі вибору постачальників, корисні для довгострокових цілей, але не включають невідкладні потреби, що виникають під час криз. В роботах [151, 411] розглядаються методи оптимізації стійкого ланцюга постачань та вибору постачальників, акцентуючи увагу на екологічних та стійких практиках, але не повністю враховуючи специфіку кризових викликів.

Інші дослідження, такі як [198, 418], зосереджуються на інтеграції SCOR 4.0 та МН для вибору постачальників, багатоцільовому рамковому виборі постачальників та оцінці ризиків у сталих ланцюгах постачань, вносячи важливий внесок у розвиток методів вибору постачальників, хоча і не можуть бути безпосередньо адаптовані до конкретних кризових умов. Особливо важливим у цьому напрямі дослідження може бути робота [60], у якій розглядається стійкий вибір постачальників під час пандемії COVID–19, безпосередньо релевантний для умов кризи, з акцентом уваги на специфічних викликах пандемії.

В сучасному світі технології відіграють значну роль у дослідженні економічних злочинів, особливо в контексті бізнесу та впливу злочинності на економіку. Наприклад, у роботі [345] розглядається використання МН та супутникових зображень для виявлення нелегального видобутку в Колумбії, що демонструє як технологічні засоби можуть використовуватися для моніторингу та аналізу незаконної економічної діяльності. Це підкреслює потенціал МН в розкритті та протидії економічним злочинам. У дослідженні [308] обговорюються можливості у віртуальному середовищі через ІКТ для запобігання фінансовій злочинності, зокрема застосування програмного забезпечення, такого як «FinTrack». Розпізнавання облич, як зазначено в дослідженні [213], відіграє центральну роль у багатьох програмах безпеки, включаючи встановлення системи соціальних кредитів для контролю населення.

Є ряд досліджень, що використовують методи МН для класифікації та аналізу економічних злочинів. У роботі [11] представлений огляд однокласних класифікаторів підтримки. Дослідження [158] розглядає використання алгоритмів стохастичного градієнтного спуску для застосувань великих даних, наприклад, для прискорення SVM або регульованої регресії у великих масштабах.

Також є роботи, що демонструють застосування різних алгоритмічних і статистичних методів у прогнозуванні та аналізі даних для визначення злочинів. У роботі [38] порівнюються методи ансамблю з моделями одиночних класифікаторів; у [386] – розглядаються стратегії багатоступеневої класифікації та класифікатори дерева рішень. Дослідження [350] використовує МН для класифікації злочинів, основний акцент робиться на аналізі різних типів злочинів, включаючи їхню інтенсивність, місцезнаходження та частоту.

Дослідження [415] зосереджено на розробці таксономії для розрізнення бізнес-моделей, що використовують МН як основу, особливо в сегменті B2B; дослідження [48] – на використанні МН для прогнозування успіху стартапів, визначаючи його як процес злиття та поглинання або первинного публічного розміщення акцій. Воно використовує великі дані з найбільшої в світі структурованої бази даних для стартапів – CrunchBase.com, для побудови прогностичної моделі. У дослідженні [290]

проведений огляд існуючих технік МН, які нещодавно зробили свій внесок у розуміння потреб стартапів, тенденцій бізнесу та можуть надати рекомендації щодо планування майбутніх стратегій для вирішення бізнес-проблем.

У сфері аналізу споживацьких відгуків в інтернет-магазинах значну увагу привертають дослідження, присвячені вивченню впливу онлайн-відгуків на продажі товарів і послуг. Наприклад, дослідження продуктів Amazon [194, 284], технічних характеристик із соціальних мереж [289] та брендів Flipkart, Snapdeal і Amazon India [75] виявили значний вплив відгуків на споживацьку поведінку та продажі. Ці дослідження підкреслюють важливість детального аналізу відгуків, який охоплює не тільки зміст, але й емоційне забарвлення та корисність інформації. Розвиток алгоритмів вебсканування, таких як Mison [185] та використання бібліотек BeautifulSoup, lxml та регулярних виразів [421], відіграє ключову роль у процесі збору та аналізу великих обсягів даних з відгуків.

Другий важливий аспект цих досліджень полягає у розробці та застосуванні інноваційних методів для аналізу емоційного забарвлення та визначення корисності відгуків. Використання глибокого навчання та згорткових нейронних мереж [199, 419] дозволяє точно ідентифікувати емоційний стан відгуків та їхній вплив на сприйняття продуктів. Такі системи можуть враховувати не тільки традиційні параметри якості та ціни, але й відображати більш глибоке розуміння споживацьких настроїв і переваг, що є ключовим для розробки ефективних маркетингових стратегій та управління продуктами.

Останнім часом велике значення у сфері електронної комерції набувають нейромережеві методи, спрямовані на оптимізацію та автоматизацію формування каталогів товарів для інтернет-магазинів. Робота [142] систематизує дослідження моди в одязі, включаючи її визнання на різних рівнях. У роботі [387] аналізується узгодженість модного стилю і реалізація персоналізованих рекомендацій щодо одягу. У роботі [439] пропонується новий метод прогнозування атрибутів одягу з використанням немаркованих даних. Всі ці дослідження роблять акценти на застосуванні глибоких нейронних мереж для аналізу і класифікації зображень одягу.

З іншого боку, дослідження фокусуються на вдосконаленні технологій візуального розпізнавання та класифікації одягу за допомогою алгоритмів глибокого навчання та генеративних змагальних мереж (GAN). У статті [266] досліджуються правила відповідності одягу за допомогою моделі GAN, а у роботі [267] розглядається атрибут-GAN з двома дискримінаторами. Робота [167] фокусується на класифікації зображень з мінімальним наглядом, а робота [104] розвиває основу глибокого навчання для передачі моделей з контрольованих зображень на вуличні зображення. У дослідженні [440] представлений новий підхід для класифікації жанрів одягу та використання орієнтирів для підвищення точності класифікації.

У ряді наукових робіт розглядаються різноманітні підходи до обробки даних та їх застосування у рекламній сфері. Робота [128] зосереджена на використанні онтологічних систем і методів обробки природними мовами для створення ефективної реклами у соціальних мережах. Водночас, у роботі [423] запропонований підхід до рекомендацій подій, що включає семантичний аналіз контенту та контекстуальний вплив. У роботах [186, 316] автори зосереджені на перетворенні нечитабельних даних у структурований формат за допомогою класифікації MN, а також аналізі настроїв із застосуванням підходу LDA. У роботі [200] оцінюється ефективність трьох методів моделювання тем: прихованого семантичного аналізу, прихованого розподілу Діріхле і корелюючого моделювання тем.

Автори [30] фокусуються на використанні інструментів MN для аналізу даних соціальних медіа, а в [404] реалізовано аналіз настроїв за допомогою алгоритмів SVM і NB. Дослідження [41, 353] відображають застосування класифікації, сегментування та кластеризації для інтернет-реклами, а також розглядають багатокритеріальну оптимізацію для вибору медіа-ресурсів і бюджетування. В роботі [141] пропонується підхід для витягування знань, що підтримують маркетингові рішення, а [266] – досліджує розвиток рекламного дизайну на основі інформаційних технологій. Автори [368, 407] застосовують алгоритм Аргіогі для виявлення впливових блогерів та взаємозв'язків у різних контекстах, тоді як у роботі [20] запропонована система рекомендацій з використанням цього алгоритму.



У дослідженні [133] обговорюються способи покращення географічного націлювання через інтеграцію різноманітних стратегій націлення, таких як поведінковий таргетинг, часове таргетування та знижки, з метою створення комплексної стратегії націлення у контексті онлайн– та офлайн–комерції. У роботах [228, 272] акцентовано на недостатності врахування бізнес–цілей у загальній практиці розробки емпіричних моделей таргетингу і підкреслено значення вибору вигідніших цільових груп. У статті [409] запропоновано двоступеневий метод на основі фільтра Гаусса та дерева рішень для видалення неефективної онлайн–реклами і забезпечення високої точності прогнозування ефективної реклами. Результати дослідження [172] демонструють значення вибору місця розміщення реклами у Facebook (бічна панель чи стрічка повідомлень) з урахуванням ступеня залучення продукту для ефективного націлювання.

Існує ряд робіт, в яких досліджується використання ШІ в генерації рекламного контенту [362]. Деякі дослідження фокусуються на використанні технологій ШІ, таких як GANs, для автоматичного створення рекламного контенту [70]. Робота [146] представляє систему, що генерує рекламні плакати. Останнє дослідження [446] пропонує глибоку генеративну модель для графічного дизайну макетів.

У статті [134] проведено аналіз підходів для боротьби з шахрайськими діями в мережі Інтернет. У роботі [397] узагальнено представлено моделювання даних про шахрайство в часі у формі виявлення аномалії на основі ймовірності або навчання та моделювання часових рядів з часовими шкалами, такими як хвилини та години. У роботі [442] запропонована ефективна модель класифікації шахрайства на основі нейронної мережі зворотного поширення, що поєднується з новим алгоритмом «Штучна колонія бджіл». У роботі [375] запропоновано класифікатор  $k$ –найближчих сусідів на основі вибору прототипу Quad Division (QDPSKNN) шляхом запровадження методу ділення чотирьох частин для обробки нерівномірного розподілу класів. Ефективність QDPSKNN оцінюється на основі виявлення шахрайства в мобільній рекламі.

Отже, в сучасних умовах динамічних ринкових змін та технологічного прогресу інтелектуальні методи управління підприємницькою діяльністю набувають все більшої значущості (Таблиця 1.2).

Таблиця 1.2

Ключові результати аналізу інтелектуальних методів управління підприємницькою діяльністю для ТГ

Дослідження	Методи та технології	Основні напрями	Ключові результати
[55, 97, 103, 112, 117, 164, 227, 288, 321, 337, 348, 363]	Теоретико–шаблонний підхід, багатовимірний аналіз	Бізнес–моделі, фінансова платоспроможність, віртуальні підприємства	Розробка концепцій бізнес–моделей, оцінка фінансової стабільності країн, створення інтегральних індексів безпеки
[178, 220, 274, 286, 320, 334, 359, 360, 371, 374, 437]	AR, VR, IoT, автоматизація, глибоке навчання	Управління підприємницькою діяльністю, міське планування, соціальне підприємництво	Оптимізація управлінських процесів, покращення навчання та командної роботи, підтримка індустріальних і соціальних підприємств
[49, 54, 81, 84, 86, 88, 106, 155, 273, 276, 287, 298, 299, 305, 314, 329, 332, 369, 370, 377, 394, 424, 448]	Чат–боти, МН, NLP, глибоке навчання	HR, рекрутинг, освіта, туризм, роздрібна торгівля	Автоматизація рекрутингу, покращення взаємодії з користувачами, підвищення ефективності HR–процесів, підтримка освітніх і туристичних платформ
[48, 75, 185, 194, 284, 290, 289, 415, 421]	Аналіз великих даних, вебсканування, NLP	Аналіз споживацьких відгуків, їхній вплив на продажі	Виявлення значущого впливу відгуків на споживацьку поведінку, розробка ефективних алгоритмів збору та аналізу даних відгуків
[142, 199, 387, 419, 439]	Глибоке навчання, CNN, GAN	Класифікація та прогнозування модних тенденцій	Підвищення точності класифікації зображень, розробка моделей для прогнозування атрибутів одягу, оптимізація рекомендаційних систем

## Продовження таблиці 1.2

Дослідження	Методи та технології	Основні напрями	Ключові результати
[128, 186, 200, 316, 423]	Обробка природної мови, LDA, класифікація МН	Рекламні стратегії, аналіз контенту	Створення ефективних рекламних кампаній, семантичний аналіз контенту для рекомендацій, покращення таргетингу реклами
[133, 172, 228, 272, 409]	Поведінковий таргетинг, регресійне моделювання, SVM	Географічне націлювання, оптимізація реклами	Покращення географічного націлювання реклами, висока точність прогнозування ефективної реклами, оптимізація розподілу рекламних ресурсів

Розробка та впровадження інноваційних бізнес-моделей, що інтегрують передові технології, такі як МН, ШІ та AR, сприяє підвищенню конкурентоспроможності підприємств та оптимізації процесів взаємодії з клієнтами. Особлива увага приділяється використанню чат-ботів у різних сферах діяльності, що дозволяє автоматизувати процеси рекрутингу, покращити взаємодію з користувачами та підвищити ефективність управління людськими ресурсами.

Крім того, інтеграція інтелектуальних технологій у міське управління та планування, а також у сферу управління ризиками та боротьбу з економічною злочинністю, демонструє значний потенціал для покращення управлінських процесів і забезпечення безпеки підприємств. Застосування методів МН для аналізу та прогнозування економічних процесів, оптимізації ланцюгів постачань, а також виявлення та протидії економічним злочинам, сприяє підвищенню ефективності і стійкості підприємств у складних умовах ринку.

Незважаючи на досягнутий прогрес, залишається необхідність у подальшій розробці та адаптації інтелектуальних методів для вирішення специфічних задач у різних аспектах управління підприємництвом. Це включає створення адаптивних і гібридних моделей, здатних враховувати специфічні потреби та особливості кожного підприємства, а також забезпечення етичності та безпеки використання

інтелектуальних систем. Таким чином, розвиток інтелектуальних методів управління підприємницькою діяльністю є критично важливим для досягнення сталого розвитку і підвищення конкурентоспроможності підприємств в сучасному світі.

Однак для досягнення максимального ефекту від впровадження інтелектуальних методів необхідно продовжувати дослідження та розробку адаптивних моделей, здатних враховувати змінні ринкові умови та специфічні потреби підприємств. Крім того, важливо приділяти увагу інтеграції різних джерел даних та забезпеченню етичності і безпеки використання інтелектуальних систем у маркетингових процесах. Таким чином, подальший розвиток інтелектуальних методів просування товарів на інтернет–ринку є ключовим фактором для підвищення конкурентоспроможності підприємств та забезпечення їхньої стійкості у динамічному цифровому середовищі.

На основі проведеного аналізу виявлено протиріччя між великим потенціалом сучасних інтелектуальних методів аналізу даних та їх недостатнім впровадженням для вирішення комплексних соціально–економічних завдань ТГ. Хоча існуючі дослідження демонструють успіхи у застосуванні методів машинного і глибокого навчання, аналізу великих даних і систем підтримки прийняття рішень, їхня інтеграція в цілісні рішення для соціально–економічного управління є фрагментарною і не завжди відповідає динамічним умовам сучасного середовища.

Основні протиріччя:

–розрізненість методів і підходів: існуючі системи не забезпечують інтеграцію всіх аспектів соціально–економічного управління, включаючи аналіз ресурсів, прогнозування ризиків, оцінку соціального впливу та прийняття обґрунтованих рішень;

–динамічність і гетерогенність даних: соціально–економічні дані ТГ характеризуються нерівномірністю, різнотипністю і швидкими змінами, що ускладнює їх ефективне використання;

–недостатня адаптивність існуючих рішень: наявні системи не завжди можуть швидко адаптуватися до зовнішніх змін чи кризових ситуацій, що знижує їхню ефективність у довгостроковій перспективі.

Для вирішення зазначених проблем пропонується створення інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ, яка повинна:

- інтегрувати дані з різних джерел для забезпечення комплексного аналізу стану інфраструктури, соціальної динаміки та економічних показників;
- розробити адаптивні моделі на основі методів машинного навчання та обробки даних для оперативного реагування на зміни.
- забезпечити модульну архітектуру системи, що включатиме методи класифікації та прогнозування.

Запропонована технологія дозволить подолати виявлені протиріччя, підвищити ефективність управління соціально–економічними процесами, забезпечити стійкість ТГ до сучасних викликів і покращити якість життя населення. У наступних підрозділах дослідження буде детально розглянуто архітектуру запропонованої системи та підходи до її реалізації.

### **1.3 Огляд методів аналізу даних, які можуть бути застосовані для інфраструктури ТГ**

Прогнозування нестационарних соціально–економічних даних є складним завданням через змінність їх статистичних характеристик у часі [152, 140]. Традиційні статистичні методи, такі як метод ковзного середнього, експоненційного згладжування, метод Холта–Вінтерса або ARIMA, часто базуються на припущенні стаціонарності даних або вимагають попереднього перетворення рядів до стаціонарного вигляду [58, 110]. Це може призводити до втрати важливої інформації та зниження точності прогнозів. Крім того, ці методи можуть бути недостатньо гнучкими для виявлення складних нелінійних залежностей і адаптації до структурних змін у даних. Через це виникає потреба у застосуванні більш адаптивних і потужних інструментів, які здатні ефективно обробляти нестационарні дані та враховувати їх динамічну природу. Методи машинного навчання, зокрема бустингові алгоритми [212], пропонують такі можливості. Вони не вимагають припущень про розподіл

даних і можуть моделювати складні нелінійні взаємозв'язки, що робить їх ефективними для прогнозування в умовах нестационарності.

Бустингові алгоритми [463], такі як XGBoost, LightGBM, CatBoost та ін., дозволяють комбінувати результати кількох слабких моделей для отримання більш точної та стійкої моделі. Це особливо корисно при роботі з соціально–економічними даними, де може спостерігатися висока варіативність і наявність шумів. Завдяки своїй адаптивності та універсальності, бустингові методи стають важливим інструментом для аналітиків та управлінців, які працюють з нестационарними даними.

Подальше дослідження буде присвячене детальному аналізу бустингових методів машинного навчання для прогнозування нестационарних даних, їх перевагам та можливостям застосування в управлінні соціально–економічними процесами ТГ.

У статті [116] наведено загальний огляд методів бустингу в МН, з фокусом на використанні слабких і сильних моделей. У дослідженні [391] розглянуто застосування методів бустингу для багатокласової класифікації на незбалансованих даних, зокрема в мультикласовому контексті. Вони порівнюють різні методи бустингу і визначають, які з них є найкращими для обробки таких даних. У статті [65] представлено загальний огляд бустингу з фокусом на методах AdaBoost та L2Boost для класифікації і регресії. Ця робота надає практичні поради щодо вибору слабких моделей і функцій втрат. У роботі [326] зроблений огляд алгоритмів МН з фокусом на застосуванні бустингу для підвищення точності передбачення і класифікації. Основний акцент робиться на алгоритмах дерев рішень та методах опорних векторів. У статті [231] розглядається проблема класифікації незбалансованих даних і застосування методів бустингу для її вирішення. Автори подають огляд різних методів бустингу для цієї задачі. У роботі [124] вивчається використання ансамблевих методів для вирішення проблеми незбалансованих класів в класифікації. Автори здійснюють порівняльний аналіз різних підходів і визначають найбільш ефективні.

Бустингові алгоритми МН є хорошим вибором для прогнозування нестационарних часових рядів, оскільки вони мають ряд переваг перед іншими, такими як методи: ковзного середнього, експоненційного згладжування, Холта–Вінтера, ARIMA та STL.

Однією з основних переваг бустингу є те, що він може компенсувати обмеження слабких моделей. Слабкі моделі, такі як дерева рішень, можуть бути ефективними для прогнозування статистичних часових рядів, однак вони можуть бути менш точними для нестационарних часових рядів. Бустинг дозволяє компенсувати це, навчаючи послідовність слабких моделей, а потім поєднуючи їх результати, щоб отримати сильну модель.

Іншою перевагою бустингу є те, що він може бути адаптивним. Бустингові алгоритми можуть навчатися на помилках попередніх моделей, щоб покращити точність прогнозів. Це може бути корисно для нестационарних часових рядів, оскільки вони можуть швидко змінюватися з часом.

Також бустингові алгоритми є досить універсальними. Вони можуть використовуватися для прогнозування різного роду часових рядів, включаючи часові ряди з трендами, сезонними коливаннями та циклічністю.

Отже, бустингові методи мають ряд переваг перед звичайними методами для прогнозування нестационарних часових рядів. Вони не вимагають припущень про розподіл даних, ефективні для нестационарних часових рядів, адаптивні та універсальні. Однак вони також більш складні і вимагають більше часу для обчислень.

Вибір конкретного алгоритму бустингового МН для прогнозування нестационарних часових рядів залежить від конкретних даних і завдань. Однак загальна рекомендація полягає в тому, щоби використовувати алгоритм, який дозволяє регулювати вагу вкладу кожної слабкої моделі.

Далі розглянемо більш детально існуючі методи бустингового МН, які можуть бути використані для прогнозування нестационарних часових рядів, а саме: XGBoost (eXtreme Gradient Boosting), LightGBM (Light Gradient Boosting Machine), CatBoost (CatBoost), GBRT (Gradient Boosted Regression Trees), AdaBoost (Adaptive Boosting), Gradient Boosting Forest.

XGBoost (eXtreme Gradient Boosting) – це алгоритм МН, який використовує метод градієнтного бустингу для побудови моделі прогнозування. XGBoost був розроблений Чень Тяньці і його командою у 2016 році [386].

XGBoost – це алгоритм градієнтного бустингу над деревами рішень. Градієнтний бустинг – це метод МН, який створює ансамбль моделей, кожен з яких називається «слабким вчителем». Слабкий вчитель – це модель, яка в середньому краще випадкової, але не обов'язково дуже хорошою.

У XGBoost кожне дерево рішень навчається так, щоб мінімізувати функцію втрат, яка визначається як сума функцій втрат для всіх навчальних прикладів:

$$L(\theta) = \sum_{i=1}^n l(y_i, f(x_i; \theta)) \quad (1.1)$$

де  $\theta$  – параметри моделі,  $l$  – функція втрат, а  $f(x_i; \theta)$  – прогноз моделі для прикладу  $x_i$ .

LightGBM (Light Gradient Boosting Machine) – це алгоритм МН, який використовує метод градієнтного бустингу для побудови моделі прогнозування. LightGBM був розроблений Microsoft Research у 2016 році [406].

Для побудови дерева LightGBM використовує метод, заснований на гістограмі. Дані розподіляються по осередках з використанням гістограмі розподілу. Цей метод можна оптимізувати і для розрідженого набору даних. Можна математично описати процедуру побудови дерева LightGBM.

Для кожного осередку  $i$  в гістограмі обчислюється коефіцієнт посилення  $G_i$ :

$$G_i = \sum_{j \in i} -y_j \nabla f(y_j) \quad (1.2)$$

де  $f(y)$  – функція втрат, а  $\nabla f(y)$  – її градієнт.

LightGBM відзначається рядом особливостей, що роблять його ефективним алгоритмом для навчання моделей МН. Зокрема, використання гістограмного методу дозволяє ефективно обробляти розріджені набори даних, що є важливим при роботі з великими обсягами інформації. Крім того, функція втрат у LightGBM може бути адаптована до різних задач МН, підвищуючи гнучкість алгоритму в застосуванні до різноманітних проблем. Нарешті, параметри регуляризації, доступні в LightGBM, дозволяють запобігати перенавчанню моделі, що сприяє покращенню її здатності до узагальнення на нових даних.



CatBoost (Categorical Boosting) [76] – це алгоритм МН, який використовує метод градієнтного бустингу для побудови моделі прогнозування. Нехай дано набір даних  $\{(x_i, y_i)\}_{i=1}^n$ , де  $x_i$  – вектор ознак розмірності  $d$ , а  $y_i$  – цілочисельний таргет. Нехай  $T$  – кількість дерев в ансамблі,  $f_t(x)$  – функція прогнозу  $t$ -го дерева, а  $w_t$  – ваговий коефіцієнт  $t$ -го дерева. Тоді функція прогнозу CatBoost для даного набору даних визначається як:

$$f(x) = \sum_{t=1}^T w_t f_t(x) \quad (1.3)$$

CatBoost має ряд переваг порівняно з іншими алгоритмами МН. По-перше, він добре працює як з числовими, так і з категоріальними ознаками, що робить його універсальним для різних типів даних. По-друге, CatBoost характеризується високою швидкістю та ефективністю, що сприяє швидкому навчанню моделей без втрати точності. Нарешті, цей алгоритм може бути використаний для вирішення різних завдань МН, таких як класифікація, регресія та оцінка, що розширює його застосування у різних сферах.

Градієнтний бустинг (GBRT) [444] – метод ансамблевого МН, який використовує послідовність слабких моделей, щоби побудувати найсильнішу модель. Слабкі моделі, які використовуються в GBRT, зазвичай є деревами рішень. Припустимо є  $n$  прикладів  $(x_1, y_1), \dots, (x_n, y_n)$ , де  $x_i \in R^d$  – вектори характеристик, а  $y_i \in R$  – цільові значення. GBRT – це ансамбль з  $T$  дерев рішень, де кожне дерево  $T_t$  має  $L_t$  листків.

Для кожного дерева  $T_t$  спочатку обирається  $L_t$  гіперпараметрів  $\theta_t = (\theta_{t,1}, \dots, \theta_{t,L_t})$ , які визначають структуру дерева. Потім, для кожного прикладу  $x_i$ , ми обчислюємо прогноз дерева  $T_t$  на цьому прикладі:

$$f_t(x_i; \theta_t) = \sum_{l=1}^{L_t} \beta_{t,l} I(x_i \in R_t^l), \quad (1.4)$$

де  $R_t^l$  –  $l$ -й лист дерева  $T_t$ , а  $\beta_{t,l}$  – ваговий коефіцієнт цього листа.

GBRT має ряд унікальних характеристик, що відрізняють його від інших алгоритмів бустингу. Зокрема, GBRT використовує метод градієнтного спуску з адаптивним темпом навчання, що дозволяє моделі швидше навчатися та досягати

вищої точності. Крім того, алгоритм застосовує регуляризацію для запобігання перенавчанню, що сприяє збереженню його універсальності та уникненню перенавчання на навчальних даних. Також GBRT має кілька вдосконалень, які підвищують його продуктивність і точність, роблячи його ефективним інструментом для різних задач МН.

AdaBoost (Adaptive Boosting) [115] – це алгоритм МН, який використовує метод адаптивного бустингу для побудови моделі прогнозування. Адаптивний бустинг – це метод ансамблевого навчання, який послідовно додає слабкі моделі до ансамблю, щоби зменшити функцію втрат. Кінцевий класифікатор визначається як сума слабких класифікаторів, зважених їхніми коефіцієнтами:

$$h(x) = \sum_{m=1}^M a_m h_m(x). \quad (1.5)$$

AdaBoost має ряд переваг, що роблять його ефективним алгоритмом МН. По-перше, він може використовувати будь-які слабкі класифікатори, що забезпечує високу універсальність та можливість адаптації до різних типів задач. По-друге, AdaBoost відзначається стабільністю та меншою схильністю до перенавчання порівняно з деякими іншими методами, що підвищує його здатність до узагальнення на нових даних. Крім того, алгоритм легко адаптується для роботи з нерівномірно розподіленими даними, що робить його ефективним у ситуаціях, де дані мають нестандартні розподіли.

Гradientний бустинг лісів (GBDT) [277] – це метод ансамблевого МН, який використовує послідовність слабких моделей, щоби побудувати сильну модель. Слабкі моделі, які використовуються в GBDT, зазвичай є деревами рішень.

Нехай  $x$  буде входом в GBDT, а  $y$  – дійсною вихідною величиною. Нехай  $f_1(x)$ ,  $f_2(x), \dots, f_T(x)$  – це  $T$  дерев рішень, які формують ансамбль. Тоді передбачення GBDT для  $x$  буде таким:

$$f(x) = \sum_{t=1}^T f_t(x) \quad (1.6)$$

Дерево  $f_t(x)$  може бути сформовано за допомогою будь-якого алгоритму формування дерева рішень, такого як ID3, C4.5 або CART. Однак, в GBDT часто

використовується алгоритм формування дерева CART, оскільки він дозволяє контролювати складність дерева за допомогою параметра  $C$ .

GBDT має кілька важливих переваг, що роблять його ефективним алгоритмом МН. По–перше, GBDT є ансамблевим алгоритмом, який об'єднує прогнози декількох моделей для отримання більш точного передбачення. По–друге, він використовує метод градієнтного спуску, що дозволяє оновлювати параметри моделі з метою мінімізації похибки. Хоча GBDT може бути схильним до перенавчання, це можна стабілізувати за допомогою різних методів, таких як регуляризація та випадковий вибір, що підвищує його здатність до узагальнення на нових даних.

Отже, усі розглянуті алгоритми базуються на деревах рішень та підтримують роботу з числовими та категорійними ознаками, що робить їх універсальними для різних типів даних. Важливо підкреслити, що ці алгоритми здатні ефективно обробляти нестационарні дані, оскільки вони не вимагають припущень про стаціонарність і можуть адаптуватися до змін у статистичних характеристиках даних. LightGBM виділяється найвищою швидкістю обробки завдяки оптимізованим методам паралелізації та ефективно працює з великими нестационарними наборами даних. CatBoost забезпечує високу точність і менш схильне до перенавчання завдяки спеціальній обробці категорійних ознак і можливості враховувати часову структуру даних, що особливо корисно при роботі з нестационарними часовими рядами. XGBoost та GBRT демонструють високу стійкість до надмірного навчання та здатність моделювати складні нелінійні взаємозв'язки в нестационарних даних, хоча XGBoost має високу складність гіперпараметрів, що може ускладнювати процес налаштування. AdaBoost та Gradient Boosting Forest пропонують загальну застосовність з високою можливістю перенавчання та високою точністю, проте можуть бути менш ефективними при обробці нестационарних даних без додаткової модифікації. Всі алгоритми підтримують регуляризацію, обробку пропусків та ітеративність, що сприяє їхній гнучкості та адаптивності до різних завдань в умовах нестационарності. Загалом, LightGBM та CatBoost виділяються своєю високою продуктивністю та ефективністю в роботі з нестационарними даними, що робить їх найбільш перспективними для широкого спектра застосувань у сфері МН.

Оцінка якості моделей є ключовим етапом в задачах інтелектуального аналізу даних, що дозволяє визначати ефективність алгоритмів та вибирати оптимальну модель для практичного застосування. Одним з основних методів оцінки є крос-валідація, яка забезпечує надійність оцінки шляхом поділу даних на декілька фолдів. Зокрема, у  $k$ -фолдовій крос-валідації вихідний набір даних поділяється на  $k$  піднаборів (фолдів), де на  $k - 1$  – піднаборі модель тренується, а на одному – перевіряється. Процедура повторюється для кожного фолду, що забезпечує усереднене значення якості моделі, незалежне від випадкового поділу даних. Іншим підходом є стратифікована крос-валідація, яка зберігає початковий розподіл класів у кожному фолді, що особливо актуально для розбалансованих даних. Метод Leave-One-Out є варіантом крос-валідації з  $k = n$ , де  $n$  – кількість об'єктів у наборі даних, і дозволяє мінімізувати похибки оцінки, проте є ресурсозатратним для великих наборів даних.

У випадках розбалансованих даних важливо застосовувати методи корекції, такі як oversampling та undersampling. Метод oversampling збільшує частоту рідкісного класу шляхом додавання або синтетичної генерації зразків, наприклад, з використанням методу SMOTE (Synthetic Minority Over-sampling Technique). Undersampling, у свою чергу, зменшує частоту переважаючого класу для досягнення балансу шляхом випадкового видалення зразків. Крім того, застосування зважених класів дозволяє компенсувати розбалансованість, що є ефективним у багатьох алгоритмах класифікації, наприклад, логістичної регресії та дерев рішень.

Метрики класифікації [78, 347] надають можливість кількісно оцінювати точність моделі за різними аспектами. Основною метрикою є точність (Accuracy), яка визначається як частка правильно класифікованих об'єктів серед загальної кількості і обчислюється за формулою:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1.7)$$

де TP – істинно позитивні, FN – хибно негативні, FP – хибно позитивні, а TNT – істинно негативні значення.

Ця метрика має значення в діапазоні  $[0, 1]$ , де значення, наближені до 1, вказують на високу точність моделі. Однак точність може бути недостатньою для розбалансованих даних, оскільки вона не враховує частоту помилкових спрацювань.

Для оцінки класифікації в умовах розбалансованості використовуються метрики прецизійності (Precision) та повноти (Recall):

$$Precision = \frac{TP}{TP+FP} \quad (1.8)$$

$$Recall = \frac{TP}{TP+FN} \quad (1.9)$$

Прецизійність показує здатність моделі уникати хибно позитивних спрацювань, тоді як повнота характеризує здатність моделі виявляти всі позитивні приклади. Для балансування прецизійності та повноти застосовується F1-бал (F1-Score), який обчислюється як гармонійне середнє між цими метриками:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (1.10)$$

F1-бал має значення в діапазоні  $[0, 1]$  і є корисним при необхідності врахування якості класифікації рідкісних класів.

Для оцінки здатності моделі розрізняти класи при різних порогах класифікації використовується метрика AUC-ROC (Area Under the Receiver Operating Characteristic Curve). ROC-крива (Receiver Operating Characteristic Curve) – це графік, що ілюструє залежність між істинно позитивною частотою (True Positive Rate, TPR) та хибно позитивною частотою (False Positive Rate, FPR) при різних порогах прийняття рішення. Кожен поріг представляє альтернативну точку ROC-кривої, а саме значення TPR та FPR можна обчислити за формулами:

$$TPR = \frac{TP}{TP+FN} \quad (1.11)$$

$$FPR = \frac{FP}{FP+TN} \quad (1.12)$$

На ROC-графіку точка  $(0, 1)$  відповідає ідеальній класифікації, де немає хибнопозитивних та хибнонегативних значень, а точка  $(1, 1)$  – повній випадковості в класифікації. Моделі, які краще відрізняють класи, розташовуються ближче до

верхнього лівого кута графіка, а лінія  $y = x$  представляє випадковий розподіл, де модель не має здатності розрізняти класи. AUC–ROC відображає загальну якість класифікації: площа під кривою AUC, що наближається до 1, вказує на високий рівень точності, тоді як AUC, наближене до 0.5, свідчить про те, що модель працює на рівні випадкового передбачення.

У задачах регресії важливими метриками [207, 211, 233] є середня абсолютна похибка (Mean Absolute Error, MAE), середня квадратична похибка (Mean Squared Error, MSE) та корінь середньої квадратичної похибки (Root Mean Squared Error, RMSE). MAE визначається як середнє абсолютне відхилення прогнозів від фактичних значень і обчислюється за формулою:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (1.13)$$

де  $y_i$  – фактичне значення;  $\hat{y}_i$  – прогнозоване значення,  $n$  – кількість зразків.

MAE дозволяє інтерпретувати середню помилку у тій самій шкалі, що і цільова змінна.

MSE і RMSE більш чутливі до великих помилок, оскільки штрафують великі відхилення. Вони обчислюються як:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (1.14)$$

$$RMSE = \sqrt{MSE}. \quad (1.15)$$

MSE та RMSE мають значення, які також знаходяться в межах шкали цільової змінної; нижчі значення метрик вказують на високу точність моделі. Крім того, коефіцієнт детермінації  $R^2$  показує частку поясненої варіації у цільовій змінній і обчислюється як:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (1.16)$$

де  $\bar{y}$  – середнє значення цільової змінної.

Значення  $R^2$  у діапазоні  $[0, 1]$  свідчить про ступінь пояснення варіації цільової змінної моделі.

У задачах кластеризації для оцінки якості [229, 309] поділу даних на кластери використовуються індекс силуетів, індекс Давіса–Болдіна та внутрішньокластерна сумарна варіація. Індекс силуетів визначає ступінь подібності кожного об'єкта до свого кластера відносно інших кластерів:

$$Silhouette = \frac{b-a}{\max(a,b)}, \quad (1.17)$$

де  $a$  – середня відстань між об'єктом та іншими об'єктами в тому самому кластері;  $b$  – середня відстань до найближчого кластера. Індекс силуетів змінюється в діапазоні  $[-1, 1]$ , де значення, ближчі до 1, свідчать про чіткі межі кластерів.

Індекс Давіса–Болдіна обчислює відношення між середньою відстанню між об'єктами в кластері та відстанню між центроїдами різних кластерів, ідеально прагнучи до мінімізації:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{s_i + s_j}{d_{i,j}} \right), \quad (1.18)$$

де  $s_i$  та  $s_j$  – середні відстані між об'єктами в кластерах  $i$  та  $j$ ,  $d_{i,j}$  – відстань між центроїдами цих кластерів.

Внутрішньокластерна сумарна варіація (Within-cluster Sum of Squares) визначає гомогенність даних в межах кластерів і прагне до мінімізації для забезпечення їх компактності. Для візуалізації структури кластерів використовуються методи зниження розмірності, такі як  $t$ -SNE та PCA, що допомагає зрозуміти розподіл даних.

Оцінка точності моделей для часових рядів [221] має свої специфічні особливості, оскільки аналізуються залежності та динамічні зміни в часі. Основними метриками для прогнозування часових рядів є середня абсолютна відсоткова похибка (MAPE), середнє абсолютне стандартне відхилення (MASE) та симетрична середня абсолютна відсоткова похибка (SMAPE).

MAPE обчислюється як середнє абсолютне відхилення між прогнозованими та фактичними значеннями, виражене у відсотках:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%. \quad (1.19)$$

Значення MAPE показує, на скільки відсотків в середньому модель відхиляється від фактичних значень, де менші значення MAPE вказують на кращу точність прогнозу.

MASE, або середнє абсолютне стандартне відхилення, використовується для порівняння прогнозів моделі з простою моделлю середнього значення і показує відносну ефективність моделі. SMAPE, або симетрична середня абсолютна відсоткова похибка, розраховується за формулою:

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\frac{|y_i| + |\hat{y}_i|}{2}} \quad (1.20)$$

SMAPE вирішує проблему асиметрії, дозволяючи більш коректно оцінити похибку для даних з великими амплітудними варіаціями. Крім того, для часових рядів важливо враховувати сезонність та тренди, що дозволяє покращити точність прогнозів та адаптувати модель до специфічних змін у даних.

Тест на стаціонарність [6, 68, 431] є інструментом у часових рядах для перевірки того, чи є середнє значення, дисперсія і коваріація даних сталі з часом. Найпоширеніший метод для цього – тест Дікі–Фуллера, зокрема його розширена версія, яка називається розширений тест Дікі–Фуллера (ADF).

Тест Дікі–Фуллера перевіряє наявність одиничного кореня в авторегресійному моделюванні часового ряду. Припустимо, що є часовий ряд  $y_t$ , і ми моделюємо його як:

$$y_t = \rho y_{t-1} + u_t, \quad (1.21)$$

де  $\rho$  – коефіцієнт при лагованому значенні  $y$ ;  $u_t$  – шум (залишки), що описуються як білий шум.

Для тесту Дікі–Фуллера розглядається таке рівняння:

$$\Delta y_t = (\rho - 1)y_{t-1} + u_t \quad (1.22)$$

або у більш загальному вигляді:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_p \Delta y_{t-p} + u_t, \quad (1.23)$$



де  $\Delta$  вказує на різницю, тобто  $\Delta y_t = y_t - y_{t-1}$ ;  $\alpha$  – константа (якщо  $\epsilon$ );  $\beta t$  – термін тренду (якщо  $\epsilon$ );  $\gamma$  – коефіцієнт перед  $y_{t-1}$  (основний інтерес тесту);  $\delta_1, \dots, \delta_p$  – коефіцієнти при лагованих різницях для включення інерції часового ряду;  $p$  – порядок лагу.

Нульова гіпотеза  $H_0$  тесту Дікі–Фуллера полягає в тому, що  $\epsilon$  є одиничний корінь ( $\gamma = 0$ ), і це відповідає нестационарності ряду. Альтернативна гіпотеза  $H_a$  припускає  $\gamma < 0$ , що вказує на стаціонарність ряду.

Тест обчислюється і на основі значення  $p$  (і порівняння тестової статистики з критичними значеннями) можна визначити, чи відкидається нульова гіпотеза. Якщо нульова гіпотеза відкидається, часовий ряд вважається стаціонарним.

Цей тест є фундаментальним інструментом для аналізу часових рядів, особливо коли йдеться про моделювання і прогнозування даних, де стаціонарність даних є критичною передумовою.

Налаштування гіперпараметрів є важливим етапом оптимізації моделей, адже значення гіперпараметрів безпосередньо впливають на продуктивність алгоритму. Одним з основних підходів є Grid Search, що реалізує повний перебір усіх можливих комбінацій гіперпараметрів із заданих діапазонів. У цьому випадку кожна комбінація значень гіперпараметрів  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$  оцінюється на основі значення цільової функції  $f(\theta)$  для задачі оптимізації, де  $k$  – кількість параметрів, а  $f$  – обрана метрика якості (наприклад, точність або AUC–ROC). Таким чином, Grid Search шукає глобальний максимум (або мінімум) цільової функції, проте потребує великої кількості обчислень, особливо для багатовимірного простору параметрів.

Іншим ефективним методом є Random Search, де комбінації гіперпараметрів обираються випадковим чином. З точки зору ймовірності, Random Search забезпечує ефективне охоплення простору параметрів при меншій кількості ітерацій. Середнє очікуване значення якості моделі для випадкових вибірок  $E[f(\theta)]$  наближається до оптимального при збільшенні кількості ітерацій, однак Random Search потребує менше обчислювальних ресурсів у порівнянні з Grid Search.

Байєсівська оптимізація використовує ймовірнісну модель, наприклад, гаусівські процеси, для побудови апроксимації функції  $f(\theta)$  на основі попередніх результатів. На кожній ітерації обирається нове значення гіперпараметрів  $\theta$  на основі максимізації показника очікуваного приросту (Expected Improvement, EI), який обчислюється за формулою:

$$EI(\theta) = E[\max(0, f(\theta) - f_{best})], \quad (1.24)$$

де  $f_{best}$  – найкраще значення функції якості, отримане на попередніх ітераціях.

Байєсівська оптимізація дозволяє зосередитись на областях з високою ймовірністю досягнення оптимального результату, знижуючи обчислювальні витрати.

Крім класичних методів, застосовуються еволюційні алгоритми, такі як генетичні алгоритми, де гіперпараметри обираються шляхом моделювання природної еволюції (вибір, кросовер, мутація). Ці алгоритми шукають комбінації гіперпараметрів, які дають змогу уникнути локальних мінімумів, досягаючи глобальної оптимізації в складних просторах параметрів.

Нурєвband є адаптивним методом, який поєднує випадковий пошук з ідеєю раннього відсіювання, забезпечуючи економію ресурсів за рахунок відсіювання менш перспективних комбінацій на початкових етапах оптимізації.

У задачах класифікації з розбалансованими класами обробка даних є критично важливою для забезпечення точності моделей, особливо у випадках, коли цільовий клас (рідкісний) має низьку частоту в наборі даних. Основні підходи до вирішення цієї проблеми включають oversampling, undersampling, зважування класів і методи, чутливі до витрат (Cost-Sensitive Learning).

Oversampling передбачає збільшення кількості зразків меншого класу для досягнення балансу в даних. Один із базових підходів – Random Oversampling, де зразки рідкісного класу випадково дублюються. Це дозволяє моделі краще розпізнавати рідкісний клас, але може приводити до перенавчання, оскільки дубльовані зразки не додають нової інформації. Більш просунутий метод – SMOTE (Synthetic Minority Over-sampling Technique), який генерує синтетичні зразки,

використовуючи інтерполяцію між наявними точками меншого класу. Формально, для кожного зразка  $x_i$  рідкісного класу SMOTE створює новий зразок  $x_{new}$  за формулою:

$$x_{new} = x_i + \delta \cdot (x_{nearest} - x_i), \quad (1.25)$$

де  $x_{nearest}$  – один з найближчих сусідів  $x_i$ ;  $\delta$  – випадкове число в діапазоні  $[0,1]$ .

Таким чином, SMOTE створює нові зразки, розподілені в межах простору меншого класу, що знижує ризик перенавчання, додаючи до набору даних різноманітні точки.

Undersampling передбачає зменшення кількості зразків більшого класу для досягнення балансу. У Random Undersampling випадково видаляються зразки більшого класу, що знижує обчислювальні витрати, але може приводити до втрати важливої інформації. Більш цілеспрямовані методи, такі як Tomek Links і Edited Nearest Neighbors (ENN), вибірково видаляють зразки, які складно класифікувати, або знаходяться на межі класів. У методі Tomek Links пара  $(x_i, x_j)$  визначається як Tomek Link, якщо  $x_i$  належить до одного класу,  $x_j$  – до іншого, і вони є найближчими сусідами. Видалення таких пар дозволяє зменшити накладання класів, що підвищує якість класифікації:

$$d(x_i, x_j) < d(x_i, x_k) \quad \forall k \neq j, \quad (1.26)$$

де  $d(x_i, x_j)$  – відстань між  $x_i$  і  $x_j$ .

Зважування класів (Class Weighting) дозволяє компенсувати дисбаланс, не змінюючи самих даних. Це здійснюється шляхом призначення вищої ваги рідкісному класу під час тренування моделі, щоб зменшити вплив дисбалансу. Наприклад, у логістичній регресії втрата моделі  $L$  змінюється на зважену втрату  $L_w$ :

$$L_w = w_{minority} \cdot L_{minority} + w_{majority} \cdot L_{majority}, \quad (1.27)$$

де  $w_{minority}$  і  $w_{majority}$  – ваги для рідкісного та домінуючого класів, відповідно;  $L_{minority}$  та  $L_{majority}$  – втрати для кожного класу.

Cost-Sensitive Learning – підхід, який мінімізує помилки для рідкісного класу, враховуючи різні витрати для кожного класу. Тут моделі навчаються з урахуванням того, що помилки в прогнозуванні рідкісного класу є більш критичними. Витрати за помилки можуть бути додані до функції втрат  $L$ , щоб модель приділяла більше уваги рідкісному класу:

$$L_{cost-sensitive} = c_{minority} \cdot L_{minority} + c_{majority} \cdot L_{majority}, \quad (1.28)$$

де  $c_{minority}$  і  $c_{majority}$  – відповідні коефіцієнти витрат.

Ансамблеві методи (Balanced Random Forest, EasyEnsemble) дозволяють додатково зменшити вплив розбалансованих даних. Balanced Random Forest виконує undersampling для кожної побудови дерева в ансамблі, що забезпечує баланс під час навчання, тоді як EasyEnsemble створює кілька підмножин меншого класу для побудови ансамблю моделей, що знижує ризик перенавчання і покращує загальну стабільність класифікації.

Ці методи обробки розбалансованих даних покращують стабільність і точність моделей, що є важливим у задачах класифікації, де наявність рідкісних класів може значно впливати на результати прогнозування.

Отже, розглянуто основні методи та метрики оцінки якості моделей, які є критичними для забезпечення точності, стабільності та надійності моделей у задачах інтелектуального аналізу даних. Метрики класифікації, такі як точність, прецизійність, повнота, F1-бал та AUC-ROC, дозволяють детально аналізувати продуктивність моделей класифікації, тоді як метрики регресії (MAE, MSE, RMSE та  $R^2$ ) забезпечують точне вимірювання похибок у прогнозуванні. Метрики для часових рядів, зокрема MAPE, MASE та SMAPE, дозволяють оцінювати моделі в умовах трендів і сезонності, а методи вибору гіперпараметрів та обробки розбалансованих даних забезпечують оптимізацію параметрів моделей і їх адаптацію до особливостей даних. Усі розглянуті методи та метрики будуть використані в наступних задачах дисертації для всебічної оцінки запропонованих моделей, що дозволить забезпечити обґрунтований вибір найефективніших підходів для підтримки соціальної стійкості, розвитку та управління в контексті інтелектуального аналізу даних ТГ.

З наведеного аналізу випливає, що традиційні статистичні методи, попри їх простоту, не здатні забезпечити адекватну точність у нестационарних умовах. Водночас, бустингові алгоритми потребують значних обчислювальних ресурсів та ретельного налаштування параметрів. Це створює протиріччя між простотою використання та потребою у високій точності.

Подолання цього протиріччя можливе через інтеграцію бустингових методів з адаптивними предикторами, які враховують специфіку соціально–економічних даних ТГ. Такий підхід дозволить:

- забезпечити високу точність прогнозування динамічних процесів;
- оптимізувати обчислювальні витрати за рахунок вибору більш точних алгоритмів.

Детальний розгляд запропонованої методології буде представлений у подальших дослідженнях, зокрема щодо її застосування для підтримки соціально–економічної стійкості ТГ.

#### **1.4 Висновки, постановка наукової проблеми та задач дослідження**

З проведеного аналізу випливає, що розробка інтелектуальної інформаційної технології аналізу соціально–економічних даних в контексті ТГ відіграватиме важливу роль у забезпеченні стабільності соціально–економічної інфраструктури. Аналіз масивів даних, зібраних з різних джерел, включаючи опитування мешканців та статистичні показники, така система дозволяє глибоко зрозуміти потреби громадян, виявити тенденції та відслідковувати зміни у соціальній, екологічній, економічній та суспільній сферах. Використання інтелектуальних методів обробки даних сприяє управлінню інфраструктурою громади, підвищенню рівня комфорту життя, зменшенню екологічного впливу та підтримці соціально–політичної стабільності, особливо в умовах сучасних викликів, таких як військова агресія.

У сучасних реаліях, коли ТГ стикаються з новими та складними викликами, такими як пандемія, економічна нестабільність та військова агресія, виникає нагальна

потреба в інструментах для підтримки стійкості та розвитку. Розробка теоретичних та прикладних засад інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ є актуальною, оскільки це дозволить оперативно реагувати на зміни, прогнозувати потенційні ризики та приймати обґрунтовані управлінські рішення.

Актуальність проблеми аналізу соціально–економічних даних ТГ обумовлена необхідністю забезпечення точного прогнозування, адаптивності, інтеграції різнорідних даних та підвищення управлінських рішень в умовах динамічних змін. Відсутність моделей, методів та засобів, які б задовольняли цим вимогам, потребує розроблення нових наукових підходів та технологій.

Вирішення окресленої наукової проблеми потребує комплексних досліджень за такими напрямками:

- 1) провести дослідження сучасного стану та визначити перспективи розвитку аналізу соціально–економічних даних у межах інфраструктури ТГ;
- 2) розробити узагальнений принцип синтезу інформаційної технології для інтелектуального аналізу соціально–економічних даних ТГ;
- 3) розробити методологію аналізу соціально–економічних даних громад, що включає адаптивні методи класифікації, кластеризації та прогнозування;
- 4) розробити новий метод формування навчальної вибірки для нестационарних даних на основі RFM та кластерного аналізу;
- 5) розробити метод ансамблевих адаптивних прогнозних моделей для багатовимірного аналізу;
- 6) розробити метод класифікаційного аналізу кількісних соціально–економічних даних;
- 7) розробити метод класифікаційного аналізу текстових соціально–економічних даних;
- 8) розробити метод кластерного аналізу соціально–економічних даних;
- 9) розробити метод прогнозування соціально–економічних даних;
- 10) розробити метод гібридного аналізу соціально–економічних даних;
- 11) оцінити ефективність запропонованих методів та інформаційної технології.

Запропоновані напрями дослідження дозволять забезпечити інтеграцію сучасних інтелектуальних методів аналізу даних, оптимізувати процес прийняття управлінських рішень та сприяти соціально–економічному розвитку ТГ в умовах динамічного середовища.

Таким чином, послідовне вирішення поставлених задач дозволить розробити інформаційну технологію інтелектуального аналізу соціально–економічних даних ТГ, що сприятиме управлінню ресурсами, підвищенню соціальної стійкості та забезпеченню стійкого розвитку ТГ в умовах динамічних соціально–економічних змін.

## РОЗДІЛ 2.

### ПРИНЦИПИ СИНТЕЗУ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ТА МЕТОДОЛОГІЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ СОЦІАЛЬНО– ЕКОНОМІЧНИХ ДАНИХ ТЕРИТОРІАЛЬНИХ ГРОМАД

#### **2.1 Узагальнений принципи синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних територіальних громад**

У кризових умовах, таких як пандемія чи війна, стійкість ТГ залежить від здатності їхньої соціально–економічної інфраструктури адаптуватися до змін та забезпечувати ключові функції. Вибір об'єктів інфраструктури є критично важливим, адже саме їхня ефективна взаємодія дозволяє підтримувати стабільність громад у складних умовах. Узагальнена структура інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ відображає системний підхід до інтеграції даних з різних джерел, аналізу їх специфіки та прийняття обґрунтованих рішень.

На рис. 2.1 представлено узагальнену структуру, що ілюструє ключові компоненти інформаційної технології, її архітектуру і взаємозв'язок між об'єктами інфраструктури, типами даних та методами інтелектуального аналізу.

Особливу увагу приділено ключовим об'єктам інфраструктури, таким як бізнес–інфраструктура, освітня, інформаційна, соціальна, логістична, економічна та культурна. Кожен з цих об'єктів виконує важливу роль у підтримці життєдіяльності громади та потребує специфічного підходу до аналізу даних, що їх характеризують.

Завдяки використанню інтегрованої системи аналізу забезпечується можливість:

- враховувати гетерогенність даних;
- обирати адаптивні методи обробки для досягнення найкращих результатів;
- підтримувати прийняття стратегічних рішень, спрямованих на підвищення стійкості громади.



Запропонована структура дозволяє формалізувати та систематизувати дані, методи аналізу та об'єкти інфраструктури, забезпечуючи надійний інструмент для вирішення завдань соціально–економічного розвитку ТГ навіть в умовах кризи.

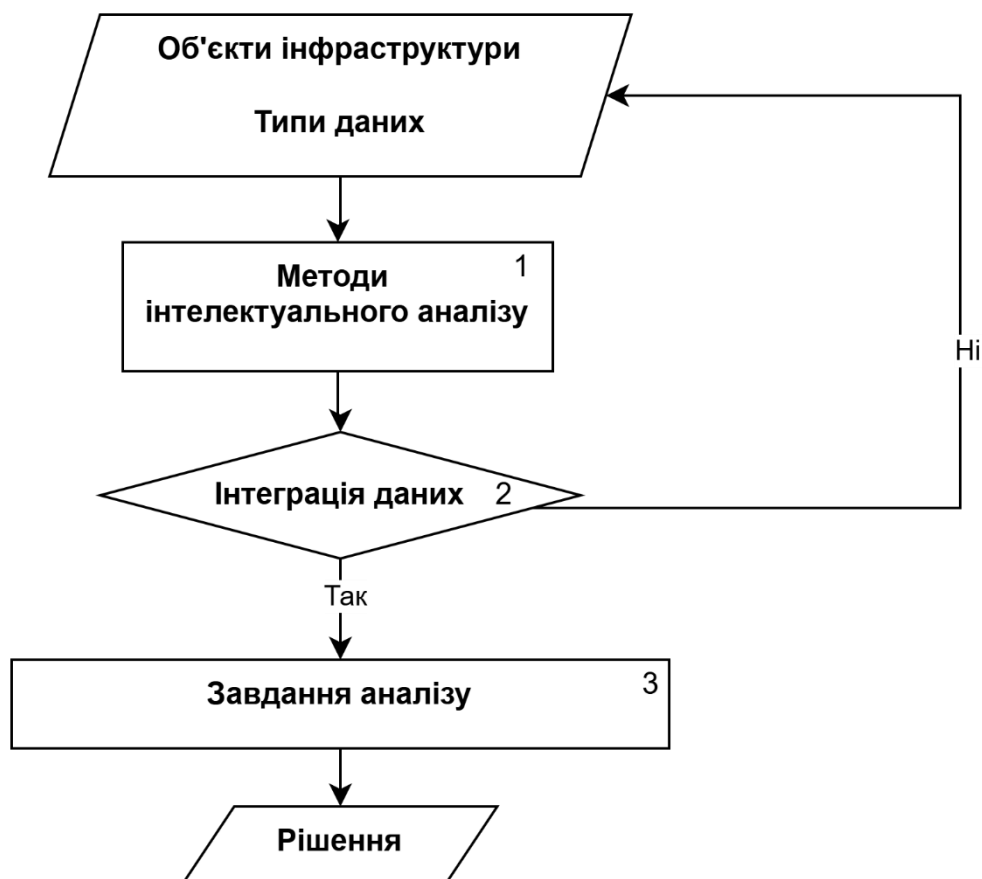


Рис. 2.1. Узагальнена структура інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ

Вибір об'єктів інфраструктури зумовлений їхньою роллю у збереженні стабільності та підтримці життєдіяльності громади. Зокрема, бізнес–інфраструктура ( $I_1$ ) є фундаментом економічного розвитку ТГ, оскільки забезпечує зайнятість, підтримує фінансові потоки та сприяє створенню локальних і регіональних ринків. В умовах кризи, особливо під час війни, важливо також виявляти фіктивні підприємства, які можуть підривати економічну безпеку громади. Освітня інфраструктура ( $I_2$ ) забезпечує підготовку кваліфікованих кадрів та сприяє

професійній перепідготовці, що стає критично важливим у періоди економічних потрясінь, коли потрібна швидка адаптація до змін на ринку праці.

Інформаційна інфраструктура ( $I_3$ ) є ключовою для боротьби з дезінформацією, яка під час війни чи пандемії може суттєво впливати на суспільну стабільність і психологічний стан населення. Соціальна інфраструктура ( $I_4$ ), зокрема її здатність аналізувати потреби мешканців, формувати споживчі кошики та моніторити настрої, допомагає забезпечити базові потреби громадян, особливо під час порушень у ланцюгах постачання чи економічних криз. У цьому контексті логістична інфраструктура ( $I_5$ ) відіграє вирішальну роль, що дозволяє швидко шукати альтернативних постачальників та адаптувати транспортні шляхи. Економічна інфраструктура ( $I_6$ ) забезпечує інструменти для оцінки ризиків і розробки стратегій збереження економічної стабільності громади, тоді як культурна та рекламна інфраструктура ( $I_7$ ) дозволяє підтримувати емоційний стан мешканців, підвищувати довіру до місцевих підприємств та залучати додаткові ресурси. Ця інтегрована система об'єктів інфраструктури забезпечує всебічну підтримку стійкості ТГ в умовах кризових ситуацій.

Отже, нехай  $I$  – множина об'єктів інфраструктури ТГ, що включає такі компоненти:

$$I = \{I_1, I_2, \dots, I_n\}, \quad (2.1)$$

де кожен  $I_j$  (для  $j = 1, 2, \dots, n$ ) представляє окремий елемент інфраструктури, наприклад:  $I_1$  – бізнес-інфраструктура, включаючи підприємства, IT-компанії, фіктивні підприємства та ринки;  $I_2$  – освітня інфраструктура, що враховує рівень кваліфікації персоналу, освітні програми та можливості для короткострокових проєктів управління;  $I_3$  – інформаційна інфраструктура, включаючи дані про дезінформацію, соціальні медіа та інформаційні системи;  $I_4$  – соціальна інфраструктура, пов'язана з формуванням споживчого кошика та аналізом споживчих настроїв;  $I_5$  – логістична інфраструктура, що включає пошук альтернативних постачальників і оптимізацію ланцюгів постачання;  $I_6$  – економічна інфраструктура, що охоплює аналіз інвестиційних ризиків, прогнозування попиту на продукцію та

вибір конкурентоспроможних товарів;  $I_7$  – культурна та рекламна інфраструктура, що включає формування та вдосконалення рекламного контенту, а також аналіз емоційного забарвлення відгуків.

Для ефективного аналізу об'єктів соціально–економічної інфраструктури ТГ необхідно враховувати різноманітність типів даних, які характеризують їх функціонування. Дані, що використовуються для аналізу соціально–економічних процесів, можуть мати різну структуру, походження і рівень організованості. Ця гетерогенність обумовлює необхідність їх класифікації та вибору відповідних методів обробки, адже лише через формалізацію типів даних можна забезпечити адекватний вибір інструментів інтелектуального аналізу, які відповідають природі та особливостям кожного виду даних.

Нехай  $D$  – множина типів даних, що відображають різні аспекти функціонування об'єктів інфраструктури та використовуються для моделювання і прогнозування. Множина даних формалізується як:

$$D = \{D_1, D_2, D_3\}, \quad (2.2)$$

де  $D_1$  – структуровані дані: числові показники, статистичні таблиці, фінансові звіти тощо;  $D_2$  – неструктуровані дані: текстові документи, медіафайли, соціальні медіа пости, зображення та відео;  $D_3$  – напівструктуровані дані: XML–, JSON–файли, які містять елементи як структурованих, так і неструктурованих даних.

Різноманітність типів даних, які характеризують об'єкти соціально–економічної інфраструктури, вимагає використання адаптивних підходів для їх аналізу. Ефективність обробки структурованих, неструктурованих та напівструктурованих даних залежить не лише від їхньої природи, але й від мети аналізу, наприклад, виявлення закономірностей, оцінка ризиків або прогнозування тенденцій. Для цього необхідно залучити інструменти, здатні працювати з такими різними типами інформації, інтегруючи їх у цілісний процес обробки і прийняття рішень.

Інтелектуальний аналіз даних ( $A$ ) включає різні методи та алгоритми (блок 1), які застосовуються до різних типів даних для виявлення прихованих

закономірностей, прогнозування тенденцій та підтримки прийняття рішень. Нехай  $A = \{A_1, A_2, \dots, A_m\}$ , де кожен  $A_k$  представляє окремий підхід або метод:

$$A = \{ \text{Машинне навчання, Обробка природної мови, Комп'ютерне бачення} \}. \quad (2.3)$$

Кожен метод  $A_k$  може застосовуватися до різних типів даних  $D_i$  залежно від специфіки задачі. Формально можна визначити функцію застосовності

$$f: D \times A \rightarrow \{0,1\}, \quad (2.4)$$

де  $f(D_i, A_k) = 1$  означає, що метод  $A_k$  застосований до типу даних  $D_i$ , і  $f(D_i, A_k) = 0$  – невідповідність.

Розглянуті об'єкти інфраструктури  $I$ , типи даних  $D$  та методи інтелектуального аналізу  $A$  становлять базис для побудови цілісного підходу до моделювання та аналізу соціально–економічних процесів у ТГ. Комплексна система інфраструктури ТГ потребує не лише ефективного управління, але й адаптивних інструментів для роботи з великими масивами даних, що характеризують різні аспекти життєдіяльності громади. Кожен об'єкт інфраструктури вимагає інтеграції даних з різних джерел і застосування специфічних методів аналізу, які враховують гетерогенність та динамічність даних.

На основі формалізації ключових компонентів  $I, D, A$ , можливо побудувати формальну модель аналізу, яка забезпечуватиме взаємозв'язок між об'єктами інфраструктури, типами даних та відповідними підходами. Така модель дозволяє не лише структурувати інформацію, але й визначати оптимальні алгоритми для обробки даних, прогнозування та підтримки прийняття рішень, спрямованих на розвиток і підвищення стійкості ТГ.

Інтеграція даних (блок 2) з різних об'єктів інфраструктури здійснюється за допомогою оператора  $\oplus$ , який об'єднує дані  $D_i$  з різних джерел:

$$D_{\text{інтегровані}} = D_1 \oplus D_2 \oplus \dots \oplus D_n, \quad (2.5)$$

де  $D_{\text{інтегровані}} \subseteq \bigcup_{i=1}^n D_i$ .

Для кожного об'єкта інфраструктури  $I_j$  визначається відповідний набір даних  $D_{I_j} \subseteq D$ . Аналіз даних для кожного  $I_j$  здійснюється за допомогою відповідного набору методів  $A_{I_j} \subseteq A$ :

$$A_{I_j} = \{ A_k \in A \mid f(D_i, A_k) = 1, \forall D_i \in D_{I_j} \}. \quad (2.6)$$

Кожне завдання (блок 3) аналізу  $Z_l$  визначається як трійка  $(I_j, D_i, A_k)$ , де  $I_j \in I$ ,  $D_i \in D$ ,  $A_k \in A$ , і  $f(D_i, A_k) = 1$ :

$$Z_l = (I_j, D_i, A_k). \quad (2.7)$$

Таким чином, множина всіх можливих завдань аналізу:

$$Z = \{ Z_l \mid Z_l = (I_j, D_i, A_k), \forall I_j \in I, D_i \in D, A_k \in A, f(D_i, A_k) = 1 \}, \quad (2.8)$$

де функція  $f(D_i, A_k) = 1$  означає, що метод аналізу  $A_k$  застосований до типу даних  $D_i$ . Тобто значення 1 свідчить про позитивну відповідність між методом і даними, а саме:

- якщо  $f(D_i, A_k) = 1$ , то метод  $A_k$  може обробляти або аналізувати дані типу  $D_i$ . Це означає, що обраний алгоритм (наприклад, класифікація, кластеризація, обробка тексту тощо) здатен працювати з певною структурою або природою даних;
- якщо  $f(D_i, A_k) = 0$ , то метод  $A_k$  не може бути застосований до даних типу  $D_i$ . Це може відбуватись через невідповідність формату даних, відсутність адаптивних механізмів у методі або інші технічні обмеження.

Далі на основі аналізу даних здійснюється прийняття рішень  $R_p$ , яке можна формалізувати як функцію  $g: Z \rightarrow R$ , де  $R$  – множина можливих рішень:

$$R_p = g(Z_l) \forall Z_l \in Z. \quad (2.9)$$

Формалізована модель представляє собою систему, де об'єкти інфраструктури  $I$  взаємодіють з різними типами даних  $D$  через застосування відповідних методів інтелектуального аналізу  $A$ , що дозволяє вирішувати різноманітні завдання аналізу  $Z$  та приймати обґрунтовані рішення  $R$ .

Таким чином, формалізована модель інтелектуального аналізу соціально–економічних даних для інфраструктури ТГ включає визначення об’єктів інфраструктури, класифікацію типів даних та відповідних підходів до їх аналізу. Ця модель забезпечує структурований підхід до інтеграції та обробки даних, сприяючи ефективному управлінню ресурсами та підтримці прийняття стратегічних рішень у контексті соціально–економічного розвитку ТГ. Результати, опубліковані в наукових роботах [457, 459].

## **2.2   Методологія інтелектуального аналізу соціально–економічних даних територіальних громад**

У сучасних умовах стрімкого розвитку інформаційних технологій та цифрової трансформації суспільства ТГ стикаються з викликами, що потребують механізмів управління ресурсами, підтримки соціальної стійкості та розвитку підприємницької діяльності. Запропонована методологія (Рис. 2.2) спрямована на вирішення цих завдань через інтеграцію передових підходів МН, адаптивних предикторів та аналітичних інструментів. Особлива увага приділяється формуванню нових методів для підвищення точності прогнозування, адаптивності управлінських рішень та оптимізації процесів у динамічному середовищі громад.

Методологія аналізу соціально–економічних даних ТГ базується на таких ключових принципах (Таблиця 2.1), що базуються на визначених протиріччях (див. розділ Вступ), які забезпечують її гнучкість, адаптивність та точність у процесі аналізу даних і прийняття управлінських рішень. Ці принципи реалізуються через послідовні етапи методології, кожен з яких враховує специфіку даних та аналітичних завдань. Усі етапи орієнтовані на досягнення високої точності та адаптивності в управлінні соціально–економічною інфраструктурою громади.

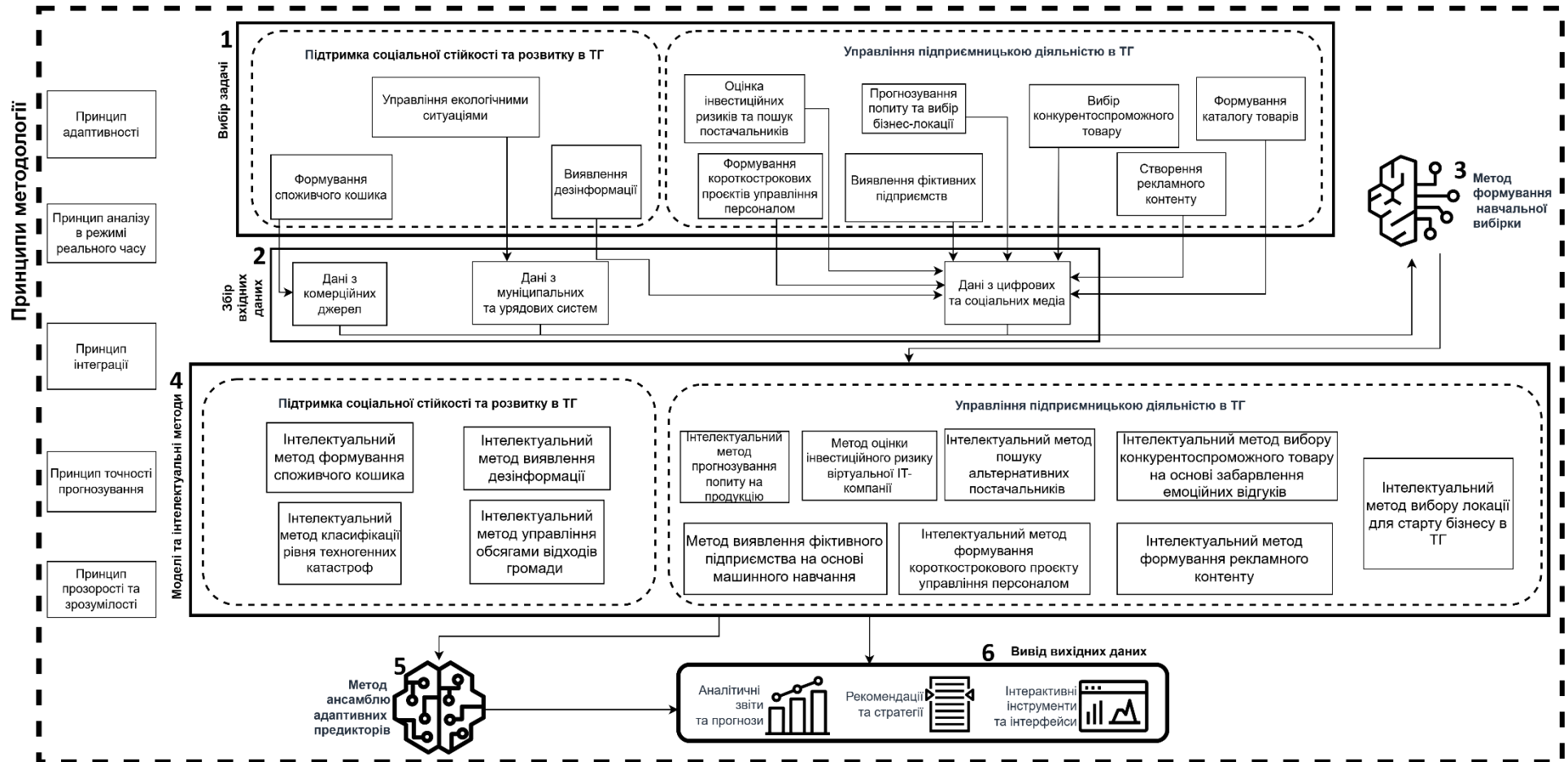


Рис. 2.2. Концепція методології аналізу соціально-економічних даних ТГ

Таблиця 2.1

Принципи методології як відповіді на протиріччя в аналізі соціально–економічних даних ТГ

Протиріччя	Принцип методології	Опис принципу
Стрімке зростання обсягів соціально–економічних даних з високим ступенем гетерогенності, нестационарності та динамічності, що ускладнює їх аналіз традиційними методами.	Принцип адаптивності	Забезпечує можливість методів та алгоритмів адаптуватися до швидких змін у соціально–економічних показниках, що дозволяє зберігати стабільну точність та надійність результатів аналізу.
Відсутність технологій, здатних забезпечити обробку даних у режимі реального часу для швидкого прийняття управлінських рішень, при зростаючій необхідності в оперативному аналізі для забезпечення стійкого розвитку ТГ.	Принцип аналізу в режимі реального часу	Охоплює розробку методів, здатних обробляти дані в реальному часі, що дозволяє швидко реагувати на зміни та приймати управлінські рішення.
Необхідність інтеграції різнорідних джерел даних для створення повної картини соціально–економічної ситуації.	Принцип інтеграції	Передбачає об'єднання різнорідних джерел даних (муніципальні, комерційні, медійні) в єдину систему для отримання повнішої картини ситуації та підвищення точності управлінських рішень.
Потреба в точних прогнозах для управління в умовах динамічних змін.	Принцип точності прогнозування	Використання передових методів МН, зокрема ансамблевих підходів, які поєднують декілька моделей для підвищення точності прогнозів.
Складність існуючих методів аналізу та прогнозування, що ускладнює інтерпретацію результатів для управлінців та інших зацікавлених сторін.	Принцип прозорості та зрозумілості	Забезпечує зрозумілу та інтуїтивно доступну форму представлення результатів, що сприяє обґрунтованому прийняттю рішень зацікавленими сторонами.



Розроблена методологію (див. Рис. 2.2) аналізу соціально–економічних даних ТГ побудована на основі кількох послідовних етапів:

Етап 1. Вибір задачі аналізу. На першому етапі визначається конкретне завдання, що підлягає вирішенню. Залежно від напрямку дослідження, задачі можуть належати до трьох основних категорій:

1) підтримка соціальної стійкості та розвитку ТГ охоплює комплексні дослідження, спрямовані на формування споживчого кошика, оцінку якості життя в містах, підтримку туристичного попиту та збереження культурної спадщини, управління екологічними ситуаціями, інтерпретацію правової інформації та виявлення дезінформації;

2) управління підприємницькою діяльністю в ТГ охоплює такі завдання, як: оцінка інвестиційних ризиків і пошук постачальників, прогнозування попиту та вибір оптимальних бізнес–локацій, формування короткострокових проєктів з управління персоналом, а також виявлення фіктивних підприємств, вибір конкурентоспроможного товару, формування каталогу товарів для інтернет–магазинів та створення рекламного контенту.

Етап 2. Збір даних. Після вибору задачі проводиться збір релевантних даних. Для кожної категорії задач використовуються відповідні джерела:

1) дані з комерційних джерел, наприклад, для формування споживчого кошика;

2) дані з муніципальних та урядових систем для задач управління підприємницькою діяльністю;

3) дані з цифрових та соціальних медіа, що особливо важливо для задач просування товарів та виявлення дезінформації.

Етап 3. Удосконалений метод формування навчальної вибірки. На цьому етапі здійснюється формування навчальних вибірок на основі методів МН з метою підвищення точності аналізу. Застосування методів сегментації та кластеризації дозволяє оптимізувати вибірку для різних категорій задач.

Етап 4. Застосування інтелектуальних методів. Цей етап є ключовим і безпосередньо залежить від вибору задачі на першому етапі. Для кожного типу

задач застосовуються відповідні інтелектуальні методи, що забезпечують аналіз:

1) для підтримки соціальної стійкості та розвитку використовуються методи оцінки якості життя, управління екологічними ситуаціями, підтримки культурної спадщини тощо;

2) для управління підприємницькою діяльністю застосовуються методи, оцінки інвестиційних ризиків, прогнозування попиту та виявлення фіктивних підприємств, також методи вибору конкурентоспроможних товарів, створення рекламного контенту та формування каталогу товарів.

Етап 5. Метод ансамблевих адаптивних предикторів. Якщо результати аналізу не досягають необхідної точності, на цьому етапі застосовується метод ансамблевих адаптивних предикторів. Цей метод забезпечує підвищення точності за рахунок поєднання прогнозів різних моделей та їх адаптивного налаштування в реальному часі.

Етап 6. Аналіз та виведення результатів. На завершальному етапі результати аналізу передаються у вигляді аналітичних звітів, прогнозів, рекомендацій та інтерактивних інструментів для подальшого використання в управлінні ТГ.

Методологія забезпечує системний підхід до аналізу соціально–економічних даних, дозволяючи адаптувати процес аналізу до специфічних умов ТГ, підвищуючи прийняття рішень та оптимізуючи управління ресурсами.

Результати, опубліковані в наукових роботах [208, 239, 242, 245, 246, 250, 251, 254 – 257, 346, 455, 458], підтверджують ефективність запропонованої методології аналізу соціально–економічних даних ТГ, яка забезпечує підвищення точності прогнозування, адаптивність аналізу в динамічних умовах, інтеграцію різномірних джерел даних і зрозуміле представлення результатів для прийняття управлінських рішень. Інтеграція передових підходів машинного навчання, адаптивних предикторів і аналітичних інструментів дозволяє не лише підвищити ефективність управління ресурсами та підтримки соціальної стійкості, але й оптимізувати процеси прийняття рішень, враховуючи специфіку динамічного середовища громад.

### 2.3 Дослідження даних в управлінні територіальними громадами на стаціонарність

Оцінка стаціонарності соціально–економічних даних ТГ є ключовою для розробки політик і стратегій управління, особливо у вимірах планування розвитку, розподілу ресурсів та кризового відновлення.

Стаціонарність даних вказує на те, що статистичні характеристики часових рядів, такі як середнє значення та дисперсія, залишаються постійними з часом, що дозволяє передбачати майбутні економічні та соціальні тенденції на основі історичних даних. Такий підхід забезпечує можливість обґрунтовано планувати бюджет, алокувати інвестиції та розробляти програми соціального захисту, виходячи з реальної динаміки змін у громаді.

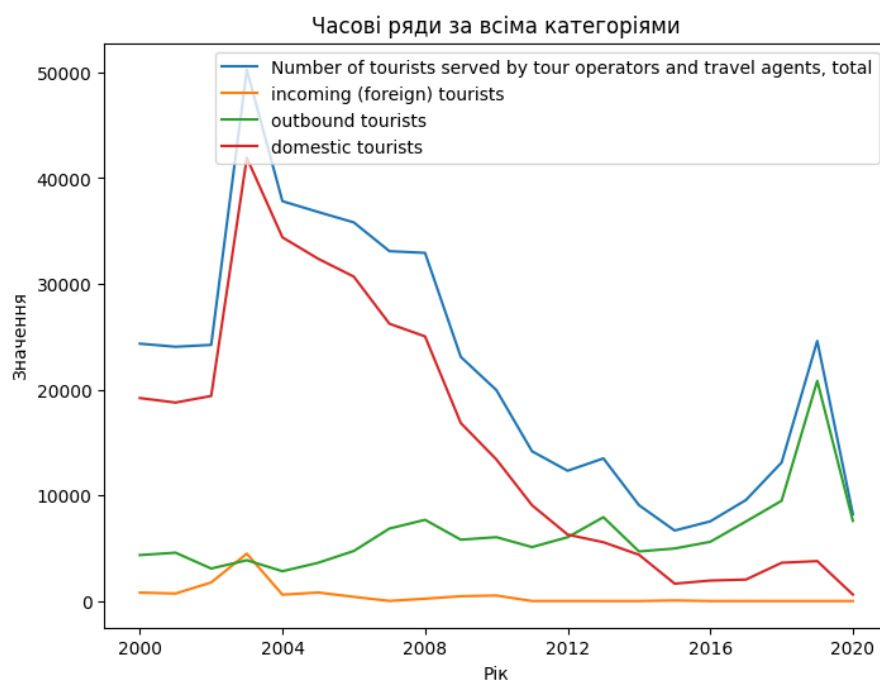
Нестаціонарність може вказувати на наявність структурних зрушень або зовнішніх шоків, які потребують адаптації управлінських рішень, тому аналіз стаціонарності слугує важливим інструментом для виявлення та прогнозування критичних змін у життєдіяльності ТГ.

Подальший аналіз включає проведення оцінки стаціонарності даних (див.підрозділ 1.3), що дозволяє визначити їхні характеристики та забезпечити застосування аналітичних інструментів для вирішення проблем, пов'язаних зі змінами у соціально–економічному середовищі ТГ.

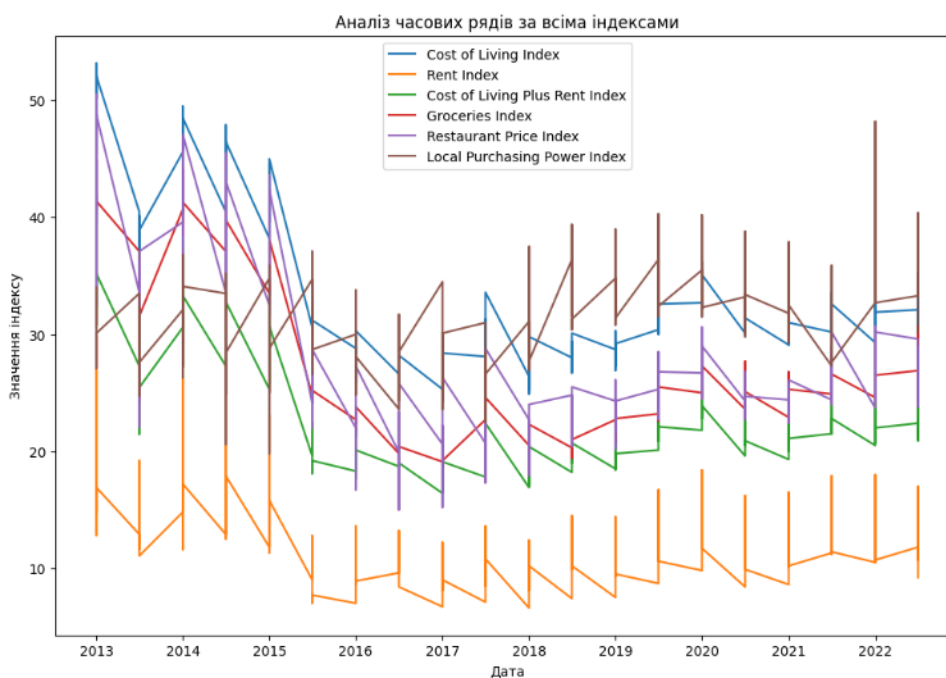
Аналіз даних соціальної стійкості у ТГ включає розгляд туристичної активності та індексів якості життя (Рис. 2.3 та табл.2.2). За даними на графіках (Рис. 2.3, *а*), чисельність туристів, обслуговуваних туроператорами, показує значне зниження після 2000 року з подальшим коливанням до 2020, яке відображено у статистиці ADF, де  $p$ -значення для цих даних (0,595) індикує їх нестаціонарність.

Аналогічно, індекси якості життя (Рис. 2.3, *б*), зокрема індекси орендної плати та місцевої купівельної спроможності, продемонстрували значні відхилення від стаціонарності, як підтверджено низькими  $p$ -значеннями та значеннями ADF, що

перевищують критичні пороги. Це вказує на значні зміни у витратах та економічних умовах, що впливають на якість життя населення в різних ТГ.



а) Розподіл туристичних даних



б) Розподіл даних за якістю життя населення

Рис. 2.3. Розподіл даних соціальної стійкості в ТГ

## Тестування на стаціонарність

Задача	Параметр	Статистика ADF	p-значення	Критичне значення 1 %	Критичне значення 5 %	Критичне значення 10 %
Соціальної стійкості	Number of tourists served by tour operators	-1,372	0,595	-3,809	-3,022	-2,651
	incoming (foreign) tourists	-2,754	0,065	-3,809	-3,022	-2,651
	outbound tourists	-2,849	0,052	-3,809	-3,022	-2,651
	domestic tourists	-0,741	0,836	-3,809	-3,022	-2,651
	Cost of Living Index	-3,192	0,020	-3,499	-2,892	-2,583
	Rent Index	-7,858	5,39E-12	-3,499	-2,892	-2,583
	Cost of Living Plus Rent Index	-4,364	3,43E-04	-3,499	-2,892	-2,583
	Groceries Index	-2,840	0,053	-3,499	-2,892	-2,583
	Restaurant Price Index	-4,979	2,44E-05	-3,499	-2,892	-2,583
	Local Purchasing Power Index	-9,759	7,64E-17	-3,499	-2,892	-2,583
Дохідність компанії	Offline_Spend	0,975	4,002	-3,448	-2,870	-2,571
	Online_Spend	0,777	4,006	-3,449	-2,870	-2,571
Інтернет ринок	Coverage	-2,589	1,003	-3,438	-2,865	-2,569
	Impressions	-3,434	1,020	-3,438	-2,865	-2,569
	Frequency	-1,241	2,36E-30	-3,438	-2,865	-2,569
	Total Spending (UAH)	-1,648	2,05E-30	-3,438	-2,865	-2,569
	CPM (Cost per 1000 Impressions)	-17,921	2,91E-30	-3,438	-2,865	-2,569
	Positive reviews	-0,647	1,002	-3,430	-2,862	-2,567
	Negative reviews	-0,801	1,008	-3,430	-2,862	-2,567

Розподіл витрат (Рис. 2.4) у компаніях між офлайн– і онлайн–режимами, демонструючи значні коливання у витратах з часом, які свідчать про потенційні сезонні впливи або відповіді на ринкові зміни.

Відповідно до табл. 2.2, тестування на стаціонарність показало, що обидва часові ряди: `Offline_Spend` та `Online_Spend` мають високі р–значення (4,002 і 4,0062, відповідно), значно перевищуючи загальноприйняті порогові рівні для значимості. Статистика ADF також є позитивною для обох рядів, що значно відрізняється від негативних критичних значень на 1 %, 5 % та 10 %. Це свідчить про наявність одиничного кореня у часових рядах, що підтверджує їх нестационарність.

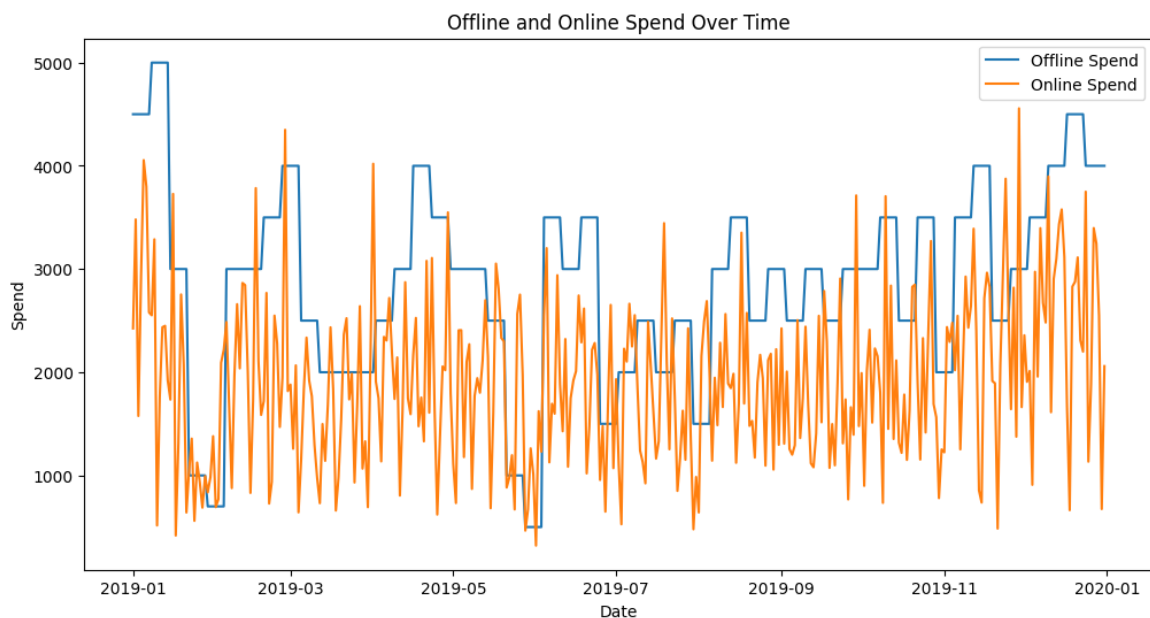
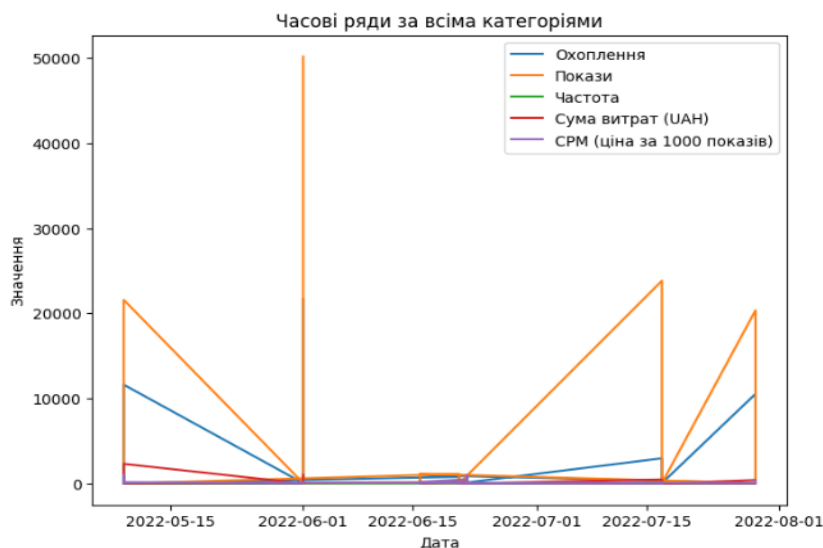


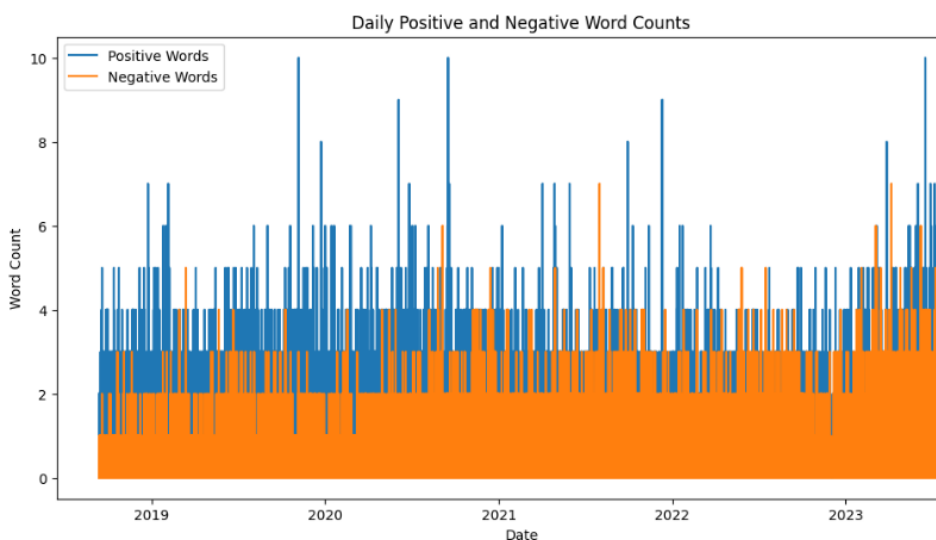
Рис. 2.4. Розподіл даних за дохідністю компаній

Рис. 2.5 ілюструє динаміку інтернет–реклами та розподілу відгуків з часом. За даними часових рядів, видно значні коливання у показниках, таких як охоплення (Coverage), покази (Impressions), частота (Frequency), загальні витрати (Total Spending) та вартість на 1000 показів (CPM). Тестування (табл.2.2) на стаціонарність для цих параметрів показує відсутність стаціонарності у більшості з них, оскільки р–значення значно вищі за критичні пороги, окрім CPM, де величина ADF (–17,921) та надзвичайно низьке р–значення (2,91E–30) свідчать про стаціонарність цього

параметра. Однак, для показників, таких як позитивні та негативні відгуки, високі  $p$ -значення та малі значення  $ADF$  також підтверджують їхню нестационарність. Ці результати вказують на складність прогнозування майбутніх трендів за цими показниками через їх високу залежність від зовнішніх умов та нестабільності у часі.



#### а) Інтернет-реклама



#### б) Розподіл відгуків

Рис. 2.5. Розподіл даних з інтернет-ринку

Дослідження стаціонарності даних у рамках управління ТГ виявляє значущі відмінності в стабільності різних часових рядів, що важливо для прийняття

обґрунтованих управлінських рішень. Аналіз, виконаний на основі табл 2.1, показує, що більшість часових рядів зазнають впливу зовнішніх факторів і є нестационарними, оскільки їхні  $p$ -значення перевищують порогові значення  $\alpha$  статистика ADF є недостатньо низькою, щоб відкинути гіпотезу про наявність одиничного кореня.

Зокрема, статистика ADF для кількості туристів, які обслуговуються туроператорами, показує значення  $-1,372$  з  $p$ -значенням  $0,595$ , що значно перевищує критичні значення на рівні  $1\%$  ( $-3,809$ ),  $5\%$  ( $-3,022$ ) та  $10\%$  ( $-2,651$ ), засвідчуючи нестационарність ряду. У випадку даних щодо індексу вартості життя (Cost of Living Index), хоча статистика ADF дорівнює  $-3,192$ , що ближче до порогових значень, ряди все ще залишаються під впливом зовнішніх чинників. Аналіз також показав, що часові ряди, які зазнають впливу зовнішніх факторів, потребують подальшого дослідження і регулярної корекції прогнозів для забезпечення більшої точності в управлінських рішеннях.

Оцінка стаціонарності в часових рядах соціально-економічних даних ТГ відкриває шлях до глибшого розуміння динаміки змін у громадах і є важливим етапом для впровадження управлінських стратегій. На базі зібраних даних і аналізу їх нестационарності, в підрозділі 2.5 буде розглянуто метод ансамблю адаптивних прогнозних моделей, який дозволяє працювати з багатовимірними нестационарними послідовностями у реальному часі, забезпечуючи тим самим можливість швидкого реагування на економічні та соціальні виклики, що постають перед ТГ.

## **2.4 Метод формування навчальної вибірки для нестационарних даних на основі RFM та кластерного аналізу**

Метод формування навчальної вибірки для нестационарних даних на основі RFM та кластерного аналізу дозволяє більш точно відбирати інформативні зразки, порівнюючи їх з аналогами, що було доведено в [187, 201, 322, 323]. Такий підхід базується на методах RFM-аналізу і кластерного аналізу, що групують схожі об'єкти за певними характеристиками. Це дозволяє підібрати найбільш репрезентативних



представників різних груп даних та їх характеристики для формування навчальної вибірки.

Застосування такого методу в контексті аналізу та обробки даних може значно покращити результати навчання моделей, зменшити вплив шуму та незначущих даних, а також зробити процес навчання більш швидким та точними. Відмінною особливістю цього підходу є його адаптивність, тобто здатність адекватно реагувати на зміни в навчальних даних і змінювати склад навчальної вибірки відповідно до нової інформації.

Удосконалено метод формування навчальної вибірки для нестационарних даних на основі RFM-аналізу та кластерного аналізу. Запропонований метод ілюструється схематично (Рис. 2.6) та представлений такими кроками:

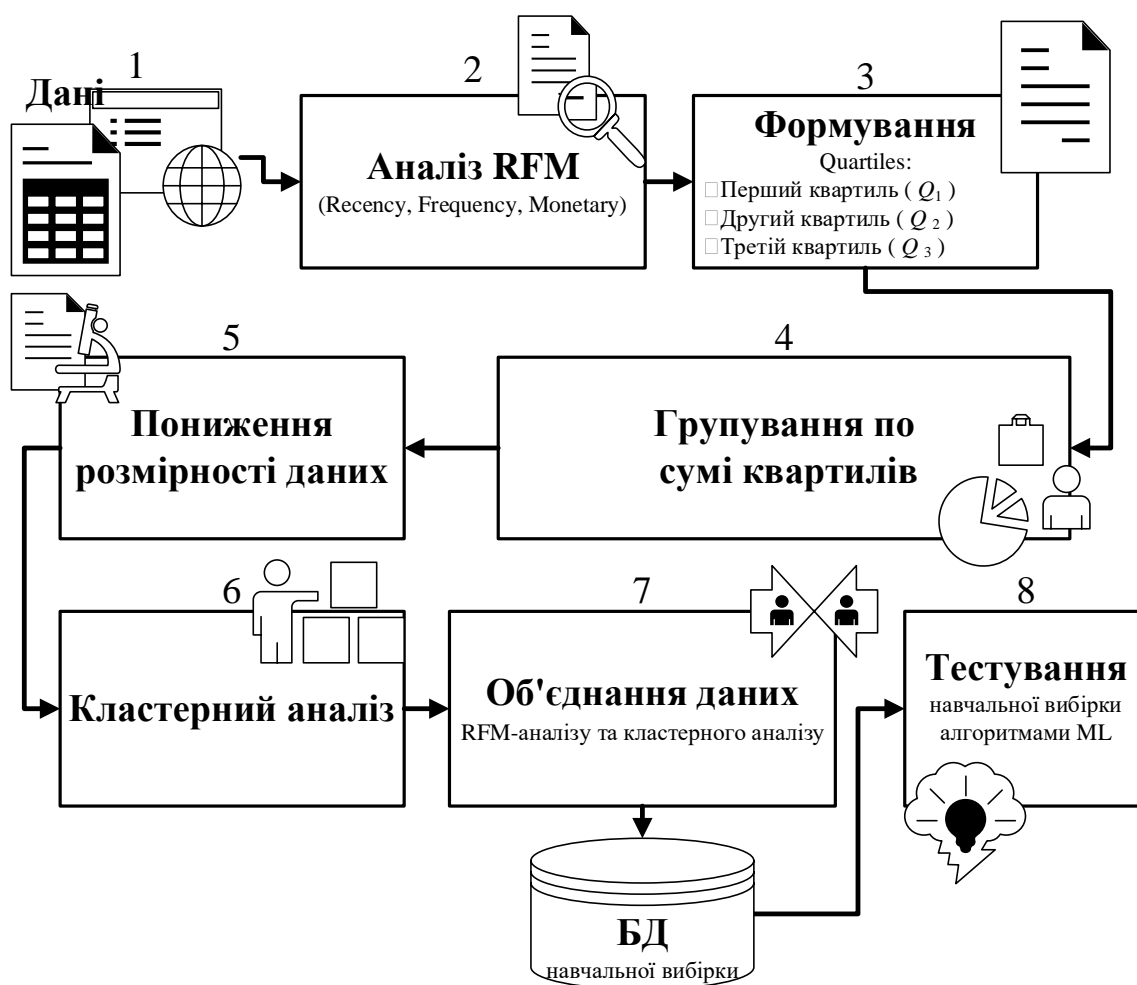


Рис. 2.6. Структура методу формування навчальної вибірки

Крок 1. Ввід вхідних даних (блок 1). На першому етапі здійснюється збирання та підготовка вихідних даних, які будуть використані для аналізу та моделювання. Це можуть бути транзакційні дані, демографічна інформація та інші релевантні ознаки. Нехай маємо набір даних:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, \quad (2.10)$$

де  $x_i$  – вектор ознак для  $i$ -го об'єкта,  $y_i$  – відповідна мітка або цільова змінна.

Крок 2. RFM-аналіз (блок 2). Аналіз RFM [27] (Recency, Frequency, Monetary) – це техніка сегментації даних. RFM-аналіз (Recency, Frequency, Monetary) використовується для сегментації об'єктів на основі їх поведінки. Для кожного об'єкта розраховуються:

– Recency (R) час з моменту останньої взаємодії,  $R_i = CurrentDate - LastInteraction_i$ .

– Frequency (F) кількість взаємодій за певний період,

$$F_i = \sum_{t \in T} \delta_i(t), \quad (2.11)$$

де  $\delta_i(t) = 1$ , якщо була взаємодія в момент часу  $t$ , і 0 якщо ні.

– Monetary (M) загальна сума витрат,

$$M_i = \sum_{k=1}^{N_i} amount_{ik}, \quad (2.12)$$

де  $N_i$  – кількість транзакцій клієнта  $i$ ,  $amount_{ik}$  – сума  $k$  транзакції.

Таким чином, для кожного клієнта формується вектор

$$RFM_i = (R_i, F_i, M_i). \quad (2.13)$$

Крок 3. Формування кватилів (блок 3). Кватилі ділять число точок даних на чотири частини, або чверті, більш-менш однакового розміру. Дані повинні бути впорядковані від найменшого до найбільшого для обчислення кватилів. Три основні кватилі такі:

– перший кватиль ( $Q_1$ ) визначається як середнє число між найменшим числом (мінімальним) та медіаною набору даних. Він також відомий як *нижчий* або 25-й

емпіричний кuartиль, оскільки 25 % даних знаходиться нижче цієї точки. Формально, для впорядкованого набору даних  $\{x_1, x_2, \dots, x_n\}$ :

$$Q1 = x_{\left(\frac{n+1}{4}\right)}, \quad (2.14)$$

де  $x_i$  –  $i$ -те впорядковане значення.

–другий кuartиль ( $Q_2$ ) є медіаною набору даних; таким чином, 50% даних лежить нижче цієї точки:

$$Q2 = x_{\left(\frac{n+1}{2}\right)}, \quad (2.15)$$

–третій кuartиль ( $Q_3$ ) – це середнє значення між медіаною та найвищим значенням (максимумом) набору даних. Він відомий як *верхній* або *75-й емпіричний* кuartиль, оскільки 75% даних лежить нижче цієї точки:

$$Q3 = x_{\left(\frac{3(n+1)}{4}\right)}. \quad (2.16)$$

Присвоюємо оцінки від 1 до 4 для кожного показника R, F та M залежно від того, в який кuartиль потрапляє значення клієнта:

- якщо значення  $x_i \leq Q1$ , то оцінка = 1;
- якщо  $Q1 < x_i \leq Q2$ , то оцінка = 2;
- якщо  $Q2 < x_i \leq Q3$ , то оцінка = 3;
- якщо  $x_i > Q3$ , то оцінка = 4.

Крок 4. Групування за сумою кuartилів (блок 4). Сумування рангів R, F та M для кожного об'єкта дає можливість створити комплексний показник цінності клієнта:

$$Score_i = R_i^{(rank)} + F_i^{(rank)} + M_i^{(rank)}, \quad (2.17)$$

де  $R_i^{(rank)}, F_i^{(rank)}, M_i^{(rank)} \in \{1,2,3,4\}$  – рангові оцінки. Далі об'єкти групуються за значеннями  $Score_i$ , що дозволяє виділити різні сегменти клієнтів з подібними характеристиками.

Крок 5. Пониження розмірності даних (блок 5). Зниження розмірності є процесом скорочення кількості випадкових змінних шляхом отримання множини

головних змінних. Для зменшення складності моделі та уникнення проблем, пов'язаних з високою розмірністю, використовується метод головних компонент (PCA).

Стандартизація даних: центрування та масштабування ознак за формулами:

$$x'_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}, \quad (2.18)$$

де  $\mu_j$  – середнє значення ознаки  $j$ ,  $\sigma_j$  – стандартне відхилення ознаки  $j$ .

Обчислення коваріаційної матриці:

$$\Sigma = \frac{1}{n-1} (X')^T X' \quad (2.19)$$

де  $X$  – матриця стандартизованих ознак.

Визначення власних векторів та власних значень:

$$\Sigma w_k = \lambda_k w_k, \quad (2.20)$$

де  $\lambda_k$  – власні значення,  $w_k$  – відповідні власні вектори.

Вибір кількості головних компонент  $k$ , що пояснюють достатню частку дисперсії  $\alpha\%$ :

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{j=1}^m \lambda_j} \geq \frac{\alpha}{100} \quad (2.21)$$

Цей підхід до зниження розмірності дозволяє зберегти найбільш інформативні ознаки, що сприяє покращенню навчання моделей.

Крок 6. Кластеризація (блок 6). Кластерний аналіз [229] – задача розбиття заданої вибірки об'єктів (ситуацій) на підмножини, які називаються кластерами, так, щоб кожен кластер складався зі схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися.

З метою отримання більш чіткої та якісної сегментації використовується комбінований підхід до кластеризації. Першим етапом цього підходу є визначення оптимального числа кластерів  $K$ . Для цього застосовуються метод «лікоть» (elbow method) або силуетний аналіз (silhouette analysis). Метод «лікоть» передбачає побудову графіка залежності суми квадратів відстаней точок до найближчих центрів

кластерів від числа  $K$ . Оптимальне значення  $K$  визначається як точка, де спостерігається суттєве зменшення темпу зменшення суми квадратів, утворюючи «лікоть» на графіку. Силуетний аналіз оцінює якість кластеризації за допомогою коефіцієнта силуету  $s(i)$ , який вимірює, наскільки добре кожна точка належить до свого кластера порівняно з найближчим сусіднім кластером. Формула для коефіцієнта силуету виглядає таким чином:

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}, \quad (2.22)$$

де  $a_i$  – середня відстань від точки  $i$  до інших точок свого кластера;  $b_i$  – мінімальна середня відстань від точки  $i$  до точок найближчого іншого кластера.

Значення  $s_i$ , близьке до 1, свідчить про добре класифіковані точки, тоді як значення близьке до  $-1$  вказує на можливі помилки кластеризації.

Наступним кроком є застосування алгоритму  $k$ –середніх з ініціалізацією  $K$ –Means, який покращує процес вибору початкових центрів кластерів для підвищення швидкості та якості збіжності. Процедура включає кілька підетапів. Спочатку ініціалізуються початкові центри кластерів  $\mu_k$  з урахуванням відстаней між точками, що знижує ймовірність потрапляння в локальні мінімуми. Далі алгоритм виконує ітераційний процес, що складається з двох основних кроків: призначення точок до найближчих центрів кластерів та оновлення центрів кластерів. Призначення точок здійснюється за формулою:

$$C_i = \left\{ x_j \mid \arg \min_k \| x_j - \mu_k \|^2 = i \right\}, \quad (2.23)$$

де  $C_i$  – множина об'єктів, призначених до кластера  $i$ ;  $\mu_k$  – центр кластера  $k$ .

Після призначення точок до кластерів, центри оновлюються за формулою:

$$\mu_i = \frac{1}{|C_i|} \sum_{x_j \in C_i} x_j, \quad (2.24)$$

де  $|C_i|$  – кількість точок у кластері  $i$ .

Цей процес повторюється, поки центри кластерів не перестануть суттєво змінюватися, тобто досягнуть збіжності. Останнім етапом є оцінка якості кластеризації за допомогою коефіцієнта силуету  $s_i$ , що дозволяє визначити, наскільки

точно точки були розподілені по кластерах. Використання цього комбінованого підходу значно підвищує якість сегментації, забезпечуючи більш однорідні та добре розмежовані кластери, що є критичним для подальшого аналізу та прийняття бізнес-рішень.

Крок 7. Об'єднання (блок 7) даних RFM-аналізу та кластерного аналізу та внесення в БД як навчальну вибірку. Результати RFM-аналізу та кластеризації комбінуються для формування збагаченої навчальної вибірки:

- формування розширених векторів ознак: об'єднуються рангові оцінки RFM та координати в просторах головних компонент.

- додавання міток кластерів: кожному об'єкту призначається мітка кластера  $y_i$ , що використовується як цільова змінна для моделей класифікації.

- збереження даних: сформована навчальна вибірка  $\{(x_i^{extended}, y_i)\}$  на основі сильних кластерів та вноситься до бази даних для подальшого використання.

Цей крок удосконалює навчальну вибірку, роблячи її більш інформативною та придатною для навчання моделей.

Крок 8. Тестування (блок 8) навчальної вибірки на основі алгоритмів класифікації МН та оцінки їх якості.

Для підтвердження розробленого методу формування навчальної вибірки на основі МН проведено його реалізацію для сегментації організаторів тендерів.

Для формування навчальної вибірки для сегментації організаторів тендерів на основі МН обрано мову Python. При цьому використано такі бібліотеки: pandas, numpy, train\_test\_split, K-Means, PCA.

В якості вхідних даних використано завершені угоди учасників тендерів в Україні з сайту ProZorro Продажі [328]. Вибірка після очищення становить 93336 значень відносно 10 параметрів. При оцінці кількісних показників виявлено 92638 аукціонів, 29164 – унікальних аукціонів та унікальних організаторів – 39747.

На основі удосконаленого методу трендів (Рис. 2.7) сформовано навчальну вибірку на основі RTF-оцінки та K-Means. На основі цих даних можна проводити передбачення груп за допомогою методів машинного навчання. Для цього

використано метод Logistic Regression [215] та Naive Bayes [216], адже ці методи мають найпростішу логіку кваліфікації та хороші результати оцінки моделювання.

ID_Orzanizer	Recency	Frequency	Monetary	Cluster	RFMScore
1551	30.0	1.0	3060000.0	3	413
695	309.0	1.0	2598260.0	3	413
954	1.0	1.0	1401975.0	4	413
598	309.0	1.0	1140000.0	3	413
579	30.0	1.0	1070856.0	3	413
...	...	...	...	...	...
1349	317.0	1.0	43.0	0	411
1754	30.0	1.0	36.0	0	411
1214	317.0	1.0	32.0	0	411
300	290.0	1.0	24.0	0	411
3	154.0	1.0	7.0	0	411

Рис. 2.7. Результат кластеризації RTF–оцінки та K–Means

Для навчання взято 70% вибірки. Проведено навчання алгоритмами Logistic Regression та Naive Bayes. Після проведення тестування, результатами оцінювання, для обох методів:

Train Set Accuracy for Power Transformed Data: 99.8 %

Test Set Accuracy for Power Transformed Data: 99.1 %

Bias Error: 0.9

Variance Error: 0.2

Результати моделювання, показують, що RTF–оцінки та K–Means дають 100–відсоткову точність групування, відповідно, на цих даних в подальшому проводити класифікацію організаторів тендерних проектів, що дає можливість визначати більш привабливих організаторів тендерів.

Отже, вибірка містить 92638 аукціонів, 29164 – унікальних аукціонів та унікальних організаторів – 39747. На основі RFM–аналізу сформовано такі групи: The

best organizers of tenders – 119; Loyal organizers of tenders – 385; Large consumers – 455; Tenders are seldom held, but for a large sum – 15; Weak tender organizers – 48. На основі кластеризації методом K–Means віднесена така кількість значень кластера: під номером 0 – 494; під номером 3 – 475; під номером 2 – 352; під номером 1 – 345; під номером 4 – 155. Після класифікації тендерів на основі створених груп за допомогою алгоритмів Logistic Regression та Naive Bayes, результати оцінки для обох методів склали: точність на тренувальній вибірці для даних з перетворенням ступеня – 99,8 %, а точність на тестовій вибірці – 99,1 %.

Результати, опубліковані в наукових роботах [264, 466], підтверджують ефективність методу формування навчальної вибірки для нестационарних даних на основі RFM–аналізу та кластерного аналізу. Застосування цього методу дозволяє покращити якість навчальних вибірок, підвищити точність та надійність прогнозування, зменшити вплив шуму в даних і адаптивно реагувати на зміни у вхідних даних. Метод демонструє високу ефективність у контексті сегментації, кластеризації та подальшої класифікації, що забезпечує покращення процесів аналізу великих масивів соціально–економічних даних.

## **2.5 Метод ансамблю адаптивних прогнозних моделей для багатовимірних нестационарних послідовностей в реальному часі**

Прогнозування багатовимірних нестационарних сигналів є складною задачею, особливо в умовах змінних умов та невизначеності. Традиційні моделі прогнозування можуть не забезпечувати достатню точність і надійність прогнозів в таких умовах. Крім того, використання окремих моделей може приводити до переоснащення або недооснащення, залежно від характеристик сигналу та вибраних параметрів моделі.

Метамоделі та ансамбль методів є способами покращення точності прогнозування, однак відсутність належного врахування нестационарності сигналів та недостатнє використання інформації з різних прогнозуючих моделей можуть створювати обмеження для таких підходів.



Метамоделю другого рівня, яка базується на ансамблі прогнозів від прогнозуючих моделей першого рівня, може допомогти збалансувати прогнози та забезпечити більш точний і надійний прогноз на основі декількох моделей.

Найбільше поширення цей підхід набув у задачах класифікації – розпізнавання образів, де досить популярним є так званий Boost-алгоритм та його численні модифікації [116, 120] (див. підрозділ 1.3). В його основі лежить ідея стекового узагальнення, коли результати кожного з членів ансамблю (стека) деяким чином об'єднуються в рамках метамоделі, параметри якої налаштовуються за допомогою тієї або іншої процедури метанавчання. Зазвичай це є зваженим усередненням, коли кожному з членів ансамблю (комітету) призначається деяка вага, отримана шляхом оптимізації прийнятого критерію навчання.

В основі бустингу лежать ідеї Байєсівського оцінювання, логістичної регресії, машин опорних векторів. Цікаво відмітити, що саме ці ідеї полягають і в основі низки штучних нейронних мереж, де також використовуються ідеї ансамблевого підходу [268] для отримання (побудови) оптимальних прогнозів. Тут ваги для кожного з членів ансамблю оцінюються за допомогою оптимізаційної процедури, що реалізується у пакетному режимі (batch mode) так, що використання відомих підходів для розв'язання задач Data Stream Mining в принципі неможливе. У роботах [58, 57] були введені рекурентні процедури налаштування параметрів метамоделі, що узагальнюють вихідні сигнали нейромереж предикторів на основі оптимізації стандартного критерію найменших квадратів за наявності тих або інших обмежень. І хоча введені процедури призначені для оцінювання в онлайн-режимі, вони не пристосовані для роботи з нестационарними часовими рядами, параметри яких змінюються непередбачуваним чином у довільні моменти часу.

Доцільним ввести у розгляд адаптивні рекурентні процедури метанавчання узагальнюючої метамоделі, що комбінує вихідні сигнали ансамблю нейропредикторів, кожен з яких може мати власну архітектуру та власний алгоритм налаштування-навчання своїх параметрів-синаптичних ваг.

Результати, опубліковані в науковій роботі [56], підтверджують ефективність запропонованого методу ансамблю адаптивних прогнозних моделей для багатовимірних нестационарних послідовностей в реальному часі.

### 2.5.1 Теоретичний базис ансамблю адаптивних прогнозних моделей для багатовимірного аналізу

Введемо у розгляд ансамбль багатовимірних предикторів  $MP_1, \dots, MP_j, \dots, MP_h$ , кожен з яких опрацьовує один і той самий багатовимірний сигнал  $x(\tau) = (x_1(\tau) \dots x_i(\tau) \dots)^T, \tau = 1, 2, \dots, T$ . При цьому оцінку, що з'являється на виході кожного з членів ансамблю, будемо позначати як  $\hat{x}_j(\tau), j = 1, 2 \dots h$ . Помітимо, що в якості членів ансамблю можуть використовуватися як традиційні моделі прогнозування на основі багатовимірної регресії, експоненційного згладжування. ARIMAX<sub>s</sub> – MIMO моделі Бокса–Дженкінса, так і нелінійні структури на основі штучних нейронних мереж від найпростіших дрібних рекурентних мереж до глибоких типу LSTM або трансформерів.

Оцінки  $\hat{x}_j(\tau)$  надходять на вхід метамоделі, що формує об'єднаний прогноз метамоделі у вигляді:

$$x^*(\tau) = \sum_{j=1}^h c_j \hat{x}_j(\tau) = \hat{x}(\tau)c \quad (2.25)$$

де  $c = (c_1, \dots, c_j, \dots, c_h)^T$ ,  $\hat{x}(\tau) = (\hat{x}_1(\tau), \dots, \hat{x}_j(\tau), \dots, \hat{x}_h(\tau))$  –  $(n \times h)$  – матриця, що сформована сигналами у виходів окремих моделей, де параметри метамоделі задовольняють умові незміщеності:

$$\sum_{j=1}^h c_j = c^T E_h = 1, \quad (2.26)$$

де  $(E_h - (h \times 1))$  – вектор, утворений одиницями).

Параметри метамоделі (вектор оцінок  $c$ ) може бути визначений за допомогою класичного методу невизначених множників Лагранжа, для чого у розгляд вводиться функція Лагранжа:

$$L(c, \lambda) = Sp(V^T(T)V(T)) + \lambda(c^T E_h - 1) =$$

$$\begin{aligned}
&= Sp(X(T) - \hat{X}(T)E_{nn} \otimes c)^T (X(T) - \hat{X}(T)E_{nn} \otimes c) + \\
&\quad + \lambda(c^T E_h - 1) = \\
&= \sum_{\tau=1}^T \|x(\tau) - \hat{x}(\tau)C\|^2 + \alpha(C^T E_h - 1),
\end{aligned} \tag{2.27}$$

де

$$x(T) = \begin{pmatrix} x^T(1) \\ \dots \\ x^T(\tau) \\ \dots \\ x^T(T) \end{pmatrix}, \tag{2.28}$$

$$\hat{x}(T) = \begin{pmatrix} \hat{x}_1^T(1) & \hat{x}_2^T(1) & \dots & \hat{x}_h^T(1) \\ \dots & \dots & \dots & \dots \\ \hat{x}_1^T(\tau) & \hat{x}_2^T(\tau) & \dots & \hat{x}_h^T(\tau) \\ \dots & \dots & \dots & \dots \\ \hat{x}_1^T(T) & \hat{x}_2^T(T) & \dots & \hat{x}_h^T(T) \end{pmatrix}, \tag{2.29}$$

$$V(T) = X(T) - \hat{X}(T)E_{nn} \otimes c, \tag{2.30}$$

$E_{nn}$  –  $(n \times n)$  одинична матриця,  $\otimes$  – символ тензорного добутку,  $Sp(\cdot)$  – символ сліду матриці,  $\lambda$  – невизначений множник Лангранжа.

Розв'язання системи рівнянь Куна–Таккера веде до оцінки [58]:

$$c = c^* + D(T) \frac{1 - E_h^T c^*}{E_h^T D(T) E_h} E_h, \tag{2.31}$$

де

$$D(T) = \left( \sum_{\tau=1}^T \hat{x}^T(\tau) \hat{x}(\tau) \right)^{-1}, \tag{2.32}$$

$$c^* = D(T) \sum_{\tau=1}^T \hat{x}^T(\tau) x(\tau) = D(T) d(T) \tag{2.33}$$

звичайна оцінка стандартного методу найменших квадратів.

У роботі [58] доведена оптимальність цієї оцінки по всій навчальній вибірці, тобто вихід метамоделей  $x^*(\tau)$  не поступається за точністю будь-якій з окремих моделей ансамблю  $\hat{x}_j(\tau)$  на інтервалі від  $\tau = 1$  до  $\tau = T$ .

Співвідношення (2.34) та (2.35) нескладно переписати у рекурентній формі за типом рекурентного методу найменших квадратів:

$$\begin{aligned}
D(T+1) &= D(T) - D(T)\hat{x}^T(T+1) \times \\
&\times (E_{nn} + \hat{x}(T+1)D(T)\hat{x}^T(T+1))^{-1} \times \\
&\quad \times \hat{x}(T+1)D(T) \\
d(T+1) &= d(T) + \hat{x}^T(T+1)\hat{x}(T+1) \\
C^*(T+1) &= D(T+1)d(T+1), \\
C(T+1) &= C^*(T+1) + D(T+1) \times \\
&\times (E_h^T D(T+1)E_h)^{-1} (1 - E_h^T C^*(T+1)) E_h
\end{aligned} \tag{2.36}$$

Використання критерію найменших квадратів пов'язане з припущенням про стаціонарність послідовностей, що обробляються, оскільки всім спостереженням від  $x(1)$  до  $x(T)$  приписуються однакові ваги. Оскільки ми припускаємо нестаціонарність контрольованих сигналів, включаючи стрибкоподібні змінення прогнозуючої моделі, оцінки побудовані на основі методу найменших квадратів, виявляються неточними. У таких ситуаціях більш придатними є предиктори, які синтезовані за допомогою процедур оцінювання на «ковзному вікні», що враховують не всю навчальну вибірку, а лише  $s$  (розмір ковзного вікна) останніх спостережень від  $x(T-s+1)$  до  $x(T)$ . При цьому при надходженні значення  $x(T+1)$  спостереження  $x(T-s+1)$  виключається з розгляду, тобто оцінка розраховується на інтервалі від  $x(T-s+2)$  до  $x(T+1)$ . При цьому процедура набуває вигляду:

$$\begin{aligned}
D(T+1) &= D(T) - D(T)\hat{x}^T(T+1) \times \\
&\times (E_{nn} + \hat{x}(T+1)D(T)\hat{x}^T(T+1))^{-1} \hat{x}(T+1)D(T) \\
D^S(T+1) &= D(T+1) + D(T+1)\hat{x}^T(T+S-1) \times \\
&\times (E_{nn} - \hat{x}(T+S-1)D(T+1)\hat{x}^T(T+S-1))^{-1} \times \\
&\quad \times \hat{x}(T+S-1)D(T+1) \\
d^S(T+1) &= d(T) + \hat{x}^T(T+1)\hat{x}(T+1) - \hat{x}^T(T+S-1)\hat{x}(T+S-1), \\
&\quad C^{*S}(T+1) \\
&= D^S(T+1)d^S(T+1)
\end{aligned} \tag{2.37}$$

Цікавою є ситуація, коли оцінювання проводиться у припущенні  $s=1$ , тобто в якості критерію оптимізації (навчання) приймається квадрат похибки оцінювання на останньому такті спостереження:

$$v^2(T) = \|x(T) - \hat{x}(T)c\|^2. \quad (2.38)$$

У цьому випадку процедури (2.39), (2.40) та (2.41) набувають простий вигляд:

$$\begin{aligned} D(T+1) &= (\hat{x}^T(T+1)\hat{x}(T+1) + \alpha E_{hh})^{-1} \\ d^s(T+1) &= \hat{x}^T(T+1)\hat{x}(T+1) \\ C^{*s}(E+1) &= D^\alpha(T+1)d^s(T+1) \\ C^*(T+1) &= C^{*\alpha}(T+1) + D^1(T+1) \times \\ &\times (E_h^T D(T+1)E_h)^{-1} (1 - E_h^T C^{*1}(T+1)) E_h \end{aligned} \quad (2.42)$$

що є узагальненням на випадок, адаптивного алгоритму ідентифікації Качмажа–Уїдроу–Хоффа (тут  $\alpha > 0$  – параметр регуляризації, який забезпечує можливість операції обернення при розрахунку  $D(T+1)$ ).

Найбільшим складним питанням тут залишається вибір розміру «вікна»  $s$ , що зазвичай реалізується, виходячи з суто емпіричних міркувань, оскільки природа можливих змін у контрольованому сигналі  $x(\tau)$  апіорі є невідомою. У цьому випадку доцільним є використання не єдиної метамоделі, а набору таких структур, побудованих при різних значеннях «ковзного вікна».

Для вибору найкращої з таких метамоделей доцільним є введення у розгляд метамоделі другого рівня, що опрацьовує виходи метамоделей першого рівня за допомогою метаалгоритму (2.43), тобто на всій навчальній вибірці при  $\tau = 1, 2, \dots, T, T+1, \dots$ .

Побудову ансамблю метамоделей, які використовують багатовимірні сигнали для виходу прогнозів, представимо докладніші кроками цього методу (Рис. 2.8):

Крок 1. Збір даних. Збір великого набору багатовимірних даних, які будуть використовуватися у ТГ.

Крок 2. Формування вхідних даних. Виходи (оцінки) кожного предиктора надходять на вхід метамоделі.

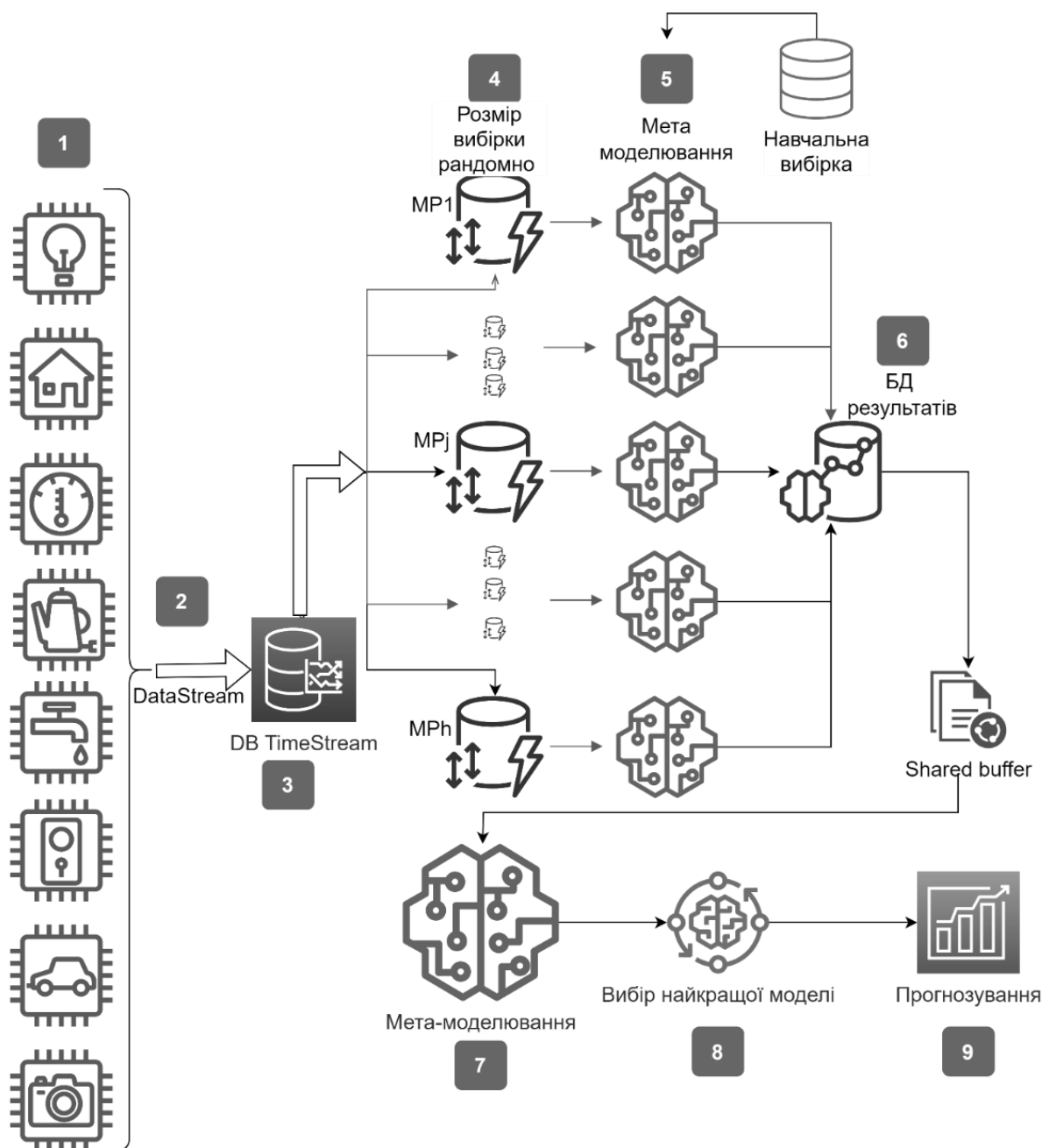


Рис. 2.8. Структурно–логічна схема методу ансамблю адаптивних прогнозних моделей для багатовимірного аналізу

Крок 3. Обробка даних. Кожен з багатовимірних предикторів ансамблю буде обробляти одні і ті самі вхідні дані декількома способами. Кожний предиктор може включати різні методи МН, такі як нейронні мережі, методи опорних векторів, градієнтний бустинг тощо.

Крок 4. Оцінювання на «ковзному вікні» рандомного значення: параметри моделей переоцінюються для кожної нової точки даних, використовуючи тільки останні  $s$  спостережень. Це забезпечує, що модель завжди оновлюватиметься з урахуванням найсвіжіших даних.

Крок 5. Синтез метамodelей. Розробка моделей, які використовують метод невизначених множників Лагранжа для визначення своїх параметрів. Це означає, що метамodelь використовує ваги з різних прогнозів для формування єдиного прогнозу. Ваги цих прогнозів визначаються за допомогою оптимізації функції Лагранжа.

Крок 6. Формування бази результатів. Оцінки метамodelі зберігаються в базі даних для подальшого аналізу.

Крок 7. Синтез метамodelі другого рівня (2.44). Розробка ще одної метамodelі, яка обробляє виходи метамodelей першого рівня (2.45). Це може допомогти зібрати інформацію з різних метамodelей і зробити більш точний прогноз.

Крок 8. Вибір найкращої метамodelі (2.46). Метамodelь другого рівня використовується для вибору найкращої метамodelі серед ансамблю на основі їхніх продуктивностей.

Крок 9. Прогнозування. Кінцева метамodelь використовується для виводу прогнозів на основі вхідних даних. Це дозволяє використовувати єдиний, оптимально зважений прогноз замість незалежних прогнозів від кожного предиктора.

Цей метод передбачає використання ансамблів моделей для роботи з багатовимірними даними і виводу прогнозів, що базуються на комбінації прогнозів від кожного предиктора.

### **2.5.2 Реалізація ансамблю адаптивних прогнозних моделей для багатовимірного аналізу**

Далі проведено апробацію запропонованого методу ансамблю метамodelей, які використовують багатовимірні сигнали для виходу прогнозів. У цьому випадку ансамлева модель була створена на основі прогнозів від моделей логістичної регресії

(LR), дерева рішень (DT), KNN, методу опорних векторів (SVM), випадкового лісу (RF) та наївного Баєса (NB), які були взяті на основі попереднього дослідження [249].

Метамодель була побудована за допомогою AdaBoostClassifier, який є алгоритмом адаптивного підсилення. Цей алгоритм працює шляхом поєднання декількох слабких моделей для створення однієї сильної моделі.

Результати метамоделі були такими (Рис. 2.9):

–точність: 0,98. Це означає, що метамодель правильно класифікувала 98 % вебсайтів;

–повернення: для класу 0 (нешахрайські сайти) становило 0,97, а для класу 1 (шахрайські сайти) – 1,00. Це означає, що метамодель виявила 97% нешахрайських сайтів та 100% шахрайських сайтів;

–F1–оцінка: для класу 0 становила 0,98, а для класу 1 – 0,95. F1–оцінка – це середнє гармонічне між точністю та поверненням, яке надає загальну оцінку моделі.

Meta-Model Performance:			
	precision	recall	f1-score
0	1.00	0.97	0.98
1	0.90	1.00	0.95
accuracy			0.98
macro avg	0.95	0.98	0.97
weighted avg	0.98	0.98	0.98

Рис. 2.9. Результати моделювання з метамоделю

Ці результати показують покращення порівняно з попередніми моделями, що були навчені окремо. Метамодель забезпечує більш точну і консистентну класифікацію вебсайтів, що робить її високоточним інструментом для виявлення шахрайських вебсайтів.

Далі розглянуто приклад використання метамоделі для прогнозування мітки для, наприклад, 16–го спостереження з тестового набору даних (Рис. 2.10). Спочатку



отримуємо це спостереження та його істинну мітку, яка для нього рівна 0, що означає, що вебсайт не є шахрайським.

```
True label: 0
Predicted label from each model: 0, 1, 0, 0, 1, 1
Predicted label from meta-model: 0
```

Рис. 2.10. Приклад застосування метамоделі для 16 рядка набору

Потім отримуємо прогнозовану мітку для цього спостереження від кожної моделі, включаючи моделі логістичної регресії, дерева рішень, KNN, методу опорних векторів, випадкового лісу та наївного Баєса. Прогнозовані мітки від цих моделей варіюються між 0 та 1, що відображає різні прогнози від різних моделей. Нарешті, отримуємо прогнозовану мітку від метамоделі для цього спостереження. Метамоделі прогнозує мітку 0, що відповідає істинній мітці. Це показує, що метамоделі здатна правильно класифікувати це спостереження, незважаючи на різні прогнози від окремих моделей. Цей результат підкреслює високоточність метамоделі в поєднанні з прогнозами від різних моделей для покращення загальної точності прогнозування.

Як видно з Рис. 2.10, при класифікації вебсайтів підтверджено високу точність метамоделі, а також те, що вона може використовувати прогнози від різних моделей для покращення загальної точності прогнозування. Отже, ці результати підтверджують, що використання метамоделі може бути високоточним підходом для покращення точності класифікації.

## 2.6 Структура потоків даних інформаційної технології інтелектуального аналізу

У сучасному світі, де технологічний прогрес та інновації відіграють ключову роль в розвитку інфраструктури ТГ, інформаційна технологія інтелектуального аналізу соціально-економічних даних стає не просто інструментом, а необхідністю

для ТГ. Ця технологія (див. підрозділ 2.1), що використовує передові методи (див. підрозділи 2.5 та 2.4), а саме: метод формування навчальної вибірки для нестационарних даних на основі RFM-аналізу та кластерного аналізу і метод ансамблів адаптивних предикторів для багатовимірних нестационарних послідовностей у рамках онлайн-навчання, є унікальною завдяки своїй здатності інтегрувати та аналізувати різноманітні потоки даних. Вона об'єднує інформацію з муніципальних систем, комерційних джерел, цифрових медіа та систем моніторингу, створюючи комплексне рішення для управління міськими ресурсами. Ця технологія сприяє покращенню якості життя громадян, реагуючи на зміни в реальному часі та пропонуючи інноваційні рішення для майбутнього розвитку ТГ.

Для розвитку інфраструктури ТГ розробка інтелектуальної інформаційної технології аналізу соціально-економічних даних (Рис. 2.11) постає не лише як корисний інструмент, а як важлива складова управління ТГ.

Загальна архітектура (Рис. 2.11) потоків даних інтелектуальної інформаційної технології для аналізу соціально-економічних даних ТГ охоплює кілька ключових етапів і компонентів, що забезпечують збирання, обробку та аналіз інформації. На початковому етапі здійснюється збір даних з різних джерел, що належать до соціальної, економічної та екологічної сфер. Ці дані акумулюються у відповідних базах даних, які є основою для подальшої обробки та аналізу.

Основні аспекти потоків даних інформаційної технології інтелектуального аналізу:

Етап 1. Вибір конкретної задачі (блок 1) для аналізу соціально-економічних даних ТГ.

Етап 2. Збір вхідних даних (блок 2). Вхідними даними запропонованої інформаційної технології є:

–дані з муніципальних та урядових систем (блок 2.1): дані з міських інформаційних систем, включаючи муніципальні бази даних і системи управління містом;

–дані з комерційних джерел (блок 2.2): інформація із супермаркетів та інших магазинів, включаючи дані про покупців та їхні покупки; дані для оцінки

інвестиційних ризиків віртуальних компаній, включаючи інформацію про ринкові умови та потенційних клієнтів.

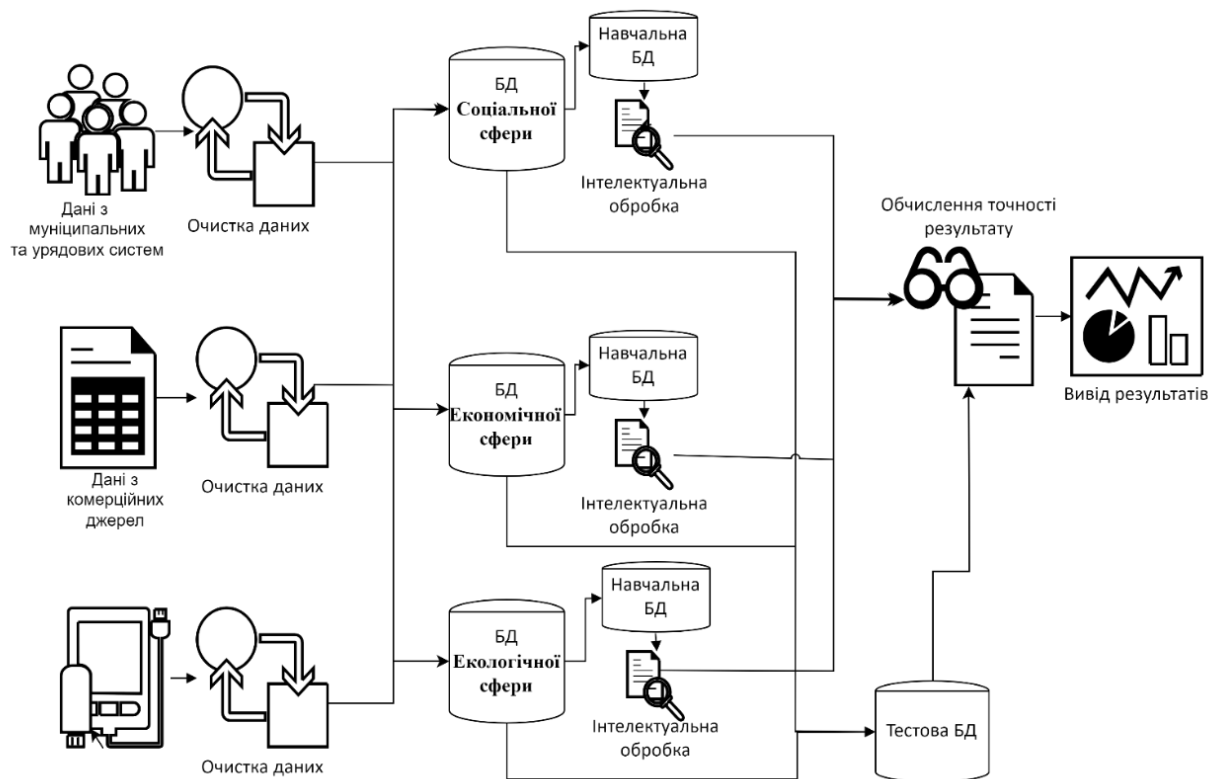


Рис. 2.11. Загальна структура потоків даних інформаційної технології інтелектуального аналізу

–дані з цифрових та соціальних медіа (блок 2.3): дані із соціальних мереж та інтернет–джерел, включаючи відгуки громадян та обговорення в соціальних мережах; дані для виявлення дезінформації, включаючи текстові та метадані з різних джерел;

–дані зі спостереження та моніторингу (блок 2.4): екологічні та техногенні дані, включаючи інформацію про відходи та стан навколишнього середовища.

Кожне з цих джерел може бути інтегроване через API, що дозволяє системі автоматично збирати, аналізувати та використовувати дані в реальному часі для різних аспектів управління міськими ресурсами та службами.

Етап 3. Перетворення даних та виділення ознак (блок 2). На цьому етапі здійснюється попередня підготовка даних для забезпечення їхньої придатності до аналізу:

- очищення даних для видалення помилкових, дубльованих або неповних записів;
- перетворення даних у формат, сумісний з методами машинного навчання;
- виділення ознак, які найбільше впливають на якість аналізу, включаючи соціальні, економічні та екологічні показники.

Етап 4. Метод формування навчальної вибірки (блок 3). Формування навчальної вибірки є ключовим для забезпечення навчання моделей (див. підрозділ 2.4).

Етап 5. Метод ансамблю адаптивних предикторів (блок 4). Використання ансамблів адаптивних предикторів дозволяє аналізувати багатовимірні нестационарні послідовності даних з високою точністю (див. підрозділ 2.6).

Етап 6. Вивід вихідних даних (блок 5). Вихідною інформацією інформаційної технології можна узагальнити у такі основні категорії:

- аналітичні звіти та прогнози (блок 5.1):
  - прогнози щодо розвитку процесів ТГ, екологічних умов та соціально-економічних тенденцій;
  - аналізи та прогнози, що базуються на адаптивних предикторах, для реагування на зміни в реальному часі;
- рекомендації та стратегії (блок 5.2):
  - рекомендації для міського планування та управління, засновані на глибокому аналізі зібраних даних;
  - стратегії для покращення якості життя, управління ресурсами та розвитку інфраструктури;
  - рішення для оптимізації комерційної діяльності, включаючи роздрібну торгівлю та туризм.
- інтерактивні інструменти та інтерфейси (блок 5.3):

- інтерактивні інструменти для візуалізації даних, що дозволяють користувачам легко інтерпретувати складні набори даних;
- інтерфейси для взаємодії з різними модулями системи, включаючи управління ресурсами, моніторинг та аналіз;
- інструменти для забезпечення зворотного зв'язку та адаптації системи до змінних умов і потреб міста.

Ці вихідні дані відображають комплексний підхід до аналізу, прогнозування та управління різними аспектами міського життя, використовуючи передові технології обробки великих даних та штучного інтелекту.

Етап 7. Оцінка якості моделей та системи (блок 6). Цей етап зосереджений на перевірці та аналізі створених моделей. Основні дії включають:

- тестування моделей на основі тестових даних;
- оцінювання точності прогнозів та їх відповідності реальним умовам.

Етап 8. Адаптація моделей: моделі налаштовуються відповідно до змінних умов, що забезпечує їхню гнучкість і надійність.

Науково обґрунтована структура ІТ дозволяє інтегрувати, обробляти та аналізувати великі обсяги даних з різних джерел.

Використання методів адаптивного прогнозування, зокрема ансамблів предикторів, забезпечує високу точність результатів навіть у реальному часі. Ця технологія є важливим інструментом у процесах прийняття рішень для розвитку ТГ, сприяючи управлінню ресурсами та підвищенню якості життя громадян.

Далі представимо ІТ у вигляді діаграми послідовностей (Рис. 2.12) представляє деталізовану інтелектуальну інформаційну технологію для аналізу соціально-економічних даних ТГ, відносно поставлених задач. Діаграма моделює взаємодію між основними компонентами системи та демонструє послідовність обробки даних на різних етапах аналізу.

Початковим кроком роботи інтелектуальної інформаційної технології є формулювання користувачем запиту на аналіз соціально-економічних даних, спрямованого на вирішення конкретного завдання, наприклад, оцінки якості життя чи управління бізнес-процесами. Після цього система автоматично інтегрує дані з різних

джерел, таких як комерційні, муніципальні, урядові системи, цифрові медіа та сенсори, створюючи повну й об'єктивну картину соціально–економічного стану ТГ. Зібрані дані проходять попередню обробку, включаючи видалення некоректних записів і кластеризацію, що дозволяє формувати навчальну вибірку для аналізу.

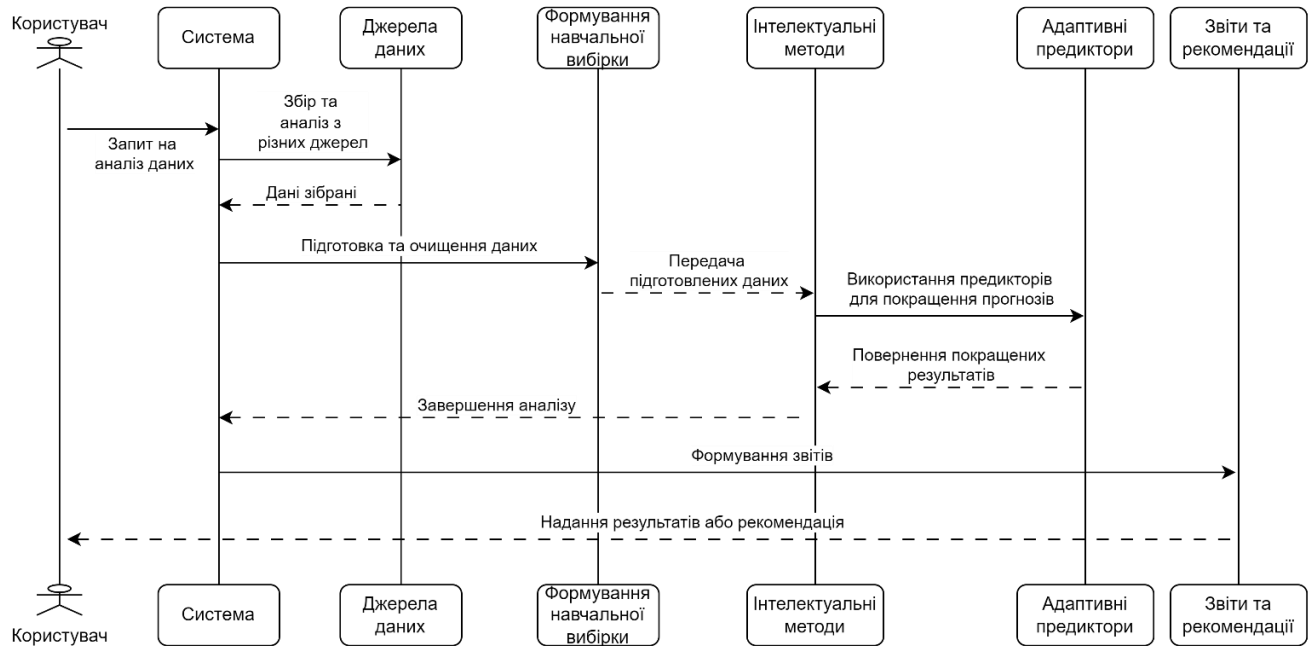


Рис. 2.12. Діаграма послідовностей деталізованої інтелектуальної інформаційної технології для аналізу соціально–економічних даних ТГ

Оброблені дані передаються до алгоритмів інтелектуального аналізу, які виконують завдання прогнозування та аналізу відповідно до поставлених цілей. У разі потреби підвищення точності прогнозів використовується метод адаптивних предикторів. На основі отриманих результатів формуються аналітичні звіти та рекомендації, які передаються користувачеві в зручному форматі, сприяючи ухваленню обґрунтованих управлінських рішень у ТГ.

Інформаційна технологія інтелектуального аналізу масивів даних в інфраструктурі ТГ, яка використовує ансамблі адаптивних предикторів для багатовимірних нестационарних послідовностей у контексті онлайн–навчання, є ключовим елементом в архітектурі ТГ. Ця технологія інтегрує дані з різноманітних

джерел, включаючи муніципальні системи, комерційні джерела, цифрові медіа та системи моніторингу, для забезпечення глибокого аналізу та прогнозування, що сприяє управлінню міськими ресурсами. Використання цієї технології дозволяє громадам реагувати на динамічні зміни, оптимізувати різні аспекти життя громади та покращувати якість життя громадян, використовуючи інноваційні підходи до обробки даних та МН.

Результати, опубліковані в наукових роботах [208, 214, 239, 241, 242, 246, 250, 251, 254, 256, 257, 259, 260, 349, 460], підтверджують ефективність інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ. Застосування цієї технології, яка базується на методах формування навчальних вибірок і ансамблів адаптивних предикторів, забезпечує інтеграцію та аналіз даних з різноманітних джерел, включаючи муніципальні системи, комерційні платформи, цифрові медіа та системи моніторингу. Вона дозволяє здійснювати глибокий аналіз і прогнозування у реальному часі, підвищуючи точність управлінських рішень, оптимізуючи використання ресурсів та сприяючи покращенню якості життя громадян через інноваційні рішення для динамічного розвитку ТГ.

## **2.7 Висновки**

У розділі 2 представлені теоретичні основи створення інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ, що охоплюють узагальнені принципи, методологію аналізу, методи прогнозування, формування вибірки для аналізу, а також концептуальну архітектуру технології.

Розроблено узагальнений принцип синтезу інформаційної технології, що включає систематизацію об'єктів інфраструктури ТГ, типів даних і методів аналізу, що дозволило визначити ключові компоненти інформаційної технології, які сприяють інтеграції різнорідних даних і забезпечують підтримку управлінських рішень навіть в умовах криз.

Розроблена методологія аналізу соціально–економічних даних, що базується на принципах адаптивності, інтеграції даних та точності прогнозування. Представлено етапи аналізу, починаючи від вибору задачі до формування аналітичних звітів, що дозволяє забезпечити гнучкість та ефективність технології у динамічному середовищі громад.

Проведено дослідження часових рядів даних ТГ, визначено їх нестационарність та запропоновано методи обробки нестационарних даних.

Удосконалено метод формування навчальної вибірки для нестационарних даних, використовуючи RFM і кластерний підхід. Це дозволило створити репрезентативні навчальні вибірки, що підвищують точність аналізу і знижують вплив незначущих даних.

Розроблено метод ансамблю адаптивних прогнозних моделей для багатовимірних нестационарних послідовностей в реальному часі, що дозволило створити основу для синтезу складних моделей, які враховують динамічність соціально–економічних процесів.

Представлено концептуальну архітектуру інформаційної технології, яка інтегрує попередньо розроблені методи та забезпечує комплексний підхід до аналізу соціально–економічних даних громад. Вона поєднує збирання, оброблення, аналіз і прогнозування даних для підвищення ефективності управління ресурсами.

Таким чином, у розділі представлено теоретичні основи створення інформаційної технології, яка інтегрує сучасні методи аналізу даних для підвищення точності прогнозів і адаптивності управлінських рішень у ТГ.



## РОЗДІЛ 3.

### МЕТОДИ ТА ЗАСОБИ КЛАСИФІКАЦІЙНОГО АНАЛІЗУ КІЛЬКІСНИХ СОЦІАЛЬНО–ЕКОНОМІЧНИХ ДАНИХ ТЕРИТОРІАЛЬНИХ ГРОМАД

#### 3.1 Метод класифікаційного аналізу кількісних соціально–економічних даних територіальних громад

На основі узагальненого принципу синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ (див. підрозділ 2.1) та визначених задач запропоновано кроки методу класифікаційного аналізу кількісних соціально–економічних даних ТГ. Розроблений метод враховує специфіку задач, таких як оцінка інвестиційних ризиків віртуальних ІТ–компаній та виявлення фіктивних підприємств.

На рис. 3.1 представлено схему методу класифікаційного аналізу кількісних соціально–економічних даних, яка деталізує ключові етапи обробки даних, їх класифікації та прийняття рішень.

Для реалізації поставлених задач визначено два основні об'єкти інфраструктури. Бізнес–інфраструктура ( $I_1$ ) охоплює віртуальні ІТ–компанії, фіктивні підприємства та інші суб'єкти підприємництва, які є основою економічного розвитку ТГ. Водночас, економічна інфраструктура ( $I_6$ ) забезпечує інструменти для оцінки ризиків, прогнозування стабільності бізнесу та підвищення ефективності управління в умовах кризових ситуацій. Включення цих об'єктів до моделі є критичним для врахування факторів ризику і стабільності функціонування громади.

Реалізація методу вимагає роботи з трьома типами даних:

- структуровані дані ( $D_1$ ) включають фінансові звіти, шкали оцінки ризиків та статистичні показники, які забезпечують кількісний базис для аналізу;
- неструктуровані дані ( $D_2$ ), такі як текстова інформація, зображення та геолокація, дозволяють розширити аналіз за рахунок якісних характеристик;

– напівструктуровані дані ( $D_3$ ) у форматах JSON або XML інтегрують зовнішні джерела інформації через API, забезпечуючи більш гнучкий доступ до додаткових даних.

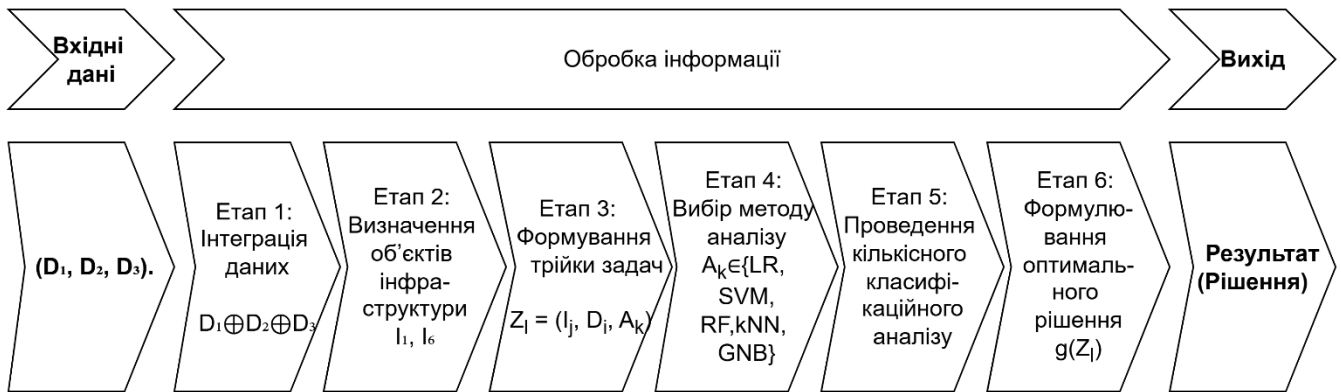


Рис. 3.1. Кроки методу класифікаційного аналізу кількісних соціально–економічних даних ТГ

Методи інтелектуального аналізу, що застосовуються у рамках розробленої технології, забезпечують адаптацію до різноманітних задач: Logistic Regression ( $LR$ ) використовується для оцінки ймовірності високих ризиків; Support Vector Machine ( $SVM$ ) дозволяє ефективно класифікувати об’єкти за рівнем ризику, тоді як Random Forest ( $RF$ ) визначає важливість факторів ризику; Gaussian Naive Bayes ( $GNB$ ) забезпечує ймовірнісне моделювання, а K-Nearest Neighbors ( $kNN$ ) підтримує класифікацію об’єктів на основі схожості у багатовимірному просторі ознак.

Реалізація методу передбачає інтеграцію даних з різних джерел, що здійснюється за допомогою оператора:

$$D_{\text{інтегровані}} = D_1 \oplus D_2 \oplus D_3, \quad (3.1)$$

де  $D_{\text{інтегровані}}$  – результуючий набір даних, отриманий шляхом поєднання структурованих, напівструктурованих та неструктурованих даних.

Формування завдань аналізу базується на трійці:

$$Z_l = (I_j, D_i, A_k), \quad (3.2)$$

де  $I_j \in \{I_1, I_6\}$  – об’єкт інфраструктури;  $D_i \in \{D_1, D_2, D_3\}$ – тип даних;  $A_k \in \{LR, SVM, RF, kNN, GNB\}$  – метод аналізу.

Такий підхід дозволяє поєднати об’єкти інфраструктури з відповідними типами даних та методами аналізу.

На основі аналізу формуються рішення, які визначаються функцією:

$$R_p = g(Z_l), \quad (3.3)$$

де  $g$  – функція вибору оптимального рішення, наприклад, для визначення найкращого інвестиційного проекту або класифікації підприємства як фіктивного чи нефіктивного.

На основі запропонованого методу класифікаційного аналізу кількісних соціально–економічних даних ТГ [459] у розділі детально представлено підходи до оцінки інвестиційного ризику віртуальної ІТ–компанії на основі методів машинного навчання [259, 465] (див. підрозділ 3.2) та метод виявлення фіктивного підприємства навчання [215, 216, 217, 263, 468] (див. підрозділ 3.3). У підрозділі 3.4 наведено експериментальні результати, що охоплюють як оцінку інвестиційного ризику для компанії (див. підрозділ 3.4.1), так і виявлення фіктивного підприємства (див. підрозділ 3.4.2). Розроблений метод класифікаційного аналізу став основою для низки наукових досліджень, зокрема в [208, 239, 245, 249, 251, 256, 262], де підтверджено його ефективність і адаптивність у вирішенні завдань управління ризиками та покращення соціально–економічної стійкості ТГ.

### **3.2 Метод оцінки інвестиційного ризику віртуальної ІТ–компанії на основі машинного навчання**

Оцінка інвестиційного ризику для віртуальних ІТ–компаній є критично важливою для розвитку економічної інфраструктури ТГ. В умовах невизначеності і швидких технологічних змін необхідні інструменти, які дозволять глибоко аналізувати та керувати ризиками. У підрозділі 3.1 представлений метод

класифікаційного аналізу кількісних соціально–економічних даних, який став основою для розробки методу оцінки інвестиційного ризику віртуальної ІТ–компанії на основі машинного навчання.

Оцінка ризику – це ключова управлінська проблема, яка полягає в аналізі і прогнозуванні потенційних збитків, оцінених за їх масштабом та імовірністю виникнення і включає аналіз збитків різних видів, як матеріальних, так і фінансових, інформаційних, робочих, часових та спеціалізованих. Відповідно до джерел [66, 177, 405] (див. підрозділ 1.2.2), потенційні збитки можуть виникати через політичні чинники, економічну та соціальну нестабільність, швидкість науково–технічного прогресу, старіння товарів та послуг, неефективність управлінських методів, некомпетентність персоналу, непорядність клієнтів, а також через недосконалу кредитно–грошову та правову політику.

Законодавство кожної країни має власні особливості, що визначають ризики для бізнесу в контексті конкретної національної юрисдикції. В Україні, наприклад, ризики включають отримання дозволів і ліцензій, попит та ринкову збутову ситуацію, операційні ризики (постачання та сировинні матеріали, недооцінка витрат, технологічні недоліки), нормативно–правові ризики, фінансові та макроекономічні ризики (наявність фінансування, фінансова спроможність приватного партнера, відсоткова ставка, інфляція, курс гривні), а також соціальні та політичні ризики, форс–мажорні обставини [473].

Специфічні ризики віртуальних підприємств охоплюють широкий спектр аспектів, серед яких можна виділити ризик упущеної фінансової вигоди, пов’язаний з можливими непрямими збитками; ресурсний ризик, пов’язаний зі зміною цін або якості ресурсів; ризик впровадження, який стосується можливості невдачі у реалізації проєктів або стратегій; бюрократичні та адміністративні ризики, які виникають внаслідок дій юридичних органів; фінансові ризики, пов’язані з можливістю втрати інвестованого капіталу; репутаційні ризики, які стосуються впізнаваності бренда та наявності офісу; ризики, пов’язані з роботою персоналу, включаючи інформаційні витоки; ризики, пов’язані зі стрімким розвитком та змінами на ринку технологій, а також ризики акумулювання і зберігання інформації.

Здійснення управління ризиками в проєктах стає надзвичайно важливим, особливо в умовах високої невизначеності. Часто при виборі стратегії необхідно знаходити баланс між її ефективністю та здійсненністю. Аналіз ризиків може спиратися на теоретичні та емпіричні моделі, включаючи кількісні методи вимірювання, які оцінюють відхилення від очікуваних результатів або потенційні збитки. Для якісної оцінки використовуються різні шкали, такі як 10-бальна або 5-бальна (табл.3.1).

Таблиця 3.1

## Характеристика компонентів ризику

Ступінь впливу – $Q(R)$			Ймовірність виникнення – $I(R)$		
5-бальна			10-бальна		
5	Дуже висока	80–100% інвестицій	Завжди	Кожен день	10
			Практично неминуче	Частіше, ніж 1 раз на місяць	9
4	Висока	60–79% інвестицій	Дуже часто	Частіше, ніж 1 раз на рік, і рідше, ніж 1 раз на місяць	8
			Часто	Частіше, ніж 1 раз на 2 роки, й рідше, ніж 1 раз на рік	7
3	Середня	40–59% інвестицій	Можливо	Частіше, ніж 1 раз на 3 роки, і рідше, ніж 1 раз на рік	6
			Випадковість, частіше викликана людським фактором	1 раз на 3 роки, і частіше, ніж 1 раз на 5 років	5
2	Низька	20–39% інвестицій	Час від часу	1 раз на 5 років, і частіше, ніж 1 раз на 7 років	4
			Дуже рідко	1 раз на 7 років, і рідше, ніж 1 раз на 5 років	3
1	Дуже низька	0–19% інвестицій	Практично неможливо	Рідше, ніж 1 раз на 7 років	2
			Неможливо	Рідше, ніж 1 раз на 10 років	1

У ході дослідження оцінювалося 20 проєктів заснування віртуальних ІТ-компаній, використовуючи бальну систему та експертні оцінки десяти експертів за

параметрами, описаними у таблиці характеристик компонентів ризику (див. табл.3.1). Рівень ризику для кожного параметра оцінювався за формулою:

$$O(R)=Q(R)\times I(R), \quad (3.4)$$

де  $O(R)$  представляє загальний рівень ризику (див. додаток Б, Табл. Б.1);  $Q(R)$  – ступінь впливу ризику;  $I(R)$ – ймовірність виникнення ризику.

Метод аналізу інвестиційних ризиків для створення віртуальної ІТ–компанії, розроблений на основі машинного навчання, що відображений на діаграмі послідовностей (див. додаток А, рис. А.1). На початковому кроці введені тестові дані (блок 2), які відображені для обробки та передаються у базу даних з тестовою вибіркою (блок 3), інтегруючись в загальну базу (блок 7). Паралельно навчальні дані формуються (блок 9) на основі експертної оцінки і також включаються до загальної бази (блок 7). Обробка даних виконується за допомогою методів, таких як Support Vector Classifier (блок 4), Random Forest Classifier (блок 5) та K-Neighbors Classifier (блок 6), які звертаються до бази даних (блок 7) для моделювання. Результати моделювання для тестової вибірки виводяться назад (блок 2), де користувач може їх оцінити, а висновки та обрані рішення інтегруються у навчальні дані (блок 8), забезпечуючи постійне вдосконалення системи на основі машинного навчання.

Розроблений метод реалізовано в експериментальних дослідженнях, представлених у підрозділі 3.4.1. Отримані результати підтверджують ефективність підходу та його практичну цінність для мінімізації ризиків і підвищення стабільності розвитку віртуальних ІТ–компаній.

### **3.3 Метод виявлення фіктивного підприємства на основі машинного навчання**

Метод виявлення фіктивного підприємства на основі машинного навчання ґрунтується на методі класифікаційного аналізу кількісних соціально–економічних даних ТГ, описаного в підрозділі 3.1. Цей підхід дозволяє ефективно ідентифікувати

фіктивні підприємства шляхом аналізу ключових показників та використання сучасних алгоритмів машинного навчання.

В аналогічному дослідженні [350] використовуються методи машинного навчання для класифікації злочинів, зокрема, для виявлення і аналізу різних типів кримінальної поведінки. Основна увага приділяється геопросторовій візуалізації, аналізу кримінальних мереж, виявленню аномалій у даних та часовому аналізу злочинних патернів. Відмінністю підходу є більш специфічне застосування машинного навчання для ідентифікації фіктивних підприємств, а не загальних злочинів, що забезпечує більшу точність та цілеспрямованість у виявленні економічних злочинів.

Суб'єкти господарювання з ознаками фіктивності можна класифікувати за певними критеріями (див. додаток А, рис. А.2). Описані в дослідженні фактори приводять до масової тінізації економіки в Україні, тому виникає нагальна потреба у розробці спеціальних державних програм протидії злочинності в економічній сфері (сфері державних закупівель, використання коштів державного бюджету, оподаткування діяльності суб'єктів господарювання, виконання робіт, надання послуг за рахунок коштів державного бюджету тощо) та окремих сферах національної економіки.

Для належної організації діяльності правоохоронних органів з виявлення, оперативного документування та розслідування злочинів, учинених із використанням можливостей суб'єктів господарювання з ознаками фіктивності, необхідно визначитися з їх характерними ознаками й особливостями змісту злочинної діяльності та різновидами таких суб'єктів господарювання. В цьому допоможе розроблений алгоритм виявлення фіктивного підприємства на основі машинного навчання (див. додаток А, рис. А.3), що дозволить оперативно приймати рішення щодо показників діяльності окремого підприємства.

На першому кроці (блок 1) користувачу потрібно подати запит на виявлення фіктивного підприємства. Для побудови алгоритму на основі машинного навчання потрібно отримати набір даних на якому буде формуватись алгоритм машинного навчання. Відповідно, потрібні дані які буде вносити (блок 2) безпосередньо

користувач в систему (Таблиця 3.2): код підприємства (генерується автоматично в системі), юридична та фізична адреси, КВЕД, прізвища керівників (може бути кілька), фото обладнання з геолокацією. Всі ці параметри можна доповнювати новими та узагальнювати існуючі, алгоритм легко адаптується.

На основі даних, внесених користувачем, буде проводитись аналіз таких параметрів (див. додаток Б, Табл. Б.2) (блок 3) та пошук відповідних значень з джерел інформації, що створюється за допомогою відповідних API.

Таблиця 3.2

## Дані, що вносить користувач

Параметр	Пояснення	Тип даних
ID	Код підприємства	int64
Company	Назва компанії	object
Address	Юридична адреса	object
FAddress1,... FAddressn	Фізична адреса	object
KVED	КВЕД	object
PIPKER1,... PIPKERn	Прізвища керівників	object
Foto	Фото обладнання з геолокацією	object.jpg

Далі відбувається порівняння даних (блок 4) між собою, а саме: FR з Foto, для визначення, чи збігається геолокація з одною з адрес компанії, і чи представлене на фото відповідне обладнання; SP з A для перевірки, чи співпадають зареєстровані авто компанії зі страховими полісами.

Далі всі дані переносять в базу сховищ (блок 5), після чого дані перетворюємо на бінарні значення (блок 6).

Додаємо параметр fit (блок 7), що показує значення чи підприємство фіктивне чи ні, тому призначаємо відповідно дискретні значення 0 або 1.

Завжди потрібно виводити основні статистичні характеристики кожної числової ознаки, щоб перевірити, чи всі дані відображаються коректно. Відповідно, виводимо (блок 8): кількість значень, середнє значення, мінімум та максимум в значеннях. Рядок std показує стандартне відхилення (яке вимірює, наскільки розсіяні значення). 25 %, 50 % і 75 % рядків показують відповідні відсотки значень у певному параметрі.



Перед дослідженням даних на основі машинного навчання слід виконати один з найважливіших кроків – це розподіл даних (блок 9). Дані розділяють на дві групи: навчальний набір 80 %, тестовий набір 20 % (блок 10).

Оцінка моделі на основі машинного навчання може бути досить складною, оскільки точність, отримана для одного тестового набору, може сильно відрізнятись від точності, отриманої для іншого тестового набору. Тому проведено перехресну перевірку (блок 11) та оцінювання (блок 12) на основі 8 різних методів та відібраного тестового набору (блок 10): Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression. Запорукою справедливого порівняння алгоритмів машинного навчання є забезпечення того, що кожен алгоритм оцінюється однаково на одних і тих самих даних.

Запропонований метод був реалізований у підрозділі 3.4.2. Результати експериментальних досліджень підтвердили його ефективність та здатність суттєво підвищити якість виявлення фіктивних підприємств, що сприяє покращенню інфраструктури ТГ та ефективному вирішенню кризових ситуацій.

### **3.4 Експериментальні результати**

За запропонованим методом (див. підрозділ 3.1) та принципом синтезу інформаційної технології (див. підрозділ 2.1) розроблено модулі інформаційної системи, які вирішують поставлені задачі оцінки інвестиційних ризиків віртуальних ІТ-компаній та виявлення фіктивних підприємств.

#### **3.4.1 Оцінка інвестиційного ризику для компаній**

Реалізація базується на методі оцінки інвестиційного ризику віртуальної ІТ-компанії на основі методів машинного навчання (див. підрозділ 3.2), що забезпечує

високий рівень точності у прогнозуванні ризиків та сприяє обґрунтованому прийняттю інвестиційних рішень.

Для аналізу даних обрано мову програмування Python. Дані розділені на дві частини навчальний набір ( $D_{train}$ ) 80 % даних використовується для тренування моделей та тестовий набір ( $D_{test}$ ) 20% даних використовується для перевірки точності моделі на нових даних.

Додавання параметра `fit` допомагає визначити оптимальний проєкт з мінімальними ризиками. Оцінка (див. підрозділ 1.3) моделі використовується на основі частоти помилок, і крос-валідація проводиться за восьмима алгоритмами класифікації, включаючи Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier та Logistic Regression, для об'єктивного порівняння їхньої точності.

З результатів крос-валідації (табл. 3.3) видно, що найкращі показники продемонстрували такі методи:

1. Метод SVC шукає оптимальну гіперплощину в багатовимірному просторі ознак, яка максимально розділяє класи ризику. Функція рішення моделі можна записати як:

$$f(x) = \text{sign}(w^T \phi(x) + b), \quad (3.5)$$

де  $X$  – вектор ознак;  $\phi(x)$  – функція перетворення в простір більшої розмірності;  $w$  – вектор ваг;  $b$  – зміщення.

Якість класифікації оцінюється за допомогою метрики ROC–AUC, для якої отримано значення AUC=0,91, що свідчить про високу дискримінативну здатність моделі.

Random Forest використовує ансамбль  $N$  дерев рішень для покращення точності та зниження варіативності. Прогноз моделі є середнім або більшістю голосів окремих дерев:

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), \dots, T_N(x)\} \quad (3.6)$$

де  $T_i$  –  $i$ -те дерево в лісі.

У цьому дослідженні використовувалося  $N=320$  дерев, що дозволило досягти середньої абсолютної похибки  $MAE=0,05$  та точності  $Score=0,98$ .

2. Метод kNN класифікує зразки на основі найближчих сусідів у просторі ознак. Для оптимального значення  $k=2$  прогноз визначається як:

$$\hat{y} = mode\{y_1, y_2\}, \quad (3.7)$$

де  $y_i$  – мітки класів двох найближчих сусідів.

Отже, для оцінки інвестиційних ризиків (табл. 3.3) використовується метод Support Vector Classifier, який оптимізує розділення гіперплощини в просторі  $R_n$ . Якість класифікації визначається метрикою ROC, а для інвестиційних ризиків отримано AUC 0,91. Далі використовується метод Random Forest для класифікації з 320 деревами, які спільно надають кращі результати. Також, метод k Nearest Neighbor використовується з оптимальним значенням  $k$ , яке визначено як 2. Для кожного методу проведено класифікацію з виведенням текстового звіту з показниками, включаючи confusion matrix, для оцінки точності на тестовій вибірці.

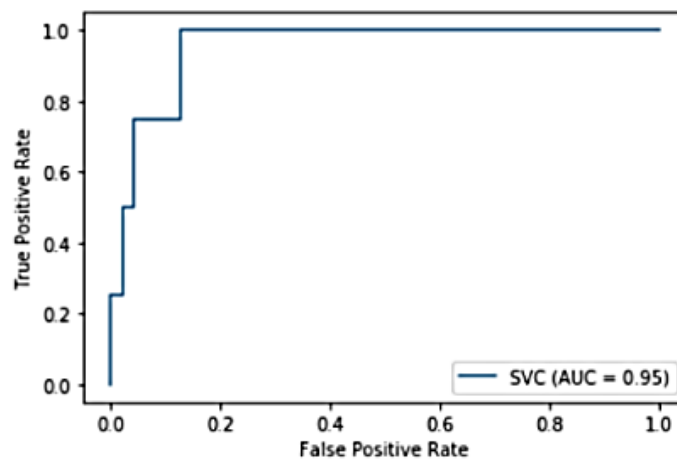
Таблиця 3.3

#### Результати перехресної оцінки

№ з/п	Метод	Оцінка прогнозу	Середнє квадратичне відхилення
1	SupportVectorClassifier	0,982	0,014
2	StochasticGradientDecentC	0,963	0,024
3	RandomForestClassifier	0,984	0,021
4	DecisionTreeClassifier	0,951	0,042
5	GaussianNB	0,962	0,023
6	KNeighborsClassifier	0,981	0,024
7	AdaBoostClassifier	0,953	0,044
8	LogisticRegression	0,984	0,027

Після внесення нової тестової вибірки отримуємо наступні результати для методу Support Vector Classifier (Рис. 3.2). У confusion matrix Support Vector Classifier визначено 47 проєктів, в яких найоптимальніші ризики, також алгоритм передбачив,

що три проєкти найбільш вигідні. З основних показників класифікації оцінки  $f_1$ , при визначенні проєктів, що не підходять для інвестування, алгоритм визначає з точністю 0,99, та проєкти, в які варто інвестувати, – з точністю 0,86. Загальна оцінка точності моделі дорівнює 0,98, що є достатньо добрим результатом. Найбільш привабливим інвестиційним проєктом для створення віртуальної ІТ-компанії система на основі Support Vector Classifier обрала проєкт з кодом P8.



#### Classification Report

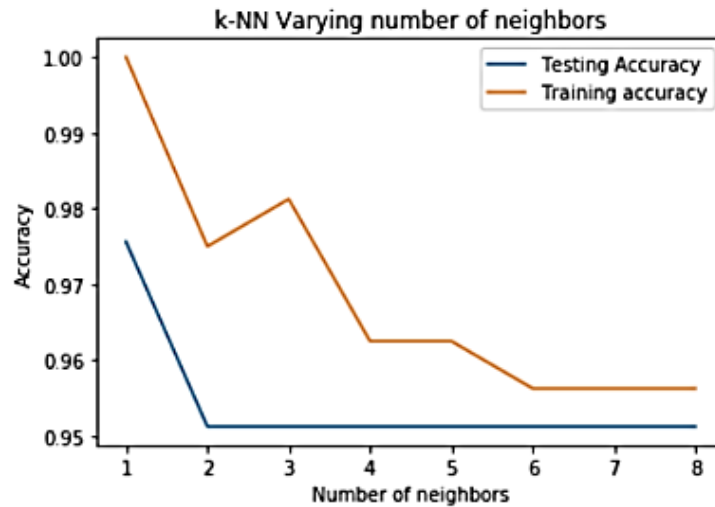
	precision	recall	f1-score	support
0	0.98	1.00	0.99	47
1	1.00	0.75	0.86	4
accuracy			0.98	51
macro avg	0.99	0.88	0.92	51
weighted avg	0.98	0.98	0.98	51

Рис. 3.2. Результати Support Vector Classifier

З результатів Random Forest Classifier середня абсолютна похибка моделювання дорівнює 0,05. Оцінка моделювання – 0,98, що є хорошим результатом. Найкращим проєктом відносно рівня ризиків система обрала на основі Random Forest Classifier також проєкт з кодом P8, як і за методом Support Vector Classifier.

У confusion matrix KNeighborsClassifier (Рис. 3.3) визначено 39 проєктів створення віртуальної ІТ-компанії, в яких ризики найоптимальніші, також алгоритм передбачив, що один проєкт – найбільш вигідний. З основних показників класифікації видно, що за оцінкою  $f_1$ , при визначенні проєктів, що не підходять для інвестування,

алгоритм визначає з точністю 0,99 та проекти, в які варто інвестувати, – з точністю 0,67.



### Classification Report

	precision	recall	f1-score	support
0	0.97	1.00	0.99	39
1	1.00	0.50	0.67	2
accuracy			0.98	41
macro avg	0.99	0.75	0.83	41
weighted avg	0.98	0.98	0.97	41

Рис. 3.3. Результати KNeighborsClassifier

Загальна оцінка точності моделі дорівнює 0,98, що є достатньо хорошим результатом. Система на основі KNeighborsClassifier обрала проект з кодом P1.

Найбільш привабливим інвестиційним проектом для створення віртуальної IT-компанії система на основі Support Vector Classifier та Random Forest Classifier обрала проект з кодом P8.

Якщо розглянути цей проект з точки зору оцінки експертів, то він потрапляє до категорії ризиків із середнім впливом на проект та з випадковістю, що частіше викликана людським фактором, і проявляється 1 раз на 3 роки.

На основі KNeighborsClassifier системи обрала проект з кодом P1. Він потрапляє до категорії ризиків із середнім впливом на проект та з можливою випадковістю ризиків і частотою: частіше, ніж 1 раз на 3 роки, і рідше, ніж 1 раз на рік. Автори вважають, що в обидва проекти створення віртуальної IT-компанії можна

інвестувати, проте Р8 менш ризиковий. Остаточне рішення приймається користувачем–інвестором.

Процес інтелектуальної оцінки інвестиційних ризиків віртуальної компанії включає в себе послідовний ряд кроків, що допомагають підприємцю приймати обґрунтовані рішення щодо інвестування у віртуальний бізнес. У цьому процесі використовуються алгоритми машинного навчання для аналізу та прогнозування ризиків на основі введених даних.

Далі розглянемо кожний крок цього процесу (Рис. 3.4) для кращого розуміння його деталей та значущості у прийнятті рішень в умовах нестабільного ринкового середовища.

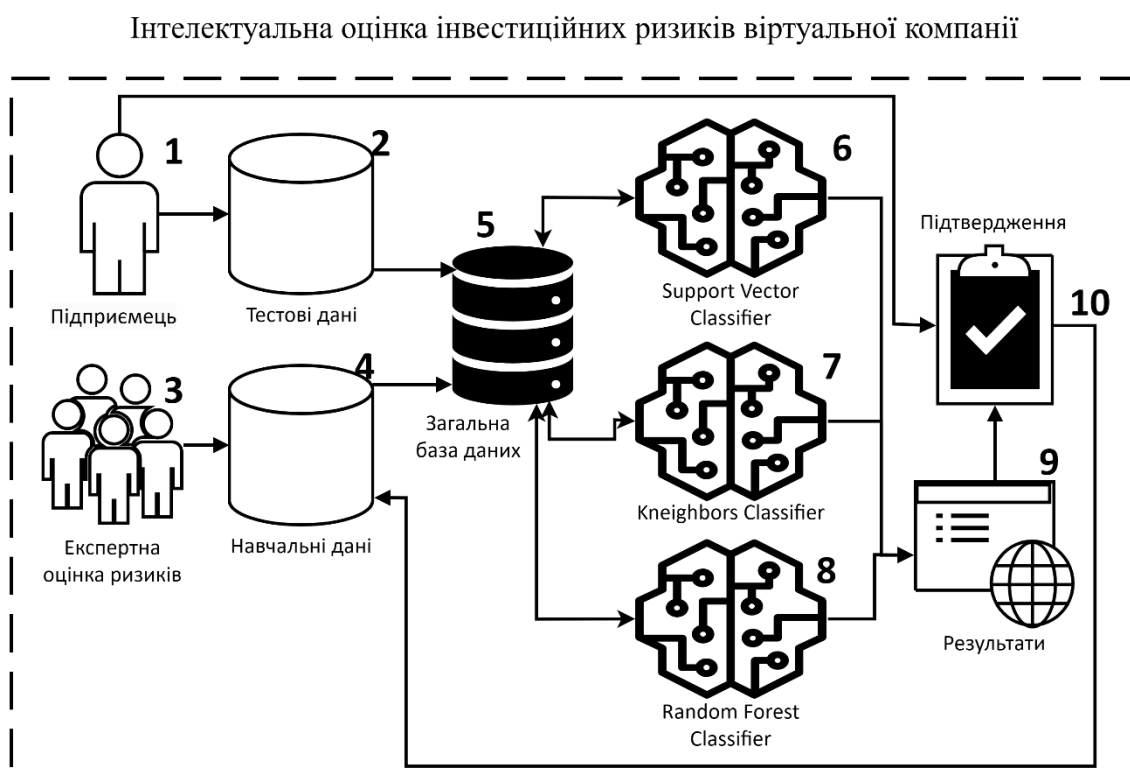


Рис. 3.4. Блок інтелектуальної оцінки інвестиційних ризиків віртуальної компанії

Для наочності розглянемо бізнес–проект «СмартМед», який спрямований на розробку та впровадження інноваційної медичної системи в громаді. Цей бізнес–проект охоплює такі складові, як телемедицина, збір та аналіз медичних даних, а

також надання рекомендацій з питань здорового способу життя і лікування за допомогою рекомендацій ШІ.

Запропонована підсистема використовує різні алгоритми машинного навчання, такі як: Support Vector Classifier, Random Forest Classifier і KNeighborsClassifier. Ці алгоритми отримують навчальні та тестові дані з бази даних для потенційних ризиків. Результати оцінки подаються підприємцю для подальшого аналізу (Рис. 3.5).

[Chat with the Municipality](#)

Data Input: **Choose (1)**
Uploaded file: 
**Run Classifier**

**Data viewing area:**

	A	B	C	D
	Parameter	Data	Parameter	Data
2	Demand and sales market	8	The risk of lost financial gain	6
3	Supply and input materials	18	Implementation risk	15
4	Underestimation of costs	5	Failure to recruit personnel with the required	0
3	Shortcomings of the technology	15	Failure to ful fill obligations by the company	7
5	Regulatory and legal risks	2	Errors in the design	9
6	Availability of financing	16	Reputation	8
7	Financial capacity of the private partner	4	Staff performance	29
8	Interest rate	3	Loss of product relevance	45
9	Force majeure circumstances	24	Accumulation and storage of information	6
10	Accounting and bookkeeping	9		

**Support Vector Classifier Report**

	Precision	Recall	Fi-score	Support
0	0.98	1.00	0.99	47
1	1.00	0.75	0.86	4
<b>Accuracy</b>			0.98	51
<b>Macro avg</b>	0.99	0.88	0.92	51
<b>Weighted avg</b>	0.99	0.98	0.98	51

The project success is: 98%

**K Neighbors Classifier Report**

	Precision	Recall	Fi-score	Support
0	0.97	1.00	0.99	39
1	1.00	0.50	0.67	2
<b>Accuracy</b>			0.98	41
<b>Macro avg</b>	0.99	0.75	0.83	41
<b>Weighted avg</b>	0.98	0.98	0.97	41

The project success is: 98%

**Random Forest Classifier Report**

	Precision	Recall	Fi-score	Support
0	0.93	1.00	0.98	40
1	1.00	0.60	0.66	3
<b>Accuracy</b>			0.95	31
<b>Macro avg</b>	0.91	0.78	0.85	31
<b>Weighted avg</b>	0.94	0.97	0.98	31

The project success is: 95%

Decision is correct

Decision is not correct

Рис. 3.5. Вікно інтелектуальної оцінки інвестиційних ризиків

Отримані результати проєкту показують досить високий рівень успішності в оцінці ризиків для віртуальної компанії. Моделі оцінки ризиків, такі як Support Vector Classifier, Random Forest Classifier і K-Neighbors Classifier продемонстрували хороші показники успішності: 98%, 95% і 98%, відповідно. Це свідчить про їхню здатність надавати надійні та точні прогнози щодо ризиків, і допомагає підприємцю приймати обґрунтовані рішення стосовно інвестицій в смарт-місті. Загальна успішність проєкту в оцінці ризиків відображає високий рівень ефективності системи та її здатність забезпечувати надійний аналіз і управління ризиками для бізнесу в нестабільному ринковому середовищі.

Результати оцінки інвестиційних ризиків демонструють значне скорочення часу ухвалення рішень за допомогою запропонованої інформаційної системи. Експерти (див. додаток Г), працюючи із системою, виконали аналіз за 1–5 годин, тоді як без системи на аналогічні завдання знадобилося від 1 до 3 днів. Таким чином, ефективність використання системи коливається в межах 99,72 %, що свідчить про її здатність суттєво підвищувати швидкість і точність ухвалення управлінських рішень в оцінці інвестиційних ризиків.

### **3.4.2 Виявлення фіктивного підприємства**

Для виявлення фіктивних підприємств запропоновано реалізацію на основі методу класифікації за допомогою машинного навчання (див. підрозділ 3.3). В якості вхідних даних використано 100 компаній, що провадили економічну діяльність в Україні, 20 з яких визначені як фіктивні. Всі дані переведені в логічні бінарні значення.

На основі даних, внесених користувачем, проводився аналіз наступних параметрів (див. Таблиця 3.2) та пошук відповідних значень з джерел інформації, створених за допомогою відповідних API.

Далі проведено перехресну перевірку (Таблиця 3.4) та оцінювання на основі 8 різних методів: Support Vector Classifier, Stochastic Gradient Decent Classifier,



Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression.

Таблиця 3.4

## Результати перехресної оцінки

№ з/п	Метод	Оцінка прогнозу	Середнє квадратичне відхилення
1	SupportVectorClassifier	0.983333	0.023570
2	StochasticGradientDecentC	0.966667	0.023570
3	RandomForestClassifier	0.966667	0.023570
4	DecisionTreeClassifier	0.950000	0.040825
5	GaussianNB	0.983333	0.023570
6	KNeighborsClassifier	0.883333	0.062361
7	AdaBoostClassifier	0.950000	0.040825
8	LogisticRegression	0.983333	0.023570

Всі методи показали (див.Таблиця 3.4 та Рис. 3.6) доволі хороший результат, проте найкращими є у Logistic Regression, Support Vector Classifier, Gaussian Naive Bayes, з показником оцінки прогнозу – 0,98 та стандартним квадратичним відхиленням на рівні 0,02, що є дуже хорошим результатом.

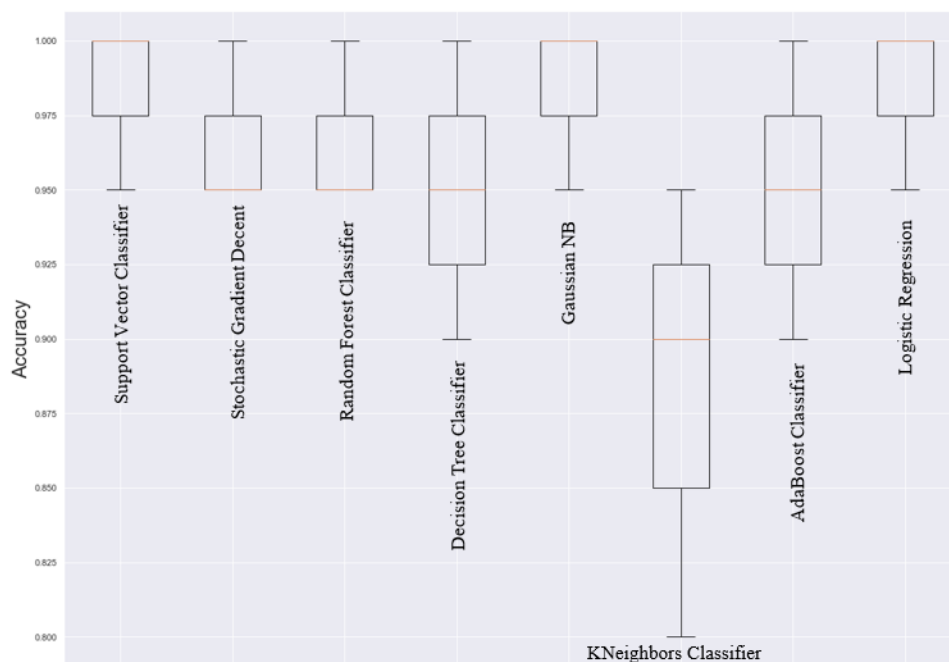


Рис. 3.6. Порівняння алгоритмів boxplot

Для наочності та глибшого розуміння механізму класифікації фіктивних підприємств за допомогою методів машинного навчання – див. додаток А (рис. А.4), який використовує логістичну регресію. Зазначений підхід детальніше розглядається та обґрунтовується у структурі моделі (Рис. 3.7).

Як видно, з отриманих результатів визначено стандартні помилки, z-оцінку та p-значення для кожного з коефіцієнтів, жоден з яких тут не є значущим, окрім K205 та EDR, що аналогічно представлено кореляцією (Рис. 3.7).

```
Call:
glm(formula = fit ~ ., family = binomial(link = "logit"), data = trainset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.5873 -0.1007 -0.0047  0.1050  2.0492

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -13.869554  1.662398  -8.343  <2e-16 ***
ID           0.020020  0.001818  11.012  <2e-16 ***
EDR          0.837101  1.328223   0.630   0.529
P            0.010901  1.139985   0.010   0.992
PO           0.130605  0.826281   1.368   0.171
K           -0.278529  0.371038  -0.751   0.453
VKK          0.056051  0.355929   0.157   0.875
L            0.247243  0.345017   0.717   0.474
K205         0.946000  0.993254   0.952   0.341
ZMI          -0.102556  0.337747  -0.304   0.761
ZD           0.527000  0.350854   1.502   0.133
TovZ        -0.513267  0.346379  -1.482   0.138
SP           0.423517  0.426886   0.992   0.321
A           -0.653820  0.510066  -1.282   0.200
`A&SP`      NA          NA          NA      NA
E           -0.232298  0.342067  -0.679   0.497
F           -0.080450  0.356129  -0.226   0.821
AR           0.182801  0.949578   0.193   0.847
ZR           0.321103  1.040714   0.309   0.758
KR           0.086097  0.832742   0.103   0.918
LB          -0.448565  0.343694  -1.305   0.192
NM          -0.062249  0.352855  -0.176   0.860
NS           0.503829  0.345578   1.458   0.145
FF           0.483070  0.352385   1.371   0.170
FC          -0.426550  0.336234  -1.269   0.205

---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1087.16  on 785  degrees of freedom
Residual deviance: 243.54  on 762  degrees of freedom
AIC: 291.54

Number of Fisher Scoring iterations: 8
```

Рис. 3.7. Оцінка значущості параметрів

Точність логістичної регресії оцінюється за певними ключовими показниками:

- AIC (Інформаційні критерії Акаїке): це еквівалент  $R^2$  у логістичній регресії.

Він вимірює придатність, коли штраф застосовується до кількості параметрів. Менші значення AIC вказують на те, що модель ближча до істини. У представленій реалізації  $AIC = 291,54$ ;

- нульове відхилення: підходить для моделі лише з перехопленням. Степінь вільності  $n-1$ . Інтерпретується як значення  $\chi$ -квадрат (пристосоване значення, яке відрізняється від перевірки гіпотези фактичного значення). Залишкові відхилення: модель із усіма змінними. Це також трактується як перевірка гіпотези  $\chi$ -квадрат;

- кількість ітерацій, що оцінюються Фішером: кількість ітерацій перед зближенням, що дорівнює 8, для поставленої задачі.

Розглянемо матрицю плутанини (Рис. 3.8) для граничної точки на 55 %, з Accuracy: 0.99.

Різні значення матриці Confusion будуть такими для навчальної вибірки:

- справжній позитив (TP) = 129 – це означає, що 129 показників даних позитивного класу правильно класифіковані моделлю;

- справжній негатив (TN) = 121 – це означає, що 121 точка даних негативного класу були правильно класифіковані моделлю;

- хибно позитивний (FP) = 5 – це означає, що 5 показників даних негативного класу були неправильно класифіковані як моделі, що належать до позитивного класу;

- хибно негативний (FN) = 7 – це означає, що 7 показників даних позитивного класу були неправильно класифіковані як моделі, що належать до негативного класу.

```
> confusion_matrix_50
      Actual
Predicted  0  1
          0 129  5
          1  7 121
```

Рис. 3.8. Confusion matrix

Для прогнозування фіктивності підприємства використана модель логістичної регресії. Відсікання даних на 55 % дало хорошу точність Accuracy: 0,99. Крива площі також дає показник точності, який, як виявилось, становив 0,99.

У контексті цього дослідження, зокрема виявлення фіктивних підприємств, далі застосовуємо метод Support Vector Machine Classification (див. додаток А, рис. А.5), який демонструє ключові кроки та принципи роботи цього класифікатора.

Моделювання Support Vector Machine здійснено за 3 підходами: лінійним (Рис. 3.9), поліноміальним (Рис. 3.10) та радіальним (Рис. 3.11). Інші підходи теж існують, але не використовуються набагато рідше.

Першим розглянемо лінійний підхід (Рис. 3.9), оцінювання моделювання проведено на тестовій та навчальній вибірках. Навчальна вибірка показала результати оцінювання на рівні 98,05 %, тестова – на рівні 97,88 %. Стандартне відхилення знаходиться на рівні 98 % для обох вибірок.

```

Train Result:
=====
Accuracy Score: 98.05%
-----
CLASSIFICATION REPORT:
              0          1 accuracy  macro avg  weighted avg
precision    1.000000    0.976077  0.980519    0.988038    0.980986
recall       0.905063    1.000000  0.980519    0.952532    0.980519
f1-score     0.950166    0.987893  0.980519    0.969030    0.980152
support      158.000000    612.000000  0.980519    770.000000    770.000000
-----
Confusion Matrix:
[[143  15]
 [  0 612]]

Test Result:
=====
Accuracy Score: 97.88%
-----
CLASSIFICATION REPORT:
              0          1 accuracy  macro avg  weighted avg
precision    1.000000    0.974545  0.978788    0.987273    0.979328
recall       0.887097    1.000000  0.978788    0.943548    0.978788
f1-score     0.940171    0.987109  0.978788    0.963640    0.978290
support      62.000000    268.000000  0.978788    330.000000    330.000000
-----
Confusion Matrix:
[[ 55  7]
 [  0 268]]

```

Рис. 3.9. Моделювання Support Vector Machine за лінійним підходом

Важливим результатом є показники матриці плутанини (Confusion Matrix). Різні значення матриці Confusion для навчальної вибірки будуть такими:

- справжній позитив (TP) = 143 – це означає, що 143 показники даних позитивного класу правильно класифіковані моделлю;
- справжній негатив (TN) = 612 – це означає, що 612 точок даних негативного класу були правильно класифіковані моделлю;
- хибно позитивний (FP) = 15 – це означає, що 15 показників даних негативного класу були неправильно класифіковані як моделі, що належать до позитивного класу;
- хибно негативний (FN) = 0 – це означає, що 0 показників даних позитивного класу були неправильно класифіковані як моделі, що належать до негативного класу.

Для тестової вивірки матриця плутанини показує, що 158 та 268 показників правильно визначені як моделі, що належать до позитивного та негативного класів, відповідно, 7 показників визначено неправильно як моделі, що належать до негативного класу та 0 показників визначено неправильно як моделі, що належать до позитивного класу.

Далі розглянемо моделювання за поліноміальним підходом (Рис. 3.10). Навчальна вибірка показала результати оцінювання на рівні 100 %, а тестова – на рівні 99,7 %. Стандартне відхилення близьке до 100 % для обох вибірок. Для навчальної вибірки матриця плутанини показує, що 158 та 612 показників правильно визначених як моделі, що належать до позитивного та негативного класів, відповідно, та 0 показників визначено неправильно. На тестовій вибірці матриця плутанини показала, що 61 та 268 показників визначено правильно як моделі, що належать до позитивного і негативного класів, відповідно, та 1 показник – неправильно.

Останній підхід – Support Vector Machine – радіальний (Рис. 3.11), який показав не дуже хороші результати. Навчальна вибірка показала результати оцінювання на рівні 79,8 %, а тестова вибірка показала оцінювання на рівні 81,21 %. Стандартне відхилення близьке до 80 % для обох вибірок. Для навчальної вивірки матриця

плутанини показує, що 158 – визначено неправильно, 612 показників – правильно визначено та 0 показників визначено правильно і неправильно. На тестовій вибірці матриця плутанини показала, що 268 показників визначено правильно, проте 62 – неправильно.

```

Train Result:
=====
Accuracy Score: 100.00%

CLASSIFICATION REPORT:
      0      1  accuracy  macro avg  weighted avg
precision  1.0    1.0    1.0    1.0    1.0
recall    1.0    1.0    1.0    1.0    1.0
f1-score   1.0    1.0    1.0    1.0    1.0
support   158.0  612.0    1.0    770.0   770.0

Confusion Matrix:
[[158  0]
 [  0 612]]

Test Result:
=====
Accuracy Score: 99.70%

CLASSIFICATION REPORT:
      0      1  accuracy  macro avg  weighted avg
precision  1.000000  0.996283  0.99697  0.998141  0.996981
recall    0.983871  1.000000  0.99697  0.991935  0.996970
f1-score   0.991870  0.998138  0.99697  0.995004  0.996960
support   62.000000  268.000000  0.99697  330.000000  330.000000

Confusion Matrix:
[[ 61  1]
 [  0 268]]

```

Рис. 3.10. Моделювання Support Vector Machine за поліноміальним підходом

При розгляді трьох підходів Support Vector Machine найкращі результати отримані при класифікації за поліноміальним підходом. Навчальна вибірка показала результати оцінювання на рівні 100 %, а тестова – на рівні 99,7 %. Також матриця плутанини показала достатньо хороші результати.

Далі представлено алгоритм виявлення фіктивного підприємства, який використовує принципи Gaussian Naive Bayes (див. додаток А, рис. А.6), ілюструючи ключові компоненти та кроки цього методу.

```

Train Result:
=====
Accuracy Score: 79.48%

CLASSIFICATION REPORT:
precision    0          1 accuracy  macro avg  weighted avg
recall      0.0      1.000000 0.794805  0.500000   0.794805
f1-score    0.0      0.885673 0.794805  0.442836   0.703937
support    158.0    612.000000 0.794805  770.000000  770.000000

Confusion Matrix:
[[ 0 158]
 [ 0 612]]

Test Result:
=====
Accuracy Score: 81.21%

CLASSIFICATION REPORT:
precision    0          1 accuracy  macro avg  weighted avg
recall      0.0      1.000000 0.812121  0.500000   0.812121
f1-score    0.0      0.896321 0.812121  0.448161   0.727921
support     62.0    268.000000 0.812121  330.000000  330.000000

Confusion Matrix:
[[ 0  62]
 [ 0 268]]

```

Рис. 3.11. Моделювання Support Vector Machine за радіальним підходом

Тепер навчимо модель, для цього поєднаємо квантильний трансформатор і найвний класифікатор Байеса за Гаусом. Після навчання моделі виводимо криву ROC (Рис. 3.12) на даних тренінгу та оцінюємо модель, обчислюючи навчальний набір AUC. Точність класифікації висока (Рис. 3.12), що підтверджено показником  $AUC \approx 0,99$ .

Далі виведемо перші 20 отриманих результатів (Рис. 3.13). Очевидно, що деякі значення не дорівнюють 1, а 0,99. Таких значень у результативній вибірці є невелика кількість, і саме вони впливають на точність отриманих результатів, яка дорівнює  $AUC \approx 0,99$ .

У підсумку, проведено дослідження за допомогою трьох методів МН: Logistic Regression, Support Vector Classifier та Gaussian Naive Bayes для виявлення фіктивних підприємств. Метод Logistic Regression показав високу точність класифікації з AUC значенням 0,99 та хорошими показниками чутливості і специфічності. Класифікація за допомогою Support Vector Machine, особливо за поліноміальним підходом, також продемонструвала хороші результати з точністю навчальної вибірки в 100 % та тестової вибірки в 99.7 % (Рис. 3.9–3.11). Нарешті, метод Gaussian Naive Bayes також показав високу точність з AUC, близькою до 0,99 (Рис. 3.12). Враховуючи ці результати, можна зробити висновок, що кожен з цих методів має свої переваги і є точним у визначенні фіктивних підприємств, при цьому поліноміальний підхід SVM та Logistic Regression виявилися особливо потужними інструментами для цієї задачі.

AUC = 0.986703135015442

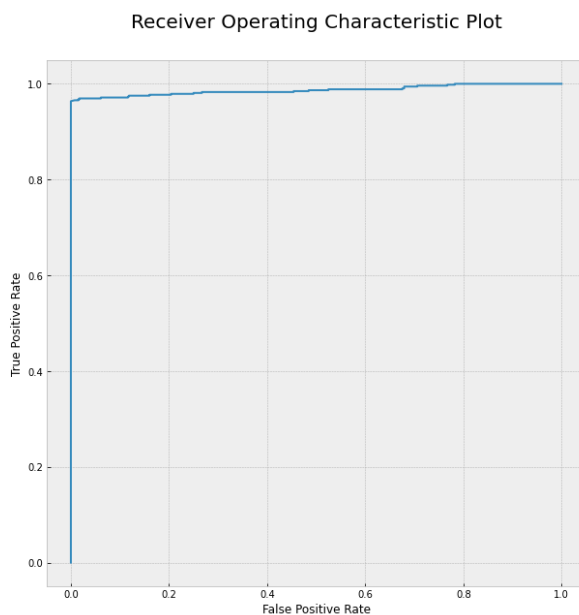


Рис. 3.12. ROC–крива

```

1 | fit
2 | 1.0
3 | 0.0
4 | 0.999999997621174
5 | 0.9999999770409362
6 | 0.0
7 | 0.0
8 | 0.9999999754939356
9 | 0.0
10 | 1.0
11 | 1.0
12 | 1.0
13 | 1.0
14 | 1.0
15 | 1.0
16 | 0.0
17 | 0.9999999988921786
18 | 1.0
19 | 1.0
20 | 1.0

```

Рис. 3.13. Отримані результати

Процес інтелектуального виявлення шахрайських практик в електронному бізнес–середовищі (Рис. 3.14) ініціюється кроком збору даних (блок 1), де користувач вводить інформацію щодо компанії чи інтернет–магазину, а також отримує дані із зовнішніх джерел (блок 2). Ці дані передаються до OLAP–бази даних (блок 3). З використанням методу класифікації (див. підрозділ 4.7) виконується інтелектуальний



аналіз (блок 4) для визначення шахрайських інтернет-магазинів та можливих фіктивних підприємств (блок 5). Результати цього аналізу виводяться (блок 6) користувачеві інформацію про ймовірність шахрайства в електронному бізнесі. Останній крок передбачає підтвердження або спростування отриманого рішення користувачем (блок 7), після цього відповідна інформація передається назад до бази даних.



Рис. 3.14. Блок інтелектуального виявлення шахрайств  
в електронному бізнес-середовищі

Отже, користувач отримує можливість перевірити, чи займається підприємство, з яким планується співпраця, шахрайською діяльністю (див. Рис. 3.14). Процес інтелектуального виявлення цього шахрайства розпочинається, коли користувач ініціює процес (Рис. 3.15), заповнюючи форму з даними, що ґрунтуються на інформації про компанію, або надаючи інформацію про конкретний сайт, який його

цікавить. Наприклад, користувач вводить назву компанії, юридичну та фізичну адреси, код КВЕД, прізвища керівників та додає фото обладнання з геолокацією у форматі .jpg.

---

**Data collection:**

user-journey.CSV  
2m ago

604KB

**Storage location**  
C: Laura\My Documents

Change

---

**Data entry form:**

<p>Company name</p> <input type="text" value="Enter"/>	<p>Company address</p> <input type="text" value="Enter"/>	<p>Photo</p> <div style="border: 1px solid #ccc; padding: 5px; text-align: center;"> <p>Drag and drop the photo</p> </div>
<p>Physical address of the company</p> <input type="text" value="Enter"/>	<p>Surname and first name managers</p> <input type="text" value="Enter"/>	
<p>KVED</p> <input type="text" value="Enter"/>	<p>Surname and first name managers</p> <input type="text" value="Enter"/>	

**Enter the link to the online store**

✓

Url address is liquid

**Run**

---

**Output area of the results:**

**Results of modelling**

Result for Online Store

The online store for which you entered the information:

Accuracy: 90%

Result: **Fake**

Result for Company

The company for which you entered information on:

Accuracy: 98%

Result: **Fake**

100%

---

Рис. 3.15. Вікно інтелектуального виявлення шахрайств в електронному бізнес-середовищі

У випадку пошуку фіктивних підприємств після введення даних система автоматично збирає інформацію з відкритих джерел України, таких як реєстри, бази даних податкової служби, реєстр декларацій та ін. Наприклад, перевіряється присутність компанії в єдиній базі реєстру юридичних та фізичних осіб, реєстрі платників ПДВ, системі єдиного податку тощо.

Щодо виявлення шахрайських інтернет–магазинів, система використовує параметри, такі як відсутність інформації про товар, його склад, виробника, гарантійний строк, адресу для повернення товару тощо.

Після збору даних система використовує алгоритми машинного навчання та класифікації для аналізу інформації. Результат обробки може бути поданий у вигляді висновку, наприклад: «Компанію визнано фіктивною з точністю 98 %» або «Інтернет–магазин визначено як шахрайський на 90 %».

Результати дослідження демонструють значне скорочення часу, необхідного для виявлення фіктивних підприємств, за рахунок використання розробленої інформаційної системи (див. додаток Г). Середній час аналізу з використанням системи становить від 2 до 6 годин, тоді як без системи цей показник коливається в діапазоні від 3 до 20 днів. Ефективність розробленого підходу оцінено на рівні 99,58 %, що свідчить про його високу дієвість у прискоренні процесу прийняття рішень та підвищенні точності аналізу.

### **3.5 Висновки**

У розділі 3 розроблено та впроваджено метод класифікаційного аналізу кількісних соціально–економічних даних ТГ, що враховує специфіку задач оцінки інвестиційних ризиків віртуальних ІТ–компаній і виявлення фіктивних підприємств.

Розробленого метод класифікаційного аналізу, який базується на узагальненому принципі синтезу інформаційної технології інтелектуального аналізу даних та інтегрує роботу зі структурованими, неструктурованими та напівструктурованими даними. Використання різних методів машинного навчання,

таких як Logistic Regression, SVM, Random Forest, Gaussian Naive Bayes та kNN, забезпечило гнучкість та адаптивність методу до різноманітних класифікаційних задач в соціально–економічній сфері. Це дозволило підвищити точність аналізу та якість прийняття рішень.

Розроблено метод оцінки інвестиційного ризику віртуальних ІТ–компаній на основі методів машинного навчання. Проведено аналіз специфічних ризиків, притаманних віртуальним підприємствам, та визначено ключові фактори, що впливають на інвестиційну привабливість. Застосування алгоритмів SVC, Random Forest та kNN дозволило досягти високої точності прогнозування інвестиційних ризиків (до 98 %), що сприяє обґрунтованому прийняттю інвестиційних рішень та підвищенню стабільності економічного розвитку ТГ.

Розроблено метод виявлення фіктивних підприємств на основі машинного навчання. Розроблено алгоритм, який використовує параметри компаній і застосовує методи Logistic Regression, SVM та Gaussian Naive Bayes для класифікації підприємств. Проведені експерименти показали високу точність виявлення фіктивних підприємств (до 99,7 %), що є критично важливим для боротьби з економічною злочинністю та забезпеченням економічної безпеки ТГ.

Наведені експериментальні результати підтверджують ефективність розроблених методів. Створено інформаційну систему, яка впроваджує запропоновані методи та дозволяє значно зменшити час прийняття рішень — від кількох днів до кількох годин. Це забезпечує оперативність і точність у вирішенні завдань щодо оцінки інвестиційних ризиків та виявлення фіктивних підприємств, що, у свою чергу, підвищує соціально-економічну стійкість ТГ.

Загалом, у розділі здійснено комплексне дослідження та розроблено ефективні методи інтелектуального аналізу соціально–економічних даних. Отримані результати мають практичну цінність та можуть бути використані для підвищення ефективності управління ризиками, покращення інвестиційного клімату та зміцнення економічної безпеки ТГ.

## РОЗДІЛ 4.

### МЕТОДИ ТА ЗАСОБИ КЛАСИФІКАЦІЙНОГО АНАЛІЗУ ТЕКСТОВИХ СОЦІАЛЬНО–ЕКОНОМІЧНИХ ДАНИХ ТЕРИТОРІАЛЬНИХ ГРОМАД

#### 4.1 Метод класифікаційного аналізу текстових соціально–економічних даних територіальних громад

На основі запропонованого узагальненого принципу синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ (див. підрозділ 2.1) було запропоновано кроки методу класифікаційного аналізу текстових соціально–економічних даних ТГ, який враховує специфіку задач, таких як виявлення дезінформації, формування короткострокових HR–проектів і вибір конкурентоспроможних товарів на основі аналізу емоційного забарвлення відгуків та інших, що використовують методи класифікації текстових даних. Вона передбачає інтеграцію гетерогенних даних, застосування сучасних методів інтелектуального аналізу та підтримку прийняття рішень для забезпечення стійкого розвитку ТГ.

На рис. 4.1 представлено етапи методу, який відображає основні кроки аналізу текстових соціально–економічних даних, від збору інформації до прийняття рішень.

Об'єкти інфраструктури формалізуються як множина  $I = \{I_1, I_2, \dots, I_n\}$ , де кожен елемент  $I_j$  представляє певний аспект соціально–економічної діяльності. У контексті поставлених задач розглядаються такі об'єкти, як інформаційна інфраструктура, що забезпечує збір і аналіз текстових даних з новинних порталів і соціальних мереж ( $I_1$ ); соціальна інфраструктура, яка відповідає за оцінку споживчих настроїв ( $I_2$ ), а також освітня інфраструктура, яка включає управління проектами з персоналом ( $I_3$ ).

Типи даних, що використовуються для аналізу, включають структуровані дані ( $D_1$ ), наприклад, метадані текстів (дата публікації, автор, джерело); неструктуровані дані ( $D_2$ ), до яких належать текстові документи, відгуки споживачів та інші текстові масиви; напівструктуровані дані ( $D_3$ ), такі як JSON або XML–файли, які інтегруються через API для перевірки та доповнення інформації.

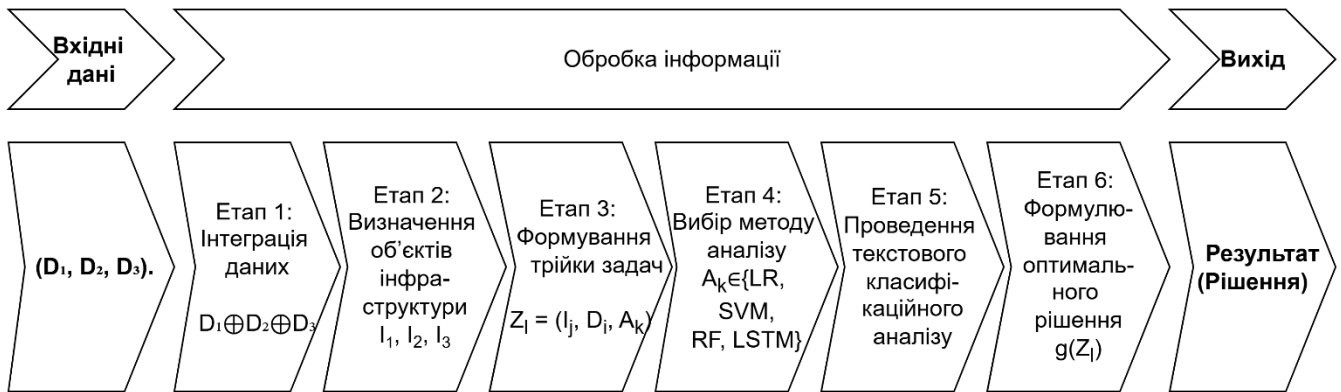


Рис. 4.1. Кроки методу класифікаційного аналізу текстових соціально–економічних даних ТГ

Для обробки цих даних застосовуються методи інтелектуального аналізу, представлені множиною  $A = \{A_1, A_2, \dots, A_m\}$ . До основних методів належать логістична регресія (Logistic Regression), яка використовується для класифікації текстів за рівнем достовірності ( $A_1$ ); метод опорних векторів (Support Vector Machine), що застосовується для аналізу емоційного забарвлення тексту ( $A_2$ ); алгоритми випадкових лісів (Random Forest) для визначення важливості факторів та прийняття рішень ( $A_3$ ); глибокі нейронні мережі, такі як LSTM і трансформери, для аналізу динамічних текстових послідовностей ( $A_4$ ).

Інтеграція даних з різних джерел здійснюється за допомогою оператора  $\oplus$ , що дозволяє об'єднати структуровані, неструктуровані та напівструктуровані дані в єдину базу:

$$D_{\text{інтегровані}} = D_1 \oplus D_2 \oplus D_3. \quad (4.1)$$

Це забезпечує повноту інформації для подальшого аналізу. Завдання аналізу визначаються трійкою параметрів:

$$Z_l = (I_j, D_i, A_k), \quad (4.2)$$

де  $I_j \in \{I_1, I_2, I_3\}$ ,  $D_i \in \{D_1, D_2, D_3\}$ ,  $A_k \in \{A_1, A_2, A_3, A_4\}$ .

Кожна трійка описує конкретну задачу, наприклад, виявлення дезінформації, класифікацію відгуків або формування команд для проєктів.

Рішення формулюються як  $R_p = g(Z_l)$ , де  $g$  – функція, що забезпечує вибір оптимального результату. Наприклад, у задачах виявлення дезінформації  $R_p$  визначає достовірність тексту, у виборі конкурентоспроможного товару – рейтинг товарів на основі аналізу відгуків.

Задача виявлення дезінформації передбачає збір текстів із джерел, їх попередню обробку (токенізація, стемінг, нормалізація), аналіз емоційного забарвлення та тональності за допомогою методів  $A_2$ , а також класифікацію текстів на достовірні та недостовірні за допомогою  $A_1$ . Для формування HR-проектів передбачено автоматизацію збору профілів кандидатів, інтерактивний відбір через чат-бот та класифікацію за допомогою  $A_3$ . Аналіз відгуків для вибору товарів включає збір текстів, очищення, векторизацію та оцінку емоційного забарвлення за допомогою  $A_2$ , після чого виконується ранжування товарів.

На основі методу класифікаційного аналізу текстових соціально-економічних даних ТГ [459] у розділі детально представлені інтелектуальні підходи до виявлення дезінформації [242, 254, 346, 455] (підрозділ 4.2) та формування короткострокового HR-проекту [261, 460] (підрозділ 4.3). У підрозділі 4.4 наведено експериментальні результати, які включають практичне виявлення дезінформації (підрозділ 4.4.1) та реалізацію HR-проектів (підрозділ 4.4.2). Метод, який інтегрує сучасні технології інтелектуального аналізу тексту, став основою для кількох наукових публікацій, зокрема в [139, 243, 245, 247, 248, 250, 254, 256, 315, 346, 347, 464, 462], що підтверджують його ефективність у вирішенні соціально-економічних завдань ТГ.

## **4.2 Інтелектуальний метод виявлення дезінформації**

У цьому підрозділі представлено інтелектуальний метод виявлення дезінформації, який базується на методі класифікаційного аналізу текстових соціально-економічних даних ТГ, описаному в підрозділі 4.1. Запропонований метод використовує передові техніки машинного навчання та нейронних мереж для глибокого аналізу текстових даних і надійного виявлення фейкової інформації. Це є

ключовим інструментом у забезпеченні достовірності інформації та підтримці інформованості громадськості, сприяючи формуванню стійкого та довірчого середовища в ТГ.

Запропонований метод відрізняється від інших підходів [9, 18, 121, 318] (див. підрозділ 1.2.1) завдяки своїй здатності працювати в режимі реального часу, точно виявляючи групи фейкової інформації. Така структура забезпечує можливість постійно реагувати на зміни у сфері дезінформації, підтримуючи точність виявлення фейкової інформації в реальному часі.

Далі представлено розроблений комплексний метод для виявлення дезінформації, який охоплює етапи від збору та попередньої обробки даних до аналізу тональності, емоцій та векторизації тексту. Метод включає в себе застосування передових технік машинного навчання і нейронних мереж, забезпечуючи глибокий аналіз текстових даних та точне виявлення фейкових новин. Структура запропонованого методу представлена у вигляді сукупності послідовних кроків (див. Додаток А, рис. А.7).

Крок 1. Збір даних (блок 1). Критично важливий етап в процесі розробки методу для виявлення дезінформації. Правильно зібрані та структуровані дані дозволяють тренувати моделі на основі машинного навчання та проводити аналітику. На першому етапі здійснюється збір даних  $D$ , які є критично важливими для тренування моделей на основі машинного навчання та проведення аналізу. Дані можна представити як сукупність текстових документів і метаданих:

$$D = \{(T_i, M_i)\}_{i=1}^N, \quad (4.3)$$

де  $T_i$  – текстовий документ  $i$ -го зразка;  $M_i$  – відповідні метадані;  $N$  – загальна кількість зібраних зразків.

Розглянемо більш детальний опис цього кроку:

### 1.1. Збір текстових даних.

1.1.1. Визначення джерел. Джерела можуть включати новинні портали, соціальні мережі, блоги, форуми та інші платформи, де користувачі можуть публікувати або ділитися інформацією. Вибір множини джерел  $S = \{s_1, s_2, \dots, s_K\}$ , де  $K$  – кількість



джерел (новинні портали, соціальні мережі тощо). Метою є отримання репрезентативного вибору текстів, що можуть містити як достовірну інформацію, так і дезінформацію.

1.1.2. Автоматизація збору даних [139]. Розробка скриптів або використання існуючих інструментів для автоматичного збору даних. Це може включати вебскрапінг, API-запити до соціальних мереж та ін. Нехай  $C$  – скрипт або програма, що здійснює функцію  $C: S \rightarrow D_T$ , де  $D_T \subset D$  – зібрані текстові дані.

1.1.3. Фільтрація та валідація. Це процес для забезпечення їх якості та відповідності вимогам дослідження. Проводиться фільтрація зібраних даних  $D_T$  для забезпечення якості; визначається підмножина  $D'_T \subset D_T$ , що відповідає критеріям якості  $Q$ :

$$D'_T = \{T_i \in D_T \mid Q(T_i) = True\}. \quad (4.4)$$

## 1.2. Збір метаданих.

1.2.1. Інформація про авторів. Збирається інформація про авторів, яку можна представити як функцію  $A: D'_T \rightarrow A$ , де  $A$  – множина профілів авторів з такими атрибутами, як кількість підписників, історія публікацій тощо.

1.2.2. Інформація про джерела. Збирається інформація про джерела текстів, що можна представити як функцію  $S: D'_T \rightarrow S$ , де  $S$  – множина атрибутів джерел (URL, дата публікації, кількість переглядів тощо).

1.2.3. Структурування метаданих. Організація зібраних метаданих у структуровані бази даних для легкого доступу та аналізу на наступних етапах дослідження. Метадані  $M_i$  для кожного тексту  $T_i$  організуються в структуровану базу даних для зручності доступу та аналізу.

Крок 2. Попередня обробка даних (блок 2). Це фундаментальний етап у процесі аналізу текстової інформації. Цей крок включає в себе різноманітні техніки та методи, які допомагають підготувати зібрані дані для подальшого аналізу. Попередня обробка даних полягає у перетворенні «сирих» текстових даних  $D'_T$  на структуру, зручну для аналізу. Для кожного тексту  $T_i$  проводяться послідовні операції:

2.1. Токенізація. Процес розбиття тексту на окремі слова, фрази, символи або інші значущі елементи, називаються токенами. Це допомагає у подальшому аналізі та обробці тексту. Розбиття тексту  $T_i$  на послідовність токенів:

$$T_i = \{w_{i1}, w_{i2}, \dots, w_{in_i}\}, \quad (4.5)$$

де  $w_{ij} - j$  – й токен у тексті  $T_i$ ;  $n_i$  – кількість токенів у тексті.

2.2. Стемінг. Процес видалення суфіксів, префіксів та інфіксів зі слів для повернення їх до їхньої основи. Це полегшує виявлення спільних тем та патернів у тексті. Застосовується функція стемінгу  $\text{stem}(\cdot)$  до кожного токена для отримання його основи:

$$w'_{ij} = \text{stem}(w_{ij}). \quad (4.6)$$

Отримуємо послідовність стемів  $\{w'_{i1}, w'_{i2}, \dots, w'_{in'_i}\}$ .

2.3. Тегування частин мови. Процес визначення частин мови кожного слова у тексті, що може бути корисним для синтаксичного аналізу та визначення семантичних відносин між словами. Визначення частин мови для кожного токена за допомогою функції:

$$t_{ij} = \text{POS}(w_{ij}), \quad (4.7)$$

де  $t_{ij}$  – тег частини мови (іменник, дієслово тощо).

2.4. Розпізнавання іменованих сутностей (Named Entity Recognition). Визначення та класифікація іменованих сутностей у тексті, таких як імена осіб, організацій, місцезнаходжень тощо.

Виявлення іменованих сутностей за допомогою функції:

$$e_{ij} = \text{NER}(w_{ij}), \quad (4.8)$$

де  $e_{ij}$  – категорія сутності (персона, організація, місце тощо).

Крок 3. Векторизація тексту (блок 3). Векторизація тексту є важливим етапом, який перетворює текстові дані в числовий формат, зручний для аналізу та обробки за допомогою методів МН. Розглянемо більш детально цей процес:

### 3.1. Вибір методу векторизації.

3.1.1. Word2Vec. Модель навчає векторні представлення слів у багатовимірному просторі таким чином, що слова, які часто зустрічаються разом, мають близькі векторні представлення.

3.1.2. GloVe. Інший підхід до векторизації слів, який використовує як локальний, так і глобальний статистичний аналіз корпусу тексту для визначення векторних представлень слів.

3.1.3. BERT. Сучасна модель, яка використовує механізм уваги для визначення відносин між словами в тексті і може вчитися глибокі контекстні представлення слів.

### 3.2. Процес векторизації.

3.2.1. Тренування або завантаження моделей. Моделі можна натренувати на даних або використати попередньо натреновані моделі для векторизації тексту.

3.2.2. Перетворення тексту. Застосування вибраної моделі для перетворення кожного слова в тексті у векторні представлення.

### 3.3. Побудова векторних представлень.

3.3.1. Векторизація слів. Отримання векторних представлень для кожного окремого слова в тексті:

$$v_{ij} = Embed(w_{ij}). \quad (4.9)$$

3.3.2. Векторизація текстів. Агрегація векторних представлень слів для отримання векторних представлень цілих текстів:

$$v_{T_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} v_{ij}. \quad (4.10)$$

Крок 4. Навчання методом онлайн-навчання з ковзним вікном для ансамблів текстових класифікаторів (блок 5). Цей крок фокусується на розробці та навчанні моделі класифікації, яка може виявляти фейкову інформацію на основі аналізу тексту та інших виявлених ознак. Детальний опис цього кроку виглядає таким чином:

#### 4.1. Створення та навчання ансамблю моделей.

4.1.1. Створення ансамблю моделей. Вибір множини моделей класифікації  $M = \{m_1, m_2, \dots, m_L\}$  (наприклад, логістична регресія, SVM, LSTM, трансформери).

4.1.2. Навчання моделей. Кожна модель  $m_l$  навчається на навчальному наборі  $D_{train}$ :

$$m_l: v_{T_i} \rightarrow \hat{y}_i^{(l)}, \quad (4.11)$$

де  $\hat{y}_i^{(l)} \in \{0,1\}$  – прогноз моделі  $m_l$  для тексту  $T_i$ .

4.1.3. Формування метамоделі. Використання алгоритму, такого як Adam, для оптимізації ваг моделей в ансамблі, забезпечує кращу інтеграцію та вибір прогнозів від кожної моделі.

Створюється метамодель  $M$ , яка об'єднує прогнози базових моделей, наприклад, за допомогою зваженого голосування:

$$\hat{y}_i = \text{sign}\left(\sum_{l=1}^L \alpha_l \hat{y}_i^{(l)}\right), \quad (4.12)$$

де  $\alpha_l$  – вага моделі  $m_l$ , яка може бути визначена шляхом оптимізації, наприклад, з використанням алгоритму Adam.

## 4.2. Впровадження методу «ковзного вікна» для онлайн–навчання

4.2.1. Імплементация «ковзного вікна»: встановлення розміру «ковзного вікна» для вибору останніх текстових документів, що будуть використовуватися для постійного оновлення та навчання моделей.

4.2.2. Онлайн–оновлення моделей: періодичне оновлення моделей ансамблю, використовуючи найсвіжіші дані, що надходять, та відкидаючи застарілі дані, для підтримання актуальності та високої точності прогнозування.

При надходженні нових даних  $D_{new}$ , оновлюється навчальний набір:

$$D_{train} = D_{train}^{old} \cup D_{new} \setminus D_{oldest}, \quad (4.13)$$

де  $D_{oldest}$  – найстаріші дані, що виходять за межі ковзного вікна.

Моделі перенавчаються на оновленому  $D_{train}$ .

4.2.3. Адаптація до змін у даних: неперервна адаптація метамоделі до змін у текстових даних, що забезпечує точнішу відповідь на динамічність і нестационарність текстових послідовностей.

Крок 5. Адаптація та перенавчання (блок 8). Цей етап має на меті забезпечення системі здатністю адаптуватися до еволюції та зміни форм дезінформації. Розглянемо деталі цього кроку:

### 5.1. Перенавчання системи.

5.1.1. Збір нових даних. Постійний збір нових даних з відкритих джерел для відображення останніх трендів та шаблонів дезінформації.

5.1.2. Оцінка потреби в перенавчанні. Аналіз поточної точності визначення, чи виникла потреба в перенавчанні на основі нових даних.

5.1.3. Перенавчання моделі. Застосування процесу навчання моделі з використанням нових даних для адаптації моделі до нових форм дезінформації.

### 5.2. Оновлення моделей та алгоритмів.

5.2.1. Аналіз нових алгоритмів та технологій. Оцінка та аналіз нових алгоритмів та технологій, які можуть бути використані для покращення точності.

5.2.2. Оновлення алгоритмів. Внесення змін у алгоритми та методи, що використовуються на основі отриманих відомостей та аналізу результатів.

5.2.3. Тестування та валідація оновлених моделей. Проведення тестування та валідації оновлених моделей для забезпечення їхньої точності.

### 5.3. Моніторинг та оцінка.

5.3.1. Моніторинг точності. Постійний моніторинг виявлення можливих проблем або областей для покращення.

5.3.2. Зворотний зв'язок та адаптація. Збір та аналіз зворотного зв'язку від користувачів та експертів для подальшого покращення та адаптації системи.

Розроблений метод виявлення дезінформації реалізовано в підрозділі 4.4, де представлені експериментальні результати його застосування. Зокрема, у підрозділі 4.4.1 демонструється ефективність методу у виявленні фейкової інформації в реальному часі, що підтверджує його практичну значущість і доцільність використання в інфраструктурі ТГ.

### 4.3 Інтелектуальний метод формування короткострокового проєкту управління персоналом

У цьому підрозділі розглядається інтелектуальний метод формування короткострокового проєкту управління персоналом, який базується на методі класифікаційного аналізу текстових соціально–економічних даних ТГ, описаному в підрозділі 4.1. Цей підхід використовує інтелектуальні інформаційні технології та аналіз масивів даних для підтримки інфраструктури ТГ, особливо в ситуаціях, що вимагають швидкого реагування на кризові моменти. Метод сприяє спрощенню процесу управління персоналом та ресурсами, оптимізуючи відбір кандидатів і зменшуючи часові та фінансові витрати на рекрутинг.

Запропонований метод формування команд для короткострокових проєктів відрізняється від аналогів [370] та [424] (див. підрозділ 1.2.2) використанням інтелектуального чат–бота для оптимізації процесу відбору, який включає розробку замкненого віртуального середовища та автоматизує вибір кандидатів, що приводить до зменшення часових та фінансових витрат на рекрутинг.

Запропонований метод має ключові кроки, які демонструються на алгоритмі (див. додаток А, рис. А.8):

Крок 1. З’являється новий проєкт (блок 1). На цьому етапі з’являється новий проєкт  $P$ , який має бути реалізований і який характеризується набором параметрів  $P = \{p_1, p_2, \dots, p_n\}$ , де  $p_i$  – конкретні характеристики, такі як тип проєкту, складність, обсяг роботи тощо.

Крок 2. Менеджер проєкту ініціює запит (блок 2). Менеджер проєкту  $M$  створює запит RRR для формування команди, який можна описати як функцію від параметрів проєкту:

$$R = f(P) = \{r_1, r_2, \dots, r_m\}, \quad (4.14)$$

де  $r_j$  – ключові вимоги до команди, такі як: необхідні навички, досвід, кількість фахівців тощо.

Крок 3. Встановлення ключових параметрів проєкту (блок 3). Менеджер визначає набір критеріїв  $C = \{c_1, c_2, \dots, c_k\}$  для відбору кандидатів, де  $c_i$  – конкретний критерій (наприклад, мінімальний досвід  $d_{min}$  наання певних технологій тощо). Також встановлюються параметри: дата початку  $T_{start}$  і дата завершення  $T_{end}$ ; загальний бюджет  $B$  або зарплатний фонд для команди; кількість фахівців  $N$ , необхідних для проєкту; набір необхідних навичок  $S = \{s_1, s_2, \dots, s_l\}$ .

Крок 4. Запит надходить у базу даних фрилансерів (блок 4). Запит  $R$  надходить до бази даних фрилансерів  $F = \{F_1, F_2, \dots, F_q\}$ , де кожен фрилансер  $F_i$  має профіль з характеристиками  $A_i = \{a_{i1}, a_{i2}, \dots, a_{ip}\}$ , де  $a_{ij}$  – параметр, наприклад, досвід, навички, рейтинг тощо.

Крок 5. Активізація чат-бота (блок 5). Чат-бот  $B$  ініціюється для проведення співбесід з потенційними кандидатами, використовуючи набір сценаріїв  $S$  та критерії  $C$ , встановлені менеджером.

Крок 6. Проведення співбесіди з кандидатами (блок 6). Для кожного кандидата  $F_i$  чат-бот проводить сесію  $I_i$ , яка складається з послідовності питань  $Q = \{q_1, q_2, \dots, q_n\}$  та відповідей  $A_i = \{a_{i1}, a_{i2}, \dots, a_{in}\}$ :

$$I_i = \{(q_j, a_{ij}) \mid j = 1, 2, \dots, n\}. \quad (4.15)$$

Крок 7. Фіксація відповідей у базі даних (блок 7). Відповіді кандидатів зберігаються у базі даних. Для кожного кандидата формується вектор ознак  $x_i \in R_m$ , де  $m$  – кількість ознак, отриманих з відповідей і профілю кандидата.

Крок 8. Конвертація відповідей у двійковий формат (блок 8). Вектори ознак  $x_i$  перетворюються на двійковий формат  $b_i \in \{0,1\}^m$ , де кожен елемент  $b_{ij}$  відповідає наявності або відсутності певної ознаки або виконанню критерію:

$$b_{ij} = \begin{cases} 1, & \text{якщо ознака або критерій виконаний,} \\ 0, & \text{інакше.} \end{cases} \quad (4.16)$$

Крок 9. Аналіз отриманих даних за допомогою алгоритмів машинного навчання (блок 9). Використовуючи матрицю двійкових ознак  $B = [b_1^T, b_2^T, \dots, b_n^T]^T$ ,

де  $n$  – кількість кандидатів, проводиться аналіз за допомогою алгоритмів машинного навчання з метою класифікації кандидатів.

Крок 10. Використання методів дерев рішень (блок 10). Будується модель дерева рішень  $T$ , яка навчається на навчальній вибірці  $D_{train} = \{(b_i, y_i)\}$ , де  $y_i \in \{0,1\}$  – мітка класу (відповідає чи не відповідає кандидат вимогам). Мета навчання – мінімізувати функцію втрат, наприклад, ентропію або критерій Джині:

$$Loss = \sum_{\text{вузли дерева}} (\text{кількість зразків} \times \text{ентропія}). \quad (4.17)$$

Крок 11. Використання готової навчальної вибірки (блок 11). Для навчання моделі  $T$  використовується готова навчальна вибірка з попередніх проєктів, що дозволяє моделі вчитися на реальних даних про кандидатів і рішеннях менеджерів.

Крок 12. Отримані результати (блок 12). Після навчання модель  $T$  застосовується до нових кандидатів для прогнозування їх відповідності:

$$\hat{y}_i = T(b_i), \quad (4.18)$$

де  $\hat{y}_i \in \{0,1\}$  – прогноз моделі для кандидата  $F_i$ .

Крок 13. Оцінка результатів менеджером проєкту (блок 13). Менеджер аналізує результати моделі і може коригувати рішення на основі додаткової інформації або власного досвіду:

$$y_i^{final} = \begin{cases} 1, \text{ якщо менеджер затверджує кандидата,} \\ 0, \text{ якщо відхиляє.} \end{cases} \quad (4.19)$$

Крок 14. Сповіщення обраних кандидатів (блок 14). Кандидатам, для яких  $y_i^{final} = 1$ , чат-бот надсилає повідомлення про прийняття у проєкт.

Крок 15. Завершення проєкту (блок 15). Після завершення проєкту збираються результати роботи команди, включаючи виконання завдань, дотримання термінів тощо.

Крок 16. Оцінка роботи членів команди менеджером (блок 16). Менеджер оцінює продуктивність кожного члена команди  $F_i$ , присвоюючи їм оцінку  $s_i$  за певною шкалою (наприклад, від 1 до 5):

$$s_i = PerformanceMetric(F_i). \quad (4.20)$$



Ці оцінки додаються до профілів фрилансерів і можуть використовуватися для майбутніх відборів.

Удосконалення майбутніх HR–проектів. Дані про продуктивність  $\{s_i\}$  використовуються для оновлення навчальної вибірки:

$$D_{train} = D_{train} \cup \{(b_i, y_i^{final}, s_i)\}. \quad (4.21)$$

Це дозволяє покращити модель  $T$ , враховуючи реальну ефективність кандидатів у проектах, що сприяє удосконаленню майбутніх HR–проектів, реалізованих з використанням машинного навчання.

Метод формування короткострокового HR–проекту, представлений у цьому підрозділі, реалізовано в експериментальних результатах підрозділу 4.4, зокрема у підрозділі 4.4.2. Це демонструє практичне застосування розробленого методу та його ефективність в оптимізації процесу відбору персоналу, що сприяє підвищенню продуктивності та успішності реалізації короткострокових проектів у ТГ.

## 4.4 Експериментальні результати

За запропонованим методом класифікаційного аналізу текстових соціально–економічних даних ТГ (див. підрозділ 4.1) та принципом синтезу інформаційної технології (див. підрозділ 2.1) розроблено модулі інформаційної системи, які вирішують поставлені задачі виявлення дезінформації та формування короткострокового проекту управління персоналом.

### 4.4.1 Виявлення дезінформації

Для реалізації запропонованого методу виявлення дезінформації (див. підрозділ 4.2) обрано набір даних, представлений на Kaggle [402] – це унікальна колекція з приблизно 60 тис. заголовків новин, зібраних з 24 лютого по 11 грудня 2022 року, що охоплює період повномасштабної російсько–української війни. Він включає в себе як перевірені, так і фальшиві новини, зібрані з українських телеграм–каналів та

російських каналів з фейками, роблячи його найбільшим відкритим джерелом відповідних даних.

Набір містить два основні атрибути: текст заголовка новини і мітку, що вказує на достовірність (True – для підтверджених новин та False – для фейкових). Загалом, він містить 4522 записів з міткою 'False'. Джерелами даних слугували такі телеграм-канали, як «СУСПІЛЬНЕ НОВИНИ», «Pererichka NEWS» та ін.

На початковому етапі проведено підготовку даних, де текстові матеріали були перетворені на числовий формат за допомогою методу TF-IDF векторизації. Далі, на основі цих даних, було треновано кілька індивідуальних моделей, включаючи логістичну регресію, SVM, випадковий ліс, градієнтний бустинг, KNN, дерево рішень, XGBoost та AdaBoost. Кожна модель була адаптована та оптимізована для вирішення задачі класифікації з використанням тренувального набору даних.

На другому етапі реалізації було сформовано метамодель на основі XGBoost, яка була тренована з використанням прогнозів, отриманих від індивідуальних моделей через механізм стекінгу. Оцінка точності класифікації здійснена на тестовому наборі даних, де результати представлені через звіт про класифікацію та матрицю помилок, візуалізовану у вигляді теплової карти.

Цей підхід демонструє високий потенціал ансамблевих методів та стекінгу для підвищення точності моделей на основі машинного навчання у складних задачах класифікації тексту, зокрема, у виявленні дезінформації.

Отримані результати класифікації представлені у форматі звіту (Рис. 4.2), який включає такі показники, як точність (precision), повнота (recall), f1-бал (f1-score) та підтримка (support) для двох класів: «False» (фейкові новини) та «True» (правдиві новини). У процесі оцінки точності моделі класифікації новин було проведено аналіз її здатності диференціювати фактично правдиві від фейкових повідомлень. Згідно з отриманими результатами, модель демонструє високу загальну точність класифікації на рівні 93 %. Для класу «False», який умовно припускається визначати фейкові новини, точність (precision) склала 0,95, що свідчить про високу вірогідність коректної класифікації новини як фейкової, коли модель так визначає. Водночас, показник відновлення (recall) для цього класу становить 0,72, що вказує на те, що

близько 28 % фейкових новин було пропущено моделлю. З іншого боку, для класу «True», який відповідає правдивим новинам, модель показала вражаючий показник відновлення на рівні 0,99, що означає успішну ідентифікацію 99 % правдивих новин. Така здатність моделі точно класифікувати правдиві новини є важливою характеристикою в контексті боротьби з дезінформацією.

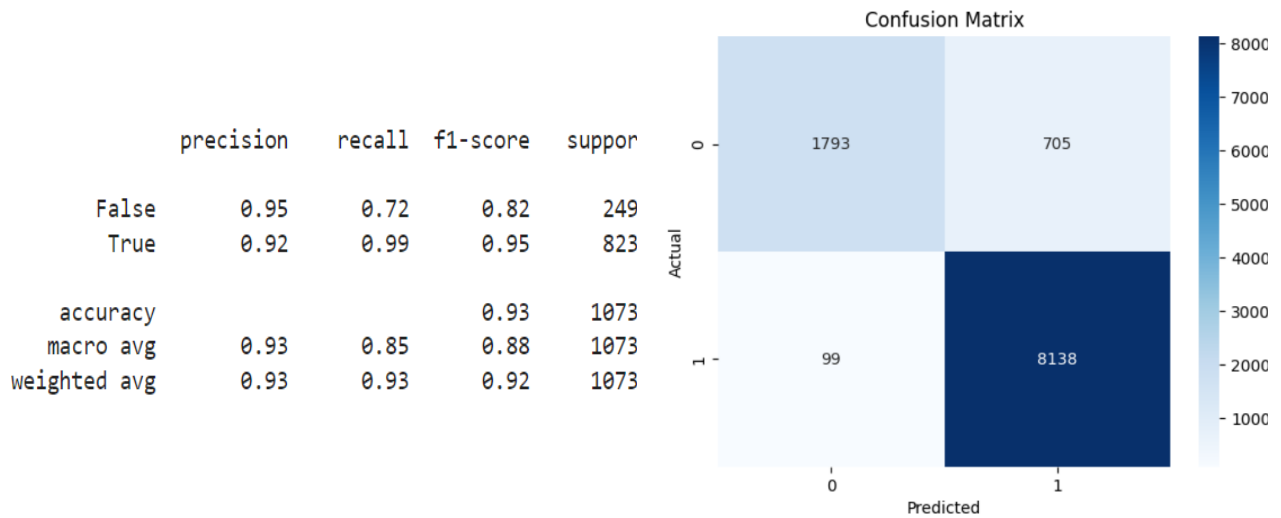


Рис. 4.2. Результати оцінювання

Аналітичний підхід до оцінки роботи моделі також охоплював аналіз матриці помилок, у якій відзначено 1793 істинно–негативних результатів (True Negative), що підтверджує коректну класифікацію фейкових новин. Однак, виявлено 705 хибнонегативних результатів (False Negative), що вказує на існування певної кількості фейкових новин, які модель помилково класифікувала як правдиві. Для класу правдивих новин зареєстровано 8138 істинно–позитивних результатів (True Positive) та 99 хибно–позитивних (False Positive), демонструючи високу точність виявлення фактично правдивого контенту. F1–оцінка для обох класів, яка гармонізує точність і відновлення, виявилась на рівні 0,82 для «False» і 0,95 для «True», що в цілому підтверджує збалансованість та надійність моделі.

Таблиця 4.1 містить результати експериментального дослідження, отримані внаслідок застосування розробленого методу та його порівняння з класичними підходами класифікації.

## Порівняння з розробленим методом

Model	Precision (False)	Precision (True)	Recall (False)	Recall (True)	F1-score (False)	F1-score (True)	Accuracy	Confusion Matrix (False, True)
Logistic Regression	0,96	0,85	0,44	0,99	0,60	0,92	86,42 %	(1091, 1407), (51, 8186)
SVM	0,90	0,93	0,75	0,99	0,85	0,96	90,61 %	(1680, 818), (68, 8169)
Random Forest	0,88	0,95	0,81	1,00	0,89	0,97	90,31 %	(2026, 472), (31, 8206)
Gradient Boosting	0,94	0,86	0,48	0,99	0,64	0,92	87,31 %	(1209, 1289), (73, 8164)
KNN	0,91	0,87	0,51	0,99	0,66	0,92	87,57 %	(1285, 1213), (121, 8116)
Decision Tree	0,81	0,95	0,83	0,94	0,82	0,94	91,45 %	(2077, 421), (497, 7740)
XGBoost	0,89	0,89	0,61	0,98	0,73	0,93	89,19 %	(1533, 965), (195, 8042)
AdaBoost	0,88	0,87	0,50	0,98	0,64	0,92	86,92 %	(1259, 1239), (165, 8072)
Запропонований метод	0,95	0,92	0,72	0,99	0,82	0,95	93%	(1793, 705), (99, 8138)

Аналізуючи ці результати, можемо зробити висновки щодо ефективності розробленого методу порівняно з традиційними підходами класифікації. Розглядаючи кожну метрику окремо, очевидно, що запропонований метод демонструє суттєві покращення за більшістю параметрів.

У контексті точності (precision) для класу «False» розроблений метод досягає значення 0,95, що є порівнянним з логістичною регресією та перевершує інші класичні методи, крім Gradient Boosting. Це свідчить про високу здатність коректно ідентифікувати фейкові новини. Щодо класу «True», то точність становить 0,92, що є конкурентоспроможним показником, особливо порівняно з SVM та XGBoost.

Показник відновлення (recall) для класу «False» у розробленому методі на рівні 0,72 значно вищий, ніж у логістичної регресії і відповідає показникам SVM, що вказує

на покращену здатність моделі виявляти фейкові новини серед негативного класу. Для класу «True» показник відновлення становить 0,99, що є спільною тенденцією для всіх розглянутих моделей, підкреслюючи їхню здатність до ідентифікації правдивих новин.

Гармонійне середнє між точністю та відновленням, визначене як F1–оцінка, для складає 0,82 для класу «False» та 0,95 – для класу «True», що підтверджує збалансованість моделі між цими двома важливими показниками. Це значно перевищує F1–оцінки логістичної регресії для класу «False» та є на рівні з найкращими результатами серед інших класичних методів.

Отже, загальна точність розробленого методу становить 93 %, що є одним з найвищих показників серед усіх розглянутих моделей, і підтверджує її сильні позиції як надійного інструмента для класифікації новин. Матриця помилок також свідчить про високу кількість правильно класифікованих екземплярів для обох класів.

Інтелектуальний модуль виявлення дезінформації (Рис. 4.3) розпочинається зі збору даних (блок 1), де використовуються різноманітні джерела для збору текстових та метаданих, необхідних для аналізу.

Далі, на кроці попередньої обробки даних (блок 2), виконується обробка природної мови, включаючи токенізацію, стемінг та розпізнавання іменованих сутностей. Крок векторизації тексту (блок 3) перетворює текстові дані у числовий формат для аналізу, використовуючи методи, такі як Word2Vec, GloVe чи BERT. Крок визначення схожості тексту (блок 4) використовує косинусну схожість та інші метрики для групування схожих текстів, а аналіз тональності та емоцій (блок 5) допомагає виявити настрої та емоційні відгуки в текстах. Ці кроки разом формують комплексний підхід до виявлення (блок 6), аналізу та класифікації дезінформації в режимі реального часу, забезпечуючи потужний інструмент для інформаційної безпеки.

Інтелектуальний модуль для виявлення дезінформації (Рис. 4.4) функціонує шляхом аналізу текстових даних, введених користувачем, з метою ідентифікації потенційно фальшивої інформації.

## Інтелектуальний модуль виявлення дезінформації

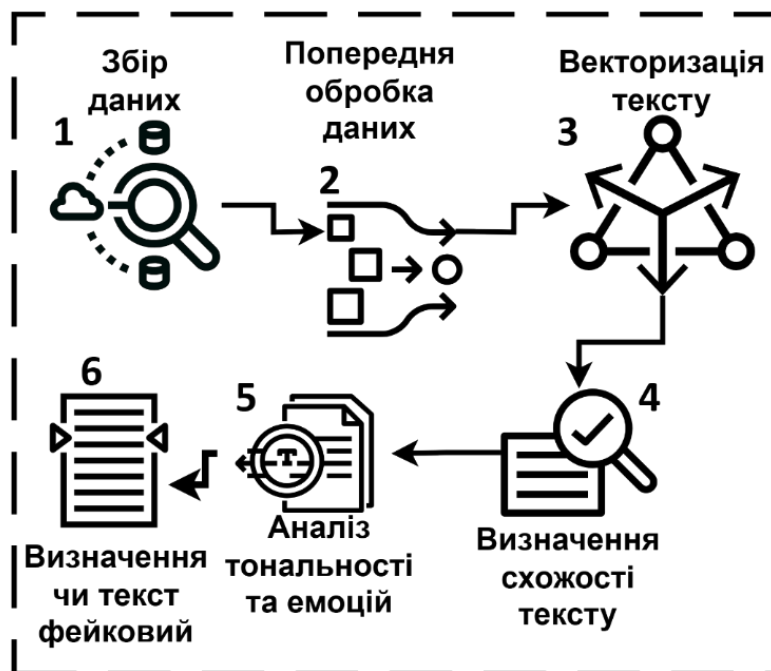


Рис. 4.3. Інтелектуальний модуль виявлення дезінформації

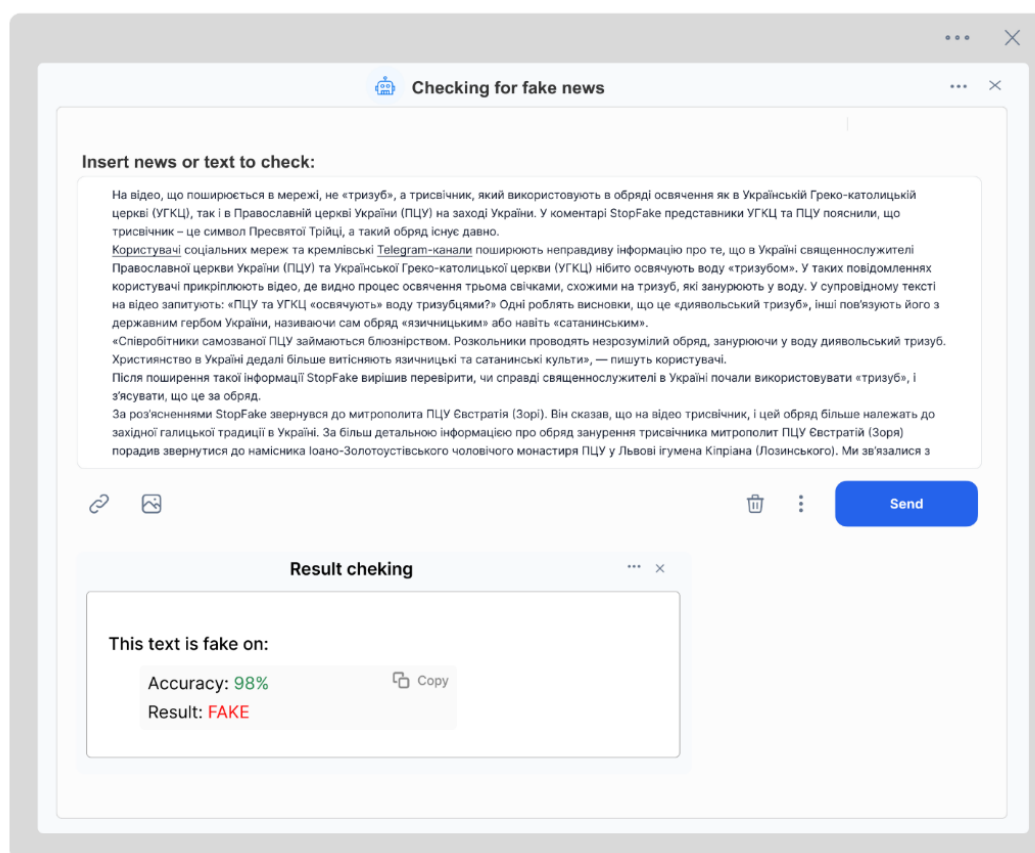


Рис. 4.4. Інтелектуальний модуль виявлення дезінформації

Система обробляє введений текст та надає висновок у формі оцінки, що вказує на ймовірність класифікації даної інформації як фейкової.

Для міської адміністрації використання такого модуля є важливим, оскільки це сприяє підтриманню інформаційної прозорості та достовірності в громадському просторі. Забезпечення надійної перевірки інформації також важливо для підвищення довіри громадян до офіційних повідомлень та зниження впливу дезінформації на громадську думку.

Результати аналізу свідчать про високу ефективність розробленої системи у виявленні дезінформації (див. додаток Г). Середній час виконання завдань із застосуванням системи становить від 1 до 2 годин, тоді як без системи цей процес займає від 1 до 5 днів. Ефективність системи оцінена на рівні 99,17 %, що підтверджує її здатність суттєво скорочувати час аналізу та підвищувати якість і точність результатів порівняно з традиційними підходами.

#### **4.4.2 Формування HR-проєкту**

Для розробки інтелектуальної функціональності чат-бота використовувалося середовище розробки DialogFlow [100], яке трансформує запити користувачів в оброблювані дані, дозволяючи чат-боту самостійно навчатися за допомогою спеціалізованих бібліотек.

Система інтерв'ювання, орієнтована на оцінку м'яких і твердих навичок кандидатів [122], включає три категорії питань (див. додаток Б, Табл. Б.3): загальні, пов'язані із загальною інформацією, м'які та тверді навички. Початковий крок взаємодії чат-бота з кандидатом передбачає збір базової інформації, такої як особисті дані та професійний досвід. Після цього чат-бот пропонує психологічні питання для оцінки м'яких навичок та на завершення адресує питання щодо рівня твердих навичок, необхідних для конкретного проєкту. Усі отримані відповіді реєструються у базі даних, як дозволяє редагування та розширення, і використовуються для розробки індивідуального алгоритму обробки відповідей для кожної HR-позиції, наприклад, розробника.

Метод дерев рішень використовується для аналізу кваліфікації претендентів, де відсутні ідеальні зразки. Тренувальна база поповнюється найбільш підходящими кандидатами, включаючи їхні дані про ефективність, оцінені менеджером після завершення проєкту. Використовуючи мову програмування Python та бібліотеки, такі як pandas, numpy, train\_test\_split та RandomForestClassifier, чат-бот та обробка даних реалізованих з використанням платформи Telegram для чат-бота з урахуванням її популярності і зручності для інтеграції та оновлення. Початкові розмови (Рис. 4.5) з претендентами формуються через базу знань DialogFlow та імпортуються в базу даних для подальшого аналізу кандидатів, де кожне ключове питання позначене спеціальними тегами.



Рис. 4.5. Діалог із заявником у чаті

Після того, як отримано доступ до бази даних і інформація перетворюється в логічний формат, зосереджуються на обробці відповідей на питання під номером 1\_4. Використовуючи словник meat\_HR, де розробник відповідає 1, а менеджер проєкту та фронт-енд – 0, далі мапуються відповіді та оновлюємо набір даних HR. Після цього видаляємо колонку '1\_4' та перетворюємо всі значення в цифровий формат. На завершення, ми використовуємо бібліотеку Random Forest Classifier зі



100 оцінювачами, навчаємо модель на даних HR та мітках, щоб побудувати дерево рішень для класифікації кандидатів (Рис. 4.6).

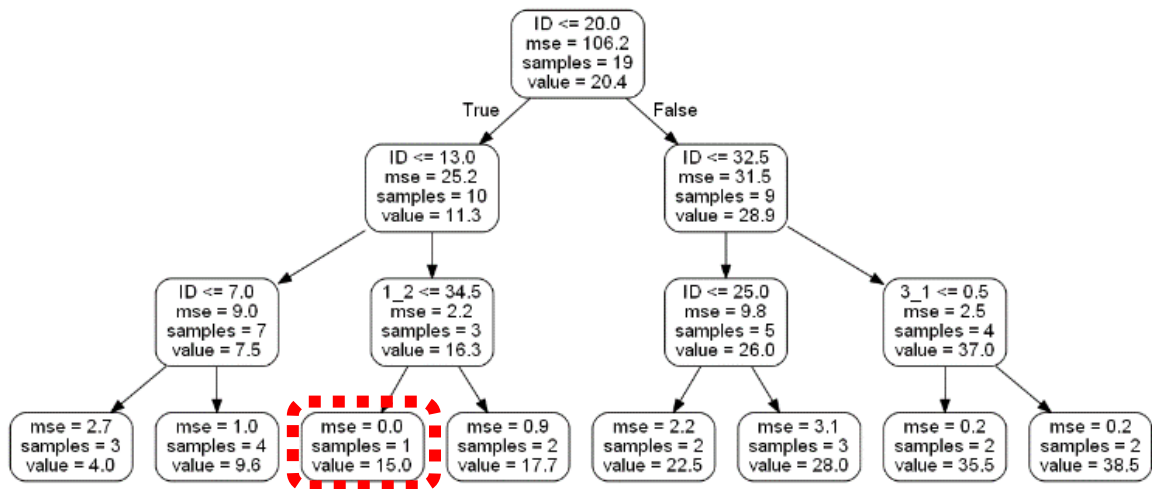


Рис. 4.6. Дерево рішень для вибору найкращого кандидата на посаду розробника

Обчислення виявили, що для сьомого елемента вибірки середньоквадратична помилка (MSE) дорівнює 9, що відповідає кандидату з ID7. Точність у визначенні найбільш підходящого кандидата складає близько 94 %.

Згідно з результатами дерева рішень, кандидат з ID має найоптимальніші якості для ролі розробника, оскільки його відповіді були найближчими до ідеалу, як показано у навчальній вибірці. Результати аналізу представляються менеджеру, який ухвалює остаточне рішення та інформує про нього всіх кандидатів за допомогою чат-бота. Відмінність розробленого методу від існуючих аналогів [370, 424] полягає в тому, що він дозволяє автоматизувати процес підбору учасників для короткострокових проєктів у віртуальному середовищі, мінімізуючи необхідність втручання HR-менеджера.

Інтелектуальний процес формування HR-команди (Рис. 4.7) проєкту на основі машинного навчання починається з надходження нового проєкту (блок 4.1), де керівник встановлює його параметри. Далі запит на створення команди та вимоги керівника надсилаються в базу фрілансерів (блок 4.3), а також проводиться парсинг резюме (блок 4.4) з інших ресурсів. Чат-бот інтелектуально проводить співбесіди з

претендентами (блок 4.5) з використанням сценаріїв та параметрів, а результати зберігаються в базі даних (блок 4.6).

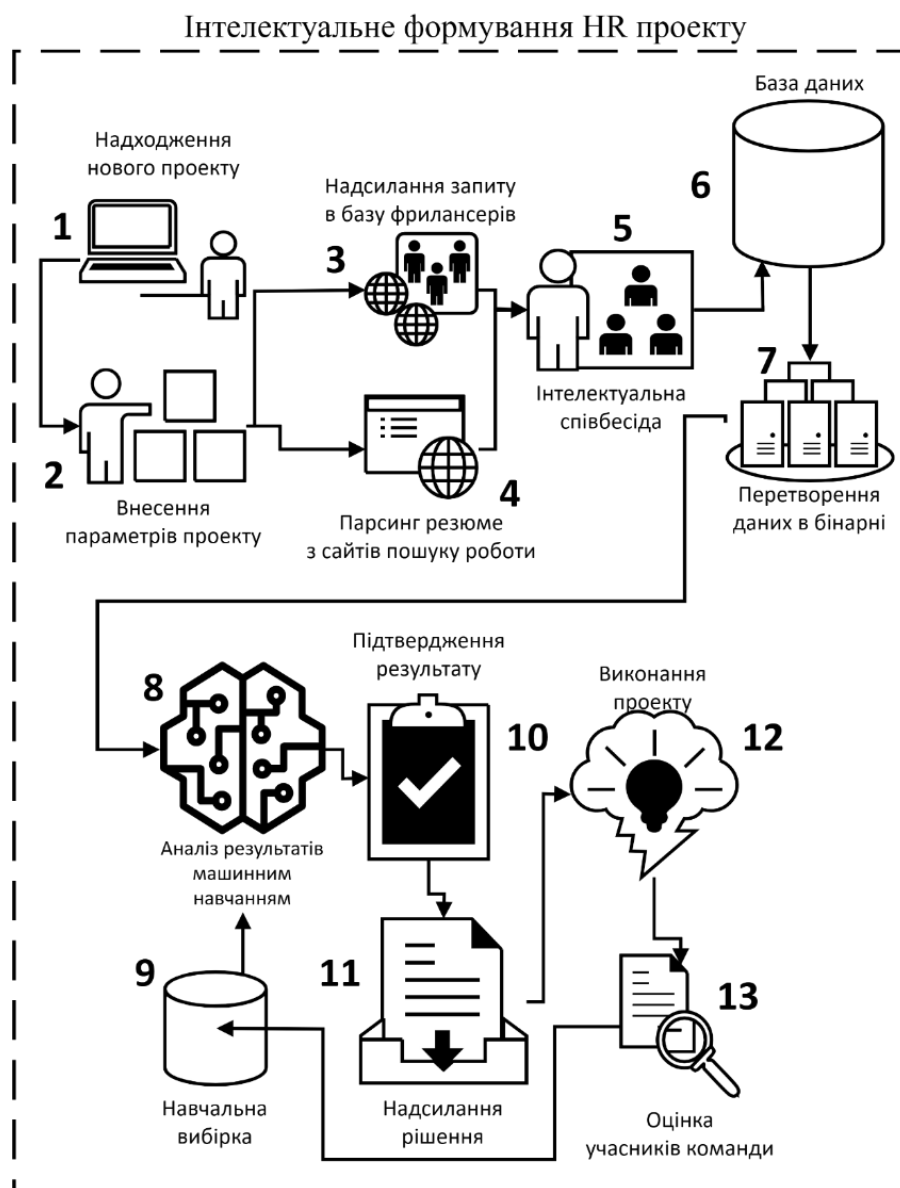


Рис. 4.7. Блок інтелектуального формування HR-команди проекту

Отримані дані перетворюються в бінарний формат (блок 4.7), після чого використовуються для аналізу (блок 4.8) за допомогою методів машинного навчання та навчальної вибірки (блок 4.9).

Керівник може підтвердити або відхилити результат (блок 4.10), а в разі позитивного рішення, чат-бот надсилає рішення погодженим претендентам (блок

4.11). Після формування команди розпочинається виконання проєкту (блок 4.12), керівник оцінює роботу учасників команди (блок 4.13), і ці дані використовуються для навчання системи (блок 4.9).

Отже, підприємець вказує, який час він виділяє для пошуку кандидатів (див. Рис. 4.8). Система запускає процес пошуку претендентів, який включає в себе парсинг резюме та відправку запитів на сайти фрилансу. Після цього система розпочинає діалог з потенційними кандидатами, досліджуючи їхні навички та досвід.

The screenshot displays the SmartMed application interface. On the left is a search form with the following fields and values:

- Download the questions and correct answers:** A file named "Applican\_questions.csv" (50 Kb) is available for download.
- Full Name and Email:** name
- Age:** 27
- Which position are you applying for?:** administration
- Have you participated in similar projects before?:** yes
- Work experience?:** 3
- If you have experience in similar projects, please provide the project name:** Project
- Do you have a medical background?:** yes
- Are you married?:** no
- Are you currently employed?:** no
- Inputs to get started:**
  - Position:** Input
  - Project:** Input
  - Search time:** Input
- Start search** button

On the right side, there are three main sections:

- Recommended employees:** A table listing candidates with their ID, PIC, Name, Message count, Portfolio, Experience, Salary, and Confirmation status.
 

ID	PIC	NAME	MESEG	PORTFOLIO	EXP	SALARY	CONFIRMATION
3352		Jane Cooper	5	Dribbble	3 Yrs	\$300-350	ACC
2363		Wade Warren	5	LinkedIn	3 Yrs	\$300-350	REJECT
1254		Esther Howard	5	LinkedIn	3 Yrs	\$300-350	REJECT
3365		Cameron Williamson	5	LinkedIn	3 Yrs	\$300-350	ACC
0376		Brooklyn Simmons	5	LinkedIn	3 Yrs	\$300-350	ACC
- Formed team:** A table listing the selected team members with their PIC, Name, Message count, Portfolio, and Job Title.
 

PIC	NAME	MESEG	PORTFOLIO	JOB TITLE
3352	Jane Cooper	5	Dribbble	Administration
3355	Cameron Williamson	5	Dribbble	Administration
0376	Brooklyn Simmons	5	Website	Administration
- Messages:** A list of recent messages from other users:
  - Andreana Viola (08:30): Hi, How are you today?
  - Francesco Long (07:00): Hi @Angel, I hope you are doing well...
  - Alexandra Michu (23/11): Hi, How are you today?
- Chat Bot:** A chat window with Alexandra Michu (Online) showing a conversation:
  - Today
  - Hi, How are you today? (9:00 AM)
  - Hello Michu, I'm fine. How about you? (9:10 AM)
  - Can you send a sample from a reference? (9:11 AM)

Рис. 4.8. Вікно інтелектуального підбору кращих працівників

На основі отриманих даних система формує вибірку найкращих претендентів, використовуючи метод `RandomForestClassifier`, та надсилає результати підприємцю для його розгляду та погодження.

Якщо результат не влаштовує підприємця, він має можливість перезапустити пошук та вибірку, щоби знайти найбільш підходящих кандидатів.

З результатів класифікації видно, що претенденти з ID: 3352, 2363, 1254, 3365, 0376 (див. Рис. 4.8) мають найкращі характеристики на посаду адміністратора. Ці претенденти дали відповіді, найближчі до ідеальних. Далі підприємець обирає тих, хто йому більше до вподоби і формує свою команду.

Результати оцінки формування HR–проєкту демонструють значне скорочення часу завдяки використанню модуля системи (див. додаток Г). За даними експертів, час на виконання завдань із системою варіювався від 1 до 120 годин, тоді як без системи цей показник коливався від 5 до 30 днів. Ефективність модуля системи досягла рівня ~80 %, що свідчить про її високу здатність оптимізувати процеси управління персоналом та зменшити часові витрати на підготовку і проведення інтерв'ю.

## 4.5 Висновки

У розділі 4 представлено метод класифікаційного аналізу текстових соціально–економічних даних ТГ. У результаті проведеної роботи вдалося розробити інтелектуальні підходи до аналізу даних, адаптовані до специфіки задач ТГ, таких як виявлення дезінформації, формування короткострокових HR–проєктів та вибір конкурентоспроможних товарів. Наведені нижче висновки детально описують досягнення кожного підрозділу та їх значення.

Розроблено метод класифікаційного аналізу текстових соціально–економічних даних ТГ, що інтегрує різномірні джерела даних та сучасні алгоритми класифікації. Це дало змогу забезпечити стійкий розвиток громад шляхом автоматизації процесів прийняття рішень на основі аналізу великих обсягів текстових даних.

Розроблено інтелектуальний метод виявлення дезінформації. Основним досягненням стало використання ансамблевих методів, таких як стекінг, для підвищення точності класифікації текстів. Отримані результати підтвердили ефективність запропонованого підходу, що дозволило досягти загальної точності моделі 93%. Це сприяє підвищенню інформаційної безпеки в громадах шляхом швидкого і надійного виявлення фейкових новин.

Розроблено метод формування короткострокових HR-проектів, який використовує інтелектуальні чат-боти для автоматизації відбору кандидатів та аналізу їх навичок. Використання дерев рішень для оцінки кандидатів дозволило значно зменшити часові витрати на рекрутинг, підвищити об'єктивність процесу та покращити якість формування команд. Цей підхід став важливим інструментом для оперативного реагування на потреби громади.

Проведено експериментальну оцінку розроблених методів, зокрема, результати класифікації дезінформації показали високу ефективність порівняно з класичними методами. У завданні формування HR-проектів час на виконання процесу скоротився з 5–30 днів до 1–120 годин, демонструючи переваги інтелектуальних рішень у практичному застосуванні.

Розроблені у розділі методи та підходи стали основою для впровадження інтелектуальних інформаційних технологій, які сприяють покращенню ефективності управління в ТГ. Запропоновані рішення забезпечують автоматизацію та точність аналізу текстових даних, що має значний потенціал для підтримки розвитку соціально-економічних процесів у громадах.

## РОЗДІЛ 5.

### МЕТОДИ ТА ЗАСОБИ КЛАСТЕРНОГО АНАЛІЗУ СОЦІАЛЬНО– ЕКОНОМІЧНИХ ДАНИХ ТЕРИТОРІАЛЬНИХ ГРОМАД

#### 5.1 Метод кластерного аналізу соціально–економічних даних територіальних громад

На основі узагальненого принципу синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ (див. підрозділ 2.1) та описаних задач запропоновано метод кластерного аналізу соціально–економічних даних ТГ для забезпечення ефективного прийняття рішень. Пропонований метод орієнтований на ідентифікацію груп об'єктів зі схожими характеристиками для покращення управління ресурсами та підтримки соціально–економічної стабільності.

На рис. 5.1 представлено кроки методу кластерного аналізу соціально–економічних даних ТГ, який деталізує етапи збору, обробки, аналізу даних та формування управлінських рішень.



Рис. 5.1. Кроки методу кластерного аналізу соціально–економічних даних ТГ

Нехай множина об'єктів інфраструктури ТГ позначається як:

$$I = \{I_1, I_2, I_3, I_4\}, \quad (5.1)$$

де  $I_1$  – бізнес–інфраструктура, включаючи підприємства та ІТ–компанії;  $I_2$  – соціальна інфраструктура, відповідальна за аналіз споживчих потреб;  $I_3$  – логістична інфраструктура, що забезпечує постачання та облік ресурсів;  $I_4$  – інформаційна інфраструктура, яка включає джерела даних про настрої населення та медіаконтент.

Соціально–економічні дані, що використовуються в аналізі, мають різну природу, що формалізується як множина типів даних:

$$D = \{D_1, D_2, D_3\}, \quad (5.2)$$

де  $D_1$  – структуровані дані, такі як фінансові звіти, демографічні показники;  $D_2$  – неструктуровані дані: текстові документи, соціальні медіа, відео;  $D_3$  – напівструктуровані дані у форматах JSON, XML, отримані через API.

Для реалізації кластерного аналізу використовується множина методів інтелектуального аналізу:

$$A = \{K - \text{Means}, \text{DBSCAN}, \text{GMM}, \text{PCA}\}, \quad (5.3)$$

де кожен метод  $A_k$  має свою область застосування залежно від характеристик даних:

– K–Means – метод кластеризації, що розділяє дані на  $k$  груп, мінімізуючи суму квадратичних відстаней між кожною точкою і центром її кластера. Використовується для виявлення схожих груп у даних.

– DBSCAN (Density–Based Spatial Clustering of Applications with Noise) – метод кластеризації, що ґрунтується на щільності даних. Дозволяє виявляти кластери довільної форми та ізолювати шуми (аномалії).

– GMM (Gaussian Mixture Model) – метод кластеризації, що використовує моделі Гауссових розподілів для представлення кластерів. Підходить для виявлення груп у даних з перекриттям між кластерами.

– PCA (Principal Component Analysis) – метод зменшення розмірності даних, який дозволяє зберегти якомога більше інформації, проєктуючи дані на нові координати, що максимізують варіацію. Використовується для візуалізації та підготовки даних перед кластеризацією.

Процес інтеграції даних, який є ключовим етапом, формалізується як:

$$D_{\text{інтегровані}} = D_1 \oplus D_2 \oplus D_3, \quad (5.4)$$

де оператор  $\oplus$  виконує злиття даних з різних джерел для формування єдиної бази даних.

Кожне завдання кластерного аналізу формується як трійка:

$$Z_l = (I_j, D_i, A_k), \quad (5.5)$$

де  $I_j \in I, D_i \in D, A_k \in A$ . Завдання передбачає вибір об'єкта інфраструктури, типу даних та відповідного методу кластерного аналізу.

Рішення, отримане на основі аналізу, позначається як:

$$R_p = g(Z_l), \quad (5.6)$$

де  $g(Z_l)$  – функція, що визначає оптимальне управлінське рішення, наприклад:

- вибір соціальних груп для цільового розподілу ресурсів;
- ідентифікація кластерів підприємств з високим інвестиційним потенціалом;
- розробка планів логістичних маршрутів для забезпечення потреб громади.

На основі методу кластерного аналізу соціально–економічних даних ТГ [457] у наступних розділах розкрито інтелектуальні підходи до формування споживчого кошика [255] (підрозділ 5.2) та пошуку альтернативних постачальників [245] (підрозділ 5.3). У підрозділі 5.4 наведено експериментальні результати, які демонструють практичну реалізацію цих підходів, зокрема у формуванні споживчого кошика (підрозділ 5.4.1) та пошуку альтернативних постачальників (підрозділ 5.4.2). Розроблений метод підтверджує свою ефективність та адаптивність до різноманітних соціально–економічних задач і вже реалізований у науковій роботі [309]. Використання цього підходу сприяє покращенню якості прийняття управлінських рішень та підтримці сталого розвитку ТГ.



## 5.2 Інтелектуальний метод формування споживчого кошика

На основі методу кластерного аналізу соціально–економічних даних ТГ, розглянутого в підрозділі 5.1, розроблено інтелектуальний метод формування споживчого кошика. Цей метод спрямований на створення оптимального набору товарів, який враховує реальні потреби різних соціальних груп та адаптується до змінних умов, таких як війна чи пандемія.

На відміну від аналогів [265, 294, 295] (див. підрозділ 1.2.1) інтелектуальний метод формування споживчого кошика дозволяє, на основі даних з мереж супермаркетів, створювати оптимальний набір товарів у споживчому кошику та визначати прожитковий мінімум. Для швидкої адаптації набору товарів у споживчому кошику авторами розроблено цей інтелектуальний метод, який ілюструється схематично (див. додаток А, рис. А.9) та представлений наступними кроками:

Крок 1. Збір та попередня обробка даних (блок 1). На першому етапі здійснюється збір інформації з супермаркетів, що включає такі дані:

1.1. Ідентифікатори покупців  $ID_i$ , де  $i = 1, 2, \dots, N$ ,  $N$  – загальна кількість покупців.

1.2. Демографічні характеристики покупців: вік  $Age_i$ , дохід  $Income_i$ .

1.3. Деталі покупок:

– ідентифікатори транзакцій:  $T_j$ , де  $j = 1, 2, \dots, M$ ,  $M$  – загальна кількість транзакцій;

– дата транзакції:  $Date_j$ ;

– список придбаних товарів:  $Price_{jk}$ , де  $k$  – індекс товару в транзакції;

– ціна товарів:  $Price_{\{jk\}}$ .

Цей процес може бути розширений шляхом включення даних з аптек та відділів продовольчих і непродовольчих товарів, що дозволить отримати більш комплексну картину споживчої поведінки.

Крок 2. Кластеризація даних

2.1. Визначення об'єктів для кластеризації (блок 2). Об'єктами кластеризації виступають покупки або транзакції. Нехай  $X = \{x_i\}$ , де  $x_i \in R^p$  – вектор ознак об'єкта  $i$ ,  $p$  – кількість ознак.

2.2. Встановлення набору змінних (блок 3). Формується набір змінних для оцінки обраних об'єктів. Для кожного покупця  $i$  визначаються такі показники:

– середня вартість покупок:

$$\bar{P}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{k=1}^{K_{ij}} Price_{jk}, \quad (5.7)$$

де  $n_i$  – кількість транзакцій покупця  $i$ ,  $K_{ij}$  – кількість товарів у транзакції  $j$ .

– частота покупок:

$$f_i = \frac{T}{n_i}, \quad (5.8)$$

де  $T$  – загальний період спостереження;

– різноманітність придбаних товарів:

$$V_i = |\{Product_{jk} \mid \forall j, k\}|, \quad (5.9)$$

де  $|\cdot|$  – потужність множини;

– демографічні характеристики:  $Age_i, Income_i$ .

2.3. Нормалізація значень змінних (блок 4). З метою усунення масштабних відмінностей між ознаками проводиться нормалізація. Використовується, наприклад, стандартна нормалізація (z-score):

$$x_{ik}^{norm} = \frac{x_{ik} - \mu_k}{\sigma_k}, \quad (5.10)$$

де  $\mu_k$  та  $\sigma_k$  – відповідно, середнє значення та стандартне відхилення  $k$  – ї ознаки в вибірці.

2.4. Вимірювання ступеня схожості між об'єктами (блок 5). Визначається міра відстані між об'єктами. Використовується евклідова відстань:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik}^{norm} - x_{jk}^{norm})^2}. \quad (5.11)$$

Альтернативно можуть використовуватися інші метрики, такі як манхеттенська відстань або косинусна схожість.

2.5. Застосування методу кластерного аналізу (блок 6). Для утворення груп схожих об'єктів застосовується метод кластерного аналізу, наприклад, алгоритм K–Means:

– ініціалізація: випадковий вибір  $K$  початкових центрів кластерів  $\{\mu_1, \mu_2, \dots, \mu_K\}$ ;

– повторення до збіжності: призначення об'єктів до найближчих центрів:  $C_k = \{x_i \mid \operatorname{argmin}_k \|x_i - \mu_k\|\}$ ;

– оновлення центрів кластерів:  $\mu_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i$ .

Мета алгоритму – мінімізувати суму квадратичних відхилень всередині кластерів:

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2. \quad (5.12)$$

2.6. Представлення результатів аналізу (блок 7). Результати кластеризації візуалізуються за допомогою методів багатовимірної візуалізації, таких як головні компоненти (PCA) або t–SNE, для подальшої інтерпретації та аналізу.

Крок 3. Опис та аналіз сформованих кластерів (блок 8). Проводиться детальний аналіз кожного кластера, на початку проведення обчислення середніх значень ознак:

$$\bar{x}_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i, \quad (5.13)$$

Далі здійснюється аналіз варіації ознак в межах кластерів та визначення характерних особливостей кожного кластера на основі демографічних та поведінкових характеристик. Ідентифіковані соціальні кластери зберігаються в окремій базі даних (блок 9), що забезпечує автоматизацію подальшої обробки та аналізу.

Крок 4. Створення споживчого кошика (блок 10) та розробка рекомендацій для його оптимізації (блок 11). На основі отриманих кластерів формується типовий споживчий кошик для кожного з них:

– ймовірність придбання товару  $l$  покупцями кластера  $k$ :

$$P_{kl} = \frac{n_{kl}}{n_k}, \quad (5.14)$$

де  $n_{kl}$  – кількість покупок товару  $l$  покупцями з кластера  $k$ ;  $n_k$  – загальна кількість покупок у кластері  $k$ ;

– очікувана витрата на товар  $l$  у кластері  $k$ :

$$E_{kl} = P_{kl} \times \bar{P}_{kl}, \quad (5.15)$$

де  $\bar{P}_{kl}$  – середня ціна товару  $l$  для кластера  $k$ .

Розроблений інтелектуальний метод формування споживчого кошика був реалізований в експериментальних дослідженнях, представлених у підрозділі 5.4.1. Отримані результати підтвердили ефективність методу у швидкій адаптації споживчого кошика, що є критично важливим у кризових умовах для прогнозування та запобігання гуманітарно–економічним кризам.

### 5.3 Інтелектуальний метод пошуку альтернативних постачальників

Базуючись на методі кластерного аналізу соціально–економічних даних ТГ з підрозділу 5.1, було розроблено інтелектуальний метод пошуку альтернативних постачальників. Цей метод покликаний підвищити стійкість та адаптивність ланцюгів постачання, особливо в умовах кризових ситуацій, шляхом всебічного аналізу критичних параметрів потенційних постачальників.

Розглянуті дослідження (див. підрозділ 1.2.2) зосереджуються в основному на оптимізації ланцюгів постачань, сталому виборі постачальників та інтеграції сучасних технологій у процес вибору. Однак, вони не враховують всебічного аналізу потенційних постачальників за широким спектром критичних параметрів, таких як ціна, якість, доступність, потужність виробництва та репутація на ринку, що є критичними для прийняття рішень у кризових умовах, де важливо швидко ідентифікувати та вибрати надійних постачальників. На відміну від аналогів [198,

60], розроблений інтелектуальний метод дозволить на основі всебічного аналізу даних ідентифікувати найбільш прийнятних постачальників, які відповідають потребам організації у кризових умовах.

Для вибору альтернативних постачальників у кризових умовах авторами розроблено інтелектуальний метод, що зосереджений на аналізі широкого спектра критичних параметрів. Метод детально представлений у вигляді послідовних кроків, які ілюструються схематично (див. додаток А, рис. А.10) та представлений кроками:

Крок 1. Збір та підготовка даних (блок 1). Завдання збору даних полягає в збиранні інформації про потенційних постачальників, що включає ряд ключових параметрів, таких як ціна, якість продукції або послуг, рівень доступності та надійності постачання, потужність виробництва, репутація компанії на ринку, а також відгуки від інших клієнтів. Ця інформація є критичною для подальшого аналізу та оцінки потенційних постачальників, що дозволяє створити об'єктивну основу для прийняття рішень у процесі відбору постачальників.

Припустимо, є  $N$  потенційних постачальників, і для кожного постачальника ми збираємо  $M$  параметрів (наприклад, ціна, якість, доступність тощо) в БД (блок 2). Це дає нам матрицю даних  $X$ , яка має розмірність  $N \times M$ .

Далі, проводимо нормалізацію та стандартизацію. Оскільки кожен параметр може мати різний масштаб, ми нормалізуємо або стандартизуємо дані для кожного параметра. Для параметра  $j$  ( $j=1,2,\dots,M$ ):

Нормалізація (Min–Max Scaling):

$$X'_{ij} = \frac{X_{ij} - \min(X_j)}{\max(X_j) - \min(X_j)}. \quad (5.16)$$

Стандартизація (Z–score Normalization) [16]:

$$X'_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j}, \quad (5.17)$$

де  $X_{ij}$  – оригінальне значення  $i$ -го постачальника за  $j$ -м параметром;  $\mu_j$  і  $\sigma_j$  – відповідно, середнє значення та стандартне відхилення  $j$ -го параметра.

В результаті отримуємо матрицю  $X'$  розмірності  $N \times M$ , де кожен рядок відповідає постачальнику, а кожен стовпець – нормалізованому або стандартизованому значенню параметра.

Крок 2. Кластеризація даних (блок 3).

2.1. Після того, як дані були підготовлені, наступним кроком є застосування алгоритмів кластеризації. Нехай розглядаються три основні методи кластеризації: K–Means, ієрархічна кластеризація та DBSCAN.

2.2. Обчислення загальний Silhouette Score [364]. Для всього набору даних він обчислюється як середнє значення  $s(i)$  за всіма об'єктами:

$$S = \frac{1}{N} \sum_{i=1}^N s(i), \quad (5.18)$$

де  $N$  – загальна кількість об'єктів.

Silhouette Score [364] варіюється від  $-1$  до  $+1$ . Високий показник означає, що об'єкти добре відокремлені у своїх кластерах. Низький або від'ємний показник може свідчити про те, що об'єкти були присвоєні неправильно.

Крок 3. Вибір основного кластера (блок 4). Після застосування обраного методу кластеризації, наступним кроком є вибір основного кластера, який базується на певних критеріях, які відповідають бізнес–цілям або специфічним потребам.

3.1. Можна використовувати вагові коефіцієнти для формулювання критеріїв. Наприклад, якщо ціль – мінімізувати ціну та максимізувати якість, можна розглянути функцію:

$$f(C) = \alpha x_{\text{ціна}} + (1 - \alpha)(1 - x_{\text{якість}}), \quad (5.19)$$

де  $\alpha$  – ваговий коефіцієнт.

3.2. Вибір кластера, що оптимізує вказану функцію:

$$C_{\text{основний}} = \arg \min_C f(C) \quad (5.20)$$

Крок 4. Вибір даних з основного кластера (блок 5). Після визначення основного кластера, наступним кроком є вибір конкретних даних (тобто об'єктів або постачальників) з цього кластера. Цей процес можна описати математично так:

4.1. Визначення міток кластера. Нехай кластери позначені мітками  $C_1, C_2, \dots, C_k$ . Основний кластер, вибраний на попередньому кроці, має мітку  $C_{\text{основний}}$ .

4.2. Обрання об'єктів. Об'єкти, які належать до основного кластера, можна ідентифікувати за їх мітками. Множина об'єктів в основному кластері:

$$O_{\text{основний}} = \{x_i: \text{мітка}(x_i) = C_{\text{основний}}\}, \quad (5.21)$$

де  $x_i$  –  $i$ -й об'єкт у наборі даних, а мітка ( $x_i$ ) – мітка кластера, до якого належить об'єкт.

Крок 5. Повторна кластеризація в обраному кластері (блок 6). Вона має на меті більш детальне вивчення його внутрішньої структури. Цей процес включає застосування алгоритму кластеризації специфічно до даних, які були виокремлені в обраному кластері на попередньому кроці. Це дозволяє виявити більш тонкі групування чи підкластери всередині основного кластера, які можуть мати унікальні характеристики чи відмінності.

5.1. Аналіз відхилень всередині кластера. Розрахуємо дисперсію параметрів всередині основного кластера для оцінки однорідності:

$$\sigma_{\text{Сосновний},j}^2 = \frac{1}{|O_{\text{основний}}|} \sum_{x_i \in O_{\text{основний}}} (X'_{ij} - \bar{X}'_{\text{Сосновний},j})^2, \quad (5.22)$$

де  $\sigma_{\text{Сосновний},j}^2$  – дисперсія  $j$ -го параметра в основному кластері;  $\bar{X}'_{\text{Сосновний},j}$  – середнє значення  $j$ -го параметра в основному кластері;  $|O_{\text{основний}}|$  – кількість об'єктів в основному кластері.

5.2. Ранжування постачальників всередині кластера. Можна обчислити зважений рейтинг для кожного постачальника всередині основного кластера:

$$R(x_i) = \sum_{j=1}^M w_j X'_{ij}, \quad (5.23)$$

де  $R(x_i)$  – рейтинг  $i$ -го постачальника;  $w_j$  – ваговий коефіцієнт для  $j$ -го параметра ( $\sum_{j=1}^M w_j = 1$ );  $X'_{ij}$  – нормалізоване або стандартизоване значення  $j$ -го параметра для  $i$ -го постачальника.

Крок 6. Аналіз результатів повторної кластеризації (блок 7). На кроці аналізу результатів повторної кластеризації здійснюється глибоке вивчення розподілу та

характеристик виявлених підгруп всередині обраного кластера. Це включає аналіз таких аспектів, як розміри підгруп, їхні основні характеристики, а також відносини та взаємодії між цими підгрупами. Метою є отримання детального розуміння внутрішньої структури кластера.

Крок 7. Вибір кращого претендента в обраному кластері (блок 8). Під час вибору кращого претендента всередині обраного кластера встановлюються та застосовуються конкретні критерії оцінки, які можуть включати фактори, як–от ціна, якість, надійність, репутація тощо. На основі цих параметрів визначаються постачальники, які найкраще відповідають вимогам та критеріям відбору, що були встановлені.

Крок 8. Вивід результатів повторної кластеризації (блок 9). Останній крок полягає у візуалізації та поданні результатів повторної кластеризації і вибору претендентів. Це включає створення графіків, таблиць та інших візуальних засобів для наочного представлення знайдених підгруп, їхніх характеристик та порівняльного аналізу вибраних постачальників. Звіт з цими результатами може слугувати основою для прийняття інформованих рішень щодо вибору постачальників (блок 10).

Реалізація запропонованого інтелектуального методу пошуку альтернативних постачальників була проведена в підрозділі 5.4.2. Експериментальні результати демонструють його здатність швидко ідентифікувати надійних і відповідних постачальників, що дозволяє організаціям ефективно реагувати на виклики кризових умов та забезпечувати безперервність постачання.

#### **5.4 Експериментальні результати**

За запропонованим методом кластерного аналізу соціально–економічних даних ТГ (див. підрозділ 5.1) та принципом синтезу інформаційної технології (див. підрозділ 2.1) розроблено модулі інформаційної системи, які вирішують поставлені задачі формування споживчого кошика та пошуку альтернативних постачальників.



### 5.4.1 Формування споживчого кошика

Для реалізації інтелектуального методу формування споживчого кошику обрано мову Python. В якості вхідних даних використано дані Customer Personality Analysis з платформи Kaggle [90].

Зазначений датасет здійснює аналіз особистості клієнта, представляючи дані як загальні витрати в доларах США за останні два роки. Інформація вибрана з цього набору даних (Таблиця 5.1). Цих даних недостатньо для повного розв'язання задачі, оскільки збір повного необхідного датасету є складним процесом, що вимагає дозволу клієнтів на використання їхніх персональних даних. Важливо отримати повний перелік товарів, куплених покупцем, для створення більш деталізованого аналізу. Проте, для застосування розробленого методу, існуючий датасет Customer Personality Analysis виявляється достатнім.

Таблиця 5.1

Структура набору даних

Параметр	count	mean	min	max
Income	2240.0	52247.25	1730.0	666666.0
Kidhome	2240.0	0.44	0.0	2.0
Teenhome	2240.0	0.51	0.0	2.0
Recency	2240.0	49.11	0.0	99.0
Wines	2240.0	303.94	0.0	1493.0
Fruits	2240.0	26.30	0.0	199.0
Meat	2240.0	166.95	0.0	1725.0
Fish	2240.0	37.52	0.0	259.0
Age	2240.0	52.19	25.0	128.0
Total_Spend	2240.0	602.25	5.0	2525.0

Перед проведенням кластеризації необхідно здійснити нормалізацію даних. Оскільки цікавлять місячні витрати, розподілимо всі витратні дані на 24 місяці. Цей крок допоможе отримати дані, готові до використання для формування споживчого кошика.

Після нормалізації було проведено кластеризацію, для чого було обрано метод K–Means [119, 188], який є одним з найпоширеніших і найпростіших методів кластеризації.

Для визначення оптимальної кількості кластерів використано метод ліктя [357]. Цей метод вимірює суму квадратів відстаней між точками та центром їх кластера для різних значень  $k$ . На графіку, який показує ці значення, вибирається точка «ліктя», де зменшення в різниці між результуючими сумами стає незначним, що вказує на найкраще значення  $k$ . Для цього аналізу ідеальна кількість кластерів була визначена як три (Рис. 5.2).

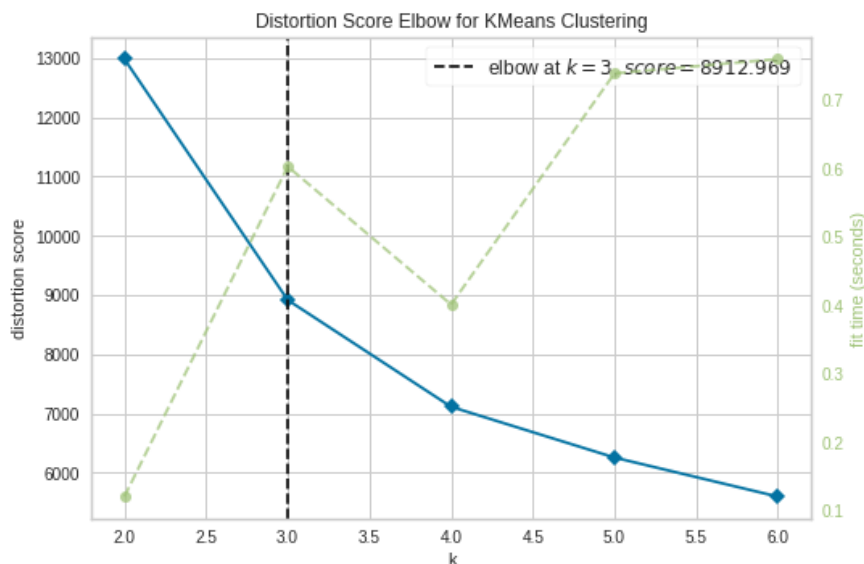


Рис. 5.2. Вибір кількості кластерів методом ліктя

Далі представлено візуалізацію трьох сформованих кластерів. За даними (Рис. 5.3а), у кластері номер нуль зосереджено найбільшу кількість клієнтів. На основі інтерпретації (Рис. 5.3б), отримані кластери характеризуються таким чином:

- кластер 2: клієнти з високими витратами та середнім доходом;
- кластер 1: клієнти з великими витратами та високими доходами;
- кластер 0: клієнти з низькими витратами та низьким доходом.

Якщо розглянути кластери з соціальної точки зору, то особливу увагу слід звернути на кластер 2 та кластер 0. Рекомендується формувати кошик згідно з

характеристиками кластера 2 для стабільних економічних умов, тоді як у кризових ситуаціях слід виходити з потреб кластера 0.

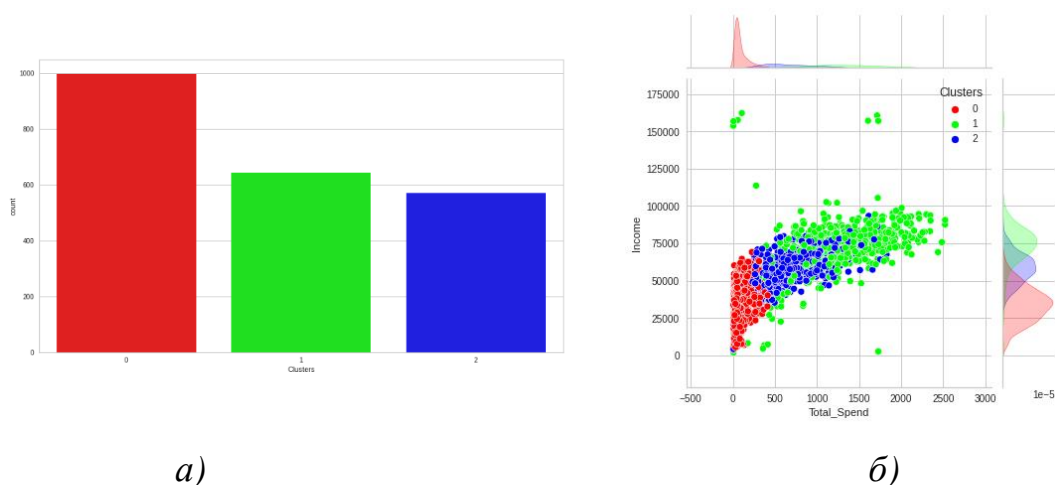


Рис. 5.3. Візуалізація кластерів: (а) кількість клієнтів в кластерах; (б) розподіл доходів та витрат клієнтів

При розгляді вікової категорії покупців (Рис. 5.4) видно, що розподіл кластерів достатньо рівномірний. Отже, вік не дуже впливає на купівельну спроможність.

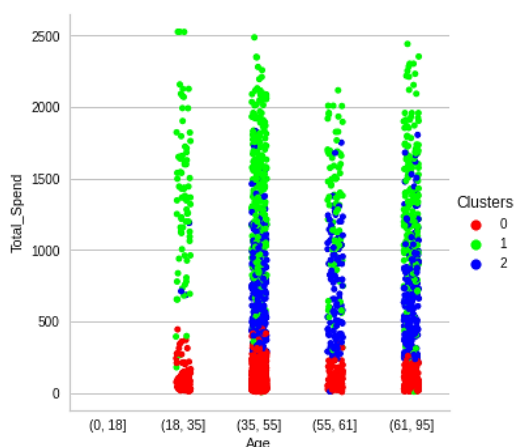


Рис. 5.4. Візуалізація кластерів: віковий розподіл

Далі розглянемо більш детально кластери з точки зору суми (в доларах США) витрат на товари протягом останніх двох років (Рис. 5.5). Діаграми типу «boxplot» найкраще ілюструють середнє значення куплених товарів у кожному кластері. З

результатів видно, що кластер 1 витрачає відносно мало на такі продукти як риба, м'ясо, солодоці та вино, тоді як кластер 2 показує більші витрати на ці категорії товарів. Що стосується технологічних товарів, то витрати між усіма покупцями приблизно однакові. Таким чином, виходячи з аналізу продажів, автори вважають, що кластер 2 містить ключові дані для формування споживчого кошика. Подробиці щодо кластера 2 та рекомендації щодо формування споживчого кошика будуть розглянуті детальніше.

Споживчий кошик (кластер 2) охоплює 27 % клієнтів з різних вікових категорій, які мають середній дохід. Щомісячні витрати цієї групи становлять: 25 дол. США на рибні продукти, 100 дол. США – на м'ясні продукти, 20 дол. США – на солодоці, 400 дол. США – на винну продукцію та 50 дол. США – на покупку технічних товарів. Також важливо зазначити, що ці показники можуть не відобразити повної картини через обмеженість доступних даних.

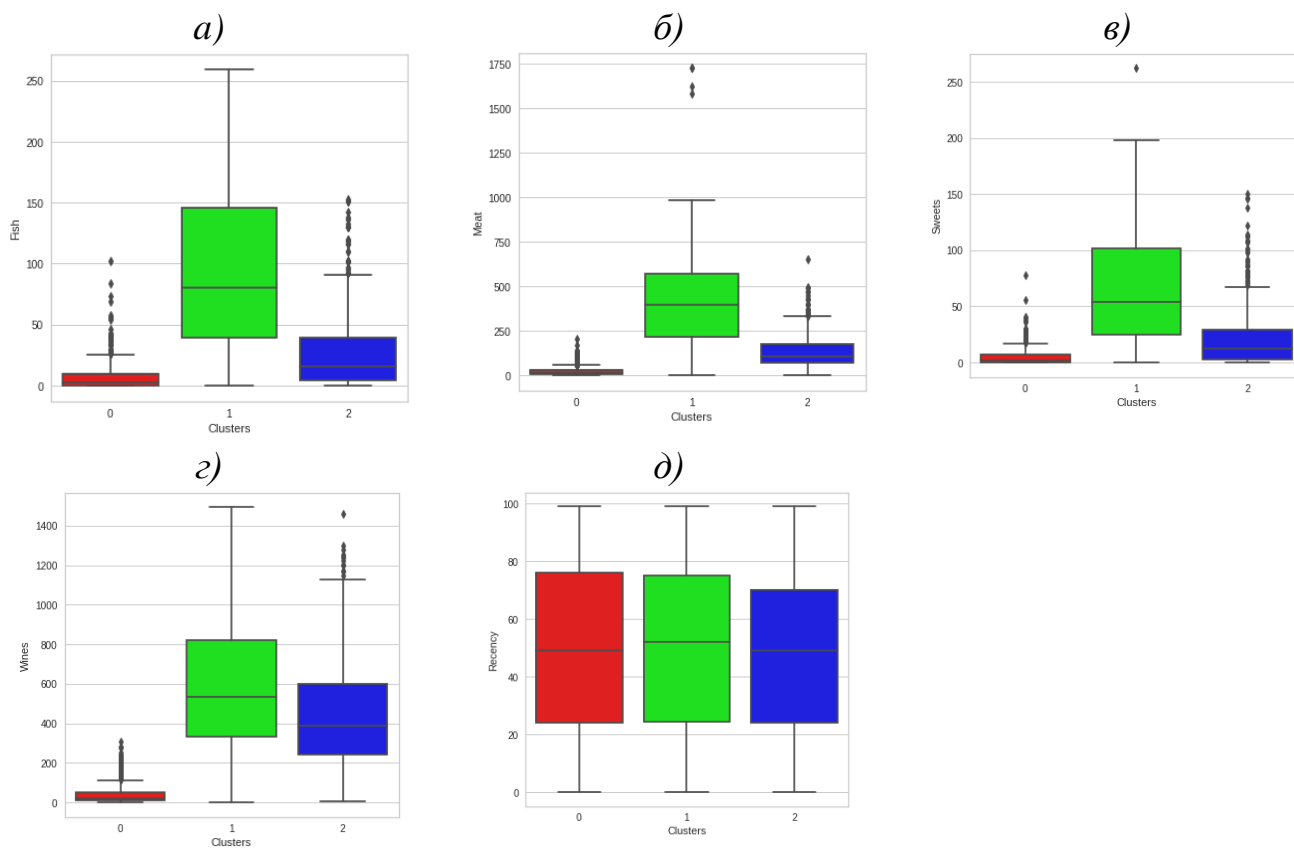


Рис. 5.5. Візуалізація кластерів. Розподіл куплених товарів: а) риба; б) м'ясо; в) солодоці; г) вино; д) технологічні товари.

Отже, за допомогою інтелектуального методу можна визначити склад споживчого кошика, виходячи з даних супермаркетів. Більш повно зібрані дані з мережевих супермаркетів можуть забезпечити найточніші результати по кластерах. Це, у свою чергу, дозволить здійснювати моніторинг купівельної спроможності споживачів у реальному часі та оперативно адаптувати споживчий кошик до поточних умов.

Для вдосконалення інтелектуального методу формування споживчого кошика було інтегровано кілька ключових етапів та інструментів, на основі інформаційної технології інтелектуального аналізу соціальних даних ТГ. Зокрема, було розширено підхід до кластеризації, використовуючи додаткові алгоритми, такі як DBSCAN та Agglomerative Clustering, щоб доповнити метод  $k$ -середніх. Це дозволило врахувати більш складні взаємозв'язки між даними, що сприяло підвищенню якості кластеризації.

Важливим удосконаленням стало додавання етапу адаптації моделей до нових даних, що включає механізм перенавчання та інтеграцію потокового машинного навчання. Раніше модель формувала споживчий кошик на основі статичних даних, що знижувало її актуальність у динамічних умовах, таких як зміна купівельної поведінки або зовнішніх економічних факторів. Завдяки нововведенням модель тепер може працювати з поточними даними в режимі реального часу, що забезпечує її релевантність і швидке реагування на зміни.

Застосування нових алгоритмів кластеризації разом із вдосконаленим механізмом нормалізації даних сприяло зменшенню втрат інформації в процесі аналізу та покращенню показників кластеризації. Наприклад, коефіцієнт силуету зріс на 1,2 %, що свідчить про покращення розділення між групами. Використання інтерактивних методів візуалізації, таких як  $t$ -SNE, забезпечило кращу інтерпретацію отриманих результатів і дозволило зробити їх більш доступними для кінцевих користувачів.

Таким чином, впроваджені зміни значно підвищили точність і адаптивність моделі, що дозволяє не лише створювати більш точні прогнози, але й оперативно адаптувати споживчий кошик до реальних умов. Завдяки цьому метод стає особливо

актуальним його використання в кризових ситуаціях, таких як економічні чи гуманітарні кризи, що вимагають швидких та обґрунтованих рішень.

Для розробки модуля (Рис. 5.6) інтелектуального формування споживчого кошику [255], потік даних розпочинається з блоку 1, де збираються дані із супермаркетів, а також можливо з аптек та інших магазинів, включаючи інформацію про покупців і їхні покупки. Далі, у блоці 2, відбувається кластеризація даних, що включає вибір об'єктів для кластеризації, визначення змінних для оцінки, нормалізація цих змінних, обчислення мір схожості між об'єктами, застосування методу кластерного аналізу та подання результатів аналізу.

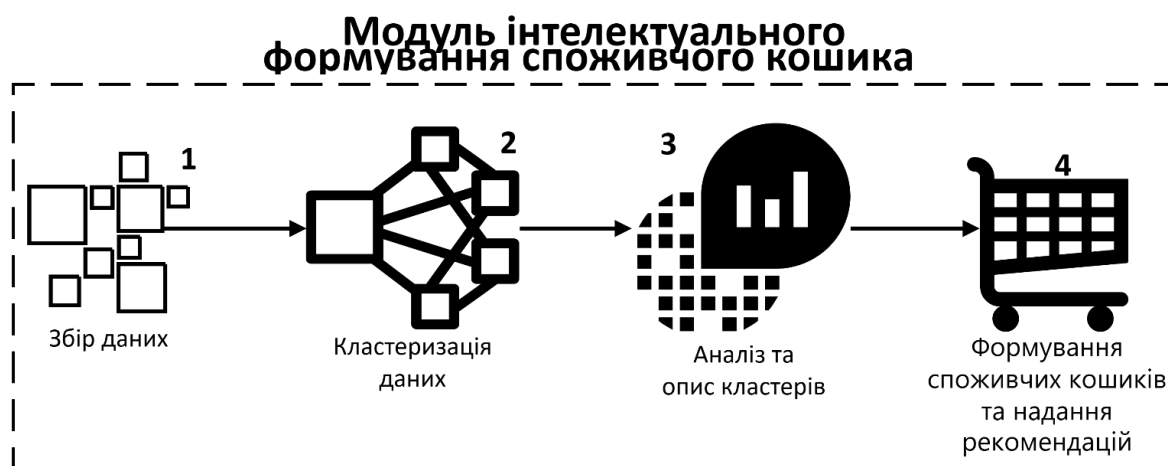


Рис. 5.6. Модуль інтелектуального формування споживчого кошика

У блоці 3 проводиться опис отриманих кластерів та формування бази даних соціальних кластерів для подальшої автоматизації процесів і наукових досліджень. Завершується процес блоком 4, де формуються споживчі кошики та надаються рекомендації, засновані на аналізі отриманих даних.

Модуль (див. Рис. 5.6) аналізу інтелектуального соціально–економічного добробуту населення розроблений з метою визначення структури споживчого кошика в ТГ. У цьому контексті користувач має можливість вибору конкретного ТГ (Рис. 5.7). Після вибору локації система здійснює збір даних про товари та ціни з мережевих супермаркетів, які можуть бути зібрані загалом або з окремих торгових мереж.


**Choosing a city:**

Select a city

**Choosing a supermarket chain :**

Select a supermarket

**Data collection:**
 Uploaded file

 Customer Personality Analysis.csv  
2m ago


604KB

**Storage location**

C:\Laura\My Documents

Parameter	count	mean	min	max
Income	2240.0	52247.25	1730.0	666666.0
Kidhome	2240.0	0.44	0.0	2.0
Teenhome	2240.0	0.91	0.0	2.0
Recency	2240.0	45.11	0.0	99.0
Wines	2240.0	303.94	0.0	1453.0
Fruits	2240.0	26.30	0.0	199.0
Meat	2240.0	166.95	0.0	1725.0
Fish	2240.0	37.52	0.0	258.0
Age	2240.0	57.19	25.0	128.0
Total Spend	2240.0	602.25	5.0	2525.0

Run

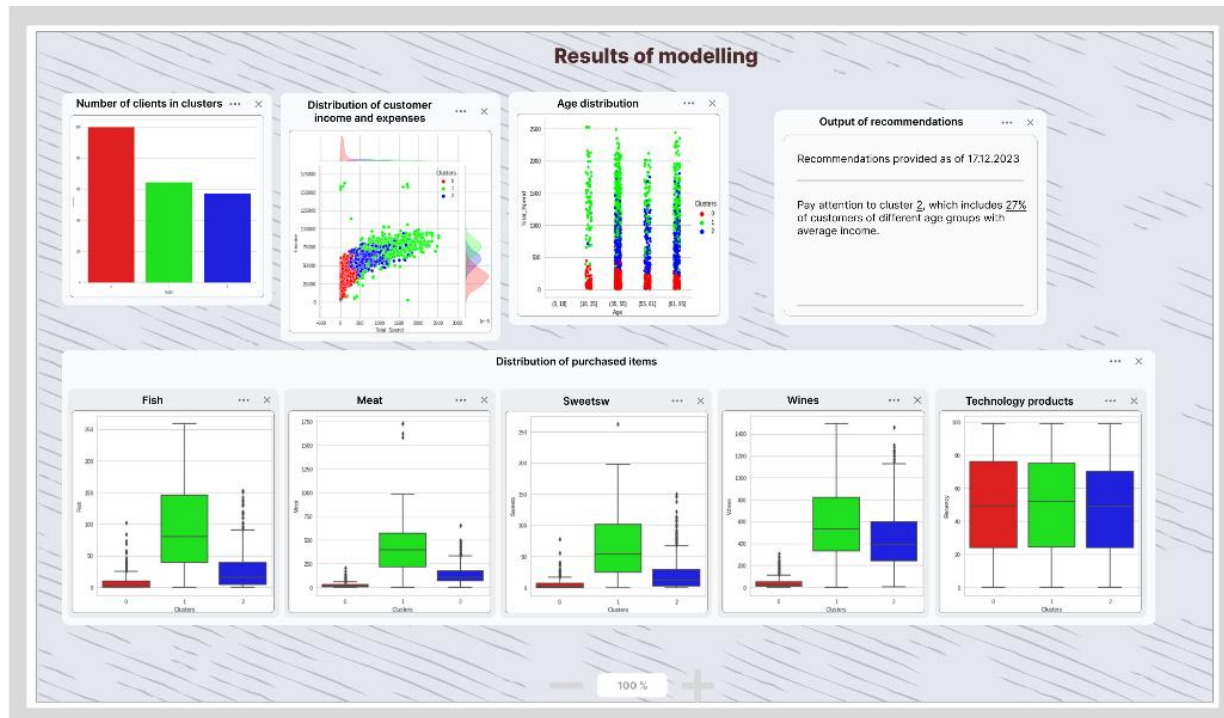
**Output area of the results:**

Рис. 5.7. Модуль інтелектуального формування споживчого кошика

Основою аналізу є формування кластерів, серед яких ключовим є кластер, що включає товари, необхідні для всіх вікових груп населення, що дозволяє здійснити всебічну оцінку споживчого кошика і відповідає потребам різних груп населення. Результати аналізу подаються за окремими товарними групами, надаючи детальну інформацію про споживчі переваги в обраному ТГ.

Це корисно для адміністрації міста, оскільки надає цінну інформацію для прийняття обґрунтованих економічних та соціальних рішень, зорієнтованих на підвищення якості життя громадян. Крім того, аналіз споживчого кошика допомагає виявити потреби різних груп населення, що сприяє ефективному плануванню ресурсів та розвитку інфраструктури громади.

Результати оцінки формування споживчого кошика демонструють суттєве зменшення часових витрат завдяки використанню модуля системи (див. додаток Г). Експертні оцінки показали, що час виконання завдань із системою складав від 1 до 3 годин, у той час як без системи цей показник коливався від 2 до 60 днів. Загальна ефективність системи досягла ~98,75 %, що свідчить про її здатність значно прискорювати процеси аналізу та складання споживчого кошика, забезпечуючи оперативність і точність у прийнятті управлінських рішень.

#### **5.4.2 Пошук альтернативних постачальників**

Для апробації методу пошуку альтернативних постачальників, обрано бізнес, який спеціалізується на ремонті автомобілів. Зважаючи на проєкт, пов'язаний з ремонтом автомобілів, основні напрями бізнесу включають у себе надання послуг автомобільного ремонту та обслуговування, а також управління запасами та постачанням необхідних запчастин. У цьому контексті особливо важливим стає процес пошуку нових постачальників під час кризових ситуацій, що дозволяє підприємствам, які займаються авторемонтом, адаптуватися до змін у попиті та ринкових умовах, шукаючи нові можливості для розвитку та підтримки бізнесу.

Підприємець ініціює збір даних з різних джерел, таких як соціальні медіа, вебсайти та інші онлайн-ресурси. Цей процес починається зі збору інформації про



потенційних постачальників, враховуючи такі параметри, як ціна, якість, доступність продукції, потужність виробництва, репутація та інші важливі фактори для бізнесу. Після цього використовується метод кластеризації K–Means, для групування постачальників за схожими характеристиками. K–Means обрано, бо Silhouette Score =0,78 (рис.5.8), що є найближчим до 1 серед інших розглянутих методів (див. підрозділ 1.2.1).

Наступним кроком (Рис. 5.9) є вибір певного кластера, який відповідає критеріям вибору, таким як найнижча ціна та висока якість. Завершальний крок включає аналіз результатів, серед яких виведення середніх значень та вибір найкращого постачальника серед обраних кандидатів.

У цьому випадку, ключовими критеріями можуть бути цінова доступність та якість продукції. Один з потенційних кандидатів може визначитися як постачальник 424 (Рис. 5.10), який відрізняється низькою ціною (10 балів) та високою якістю (5 балів). Додатково можуть бути важливі: швидка доставка (3 дні), добра репутація (5 балів) та позитивні відгуки (5 балів). Цей кандидат виділяється серед інших завдяки конкурентоспроможному прайс–листу та позитивній репутації на ринку, що робить його привабливим вибором для бізнесу.

Cluster	Pricing	Quality	Availability	Production Capacity	Reputation and History	Feedback
0	19.6944	2.93056	3.36111	0.5	3.48611	3.93056
1	11.2326	3	4.37209	0.395349	3.18605	1.53488
2	28.3514	3.32432	8.35135	0.432432	2.91892	4.21622
3	15.6364	3.25758	8.57576	0.575758	2.72727	3.95455
4	16.5833	3.23611	3.77778	0.388889	2.11111	2.70833
5	11.6667	2.93333	8.63333	0.455556	2.71111	3.03333
6	28.3148	2.87037	8.27778	0.518519	2.98148	1.44444
7	28.8382	2.91176	3.27941	0.514706	3.32353	3
8	23.7333	2.76667	8.78333	0.416667	2.45	2.88333
9	22.4603	3	4.03175	0.492063	2.20635	1.69841
10	14.871	2.12903	5.46774	0.467742	3.77419	2.27419
11	25.875	2.66667	4.52778	0.5	2.68056	3.43056
12	19.6667	3.2	8.26667	0.533333	2.97333	2.90667
13	22.7414	3.93103	6	0.517241	4.08621	3.53448
14	11.7903	3.06452	3.54839	0.629032	3.01613	4.03226

Рис. 5.8. Середні значення для вибраного кластера на кроці 1

Selected_Cluster	Pricing	Quality	Availability	Production Capacity	Reputation and History	Feedback	Cluster
0	11.125	2.83333	5.25	0.416667	3.41667	1.54167	1
1	11.275	2.95	2.475	0.575	2.675	3	1
2	11.1304	3.73913	4.82609	0.608696	2.73913	4.34783	1

Рис. 5.9. Середні значення для вибраного кластера на кроці 2

	Pricing	Quality	Availability	Production Capacity	Reputation and History	Feedback	Cluster	Selected_Cluster
804	10	3	5	1	5	5	1	2
797	10	3	6	1	1	4	1	2
779	10	3	6	0	2	4	1	2
369	10	3	4	1	3	4	1	2
424	10	5	3	1	5	5	1	2
440	10	2	5	1	1	4	1	2
550	10	2	4	0	4	5	1	2
117	11	4	5	1	3	4	1	2
696	11	5	5	0	3	4	1	2
675	11	4	5	1	4	5	1	2

Рис. 5.10. Перші 10 значень відсортованих за ціною

Процес інтелектуального пошуку альтернативних клієнтів/постачальників (Рис. 5.11) розпочинається зі збору даних (блок 1) з різних джерел, включаючи соціальні медіа та вебсайти, і подальшого зберігання їх у OLAP базі даних (блок 2).

Потім дані проходять інтелектуальний аналіз (блок 3) за допомогою аналітичних методів та алгоритмів машинного навчання для виявлення потенційних кандидатів.

Вибрані кандидати піддаються оцінці (блок 4), включаючи аналіз ризиків та можливостей, і на основі цього проводиться вибір (блок 5).

Завершальний крок включає вибір та інтеграцію обраних альтернатив (блок 6), готовність до подальшої співпраці та використання цих даних для прийняття стратегічних рішень.

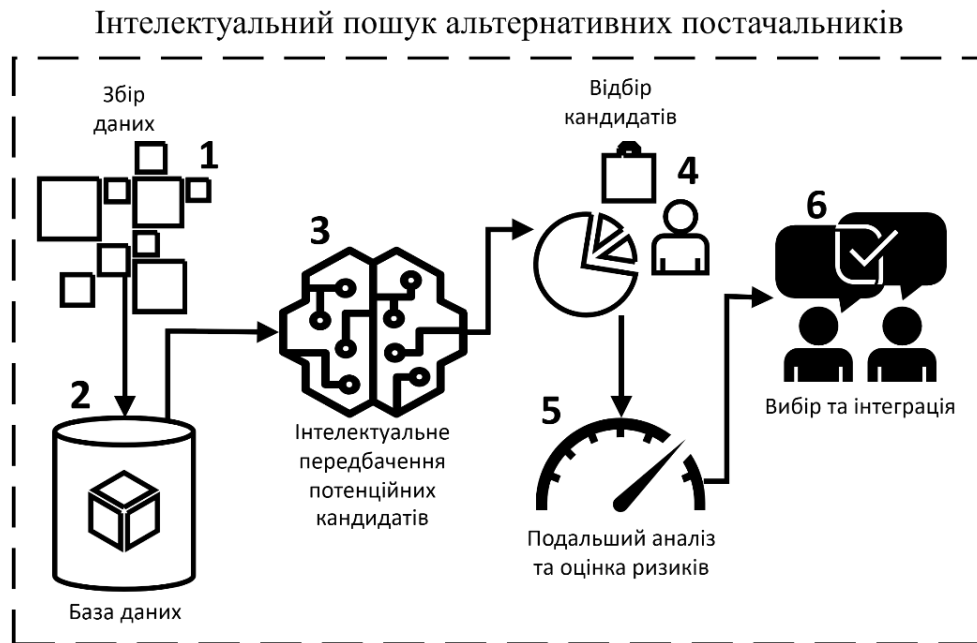


Рис. 5.11. Блок інтелектуального пошуку альтернативних клієнтів/постачальників

Отже, підприємець дає запит (Рис. 5.12) на збирання даних з різних джерел, таких як соціальні медіа, вебсайти та інші онлайн-ресурси. Отже, цей процес починається зі збору даних про потенційних постачальників, включаючи параметри, такі як: ціна, якість, доступність продукції, потужність виробництва, репутація та інші фактори, що є важливими для бізнесу. Потім застосовується метод кластеризації, наприклад, K-Means [35], для групування постачальників у кластери зі схожими характеристиками. Після цього обирається певний кластер, який відповідає критеріям вибору, таким як найнижча ціна та висока якість. Завершальний крок включає аналіз результатів, включаючи виведення середніх значень та вибір найкращого постачальника серед обраних кандидатів.

У цьому випадку (див. Рис. 5.12) важливими критеріями можуть бути цінова доступність та якість продукції. За цими параметрами одним з потенційних кандидатів може бути постачальник 424 з низькою ціною (10 балів) та відмінною якістю (5 бали). Додатково, важливими можуть бути швидка доставка (3 дні), добра репутація (5 балів) та позитивні відгуки (5 балів). Цей кандидат вирізняється серед інших завдяки своєму конкурентоспроможному прайс-листу та позитивній репутації на ринку, що може зробити його привабливим вибором для бізнесу.

Chat with the Municipality



Data output:

Request for data collection:

APP\_Data.csv  
50 Kb

	Pricing	Quality	Availability	Production Capacity	Reputation and History	Feedback
0	12	4	10	0	2	2
1	16	4	2	0	1	5
2	15	5	10	1	4	5
3	26	1	9	1	3	3
4	26	1	6	1	2	5

Run

Clustering

Average values for the selected cluster

Cluster	Pricing	Quality	Availability	Production Capacity	Reputation and History	Feedback
0	19.6944	2.93056	3.36111	0.5	3.48611	3.93056
1	11.2326	3	4.37289	0.395249	3.18695	1.53468
2	28.3514	3.32432	0.35135	0.432432	2.91892	4.21622
3	15.6364	3.25758	0.57576	0.575758	2.72727	3.95455
4	16.5833	3.23611	3.77778	0.388889	2.11111	2.70833
5	11.6667	2.93333	0.63333	0.455556	2.71111	3.03333
6	28.3148	2.87037	0.27778	0.518519	2.98148	1.44444
7	28.8382	2.91176	3.27941	0.514706	3.32353	3
8	23.7333	2.76667	0.78333	0.416667	2.45	2.08333
9	22.4003	3	4.03175	0.492063	2.20635	1.69841
10	14.071	2.12903	5.46774	0.467742	3.77419	2.27419
11	25.875	2.66667	4.52778	0.5	2.68856	3.43856
12	19.6667	3.2	0.26667	0.533333	2.97333	2.90667
13	22.7414	3.93109	6	0.517241	4.08621	3.53448

Cluster Distribution by Pricing



Selecting a cluster

1

Run

Average values for the selected cluster

Selected_Cluster	Pricing	Quality	Availability	Production Capacity	Reputation and History	Feedback	Cluster
0	11.125	2.83333	5.25	0.416667	3.41667	1.54167	1
1	11.275	2.95	2.475	0.575	2.675	3	1
2	11.1304	3.73913	4.82689	0.608696	2.73913	4.34783	1

Selecting a cluster

1

Run

First 10 values sorted by price

	Pricing	Quality	Availability	Production Capacity	Reputation and History	Feedback	Cluster	Selected_Cluster
804	10	3	5	1	5	5	1	2
797	10	3	6	1	1	4	1	2
779	10	3	6	0	2	4	1	2
369	10	3	4	1	3	4	1	2
424	10	5	3	1	5	5	1	2
440	10	2	5	1	1	4	1	2
550	10	2	4	0	4	5	1	2
117	11	4	5	1	3	4	1	2
696	11	5	5	0	3	4	1	2

Selecting a ID

424

Show

MEDIC TERNOPIL  
Ternopil, Lvivska str. 6,  
tel. 0945896265

Рис. 5.12. Вікно інтелектуального пошуку альтернативних постачальників

Оцінка ефективності пошуку альтернативних клієнтів та постачальників показала значне скорочення часу виконання завдань завдяки використанню модуля системи (див. додаток Г). Час із системою варіювався від 1 до 5 годин, тоді як без системи пошук займав від 7 до 60 днів.

Ефективність системи досягла ~99,61 %, підтверджуючи її здатність значно оптимізувати процеси пошуку, знижуючи часові витрати та покращуючи точність у виборі клієнтів і постачальників.

## 5.5 Висновки

У розділі 5 розглянуто розробку та впровадження методу кластерного аналізу соціально–економічних даних ТГ, спрямованого на підтримку ефективного прийняття управлінських рішень. Здійснено детальний аналіз кожного підрозділу, що дозволяє оцінити результати та наукову значущість проведених досліджень.

Розробленого метод кластерного аналізу соціально–економічних даних, що враховує різноманітність типів даних (структурованих, напівструктурованих, неструктурованих) та інтеграцію даних з різних джерел. Впроваджено використання алгоритмів кластеризації (K–Means, DBSCAN, GMM, PCA) для групування об'єктів зі схожими характеристиками. Це дозволило створити базу даних соціальних кластерів, яка сприяє автоматизації аналізу та підтримці сталого розвитку громад. Запропонований підхід забезпечує ідентифікацію груп з високим інвестиційним потенціалом, оптимізацію ресурсів та розробку логістичних маршрутів, що сприяє підвищенню ефективності управлінських рішень.

Розроблено інтелектуальний метод формування споживчого кошика, який базується на даних супермаркетів. Проведено кластеризацію споживачів на основі демографічних та економічних показників з використанням алгоритму K–Means. Це дало змогу визначити соціально–економічні групи, кожна з яких має унікальний споживчий кошик. Метод дозволив адаптувати набори товарів до потреб окремих груп населення та в кризових умовах, таких як економічні чи гуманітарні кризи,

швидко реагувати на зміни купівельної поведінки. Удосконалення моделі шляхом інтеграції потокового машинного навчання забезпечило актуальність результатів у динамічних умовах.

Розроблено інтелектуальний метод пошуку альтернативних постачальників, що використовує кластеризацію для оцінки широкого спектра характеристик постачальників (ціна, якість, доступність, репутація). Реалізовано підхід з використанням алгоритму K–Means для групування постачальників та подальшої оптимізації вибору.

Представлено експериментальні результати, які підтвердили ефективність запропонованих методів. Показано, що використання кластерного аналізу значно скоротило часові витрати та підвищило якість прийняття управлінських рішень. Запропоновані методи виявилися придатними для широкого спектра задач, від визначення складу споживчого кошика до пошуку надійних постачальників, і забезпечили адаптивність до змінних умов.

Запропоновані методи кластерного аналізу підтвердили свою ефективність у вирішенні соціально–економічних задач ТГ. Результати досліджень свідчать про можливість застосування розроблених підходів для підтримки сталого розвитку, підвищення якості управлінських рішень та адаптації до кризових ситуацій.

## РОЗДІЛ 6.

### МЕТОДИ ТА ЗАСОБИ ПРОГНОЗУВАННЯ СОЦІАЛЬНО–ЕКОНОМІЧНИХ ДАНИХ ТЕРИТОРІАЛЬНИХ ГРОМАД

#### 6.1 Метод прогнозування соціально–економічних даних територіальних громад

На основі запропонованих узагальненого принципу синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ (див. підрозділ 2.1) та описаних задач деталізовано метод прогнозування соціально–економічних даних ТГ для забезпечення ефективного прийняття рішень. Цей метод спрямований на інтеграцію різноманітних джерел даних, вибір відповідних методів аналізу та формування прогнозів, що сприяють прийняттю релевантних управлінських рішень. Її основною метою є підтримка соціально–економічної стабільності та покращення управління ресурсами громад.

На рис. 6.1 наведено кроки методу прогнозування соціально–економічних даних, що відображає ключові етапи інтеграції, обробки, аналізу та прийняття рішень.



Рис. 6.1. Кроки методу прогнозування соціально–економічних даних ТГ

Нехай  $I$  – множина об'єктів інфраструктури ТГ, які залучені до аналізу та прогнозування:

$$I = \{I_1, I_2, I_3, I_4\}, \quad (6.1)$$

де  $I_1$  – бізнес–інфраструктура, яка включає підприємства, малі бізнеси та ІТ–компанії;  $I_2$  – соціальна інфраструктура, що охоплює послуги для мешканців, аналіз потреб і моніторинг настроїв;  $I_3$  – логістична інфраструктура, що відповідає за транспортування ресурсів та постачання товарів;  $I_4$  – інформаційна інфраструктура, яка містить джерела даних, медіаконтент та соціальні мережі.

Соціально–економічні дані, що використовуються для прогнозування, формалізуються як множина типів даних:

$$D = \{D_1, D_2, D_3\}, \quad (6.2)$$

де  $D_1$  – структуровані дані: числові показники, фінансові звіти, демографічні таблиці;  $D_2$  – неструктуровані дані: текстові документи, пости у соціальних мережах, зображення, відео;  $D_3$  – напівструктуровані дані, зокрема у форматах JSON або XML, отримані через API.

Для прогнозування використовуються методи інтелектуального аналізу даних  $A$ , які формалізуються як:

$$A = \{LSTM, ARIMA, RF, XGBoost\}, \quad (6.3)$$

де кожен метод  $A_k$  застосовується до конкретного типу даних  $D_i$  залежно від характеристик задачі:  $LSTM$  – використовується для аналізу та прогнозування часових рядів, таких як продажі чи демографічні показники;  $ARIMA$  – модель авторегресії для прогнозування тенденцій на основі історичних даних;  $RF$  (Random Forest) – метод дерев рішень для аналізу структурованих даних;  $XGBoost$  – градієнтний бустинг для високоточних прогнозів.

Процес інтеграції даних з різних джерел формалізується як:

$$D_{\text{інтегровані}} = D_1 \oplus D_2 \oplus D_3, \quad (6.4)$$



де оператор  $\oplus$  виконує злиття даних різних типів у єдину базу, що забезпечує основу для подальшого аналізу.

Кожне завдання прогнозування формулюється як трійка:

$$Z_l = (I_j, D_i, A_k), \quad (6.5)$$

де  $I_j \in I, D_i \in D, A_k \in A$ . Ці задачі включають визначення ключового об'єкта аналізу, тип даних та метод прогнозування.

На основі результатів аналізу формуються управлінські рішення  $R_p$ , які можна описати функцією:

$$R_p = g(Z_l), \quad (6.6)$$

де  $g(Z_l)$  – функція, що визначає оптимальне управлінське рішення. Наприклад:

- прогнозування попиту на товари для планування постачання;
- виявлення регіонів з високим ризиком економічної нестабільності;
- формування стратегій розвитку на основі прогнозованих трендів.

На основі методу прогнозування соціально–економічних даних ТГ [457] у наступних підрозділах деталізовано інтелектуальні підходи до прогнозування попиту на продукцію [240] (підрозділ 6.2) та вибору оптимальної локації для старту бізнесу в ТГ [253, 456] (підрозділ 6.3). У підрозділі 6.4 наведено експериментальні результати, що підтверджують ефективність цих методів, зокрема у задачах прогнозування попиту на продукцію (підрозділ 6.4.1) та вибору локації для бізнесу (підрозділ 6.4.2). Розроблений метод демонструє значний потенціал для підтримки стратегічного планування та прийняття управлінських рішень у громадах, забезпечуючи адаптивність і точність прогнозування. Цей метод успішно реалізовано у низці наукових робіт, зокрема [207, 211, 214], що підкреслює його актуальність та практичну значущість.

## 6.2 Інтелектуальний метод прогнозування попиту на продукцію

У цьому підрозділі представлено інтелектуальний метод прогнозування попиту на продукцію, який базується на методі прогнозування соціально–економічних даних ТГ, розробленому в підрозділі 6.1. Запропонований метод використовує сучасні технології машинного навчання для більш точного й адаптивного прогнозування попиту в умовах кризи.

Запропонований метод розширює підходи, застосовані в дослідженнях (див. підрозділ 1.2.2), як–от використання LSTM мереж для аналізу попиту в управлінні ланцюгами поставок [280] та прогнозування кількості пацієнтів в екстрених медичних установах [434], інтегруючи більш широкий спектр джерел даних та різноманітних технік МН. Це не лише забезпечує більшу гнучкість і точність прогнозування, але й робить цей метод більш адаптованим до різноманітних умов і секторів, особливо в умовах кризи.

У сучасному світі, де дані є критично важливим ресурсом для прийняття обґрунтованих бізнес–рішень, прогнозування часових рядів відіграє ключову роль. Це особливо важливо в умовах, де змінність ринку та потреби клієнтів постійно еволюціонують, особливо в умовах кризи. Прогнозування даних не лише дає можливість краще розуміти поточні тренди, але й передбачати майбутні зміни, що є надзвичайно цінним для стратегічного планування та оптимізації ресурсів. В цьому контексті, розроблено інтелектуальний метод прогнозування попиту товарів/послуг в умовах кризи (див. додаток А, рис. А.11), який охоплює етапи від вибору та підготовки даних до визначення оптимальної моделі прогнозування та візуалізації результатів.

Опис інтелектуального методу прогнозування попиту товарів/послуг в умовах кризи представлено такими кроками:

Крок 1. Збір даних (блок 1). Збираються дані з різних джерел:

–відгуки клієнтів з онлайн–джерел. Це можуть бути текстові дані, які можна представити як множину  $F = \{f_1, f_2, \dots, f_m\}$ , де  $f_i$  – окремий відгук клієнта;

– дані з роздрібних точок. Інформація про продажі в різних магазинах, позначена як  $R = \{r_1, r_2, \dots, r_k\}$ , де  $r_j$  – запис про продаж у конкретній точці;

– історичні дані продажів за попередні періоди. Ці дані формують часовий ряд  $S = \{(t_i, s_i)\}$ , де  $t_i$  – час,  $s_i$  – кількість проданих товарів/послуг у момент часу  $t_i$ .

Крок 2. Інтеграція та підготовка даних (блок 2): об'єднання даних у OLAP бази для поліпшення структуризації; очистка, нормалізація та трансформація даних перед аналізом. Позначимо набір даних як  $D$ , де  $D = \{d_1, d_2, \dots, d_n\}$  та кожен елемент  $d_i$  містить параметри: час  $t$ , кількість послуг/товарів  $s$ , продажі послуг/товарів  $p$ , позитивні відгуки про послуг/товарів  $f_p^s$ , негативні відгуки про послуг/товарів  $f_n^s$ . Нормалізація даних може проводитися за допомогою методів, таких як мін–макс скейлінг:

$$s_i^{norm} = \frac{s_i - s_{min}}{s_{max} - s_{min}}, \quad (6.7)$$

де  $s_{min}$  та  $s_{max}$  – мінімальне та максимальне значення кількості товарів/послуг у наборі даних.

Крок 3. Розділення даних (блок 3).

3.1. Створення тренувального та тестового наборів. Дані поділяються на тренувальний і тестовий набори, зазвичай на основі часових інтервалів для забезпечення об'єктивності оцінювання.

$$\text{Тренувальний набір: } D_{train} = \{d_1, \dots, d_{t-0.2 \times t}\}. \quad (6.8)$$

3.2. Визначення періоду тренування та тестування. Історичні дані за певний період використовуються для тренування, а останні дані – для тестування.

$$\text{Тестовий набір: } D_{test} = \{d_{t-0.2 \times t}, \dots, dt\}. \quad (6.9)$$

Крок 4. Вибір кращої моделі для прогнозування часових рядів (блок 4).

4.1 Вивчення різних моделей. Розгляд різних моделей МН, зокрема, ансамблевих методів, а також спеціалізованих моделей для часових рядів. Розглядаємо набір моделей  $M = \{m_1, m_2, \dots, m_k\}$ , де кожна  $m_i$  є потенційною моделлю для прогнозування.

4.2. Крос–валідація та оцінка моделей. Для кожної моделі  $m_i$  виконуємо крос–валідацію на  $D_{train}$  та обчислюємо метрики (див. підрозділ 1.3), такі як RMSE, MAE [13]. Вибір кращої моделі  $m_{best}$  заснований на мінімізації цих метрик.

Крок 5. Побудова комплексного прогнозу.

5.1. Тренування вибраної моделі (блок 5). Використання  $m_{best}$  для прогнозування окремих параметрів  $\hat{p}, \hat{f}_{ps}, \hat{f}_{ns}, \hat{f}_{pp}, \hat{f}_{np}$  на  $D_{test}$ . Параметри моделі  $m_i$  налаштовуються шляхом мінімізації функції втрат:

$$L_i = \sum_{j=1}^{N_{train}} l(y_j, \hat{y}_j^{(i)}), \quad (6.10)$$

де  $l$  – функція втрат (наприклад, квадратична похибка);  $y_j$  – фактичне значення;  $\hat{y}_j^{(i)}$  – прогноз моделі  $m_i$ .

Обирається модель  $m^*$  з найкращими показниками:

$$m^* = \arg \min_{m_i \in M} RMSE_i. \quad (6.11)$$

5.2. Інтеграція прогнозів (блок 6). Використання прогнозованих значень ключових параметрів для формування кінцевого прогнозу. Комбінування прогнозів окремих параметрів для побудови фінального прогнозу  $\hat{S}_{final}$  кількості послуг.

Крок 6. Обчислення інтервалу похибки (блок 7).

6.1. Застосування Методу Bootstrap [14]: створення кількох варіантів  $B$  навчальних наборів для оцінки стабільності моделі  $D_{train}$ .

6.2. Обчислення інтервалу надійності на основі розподілу прогнозів  $\{\hat{S}^{(b_1)}, \dots, \hat{S}^{(b_B)}\}$ .

Крок 7. Візуалізація прогнозу з інтервалом похибки (блок 8).

7.1. Візуалізація фактичних  $s$  та прогнозованих  $\hat{S}_{final}$  значень.

7.2. Відображення фактичних  $s$  та прогнозованих  $\hat{S}_{final}$  значень.

7.3. Відображення інтервалу надійності як затіненої області навколо  $\hat{S}_{final}$

Розроблений інтелектуальний метод успішно реалізований та протестований в експериментальних дослідженнях, результати яких представлені в підрозділі 6.4.1. Отримані результати підтверджують ефективність та практичну цінність

запропонованого підходу для підтримки стратегічного планування та оптимізації ресурсів у ТГ.

### **6.3 Інтелектуальний метод вибору локації для старту бізнесу в територіальних громадах**

У цьому підрозділі описано інтелектуальний метод вибору локації для старту бізнесу в ТГ, що базується на методі прогнозування соціально–економічних даних, представленому в підрозділі 6.1. Метод використовує технології машинного навчання для визначення оптимальних місць для відкриття нових бізнес–проектів, враховуючи географічні та ринкові особливості, що є ключовим для розвитку економічної інфраструктури громади.

Дослідження [48] (див. підрозділ 1.2.2), яке використовує алгоритми машинного навчання для прогнозування успіху стартапів, є значущим аналогом цього дослідження. Проте, поточне дослідження розширює сферу застосування МН, зосереджуючись на аналізі оптимальних локацій для запуску бізнес–проектів. Відмінність полягає у підході, що враховує географічні та місцеві ринкові особливості, які можуть відігравати критичну роль у визначенні потенціалу успіху стартапу. Таке фокусування на географічному контексті та специфіці ринку відкриває нові перспективи для розуміння динаміки успішності стартапів.

Інтелектуальний метод вибору локації для старту бізнесу в ТГ – це комплексний підхід, який використовує технології машинного навчання для визначення оптимального місця для відкриття нового бізнесу. Ось основні кроки цього методу та його структура (див. додаток А, рис. А.12):

Крок 1. Збір даних. Цей процес включає збір даних з різних джерел:

1.1. Відеокамери. Камери відеоспостереження, розміщені в магазинах та на вулицях міста, можуть надавати цінну інформацію про поведінку покупців та пішоходів, їх стать та вікову категорію. Вони можуть виявити, які магазини або

вулиці приваблюють найбільше людей, в який час дня та дні тижня. Зібрані дані можна представити як множину спостережень:

$$D_{\text{відео}} = \{ (x_i, y_i, t_i, s_i, v_i) \mid i = 1, 2, \dots, N_{\text{відео}} \}, \quad (6.12)$$

де  $x_i, y_i$  – координати місцезнаходження спостереження;  $t_i$  – час спостереження;  $s_i$  – стать особи (наприклад,  $s_i \in \{\text{чоловік, жінка}\}$ );  $v_i$  – вікова категорія особи;  $N_{\text{відео}}$  – загальна кількість спостережень з відеокамер.

1.2. Мобільні додатки та сервіси геолокації. З мобільного додатка ТГ можна використовувати вбудовані інструменти аналітики для збору даних про поведінку користувачів. Деякі сервіси, такі як Google Maps чи відповідні мобільні додатки, надають API чи геолокацію, які дозволяють відстежувати місцезнаходження користувачів (за їхньою згодою). Це може бути корисно для визначення популярних місць і маршрутів. Мобільні додатки ТГ та сервіси геолокації надають дані про поведінку користувачів:

$$D_{\text{гео}} = \{ (x_j, y_j, t_j, u_j) \mid j = 1, 2, \dots, N_{\text{гео}} \}, \quad (6.13)$$

де  $x_j, y_j$  – координати користувача;  $t_j$  – час спостереження;  $u_j$  – унікальний ідентифікатор користувача;  $N_{\text{гео}}$  – загальна кількість спостережень з геолокаційних сервісів.

Крок 2. Обробка та аналіз даних. Зібрані дані потім обробляються та аналізуються за допомогою алгоритмів МН. Це може включати розпізнавання образів, визначення шаблонів, виявлення трендів та визначення ключових впливових факторів.

## 2.1. Обробка даних.

2.1.1. Розпізнавання образів [425]. Використовуючи алгоритми МН, дані з відеоаналітики проходять процес розпізнавання образів. Це може включати в себе визначення облич, об'єктів, відстеження руху та інші техніки, які допомагають визначити стать та вік людей. Дані з відеоаналізу проходять процес розпізнавання образів за допомогою алгоритмів машинного навчання:

– виявлення облич та об’єктів. Використовуються конволюційні нейронні мережі (CNN) для виявлення облич та інших об’єктів на зображеннях. Математично, для зображення  $I_i$ :

$$f_{\text{обличчя}}(I_i) = \begin{cases} 1, & \text{якщо обличчя виявлено} \\ 0, & \text{якщо обличчя не виявлено} \end{cases}; \quad (6.14)$$

– визначення статі та віку. На основі виявлених облич моделі прогнозують стать  $s_i$  та вікову категорію  $v_i$ :

$$s_i = f_{\text{стать}}(I_i), v_i = f_{\text{вік}}(I_i), \quad (6.15)$$

де  $f_{\text{стать}}$  та  $f_{\text{вік}}$  – моделі класифікації.

2.1.2. Попередня обробка числових масивів даних [257]. Обробка даних включає в себе перевірку на наявність відсутніх або некоректних даних, нормалізацію даних для забезпечення їх сумісності та видалення шуму або викидів, які можуть спотворити результати аналізу.

2.2. Сегментація даних [248]. Після обробки даних вони аналізуються за допомогою різних алгоритмів МН, що може включати групування, класифікацію чи кластеризацію.

2.3. Візуалізація даних. Для кращого розуміння результатів аналізу даних їх часто візуалізують. Візуалізація може включати в себе графіки, діаграми, теплові карти та інші графічні зображення, які допомагають інтерпретувати складні набори даних.

Крок 3. Моделювання та прогнозування. На основі аналізу даних створюються моделі, які можуть прогнозувати потенційну успішність бізнесу в різних локаціях для певних сегментів [214].

Прогнозування потенційного обсягу продажів. Використовуються регресійні моделі, такі як лінійна регресія:

$$y = \alpha_0 + \sum_{i=1}^n \alpha_i x_i + \epsilon, \quad (6.16)$$

де  $y$  – прогнозований обсяг продажів;  $x_i$  – фактори впливу;  $\alpha_i$  – коефіцієнти моделі;  $\epsilon$  – похибка.

Розраховується індекс конкурентності  $C(L)$  для локації  $L$ :

$$C(L) = \sum_{j=1}^M \frac{1}{d(L, L_j)}, \quad (6.17)$$

де  $M$  – кількість конкурентів;  $L_j$  – локація конкурента  $j$ ;  $d(L, L_j)$  – відстань між локаціями.

Крок 4. Вибір локації. На основі результатів моделювання та прогнозування вибирається найбільш оптимальна локація для відкриття нового бізнесу. На основі результатів моделювання вибирається оптимальна локація  $L^*$ :

$$L^* = \arg \max_{L \in \mathcal{L}} [P(L) - C(L) - O(L)], \quad (6.18)$$

де  $L$  – множина доступних локацій;  $P(L)$  – прогнозований прибуток у локації  $L$ ;  $C(L)$  – оцінка конкуренції в локації  $L$ ;  $O(L)$  – операційні витрати в локації  $L$ .

Запропонований інтелектуальний метод реалізований та протестований, результати експериментів представлені в підрозділі 6.4.2. Експериментальні дані підтверджують ефективність методу у визначенні оптимальних локацій, що забезпечують високу ймовірність успіху нового бізнесу в ТГ.

## 6.4 Експериментальні результати

За запропонованим методом прогнозування соціально–економічних даних ТГ (див. підрозділ 6.1) та принципом синтезу інформаційної технології (див. підрозділ 2.1) розроблено модулі інформаційної системи, які вирішують поставлені задачі прогнозування попиту на продукцію та вибору локації для старту бізнесу в ТГ.

### 6.4.1 Прогнозування попиту на продукцію

Для апробації запропонованого методу обрано проєкт «СмартМед», який спрямований на розробку інноваційних медичних рішень в контексті смарт–міст. Проєкт охоплює широкий спектр активностей, включаючи телемедицину, аналіз



медичних даних та інтелектуальні рекомендації для підтримки здоров'я. За допомогою аналізу даних, було визначено ключові параметри для прогнозування, серед яких: paracetamol sales, positive feedback about the service, negative feedback about the service, positive feedback on paracetamol, negative feedback on paracetamol, number of services. На основі критеріїв RMSE та MAE було вибрано модель HistGradientBoostingRegressor як найбільш точну для прогнозування цих параметрів. Результати порівняння цієї моделі з іншими алгоритмами машинного навчання (Таблиця 6.1).

Таблиця 6.1

## Результати оцінювання моделей

Параметр	HistGradientBoosting Regressor [15]		XGBoost [16]		CatBoost [17]		LightGBM [18]	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
paracetamol sales	10.2	8.6	11.2	9.3	15.3	9.7	11.2	12.5
positive feedback about the service	12.3	10.1	22.3	20.3	34.4	39.1	11.2	12.4
negative feedback about the service	14.8	11.9	12.2	13.9	12.3	14.9	15.3	13.2
positive feedback on paracetamol	7.2	6.5	22.3	26.5	34.2	37.6	35.2	34.2
negative feedback on paracetamol	13.5	10.8	16.6	18.9	53.4	60.5	53.2	62.3
number of services	9.5	8.2	19.5	28.2	59.1	45.9	39.1	48.7

Для прогнозування параметра number of services було застосовано метод XGBoost, який продемонстрував вищу точність за  $R^2$  (Таблиця 6.2). Враховуючи отримані прогнози та їх взаємозв'язки, було сформовано остаточний прогноз для number of services (Рис. 6.2).

Точність для «number of services»  $R^2$  відносно інших параметрів

Модель	$R^2$
HistGradient BoostingRegressor	0,91
XGBoost	0,98
CatBoost	0,89
LightGBM	0,85

Далі побудовано модель та проведено аналіз точності моделі за допомогою показників RMSE та MAE, що підтвердив її точність (Рис. 6.2). Значення RMSE склало приблизно 10.35, вказуючи на середньоквадратичну помилку, а MAE, який склав 8.97, відображає середню абсолютну помилку прогнозів.

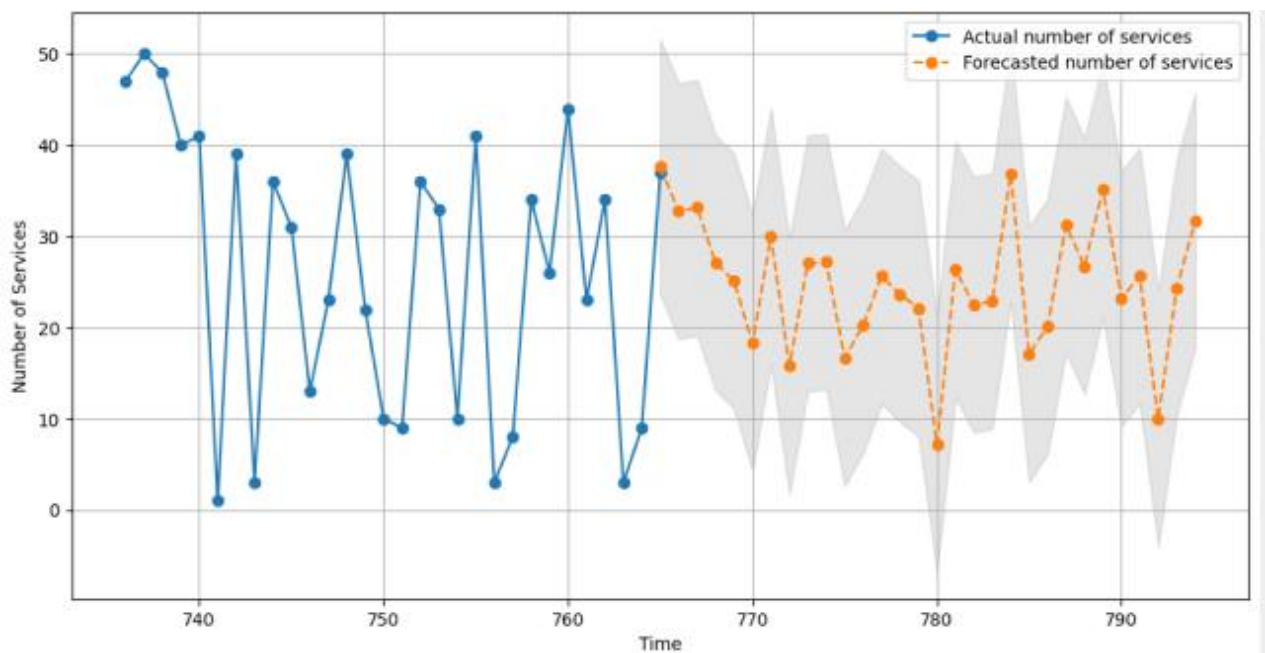


Рис. 6.2. Інтелектуальне прогнозування попиту

Отже, результати апробації свідчать про практичну застосовність запропонованого підходу до прогнозування в межах проєкту «СмартМед», зокрема для управління ресурсами в сфері охорони здоров'я.

В умовах кризи можуть відбуватися суттєві зміни попиту на товари та послуги, що ускладнює завдання прогнозування для бізнесу. З метою реагування на такі зміни

та забезпечення ефективного управління запасами і виробництвом розроблено підхід інтелектуального прогнозування попиту (Рис. 6.3). Процес прогнозування починається з парсингу відгуків клієнтів (блок 1) чи збору даних зі супермаркетів (блок 2). Також у базу передаємо дані щодо продажу товарів чи послуг за попередні роки (блок 3). Ці дані зливаються в OLAP [26] базу даних (блок 4), де вони готуються для подальшого аналізу. Далі система виконує запити за назвами товарів (блок 5) та збирає дані [27] про їхній попит в розрізі часу, які подаються в таблицю (блок 6).

На основі цих даних інтелектуальна підсистема (блок 7) розробляє багаторівневий прогноз, враховуючи зміни в споживчому попиті, що можуть виникнути в умовах кризи. Результати прогнозування (блок 8) надають бізнесу можливість попередньо планувати виробництво та управляти запасами таким чином, щоб відповідати змінним потребам ринку. Такий підхід допомагає знизити ризики та оптимізувати бізнес–процеси в умовах економічної нестабільності.

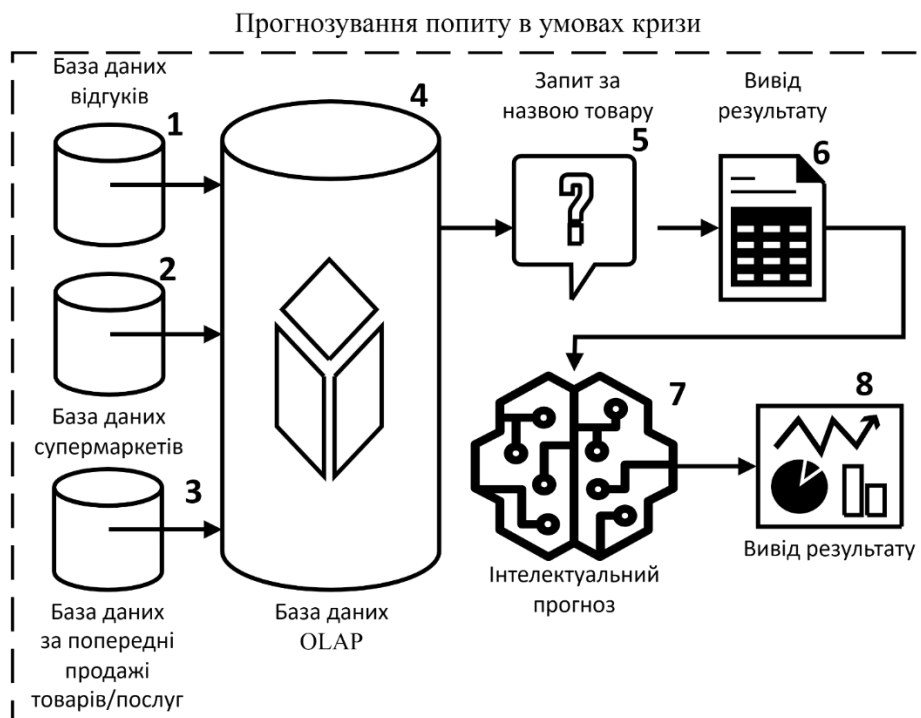


Рис. 6.3. Блок прогнозування попиту в умовах кризи

Далі для «СмартМед» представлено інтелектуальне прогнозування попиту (див. Рис. 6.3), враховуючи кризові ситуації (Рис. 6.4), що дозволяє ефективно реагувати на

такі зміни та оптимізувати управління запасами та виробництвом. Для підприємця процес прогнозування розпочинається із запиту на збір даних з різних джерел (Рис. 6.4). Система сама парсить відгуки клієнтів з онлайн-джерел (наприклад, вебсайтів, соціальних мереж) та отримує дані із супермаркетів та аптек про медичні товари. Також у систему передаються дані про продажі товарів і послуг за попередні роки. На основі методів прогнозування виводяться результати, що надають бізнесу можливість стратегічно управляти компанією.

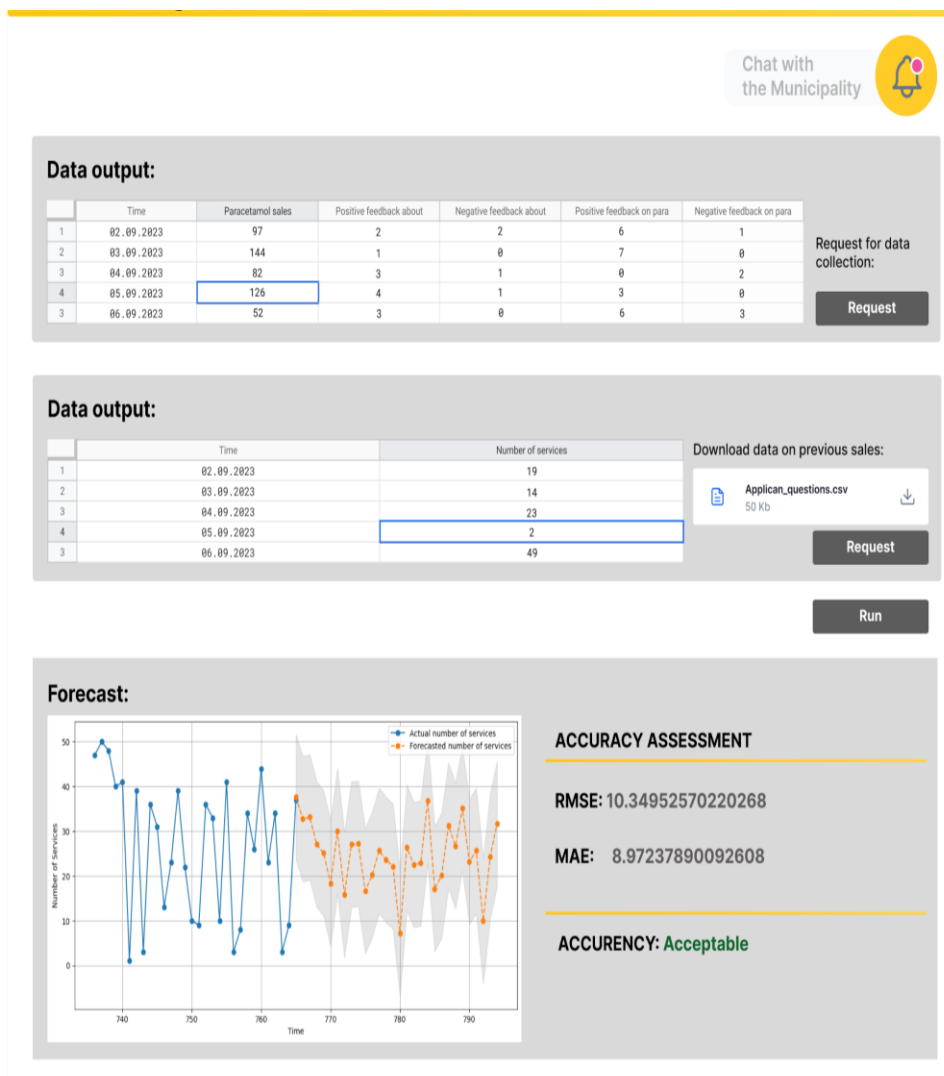


Рис. 6.4. Вікно інтелектуального прогнозу попиту

Отже, у цьому прикладі зроблено прогнозування параметра number of services на основі інших параметрів, таких як: paracetamol sales, positive feedback about the service, negative feedback about the service, positive feedback on paracetamol i negative

feedback on paracetamol, використовуючи модель HistGradientBoostingRegressor. Після навчання цієї моделі на даних зроблено прогноз для кожного з параметрів paracetamol sales, positive feedback about the service, negative feedback about the service, positive feedback on paracetamol і negative feedback on paracetamol методом XGBoost на 10 днів у майбутньому. Після отримання цих прогнозів побудовано прогноз для параметра number of services на такий самий період, використовуючи раніше отримані прогнози з врахуванням взаємозв'язків між цими параметрами.

Отримані (див. Рис. 6.4) значення RMSE (Root Mean Square Error) та MAE (Mean Absolute Error) вказують на точність прогнозу моделі. RMSE, який становить приблизно 10,35, вимірює середньоквадратичну помилку між фактичними та прогнозованими значеннями. Вище значення RMSE вказує на те, що середньоквадратична помилка становить близько 10,35 одиниць. MAE, який складає приблизно 8,97, вимірює середню абсолютну помилку між фактичними та прогнозованими значеннями. Більш низьке значення MAE вказує на меншу середню абсолютну помилку, що свідчить про кращу точність прогнозу моделі. Отже, отримані результати свідчать про те, що модель демонструє прийнятну точність у прогнозуванні кількості послуг.

Прогнозування попиту в умовах кризи демонструє високу ефективність роботи системи (див. додаток Г). Час із системою склав від 1 до 5 годин, що відповідає аналогічним витратам часу без її використання, однак система забезпечила оптимізацію процесу прогнозування, досягаючи ефективності ~95,83%. Це свідчить про значне підвищення точності і надійності прогнозів, особливо в умовах нестабільності, що є критично важливим для прийняття оперативних управлінських рішень.

## 6.4.2 Вибір локації для старту бізнесу в ТГ та прогнозування потоку відвідувачів

Отже, для апробації запропонованого методу використано відео [101] з YouTube-каналу, що демонструє людський потік у торговій зоні міста протягом майже 29 хвилин.

У першу чергу, відео з YouTube завантажується за допомогою бібліотеки `pytube` та відкривається в `OpenCV`. Потім використовуються нейронні мережі, завантажені з `.pb`, `.pbtxt` та `.caffemodel` файлів, для виявлення облич та оцінки віку та статі.

Виявлення облич на відео здійснюється за допомогою нейронної мережі. Кожен кадр відео проходить через мережу, і якщо впевненість в тому, що обличчя виявлено, більше 60 %, то обличчя вирізається з кадру.

Для кожного обличчя створюється `blob`, який передається в моделі для визначення віку та статі.

На завершення, деталі про вік та стать кожної особи зберігаються у список, перетворюються в `DataFrame` та зберігаються в `csv`-файлі.

Згідно з вихідними даними (Рис. 6.5), загалом було виявлено 290 осіб. Найбільша кількість людей належить до вікової групи 25–32 років – 93 особи. За статевим розподілом більшість осіб – жінки (168 осіб), у порівнянні з чоловіками (122 особи). Менше всього людей було з вікової групи 60–100 років – всього 5 осіб.

У подальшому кроці було сформовано словник, що поєднує різні комбінації віку та статі з відповідними видами бізнесу, які найімовірніше будуть релевантними для цих демографічних груп. За допомогою цього словника було створено новий стовпець 'Best Business' в `DataFrame`. Він заповнюється шляхом поєднання значень віку та статі з кожного запису та маппінгу цієї комбінації на відповідну бізнес-категорію. Результатом є `DataFrame`, який включає рекомендовану бізнес-категорію для кожного індивіда на основі їх віку та статі, що може слугувати корисним інструментом при визначенні бізнес-стратегій та маркетингових підходів.

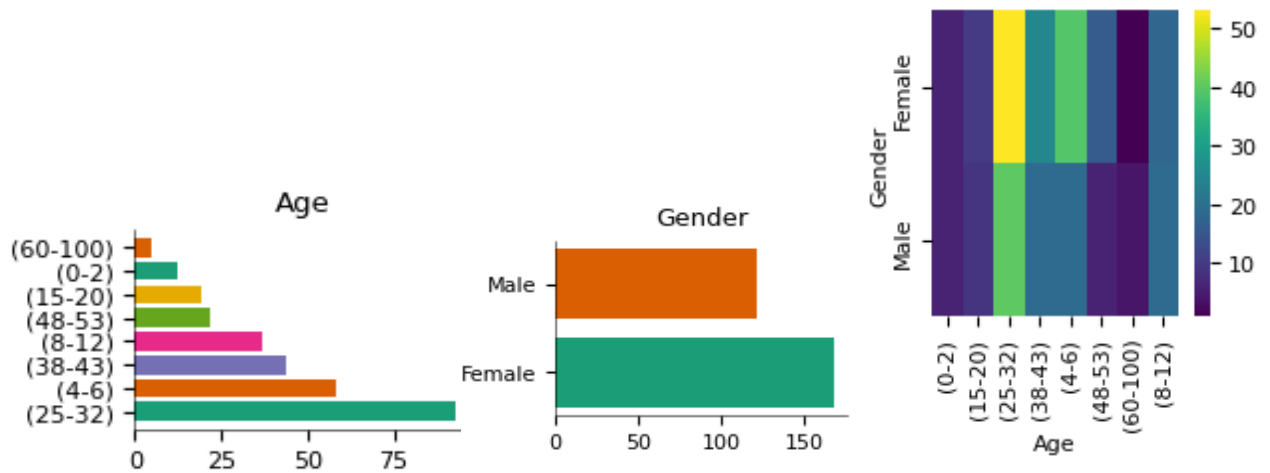


Рис. 6.5. Статистика розпізнаних осіб

В наступному кроці Best Business розбивається на індивідуальні бізнес-категорії [317]. Кожен запис в Best Business, що раніше був списком, тепер розбитий на окремі категорії. Ці розділені категорії згортаються в стек, дозволяючи підрахувати частоту кожної категорії в даних.

Кінцевим результатом (Рис. 6.6) є відображення категорій бізнесу та їх кількості, що показує, як часто кожна категорія зустрічається у вхідних даних. Ця інформація може надати корисний огляд популярних категорій бізнесу в різних вікових та статевих групах.

Books	158
Technology	118
Fashion	105
Cosmetics	104
Home Goods	102
Toys	70
Children Clothes	58
Children Books	58
Sport Goods	40
Clothes	40
Educational Games	37
Sportswear	28
Health Products	27
Cars	25
Fashion Clothes	18
Baby Clothes	12
Baby Food	12
Accessories	10
Electronics	9
Video Games	9

Рис. 6.6. Вивід найкращого товару для відповідної локації

Books (див.Рис. 6.6) знаходяться на чолі списку з 158 випадками, що показує, що ця категорія є надзвичайно популярною серед людей, які найчастіше проходять повз камеру. Technology та Fashion, з відповідно 118 та 105 згадками, також виявилися дуже затребуваними категоріями. Отже, у даному місці варто розглянути введення бізнесу, що стосується цих трьох категорій. Ці дані можуть допомогти в належний час скерувати бізнес-стратегії, створення продуктів та маркетингові заходи, враховуючи інтереси різних вікових та гендерних груп.

Можна далі розширити цю систему, встановивши більше камер по ТГ і аналізуючи дані з кожної окремої локації. Це дозволить отримати більш детальну карту демографічних характеристик різних районів міста і, можливо, допоможе виявити патерни, які можуть бути використані для оптимізації бізнес-операцій.

Також можливим є аналіз відеоданих тривалішого періоду, що надасть більше інформації про тенденції та зміни в демографічній структурі.

Нарешті, важливо врахувати, що якість вхідних даних має велике значення для цього процесу. Камери повинні бути встановлені таким чином, щоб чітко було видно обличчя людей, що забезпечує найкращі результати визначення віку та статі.

Результати моделювання потоку відвідувачів за допомогою SARIMA демонструють динаміку відвідувань протягом восьмигодинного періоду вихідного дня. Графік (Рис. 6.7) вказує на те, що реальні дані характеризуються поступовим зростанням кількості відвідувачів у перші п'ять годин з помітним піком на шостій та сьомій годинах (відповідно 20 % і 30 % від загального обсягу відвідувачів). Прогнозовані результати, отримані методом SARIMA, загалом відображають цю тенденцію, демонструючи кореляцію з реальними спостереженнями та забезпечуючи основи для прийняття рішень.

Для оцінки точності моделі прогнозування було використано показники середньої абсолютної похибки (MAE) та середньоквадратичної похибки (RMSE). Значення MAE склало 9,8, що свідчить про незначні середні відхилення між реальними та прогнозованими даними, тоді як  $RMSE = 12,3$  – демонструє непогану точність моделі в умовах значних змін у потоці відвідувачів.



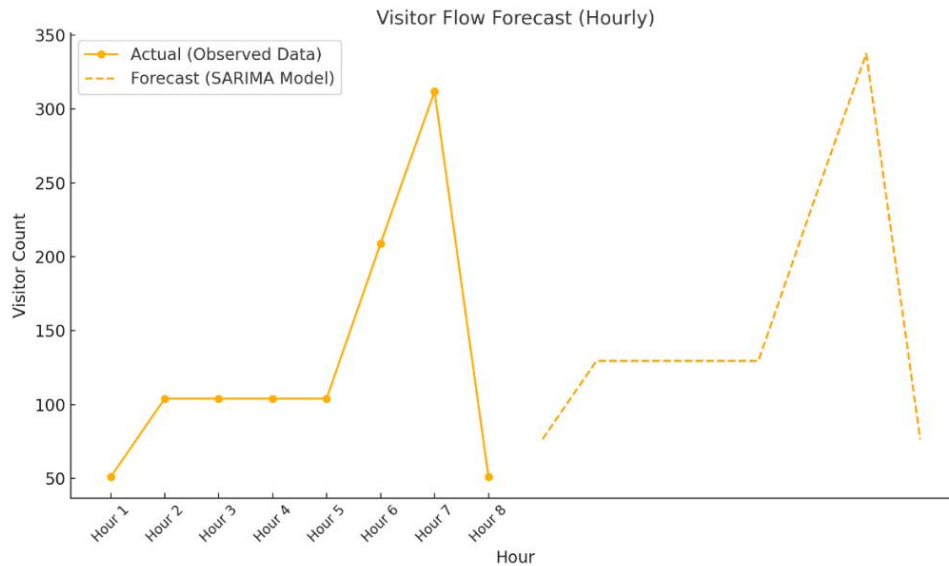


Рис. 6.7. Прогноз потоку відвідувачів за допомогою SARIMA

Результати аналізу (див. Рис. 6.6) підтверджують, що основну кількість ресурсів (персонал, маркетинг, матеріально–технічну базу) слід сконцентрувати на шостій та сьомій годинах, коли кількість відвідувачів досягає пікового рівня. На останній годині спостерігається помітне зниження активності, що свідчить про доцільність оптимізації операційних витрат у цей період. Таким чином, використання SARIMA дозволяє прогнозувати потік відвідувачів з високою точністю, забезпечуючи інструменти для стратегічного планування бізнес–процесів.

Модуль (Рис. 6.8) інтелектуального вибору локації для старту бізнесу представлений як ключовий елемент системи, який виконує ітеративний процес аналізу даних для визначення оптимального місця розташування нового бізнесу. На етапі збору даних модуль отримує інформацію з відеоаналітики (блок 1), що аналізує потоки пішоходів та визначає демографічні показники, а також з мобільних сервісів (блок 2), які надають дані про геолокацію користувачів і поведінкові патерни. Додатково враховується інформація про конкуренцію та інфраструктуру. Обробка даних включає попередню підготовку, очищення, стандартизацію та розпізнавання образів для сегментації за віком і статтю. Далі застосовуються алгоритми машинного навчання для аналізу шаблонів і трендів, а також візуалізації результатів.

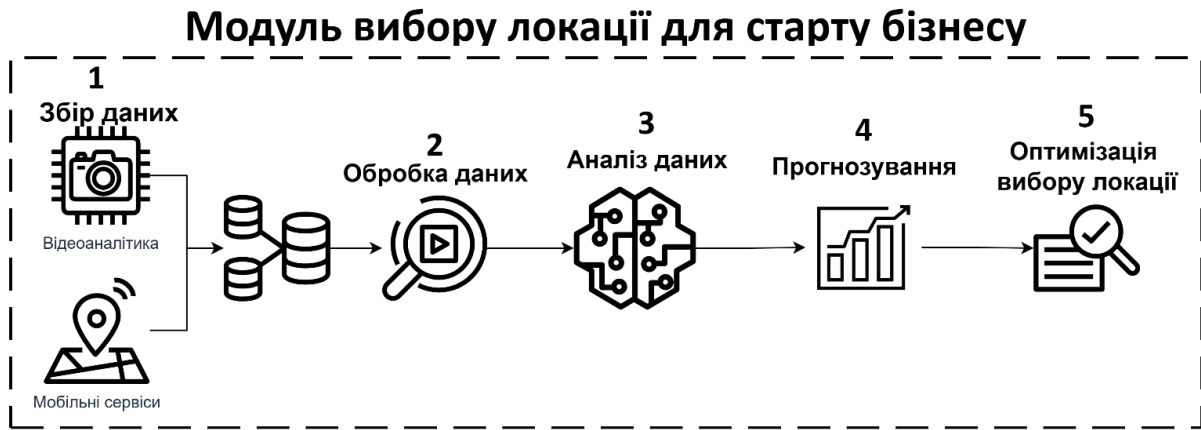


Рис. 6.8. Блок вибору локації для старту бізнесу

На основі цього здійснюється прогнозування можливого обсягу продажів та конкурентного середовища за допомогою регресійних і SARIMA–моделей. На фінальному етапі модуль забезпечує оптимізацію вибору локації (вихід на блок 4) шляхом інтеграції прогнозованих показників прибутку, операційних витрат та конкурентності. Це дозволяє сформувані обґрунтовані рішення для ефективного старту бізнесу та сприяє його сталому розвитку.

Представлене вікно (рис.6.9) модуля інтелектуального вибору локації для старту бізнесу забезпечує інтерактивний інтерфейс для аналізу вхідних даних і надання прогнозів.

У верхній частині вікна (див.рис.6.9) користувач може завантажити посилання на відеодані, що аналізуються для збору інформації про поведінкові та демографічні патерни. Центральна частина інтерфейсу включає поля введення для посилання на онлайн–магазин, що дозволяє інтегрувати додаткові дані про діяльність компанії. У нижній частині представлено блок результатів аналізу, включаючи графіки статевого і вікового розподілу відвідувачів, демографічні сегменти та прогноз потоку відвідувачів за SARIMA–моделлю. Цей інтегрований підхід забезпечує візуалізацію результатів у вигляді графіків і таблиць, що дозволяє користувачам ефективно інтерпретувати отриману інформацію та приймати обґрунтовані рішення щодо вибору локації.

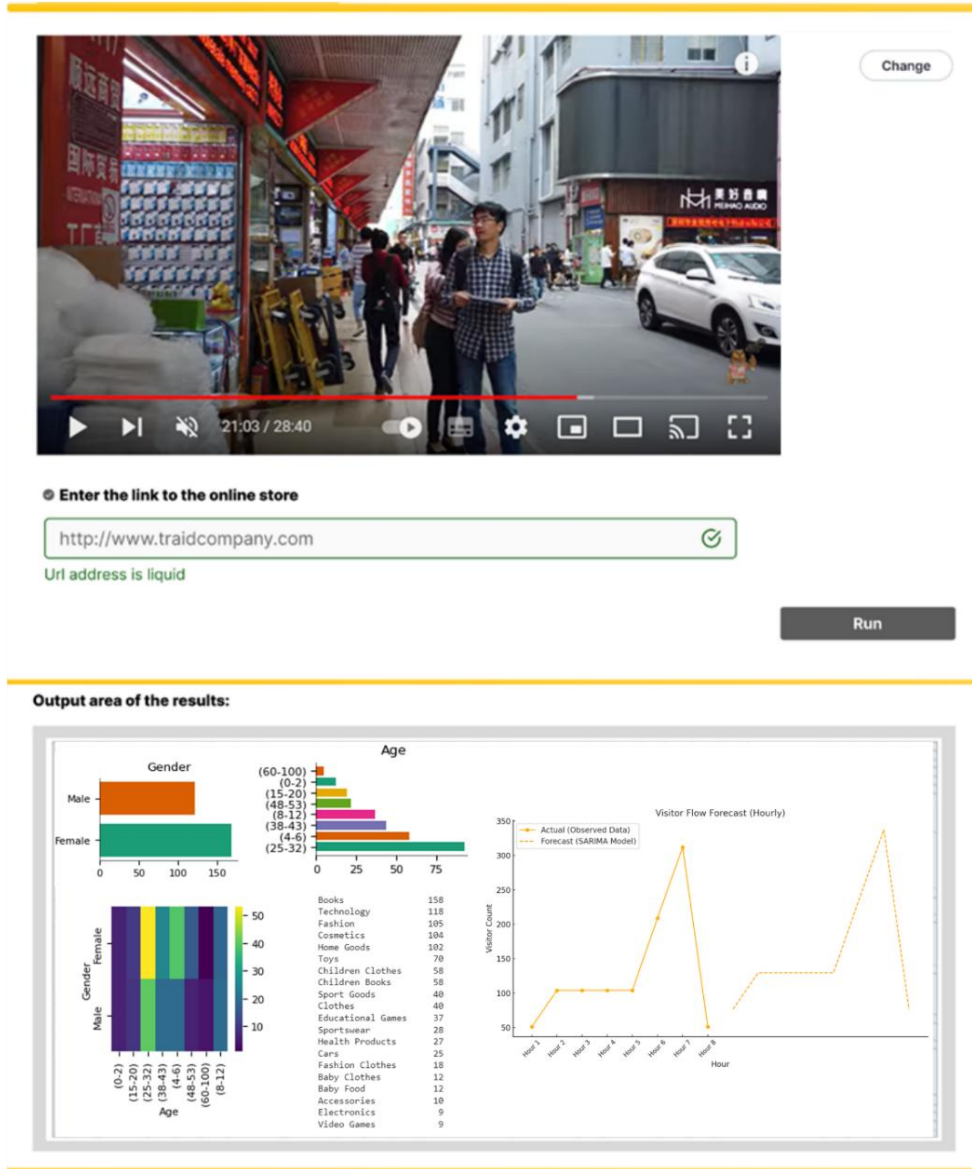


Рис. 6.9. Вікно вибору локації для старту бізнесу

Вибір локації для старту бізнесу та прогнозування потоку відвідувачів демонструє високу ефективність роботи системи (див. додаток Г). Час роботи із системою варіюється від 1 до 5 годин залежно від складності аналізу, що суттєво скорочує час порівняно з традиційними методами, які можуть займати від 1 до 10 днів. Ефективність методу досягає  $\sim 98,75\%$ , що підтверджує його здатність значно підвищувати точність і швидкість прогнозування. Це особливо важливо для обґрунтованого прийняття рішень у динамічних бізнес-умовах, знижуючи ризики та забезпечуючи максимальну ймовірність успіху нового бізнесу.

## 6.5 Висновки

У розділі 6 представлено розробку методу прогнозування соціально–економічних даних ТГ, який базується на інтеграції даних з різних джерел та застосуванні методів інтелектуального аналізу. Результати дослідження деталізовано за кожним підрозділом, що дозволяє оцінити досягнення та значення проведених досліджень.

Розроблено метод прогнозування соціально–економічних даних ТГ, який враховує специфіку даних різних типів (структурованих, напівструктурованих, неструктурованих) та використовує сучасні методи аналізу (LSTM, ARIMA, RF, XGBoost). Розроблений підхід дозволяє створювати точні прогнози для стратегічного планування, включаючи виявлення ризиків, прогнозування попиту на товари чи послуги та визначення зон з високим ризиком економічної нестабільності. Це сприяє підвищенню точності управлінських рішень та забезпечує соціально–економічну стабільність громад.

Розроблено інтелектуальний метод прогнозування попиту на продукцію, зокрема в умовах криз. Запропоновано підхід, який враховує дані продажів, відгуки клієнтів та часові ряди для забезпечення гнучкості і точності прогнозів. Метод показав свою ефективність у реальних умовах (на прикладі проекту «СмартМед»), демонструючи високу точність прогнозування параметрів, таких як попит на медичні товари. Результати підтверджують його практичну цінність для планування ресурсів у кризових умовах.

Розроблено інтелектуальний метод вибору оптимальної локації для запуску бізнесу, що базується на аналізі геолокаційних даних, демографічних характеристик та поведінкових індикаторів. Метод використовує технології машинного навчання для аналізу потоків людей, визначення популярних локацій та оцінки конкурентного середовища. Результати апробації засвідчили його здатність забезпечити точність у визначенні найбільш перспективних локацій для стартапів, що сприяє оптимізації бізнес–стратегій та залученню інвестицій.

Експериментальні результати демонструють практичну реалізацію запропонованих методів у задачах прогнозування попиту та вибору локацій. Зокрема,

використання HistGradientBoostingRegressor для прогнозування попиту та застосування аналізу відеоданих для вибору бізнес–локацій підтвердили їхню ефективність і адаптивність. Показники точності (RMSE та MAE) засвідчили високу якість моделей, а скорочення часу виконання задач забезпечило оптимізацію процесів.

Запропоновані методи прогнозування підтвердили свою ефективність у вирішенні стратегічних соціально–економічних задач ТГ. Вони забезпечують гнучкість і точність прогнозування, підвищують якість управлінських рішень та адаптуються до кризових ситуацій. Отримані результати свідчать про доцільність впровадження цих підходів у практичну діяльність громад, що сприятиме сталому розвитку.

## РОЗДІЛ 7.

### МЕТОДИ ТА ЗАСОБИ ГІБРИДНОГО АНАЛІЗУ СОЦІАЛЬНО– ЕКОНОМІЧНИХ ДАНИХ ТЕРИТОРІАЛЬНИХ ГРОМАД

#### 7.1 Метод гібридного аналізу соціально–економічних даних територіальних громад

Зростання складності соціально–економічних процесів у ТГ та необхідність обробки великих обсягів різномірних даних вимагають впровадження нових підходів до їх аналізу. На основі узагальненого принципу синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ (див. підрозділ 2.1) запропоновано метод гібридного аналізу, який поєднує різні методи інтелектуального аналізу даних, зокрема класифікацію, кластеризацію та прогнозування. Цей підхід дозволяє забезпечити гнучкість, адаптивність і комплексність у вирішенні управлінських задач. Використання гібридного аналізу передбачає два рівні (I рівень і II рівень) обробки даних: попереднє групування на першому рівні та детальну обробку на другому, що сприяє ефективному прийняттю стратегічних рішень. Метод враховує як кількісні показники, так і неструктуровані текстові дані, забезпечуючи всебічний підхід до аналізу.

На рис. 7.1 представлено кроки методу гібридного аналізу соціально–економічних даних ТГ, яка демонструє ключові компоненти, етапи інтеграції та обробки даних, а також методи аналізу.

Для гібридного аналізу виділено множину об'єктів інфраструктури ТГ:

$$I = \{I_1, I_2, I_3, I_4\}, \quad (7.1)$$

де  $I_1$  – бізнес–інфраструктура, що включає підприємства, малі бізнеси та ІТ–компанії;  $I_2$  – соціальна інфраструктура, яка займається моніторингом потреб населення та аналізом настроїв;  $I_3$  – логістична інфраструктура, відповідальна за управління

ресурсами та оптимізацію маршрутів;  $I_4$  – інформаційна інфраструктура, що включає джерела даних з медіа та соціальних мереж.

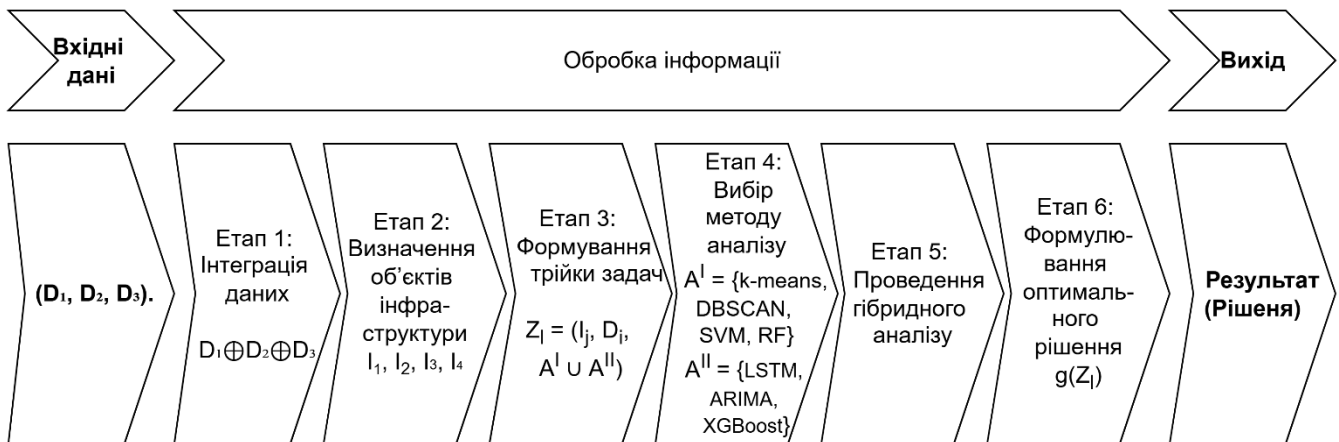


Рис. 7.1. Кроки методу гібридного аналізу соціально–економічних даних ТГ

Множина типів даних, що використовуються в аналізі, описується як:

$$D = \{D_1, D_2, D_3\}, \quad (7.2)$$

де  $D_1$  – структуровані дані: фінансові показники, демографічна статистика;  $D_2$  – неструктуровані дані: текстові документи, публікації в соціальних мережах, відео;  $D_3$  – напівструктуровані дані, такі як JSON чи XML, отримані через API.

Для першого рівня аналізу (рівень I) використовуються базові методи класифікації та кластеризації:

$$A^I = \{k\text{-means, DBSCAN, SVM, RF}\}, \quad (7.3)$$

де: K–Means – розподіляє об'єкти на кластери на основі близькості до центроїдів; DBSCAN – визначає кластери на основі щільності розподілу даних, що дозволяє враховувати шуми; SVM (Support Vector Machine) – використовується для класифікації структурованих даних; Random Forest (RF) – метод для аналізу як кількісних, так і текстових даних.

Другий рівень аналізу (рівень II) передбачає детальний прогноз на основі результатів першого рівня:

$$A^{\text{II}} = \{LSTM, ARIMA, XGBoost\}, \quad (7.4)$$

де: LSTM (Long Short–Term Memory) – модель для аналізу часових рядів; ARIMA – модель для прогнозування трендів на основі історичних даних; XGBoost – модель для точного прогнозування обсягів чи категорій.

Процес гібридного аналізу базується на інтеграції даних, формалізованій як:

$$D_{\text{інтегровані}} = D_1 \oplus D_2 \oplus D_3, \quad (7.5)$$

де оператор  $\oplus$  виконує об'єднання даних різних форматів в єдину базу для подальшого аналізу.

Задачі для аналізу формуються як трійки:

$$Z_l = (I_j, D_i, A_k), \quad (7.6)$$

де  $I_j \in I, D_i \in D, A_k \in A^{\text{I}} \cup A^{\text{II}}$ .

Результати аналізу формуються як:

$$R_p = g(Z_l), \quad (7.7)$$

де  $g(Z_l)$ – функція, що визначає управлінське рішення на основі аналізу, наприклад:

- класифікація регіонів за рівнем економічної стабільності;
- прогнозування потреб громади у ключових ресурсах;
- визначення кластерів з високим інвестиційним потенціалом.

Запропонований метод гібридного аналізу соціально–економічних даних ТГ поєднує класифікацію, кластеризацію та прогнозування для комплексного вирішення управлінських задач. Використання двох рівнів аналізу забезпечує гнучкість і адаптивність технології. Такий підхід дозволяє ефективно працювати з гетерогенними даними, враховувати динамічні зміни та формувати стратегічні рішення, спрямовані на соціально–економічний розвиток громад.

На основі методу гібридного аналізу соціально–економічних даних ТГ [457] детально розглянуто інтелектуальні методи для класифікації рівня техногенних катастроф [257, 469] (підрозділ 7.2) та прогнозування обсягів відходів громади [246] (підрозділ 7.3). У підрозділі 7.4 представлено експериментальні результати, що



включають класифікацію рівнів техногенних катастроф (підрозділ 7.4.1) та прогнозування обсягів відходів (підрозділ 7.4.2). Запропонований підхід забезпечує всебічний аналіз і прийняття стратегічних рішень у кризових ситуаціях, а також дозволяє ефективно вирішувати екологічні та управлінські задачі. Розроблений метод успішно реалізовано у низці наукових робіт, зокрема [245, 256, 461, 467], що підтверджує його актуальність і практичну значущість для соціально–економічного розвитку ТГ.

## **7.2 Інтелектуальний метод класифікації рівня техногенних катастроф**

У підрозділі 7.1 були закладені основи методу гібридного аналізу соціально–економічних даних ТГ, що створюють базу для розробки інтелектуальних підходів до вирішення актуальних проблем. Спираючись на ці принципи, у цьому підрозділі пропонується інтелектуальний метод класифікації рівня техногенних катастроф, який дозволяє оперативно та точно визначати потенційні ризики в регіоні, особливо в умовах війни.

Незважаючи на значний прогрес у розробці різноманітних інструментів для аналізу та автоматизованого реагування на техногенні катастрофи, що відображено в дослідженні [192] (див. підрозділ 1.2.1), прийняття рішень у цій області все ще залишається складним завданням. Виникає потреба в більш точному і оперативному підході, що забезпечував би краще реагування на подібні кризові ситуації. У відповідь на цей виклик, потрібно розробити інноваційний інтелектуальний метод для класифікації рівнів наслідків техногенних катастроф. Цей підхід не тільки дозволить реагувати на подібні події, але й сприятиме запобіганню можливих гуманітарно–економічних катастроф, забезпечуючи більш надійне та обґрунтоване рішення.

На кроці аналізу даних важливо зменшити часові витрати на визначення рівня потенційної небезпеки в регіоні, тому авторами розроблено інтелектуальний метод класифікації рівня техногенних катастроф. Запропонований метод ілюструється схематично (див. див. додаток А, рис. А.13) та представлений наступними кроками:

Крок 1. Збір даних (блок 1). На першому етапі здійснюється збір даних  $D = \{(x_i, y_i)\}_{i=1}^N$ , де  $x_i$  – текстові описи аварій, отримані від свідків надзвичайних ситуацій, а  $y_i \in \{1, 2, \dots, C\}$  – мітки класів аварій, де  $C$  – кількість класів.

Крок 2. Дослідницький аналіз даних (EDA) (блок 2). EDA [209] використовується для відповіді на запитання, перевірки бізнес-припущень та створення гіпотез для подальшого аналізу. Для кожного текстового опису  $x_i$  обчислюються такі показники, як середня довжина тексту:

$$l = \frac{1}{N} \sum_{i=1}^N l_i, \quad (7.8)$$

де  $l_i$  – кількість слів в описі  $x_i$ .

Процес EDA включає наступні етапи: дослідження даних (блок 2.1), очищення даних (блок 2.2), побудова моделі (блок 2.3) та представлення результатів (блок 2.4).

Крок 3. Класифікація на основі текстових даних. Швидке збирання даних можливе через описи аварій від очевидців.

Крок 3.1. Використання текстових даних для цього етапу (блок 3). Текстові дані  $\{x_i\}_{i=1}^N$  перетворюються в числові вектори ознак  $\{v_i\}_{i=1}^N$  за допомогою методів векторизації, таких як TF-IDF або вбудовування слів (word embeddings). Кожен вектор  $v_i \in R^d$ , де  $d$  – розмірність простору ознак.

Крок 3.2. Очищення текстових даних (блок 4) [232]. Очищення текстових даних включає такі етапи:

– видалення стоп-слів (блок 4.1): усунення загальних слів  $S = \{w_j\}$ , що не несуть важливої інформації, з тексту  $x_i$ ;

– нормалізація тексту (блок 4.2): перетворення тексту до нижнього регістру та видалення пунктуації, тобто  $x_i' = \text{normalize}(x_i)$ ;

– лематизація (блок 4.3): приведення слів до їх базової форми  $\text{lemma}(w)$  для зменшення розмірності словника;

– токенізація (блок 4.4) – розбиття тексту на послідовність токенів  $T_i = \{w_{i1}, w_{i2}, \dots, w_{in}\}$ .

Крок 3.3. Аналіз та візуалізація N-грам (блок 5 і блок 6) [244]. Для набору токенизованих текстів  $\{T_i\}$  обчислюються частоти N-грам  $freq(g_j)$ , де  $g_j$  – N-грама. N-грами формуються як послідовності  $(w_{ik}, w_{i(k+1)}, \dots, w_{i(k+N-1)})$ . Результати аналізу візуалізуються у вигляді графіків або хмар слів для ідентифікації найбільш поширених фраз.

Крок 3.4. Тегування слів (блок 7). Тегування слів полягає у визначенні частин мови для кожного слова в тексті. Для кожного токена  $w_{ij}$  визначається тег  $t_{ij} = POS(w_{ij})$ , де POS – функція призначення частини мови. Це дозволяє покращити моделювання контекстуальних зв'язків між словами.

Крок 3.5. Розподіл даних на навчальний (блок 8.1) та тестовий (блок 8.2) набори. Дані розподіляються на навчальний набір:

$$D_{train} = \{(v_i, y_i)\}_{i=1}^{N_{train}}, \quad (7.9)$$

де  $N_{train} = \lfloor 0.8N \rfloor$ , тобто 80 % даних використовуються для навчання, а 20 % – для тестування.

Крок 3.6. Synthetic Minority Oversampling TEchnique (SMOTE) (блок 9) [147], балансування розподілу класів у наборі даних шляхом генерації нових синтетичних зразків менш представленого класу. Для зразка  $v_i$  меншого класу обирається випадковий сусід  $v_k$ , і новий зразок генерується як:

$$v_{new} = v_i + \lambda(v_k - v_i), \quad (7.10)$$

де  $\lambda \sim Uniform(0,1)$ .

Крок 3.7. Класифікація даних (блок 10) за допомогою ансамблевих методів машинного навчання з підходом бустингу  $D_{train}$ . Для моделювання обрані алгоритми AdaBoost (блок 10.1), GradientBoost (блок 10.2), XGBoost (блок 10.3), CatBoost (блок 10.4) та LightGBM (блок 10.5). Моделі оптимізують функцію втрат:

$$L(\theta) = \sum_{i=1}^{N_{train}} l(y_i, f(v_i; \theta)), \quad (7.11)$$

де  $l$  – функція втрат (наприклад, логістична втрата);  $\theta$  – параметри моделі;  $f$  – прогноз моделі.

Крок 3.8. Візуалізація результатів моделювання та вибір найкращого методу (блок 11). Найкращий метод вибирається на основі показників моделювання, таких як точність (Accuracy), прецизійність (Precision), повнота (Recall) та F1-бал (F1-Score) [78] (див. підрозділ 1.3).

Крок 3.9. Класифікація на основі тестового набору даних (блок 12). Обрана модель застосовується до тестового набору  $D_{test}$  для прогнозування міток  $\hat{y}_i = f(v_i; \theta^*)$ , де  $\theta^*$  – оптимальні параметри моделі. Оцінюється продуктивність моделі на невідомих даних.

Крок 4. Класифікація з використанням кількісних даних. Цей крок передбачає використання кількісних ознак:

$$\{u_i\}_{i=1}^N, \quad (7.12)$$

де  $u_i \in R^m$ , для побудови моделі класифікації. Мета – знайти функцію  $g: R^m \rightarrow Y$ , яка прогнозує мітки класів на основі кількісних даних.

Крок 4.1. Індекссування даних (блок 13). Дані впорядковуються та індексуються для забезпечення швидкого доступу. Кожному зразку призначається унікальний індекс  $i$ , і дані зберігаються у структурі:

$$D = \{(u_i, y_i)\}_{i=1}^N. \quad (7.13)$$

Крок 4.2. Розподіл даних на навчальний (блок 14.1) та тестовий (блок 14.2) набори. Дані розбиваються на навчальний набір  $D_{train}$  та тестовий набір  $D_{test}$  з тим же співвідношенням, що й раніше, забезпечуючи узгодженість експерименту.

Крок 4.3. Класифікація з використанням  $g(u; \theta)$  методів AdaBoost (блок 15.1), GradientBoost (блок 15.2), XGBoost (блок 15.3), CatBoost (блок 15.4) та LightGBM (блок 15.5). Кожна модель оптимізує функцію втрат:

$$L(\theta) = \sum_{i=1}^{N_{train}} l(y_i, g(u_i; \theta)). \quad (7.14)$$

Крок 4.4. Візуалізація результатів класифікації (блок 16). Результати моделювання відображаються за допомогою графіків, таких як ROC-криві, які показують залежність між істинно позитивною та хибно позитивною частотою при

різних порогах класифікації. Також можуть використовуватися матриці невідповідностей для детального аналізу помилок моделі.

Крок 4.5. Застосування класифікації до тестового набору даних (блок 17). Обрана модель застосовується до тестового набору  $D_{test}$  для прогнозування міток  $\hat{y}_i = g(u_i; \theta^*)$ , після чого обчислюються метрики якості для оцінки моделі на невідомих даних.

Крок 5. Вивід оцінок (блок 18).

Крок 6. Вивід результатів (блок 19).

Розроблений інтелектуальний метод класифікації рівня техногенних катастроф забезпечує ефективно та швидко визначення рівня потенційної небезпеки в регіоні. Метод було реалізовано та протестовано в підрозділі 7.4.1, де представлено експериментальні результати його застосування на реальних даних.

### **7.3 Інтелектуальний метод прогнозування обсягів відходів громади**

У підрозділі 7.1 викладені основи методу гібридного аналізу соціально-економічних даних ТГ. Враховуючи ці принципи, у цьому підрозділі пропонується інтелектуальний метод прогнозування обсягів відходів громади, який враховує специфіку утворення та складу відходів для більш точного та ефективного управління ними.

Порівняно (див. підрозділ 1.2.1) з аналізом подібних досліджень у галузі управління відходами громади пропонується новий інтелектуальний підхід до управління міськими відходами. Розроблено метод, який включає класифікацію та кластеризацію відходів, моделювання прогнозів обсягів відходів за допомогою різних моделей та оцінку їх результатів. Підхід сприяє оптимальному управлінню відходами та покращенню екологічної ситуації у громаді, відносно аналогів [109, 143, 170, 436].

Метод використовує класифікацію відходів за типом та кластеризацію місць їхнього утворення для прогнозування обсягів відходів у майбутньому. Далі

розглянемо більш детально кроки методу та його структуру у вигляді алгоритму (див. додаток А, рис. А.14):

Крок 1. Класифікація відходів. Використовуючи дані про склад та кількість відходів, класифікуються за типами відходів. Для цього можна використовувати різні параметри, такі як склад, кількість, розмір, форма тощо.

На цьому етапі використовується набір даних:

$$D = \{(x_i, y_i)\}_{i=1}^N, \quad (7.15)$$

де  $x_i$  – вектор характеристик  $i$ -го зразка відходів, що включає такі параметри, як склад, кількість, розмір, форма тощо;  $y_i$  – мітка класу відходів.

Метою є побудова функції класифікації  $f(x_i; \theta)$ , де  $\theta$  – параметри моделі, яка прогнозує клас відходів  $\hat{y}_i = f(x_i; \theta)$ . Для цього можна використовувати алгоритми машинного навчання, такі як логістична регресія, дерева рішень або метод опорних векторів. Навчання моделі здійснюється шляхом мінімізації функції втрат  $L(\theta)$ , наприклад, крос-ентропійної втрати:

$$L(\theta) = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_{ik} \log f_k(x_i; \theta), \quad (7.16)$$

де  $K$  – кількість класів відходів;  $y_{ik}$  – бінарна змінна, яка дорівнює 1, якщо зразок  $i$  належить до класу  $k$ , і 0 – в іншому випадку;  $f_k(x_i; \theta)$  – ймовірність того, що зразок належить до класу  $k$ .

Крок 2. Кластеризація місць утворення відходів. Використовуючи дані про місце та час утворення відходів, проводиться кластеризація [248] місць їхнього утворення за допомогою алгоритмів кластеризації, таких як K-Means, Agglomerative Clustering, DBSCAN, Birch, OPTICS, Spectral Clustering. Алгоритми кластеризації, краще використовувати для групування місць утворення відходів.

Далі представимо загальний математичний підхід до кластеризації. Нехай  $S = \{s_1, s_2, \dots, s_n\}$  – множина місць утворення відходів, де кожне місце  $s_i$  описується координатами  $(x_i, y_i)$  та часом утворення  $t_i$ .

Метою кластеризації є розбиття множини  $S$  на  $C$  кластерів  $\{C_1, C_2, \dots, C_C\}$  таким чином, щоб місця в одному кластері були близькими за певною метрикою відстані

або схожості. Формально, шукаємо таке розбиття, що мінімізує внутрішньокластерну дисперсію:

$$\min_{C_1, \dots, C_C} \sum_{j=1}^C \sum_{s_i \in C_j} \|s_i - \mu_j\|^2, \quad (7.17)$$

де  $\mu_j$  – центр кластера  $C_j$ , який можна обчислити як середнє значення координат точок у кластері:

$$\mu_j = \frac{1}{|C_j|} \sum_{s_i \in C_j} s_i. \quad (7.18)$$

Крок 3. Моделювання прогнозів. Для кожного кластера використовуються різні моделі прогнозування, такі як ARIMA, DNN, XGBoost (див. підрозділ 2.1), для прогнозування обсягів відходів у майбутньому. Для кожного кластера відходів моделюються прогнози обсягів відходів на певний період майбутнього. Для цього можна використовувати різні методи МН, такі як авторегресійні інтегровані змінні середнього руху (ARIMA), глибокі нейронні мережі (DNN), градієнтний бустинг над деревами рішень (XGBoost) та ін. Для кожного кластера відходів використовується окрема модель прогнозування, оскільки кожен кластер може мати свої унікальні особливості, такі як час утворення відходів, їх склад та кількість. Наприклад, відходи, що утворюються в певному районі міста в нічний час, можуть мати інший прогноз, ніж відходи, що утворюються в тому самому районі вдень.

Для кожного кластера  $C_j$  моделюються прогнози обсягів відходів на певний період майбутнього. Нехай  $\{w_{j,t}\}_{t=1}^T$  – часовий ряд обсягів відходів у кластері  $C_j$  за  $T$  періодів. Метою є побудова моделі прогнозування  $M_j$ , яка прогнозує майбутні значення  $w_{j,T+1}, w_{j,T+2}, \dots$ . Далі представимо кожену модель більш детально.

ARIMA модель описується рівнянням:

$$\phi(B)(1 - B)^d w_{j,t} = \theta(B)\varepsilon_t, \quad (7.19)$$

де  $\phi(B)$  – поліном авторегресії порядку  $p$ ;  $\theta(B)$  – поліном ковзного середнього порядку  $q$ ;  $B$  – оператор зсуву назад:  $Bw_{j,t} = w_{j,t-1}$ ;  $d$  – порядок диференціювання для досягнення стаціонарності;  $\varepsilon_t$  – випадкова похибка.

Глибокі нейронні мережі моделюють нелінійні залежності у даних. Модель може бути представлена як:

$$w_{j,t} = f_{\theta}(w_{j,t-1}, w_{j,t-2}, \dots, w_{j,t-p}), \quad (7.20)$$

де  $f_{\theta}$  – нелінійна функція, що параметризується вагами нейронної мережі  $\theta$ ;  $p$  – кількість попередніх часових кроків (вікно спостереження).

XGBoost – це метод градієнтного бустингу над деревами рішень. Прогнозна модель представляється як сума  $K$  базових моделей:

$$\hat{w}_{j,t} = \sum_{k=1}^K f_k(x_{j,t}), \quad (7.21)$$

де  $f_k$  – дерево рішень;  $x_{j,t}$  – вектор ознак для часу  $t$  (може включати попередні значення  $w_{j,t-1}, w_{j,t-2}, \dots$  та інші релевантні змінні).

Модель навчається шляхом мінімізації об'єктивної функції:

$$L(\theta) = \sum_{t=1}^T l(w_{j,t}, \hat{w}_{j,t}) + \sum_{k=1}^K \Omega(f_k), \quad (7.22)$$

де  $l$  – функція втрат (наприклад, квадратична втрата);  $\Omega(f_k)$  – регуляризаційний термін для контролю складності моделі.

Після навчання моделі  $M_j$  для кожного кластера здійснюється прогноз обсягів відходів на майбутні періоди  $t = T + 1, T + 2, \dots$ :

$$\hat{w}_{j,t} = M_j(w_{j,t-1}, w_{j,t-2}, \dots). \quad (7.23)$$

Крок 4. Оцінка результатів. Результати прогнозів для кожного кластера оцінюються за допомогою метрик, таких як MAE, MSE (див. підрозділ 1.3) для визначення точності моделі.

Крок 5. Використовуючи найкращу модель (крок 4), зробити прогноз на майбутні роки для кожного найкращого кластера, окремо для кожного значення заощадження енергії та сирого нафти.

Запропонований інтелектуальний метод прогнозування обсягів відходів громади сприяє оптимізації управління відходами та покращенню екологічної ситуації. Метод було реалізовано в підрозділі 7.4.2, де представлено



експериментальні результати його застосування з використанням різних моделей прогнозування.

## **7.4 Експериментальні результати**

За запропонованим методом гібридного аналізу соціально–економічних даних ТГ (див. підрозділ 7.1) та принципом синтезу інформаційної технології (див. підрозділ 2.1) розроблено модулі інформаційної системи, які вирішують поставлені задачі класифікації рівня техногенних катастроф та прогнозування обсягів відходів громади.

### **7.4.1 Класифікація рівня техногенних катастроф**

Для тестування розробленого інтелектуального методу оцінки впливу техногенних катастроф вибрано мову Python. В якості джерела даних використовувались записи про аварії на 12 різних підприємствах у трьох країнах, де кожен запис включає деталі про інцидент, такі як: дата та час, місце події, тип індустрії, причини та наслідки аварії, інформацію про кількість жертв, рівень серйозності аварії та потенційний рівень катастрофи [166] (табл. 7.1).

Для оцінки точності класифікації було використано крос–валідацію, що дозволило отримати високу точність моделі. Після навчання та перевірки моделі на валідаційній вибірці, вона була застосована для оцінки наслідків техногенних катастроф на інших підприємствах та в інших регіонах.

В процесі розвідувального аналізу даних для гірничодобувної промисловості було здійснено візуальний огляд ключових показників, що підкреслює значно вищі ризики аварій у цій галузі порівняно з металургійною промисловістю. Було виявлено, що місце Local\_03, має найвищий відсоток аварій у країні. Основна частина критичних ризиків класифікується як «Інші», що вказує на необхідність більш точної класифікації. Також було відзначено, що легкі інциденти (рівень I) є найбільш поширеними, але існує значна кількість помірно тяжких аварій (рівень IV). Цікаво,

що співвідношення між реальним і потенційним рівнем аварійності варіюється залежно від сектора промисловості, з вищими рівнями серйозності в гірничодобувному секторі.

Таблиця 7.1

## Опис даних

№ з/п	Параметр	Опис	Non-Null, Count, Dtype
1	Data	Часова мітка або дата/час події	411, non-null, object
2	Countries	Країна, де сталася подія (анонімізовано)	
3	Local	Місцезнаходження заводу (анонімізовано)	
4	Industry sector	Сектор економіки, до якого належить завод	
5	Accident level	Ступінь тяжкості аварії від I до VI (де I – легкий, VI – дуже тяжкий)	
6	Potential Accident Level	Оцінка потенційної тяжкості аварії, заснована на різних факторах	
7	Genre	Стать постраждалої особи	
8	Employee or Third Party	Статус постраждалої особи: працівник чи третя сторона	
9	Critical Risk	Опис ризику, асоційованого з аварією	
10	Description	Детальний опис обставин нещасного випадку	

Кореляційна матриця (Рис. 7.2) підтвердила помірний зв'язок між цими змінними, з найвищою кореляцією між рівнями II та III потенційної небезпеки та рівнем I явної небезпеки.

Після попередньої обробки датасету здійснено аналіз текстових записів, що документують аварійні ситуації. Це дозволило класифікувати рівень техногенних катастроф на основі описів від очевидців.

В ході застосування ансамблевих методів машинного навчання з бустинговим підходом, де використані параметри  $n\_estimators = 50$  та  $learning\_rate = 0,05$ , ми отримали наступні результати (Таблиця 7.2): AdaBoost з показником 0,59; Gradient Boost – 0,92; XGBoost – 0,90; CatBoost – 0,79; LGBost – 0,98. Найкращий результат

демонструє метод LGBost, що вказує на його високу точність в швидкому оцінюванні.

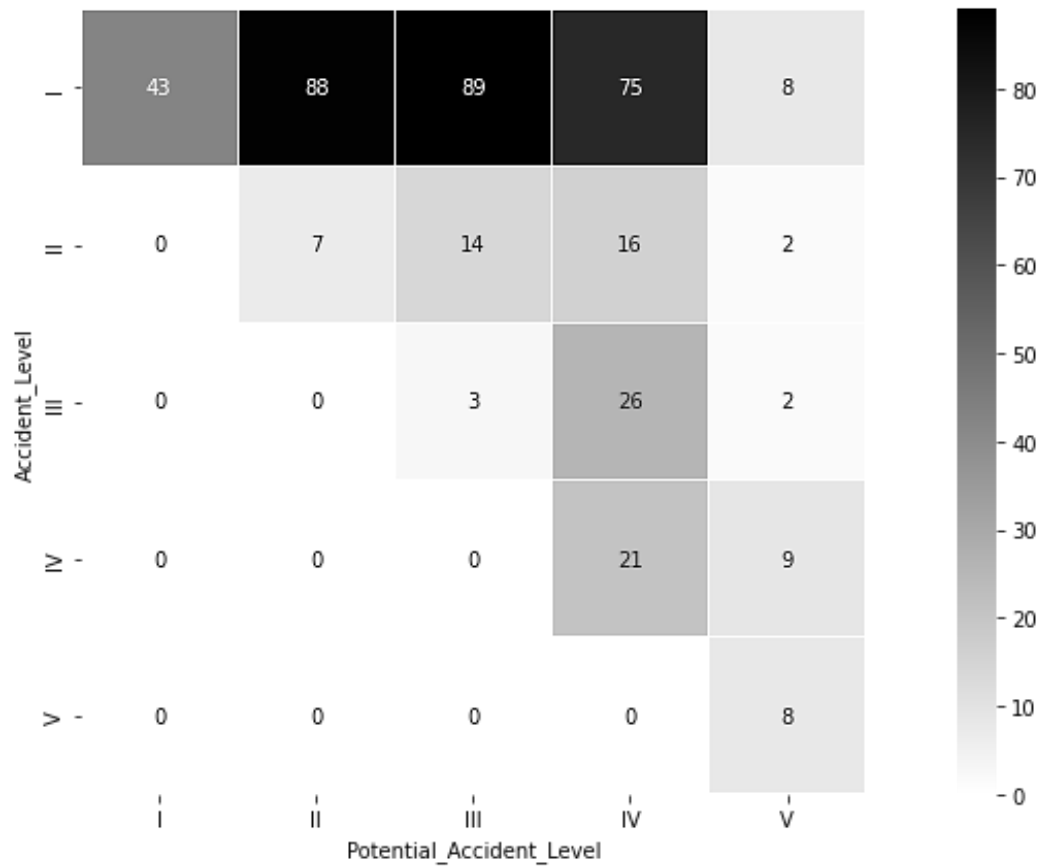


Рис. 7.2. Кореляція між рівнем потенційної аварії та явним рівнем аварії

Таблиця 7.2

Результати моделюванні ансамблевими методами МН.

Модель	Точність
AdaBoost	0,595376
Gradient Boost	0,924855
XGBoost	0,901734
CatBoost	0,988439
LGBost	0,791908

Після проведення моделювання на тестовій вибірці, сформовано звіт у вигляді Confusion Matrix [215] (Рис. 7.3). Проаналізувавши наведену на рисунку матрицю невідповідностей, можна зробити висновок, що модель здатна правильно

класифікувати більшість тестових точок даних до фактичних. Правильно класифіковано 34 значень рівня загрози I з 173 правильно визначених значень, а рівень небезпеки V при аварії правильно класифіковано 33 рази.

Аналіз показників точності виявляє, що найкраща класифікація спостерігається для потенційних рівнів небезпеки V та I в аварійних ситуаціях з точністю 0,97 та 0,94, відповідно. Загальна середня точність усіх класифікацій становить 0,68, що вказує на різницю в здатності моделі точно розпізнавати різні рівні небезпеки.

```

*****Classification Report*****
              precision    recall  f1-score   support

     I         0.97         0.81         0.88         42
     II        0.59         0.59         0.59         34
     III       0.62         0.55         0.58         38
     IV        0.29         0.45         0.35         22
     V         0.94         0.89         0.92         37

 accuracy                   0.68         173
 macro avg                 0.68         173
 weighted avg              0.73         173

```

Confusion Matrix between predictions and actuals

I	34	1	0	0	0
II	0	20	7	6	1
III	2	5	21	6	0
IV	5	8	9	10	3
V	1	0	1	0	33
	I	II	III	IV	V

Рис. 7.3. Матриця плутанини між прогнозами та фактами

Таким чином, переходимо до прогнозування рівня небезпеки на основі тексту з тестової вибірки, який включає такі характеристики: країна Country\_02, локація Local\_02, сектор Mining, рівень аварії Accident level I, потенційний рівень небезпеки Potential Accident Level V, стать Male, статус Employee, тип учасника Others, і поданий текст:

*«When performing cleaning with LHD in block F 9970 at level 420, the operator was surprised by a rock block displacement of the side of the gallery, reaching his right leg causing him superficial injury»*

Система (Рис. 7.4) передбачила рівень потенційної небезпеки правильно.

```
predict('When performing cleaning with LHD in block F 9970 at level 420, the operator was surprised  
array([[ 'V' ]], dtype=object)
```

Рис. 7.4. Результат передбачення

Процес класифікації, який базується на кількісних показниках, починається з перетворення сирих даних в індексовані значення. Після цього дані поділяються на дві основні частини – вибірки: навчальну та тестову.

Проведемо класифікацію Potential Accident Level на основі ансамблевих методів: AdaBoost, Gradient Boost, XGBoost та CatBoost. Найкращі результати моделювання [257] отримано на основі методу класифікації CatBoost (Рис. 7.5).

Клас «Employee or Third Party» (клас №2) був класифікований з абсолютною точністю 100 %, з яких 25 % випадків були правильно прогнозовані моделлю. Це підтверджується оцінкою f1–score, яка становить 0,40 для 28 випадків, належних до класу «Employee or Third Party». Подальший аналіз класу «Industry Sector» (клас №0) показав, що 80 % значень було успішно розпізнано, з них 99% було коректно передбачено моделлю. Це підтверджується оцінкою f1–score, яка дорівнює 0,89 для 282 випадків, вказуючи на високу точність. В середньому, точність прогнозів для класифікації «Potential Accident Level» склала 81 % для вибірки, що містить 382 значення, демонструючи хороші результати.

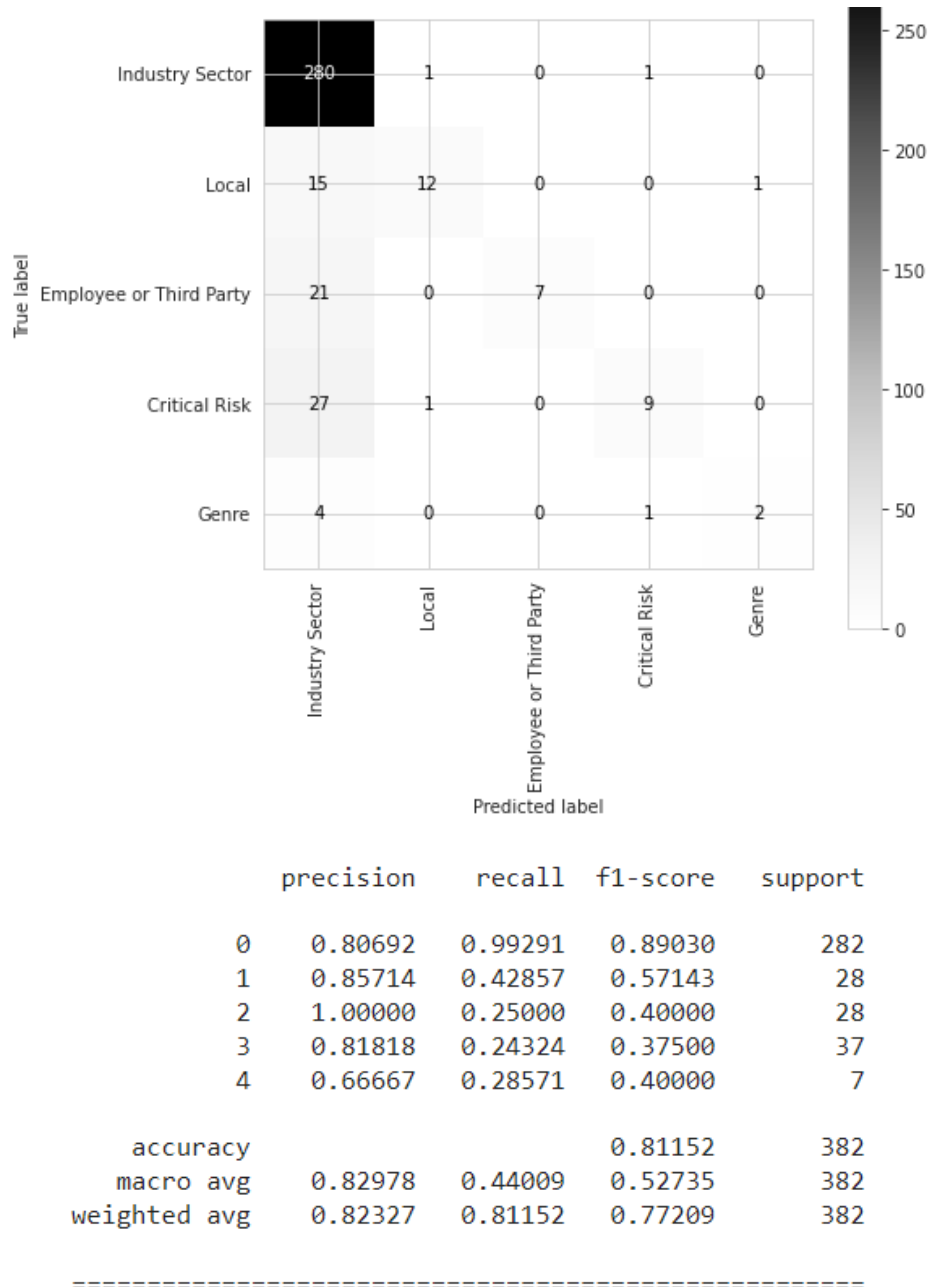


Рис. 7.5. Оцінка моделювання CatBoost

Після завершення моделювання виявлено, що метод CatBoost є наочним для класифікації та визначення рівня потенційної небезпеки (Potential Accident Level). Таким чином, цей метод буде застосовано до тестової вибірки.

Результати моделювання на даних, невідомих системі (Рис. 7.6), показали таке: при класифікації потенційного рівня небезпеки в контексті сектору промисловості (клас №0) було ідентифіковано 80 % значень, з яких 97 % були правильно передбачені моделлю. Оцінка f1-score становить 0,88 для 34 значень, що свідчить про достатньо

високу точність. Середня точність прогнозних значень класифікації рівня потенційної небезпеки склала 77 % для вибірки з 43 значень, що є позитивним результатом.

	precision	recall	f1-score	support
0	0.805	0.971	0.880	34
1	0.000	0.000	0.000	2
2	0.000	0.000	0.000	3
3	0.000	0.000	0.000	3
4	0.000	0.000	0.000	1
accuracy			0.767	43
macro avg	0.161	0.194	0.176	43
weighted avg	0.636	0.767	0.696	43

CatBoost Test Accuracy - 76.74418604651163

Рис. 7.6. Оцінка моделювання CatBoost на тестовій вибірці

Система передбачила рівень (Рис. 7.7) потенційної небезпеки майже правильно (окрім рядка 4) для перших 5 рядків тестової вибірки.

y_test		test_preds[:5]
256	0	array([[0],
56	0	[0],
72	0	[0],
389	4	[3],
131	0	[0]])

а) тестова вибірка      б) результат передбачення

Рис. 7.7. Результат передбачення рівня потенційної небезпеки

У цьому підрозділі представлено розроблений інтелектуальний метод класифікації рівня техногенних катастроф, що дозволяє точно оцінювати потенційні ризики та надавати своєчасні рекомендації при кризових ситуаціях у ТГ.

Основою покращеного підходу стало включення нових етапів, що забезпечують підвищену точність класифікації та адаптивність системи. Зокрема, було реалізовано формування навчальної вибірки на основі RFM-аналізу та кластерного аналізу, що дозволило краще структурувати дані для навчання. Ці підходи забезпечили диференціацію даних за ключовими показниками, такими як частота подій, інтенсивність їх наслідків і тривалість між аваріями. Застосування кластерного аналізу дозволило групувати події за схожими характеристиками, що значно покращило якість навчальної вибірки.

Для підвищення точності аналізу було розроблено ансамбль адаптивних прогнозних моделей, який включає два рівні:

1) моделі першого рівня, такі як XGBoost, CatBoost та LightGBM, які виконують класифікацію даних на базі різних алгоритмів;

2) метамодель другого рівня Stacking агрегує результати базових моделей, забезпечуючи їх інтеграцію та підвищуючи узгодженість прогнозів. Завдяки цьому вдалося зменшити похибки, характерні для окремих моделей, і суттєво покращити загальну точність класифікації.

Застосовано нову стратегію адаптації та перенавчання ансамблю адаптивних моделей. Цей етап включав регулярне оновлення параметрів моделей на основі нових даних, що надходять з різних джерел. Динамічне перенавчання дозволило враховувати зміну тенденцій у даних та специфіку нових кризових ситуацій. Це забезпечило високу актуальність та точність результатів навіть за умов змінних характеристик даних.

Завдяки запропонованим покращенням точність класифікації рівня техногенних катастроф зросла на 5%, що підтверджено результатами тестування на реальних даних (Рис. 7.8). Використання RFM-аналізу та кластеризації значно підвищило якість підготовки даних, що, у свою чергу, сприяло кращому навчанню моделей. Ансамбль адаптивних моделей продемонстрував стабільну продуктивність, забезпечивши високі показники точності, F1-міри та узгодженості прогнозів. Адаптивне перенавчання гарантувало точність системи в динамічних умовах, що є критичним для управління кризами.



	precision	recall	f1-score	support
0	0.85	1.00	0.92	34
1	0.50	0.50	0.50	2
2	0.67	0.67	0.67	3
3	0.00	0.00	0.00	3
4	0.00	0.00	0.00	1
accuracy			0.82	43
macro avg	0.40	0.43	0.42	43
weighted avg	0.74	0.82	0.77	43

Stacking Model Test Accuracy - 81.744%

Рис. 7.8. Оцінка моделювання Stacking на тестовій вибірці

Для реалізації запропонованого інтелектуального методу розроблено мобільний додаток (Рис. 7.9), орієнтований на використання свідками аварій і може бути інтегрований у систему «розумного міста».

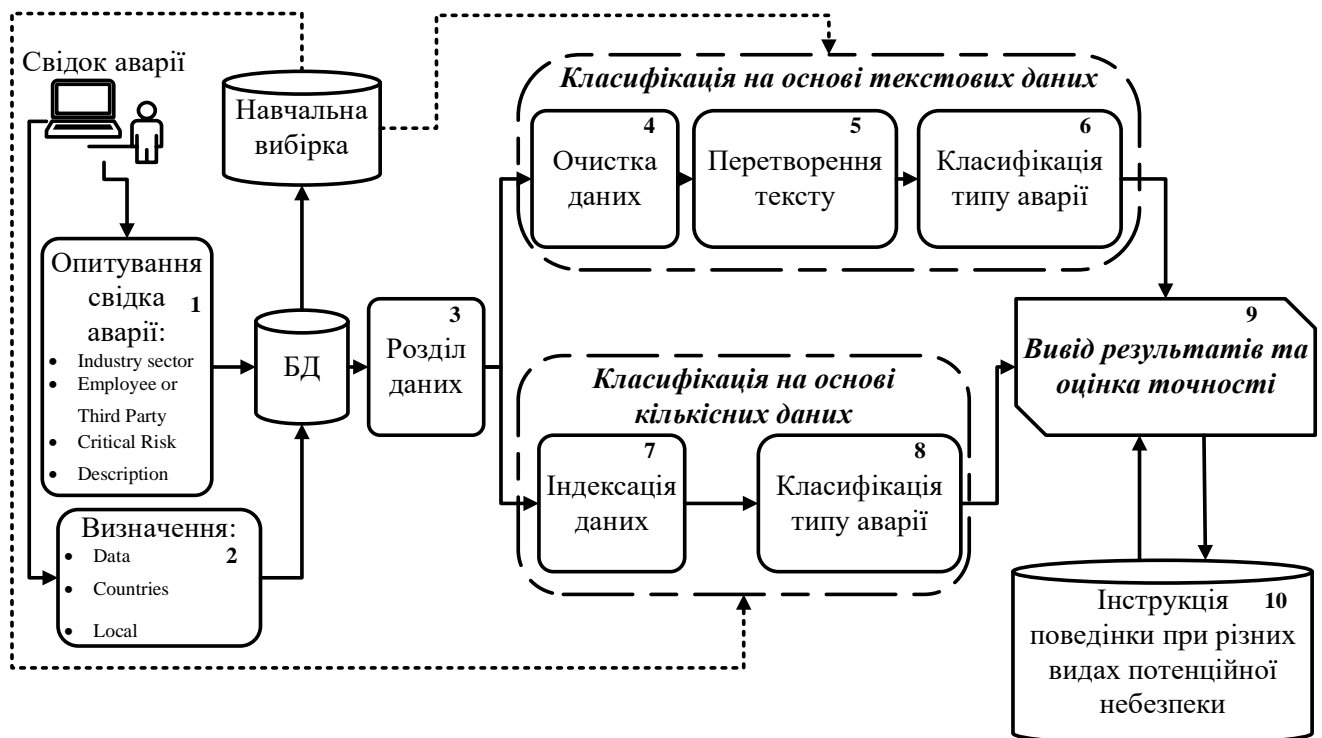


Рис. 7.9. Архітектура мобільного додатка

На першому етапі (див.Рис. 7.9) користувач додатка заповнює коротку анкету, де вказує галузь (Industry Sector), статус свідка (Employee or Third Party) та опис критичного ризику (Critical Risk Description) – блок 1. Паралельно, додаток автоматично збирає інформацію про дату (Data), країни (Countries) і локалізацію (Local) – блок 2. Отримані дані зберігаються в базі даних, де вони можуть бути використані для створення навчальної вибірки з метою перенавчання системи. Потім дані передаються до наступного етапу, де виконуються їх розподіл – блок 3, який включає класифікацію потенційної аварійної небезпеки на основі текстових та кількісних даних. Обробка цих двох типів даних здійснюється паралельно.

Під час обробки текстових даних спочатку виконуються очищення даних (Data Cleaning) – блок 4 та їх текстове перетворення (Text Transformation) – блок 5, після чого здійснюється класифікація (Classification) – блок 6. У свою чергу, кількісні дані індексуються (Indexing) – блок 7, а потім проходять класифікацію рівня потенційної небезпеки (Hazard Level Classification) – блок 8. На завершальному етапі результати обробки виводяться на екран – блок 9, включаючи рівень потенційної небезпеки та рекомендації щодо заходів безпеки або відповідної поведінки у конкретній ситуації. Дані рекомендації формуються на основі класифікації рівня потенційної небезпеки.

Спершу визначається геолокація користувача (Рис. 7.10а), яка дозволяє встановити точне місце виникнення техногенної катастрофи. Після цього користувач додає голосовий опис події, обирає відповідний сектор промисловості, де сталася аварія, та зазначає свій статус у ситуації (співробітник чи стороння особа) (Рис. 7.10б).

Загалом, мобільний додаток забезпечує автоматизацію процесу оцінки потенційного рівня небезпеки техногенних катастроф шляхом аналізу текстових даних і кількісних показників.

Використання методів машинного навчання та текстової класифікації дозволяє ефективно визначати категорію небезпеки на основі опису ситуації. Водночас, слід зазначити, що додаток не є повноцінною заміною професійних експертів та фахівців, які мають досвід і знання для детального аналізу та оцінки ризиків у конкретних умовах. Також можливе існування додаткових чинників, що не враховані у додатку, які можуть впливати на рівень небезпеки в окремих випадках.

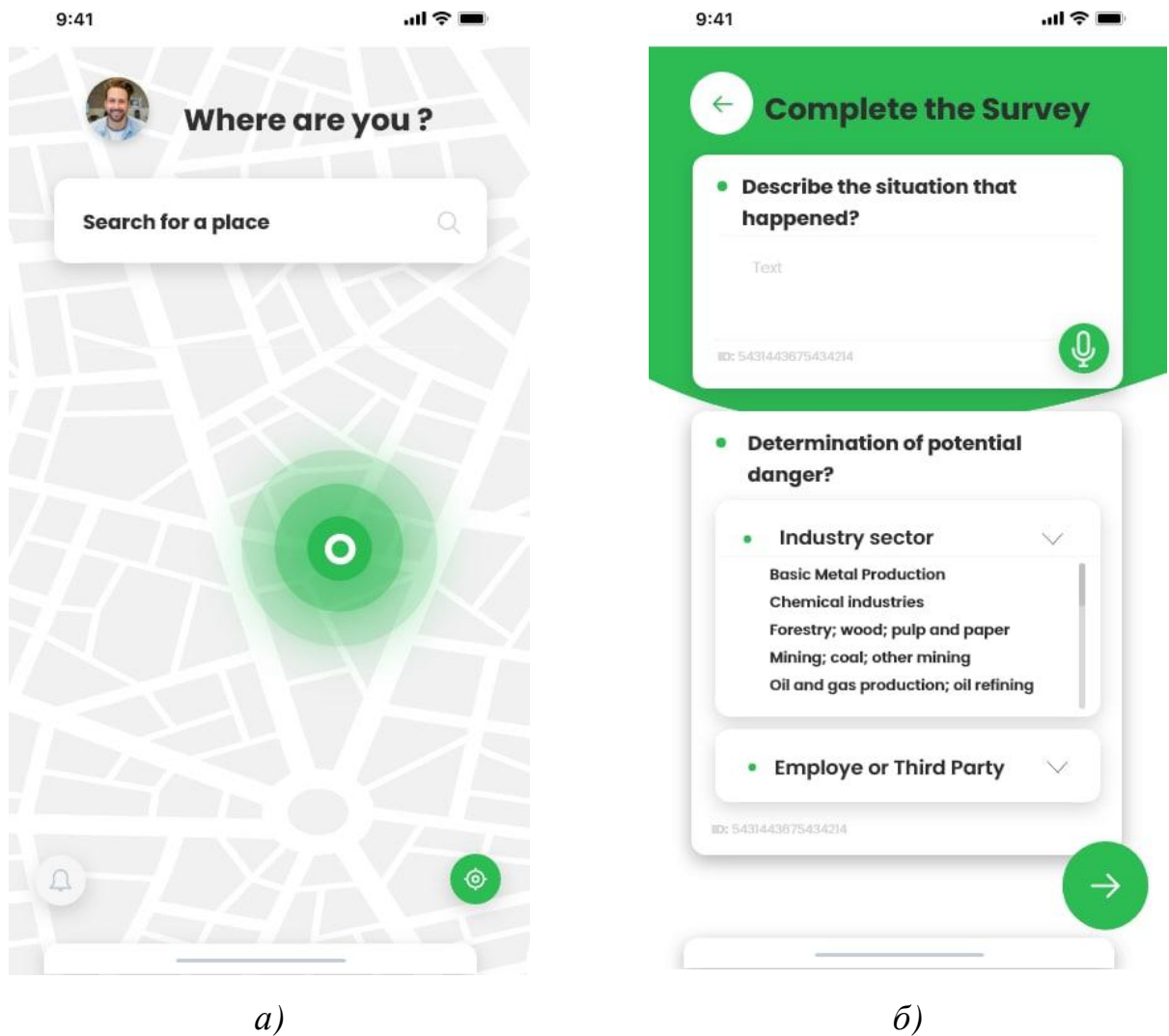


Рис. 7.10. Введення вхідних даних

На основі отриманих результатів (Рис. 7.11) формується інформація про рівень потенційної небезпеки, що сприяє оперативній роботі служб надзвичайних ситуацій, завдяки чому вони можуть швидше зорієнтуватися на місці аварії та ще під час руху підготувати необхідне обладнання й ресурси. Це, у свою чергу, дозволяє значно скоротити час реагування та підвищити ефективність ліквідації наслідків аварії.

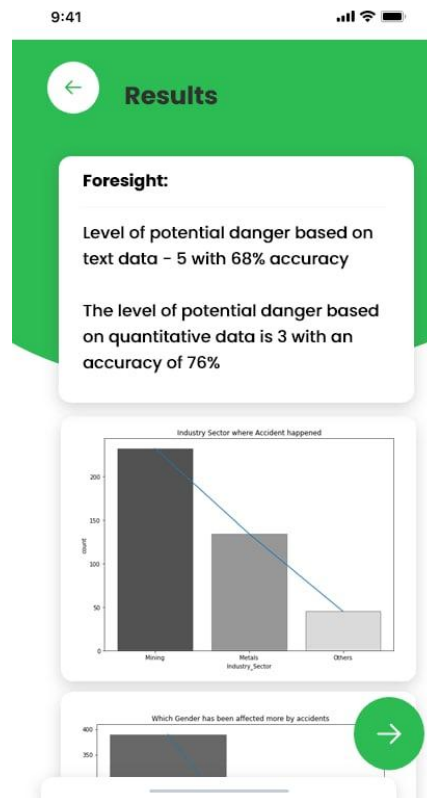


Рис. 7.11. Виведення отриманих результатів

Виявлення техногенних катастроф показало високу ефективність системи (див. додаток Г), яка дозволяє значно оптимізувати час прийняття рішень. Час із системою варіювався від 1 до 4 годин, що відповідає аналогічним показникам без її використання, проте система забезпечила точність і оперативність процесу, досягнувши ефективності  $\sim 95,83\%$ . Це підкреслює важливість впровадження інтелектуальних технологій для ефективного моніторингу та аналізу критичних ситуацій.

#### 7.4.2 Прогнозування обсягів відходів громади

Апробація методу проведена на даних порталу kaggle [373]. Цей датасет містить річні дані про кількість відходів, згенерованих та перероблених в Сінгапурі з 2003 по 2020 рік. Датасет поділений на дві частини: дані з 2003 по 2017 рік та дані з 2018 по 2020 рік. Кожен запис включає інформацію про тип відходів (наприклад, папір, скло,

метал, пластик), кількість відходів, що згенеровані та кількість відходів, що перероблені.

Для виявлення схожих характеристик різних груп матеріалів щодо кількості відходів та енергозбереження проведено кластеризацію методами: K–Means, Agglomerative Clustering, DBSCAN, Birch, OPTICS, Spectral Clustering (Рис. 7.12).

Для управління відходами, потрібно зрозуміти, які групи (кластери) відходів можна розглядати разом (Рис. 7.12), тому буде проведено прогнозування значень кількості відходів по кожному кластеру, використовуючи методи МН, такі як ARIMA, DNN та XGBoost. Оцінка різних моделей для всіх видів кластеризації може допомогти вибрати найкращу модель для прогнозування кількості відходів та краще зрозуміти, які групи відходів можна об'єднати для більш простого управління відходами.

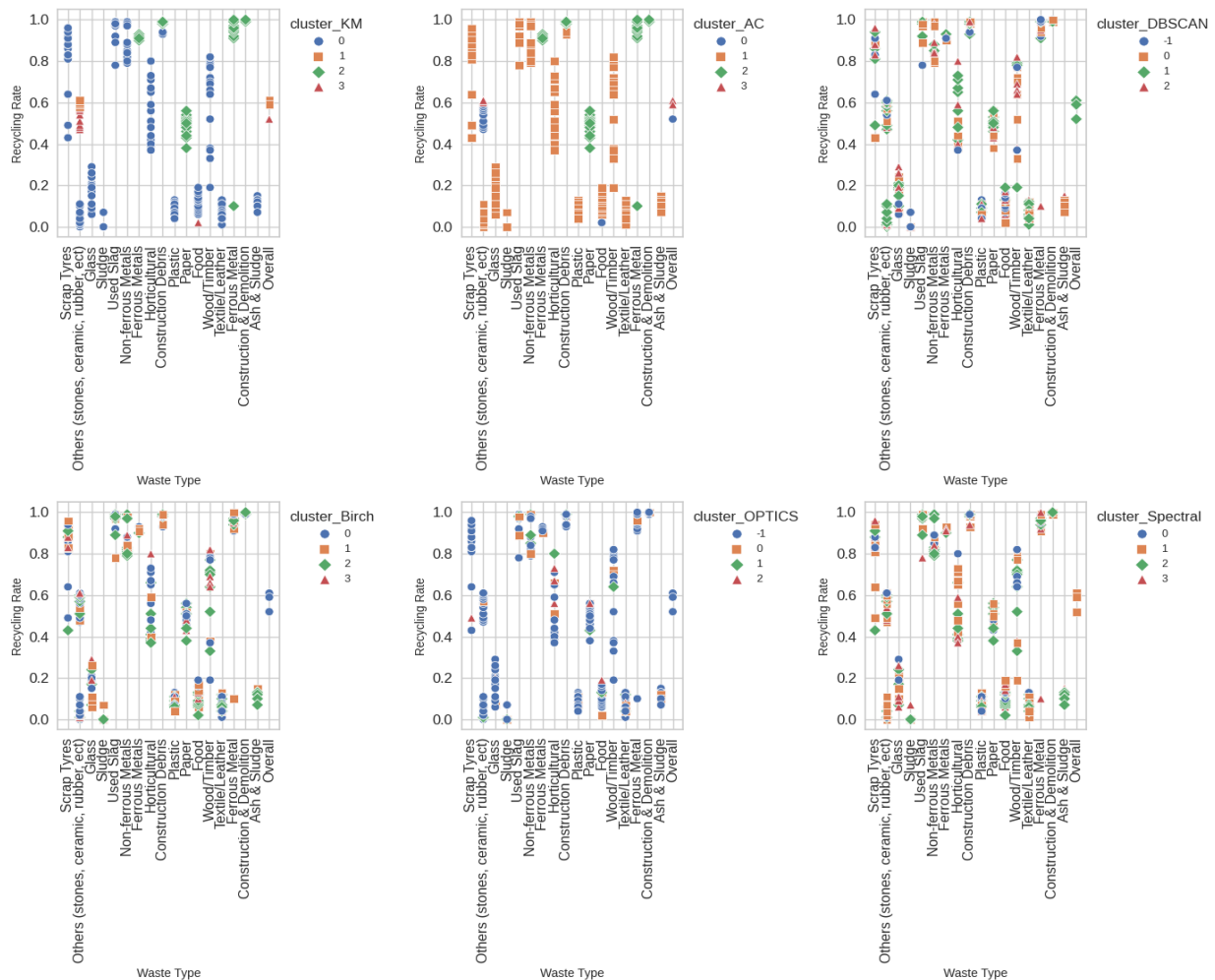


Рис. 7.12. Кластеризація відходів за типом та ступенем переробки

Аналіз проведених досліджень [246] показує, що кращі результати прогнозування кількості перероблених відходів можна отримати за допомогою моделей ARIMA, DNN та XGBoost, проте їх точність залежить від конкретного кластера та методу кластеризації. Зокрема, метод K–Means дав кращі результати за метриками MAE та MSE для всіх трьох моделей на кластерах з номерами 0 та 1.

Діаграма прогнозованих значень кількості перероблених відходів за роками демонструє високу точність підходу Xgboost у прогнозуванні для різних типів кластеризації, хоча є деякі відхилення для деяких кластерів, наприклад K–Means, Agglomerative Clustering та OPTICS. А DBSCAN, Birch та Spectral Clustering показали хорошу точність прогнозу. Тому при управлінні відходами міста варто використовувати DBSCAN, Birch та Spectral Clustering для групування та Xgboost для прогнозування даних.

Далі проведемо прогнозування заощаджень енергії на майбутні три роки з використанням методу XGBoost для типів кластерів DBSCAN, Birch та Spectral Clustering (рис.7.13)

Отже, важливо підкреслити, що основна відмінність запропонованого дослідження від попередніх робіт, таких як [26, 39, 40], полягає у використанні моделі XGBoost для прогнозування обсягів відходів. Ця модель забезпечує високу точність прогнозування, досягаючи до 98%. Крім того, були застосовані різні методи кластеризації, такі як DBSCAN, Birch та Spectral Clustering, для покращення якості прогнозів при управлінні відходами. Цей підхід підкреслює важливість інтеграції різноманітних методів для досягнення високої точності в управлінні відходами громади, що представляє інноваційний внесок у цю дослідницьку область.

На основі запропонованого підходу було використано комплексну інтеграцію методів лінійної регресії, ARIMA, XGBoost та CatBoost для об'єднання прогнозів. Така комбінація моделей дозволила максимально використати сильні сторони кожного з них, забезпечуючи більш точні та надійні прогнози обсягів відходів.

У результаті впровадження цих вдосконалень, система прогнозування обсягів відходів стала більш точною, що позитивно вплинуло на екологічну ситуацію в громаді та оптимізацію ресурсів для збору та переробки відходів. Підвищення

точності прогнозування сприяло кращому плануванню та розподілу ресурсів, що є важливим фактором для сталого розвитку ТГ.

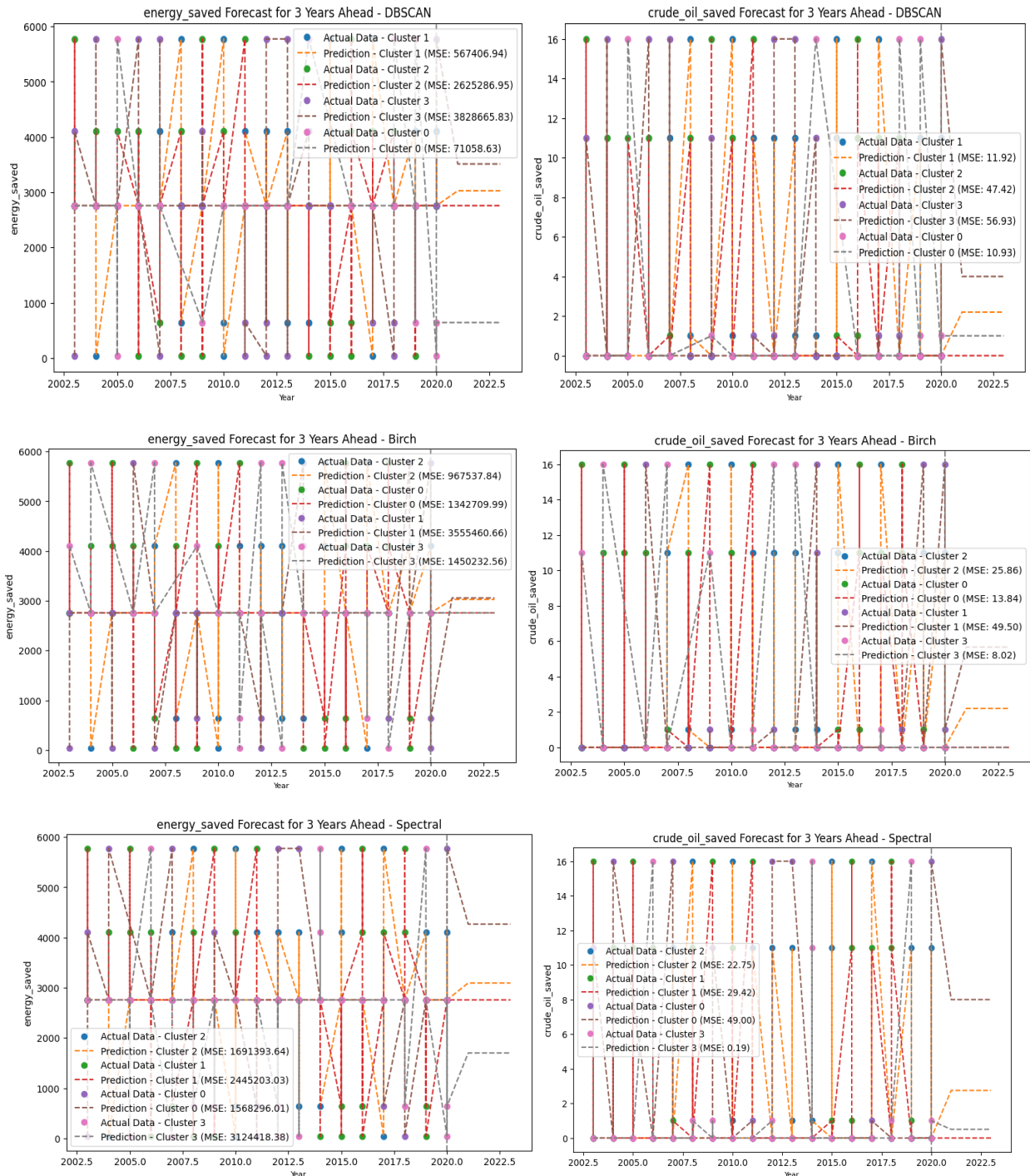


Рис. 7.13. Прогнозування заощаджень енергії на майбутні три роки з використанням методу XGBoost для різних типів кластерів (DBSCAN, Birch та Spectral Clustering)

Інтелектуальний модуль екологічного [208, 246] та техногенного моніторингу [257] для ТГ починається (Рис. 7.14) з блоку 1 збору даних, де відбувається збір необхідної інформації, включаючи супутникові знімки для моніторингу транспорту, текстові дані для аналізу техногенних катастроф та дані про відходи. Далі, в блоку 2 тренування моделей МН, зібрані дані піддаються глибокому аналізу та класифікації за допомогою алгоритмів МН, що охоплюють виявлення автомобілів, кластеризацію відходів, а також класифікацію інформації про техногенні катастрофи.



Рис. 7.14. Інтелектуальний модуль екологічного та техногенного моніторингу

Потім, у блок 3 прогнозування та моделювання, на основі аналізу розробляються прогнози, зокрема щодо майбутніх обсягів відходів та рівня транспортних викидів. Нарешті, всі ці дані та аналітичні висновки об'єднуються у блоці 4 інтегрованого виведення результатів, який забезпечує цілісний погляд на стан ТГ, дозволяючи ефективно відстежувати та управляти його екологічними та техногенними аспектами.

Інтелектуальний модуль (див. Рис. 7.14) екологічного та техногенного моніторингу призначений для комплексного аналізу даних (Рис. 7.15), пов'язаних з управлінням відходами, реагуванням на технологічні катастрофи та моніторингом



викидів CO<sub>2</sub> від автотранспорту. Цей інструмент дозволяє систематизувати та аналізувати великі обсяги даних, що сприяє виявленню тенденцій і формуванню прогнозів щодо зазначених аспектів управління громади.

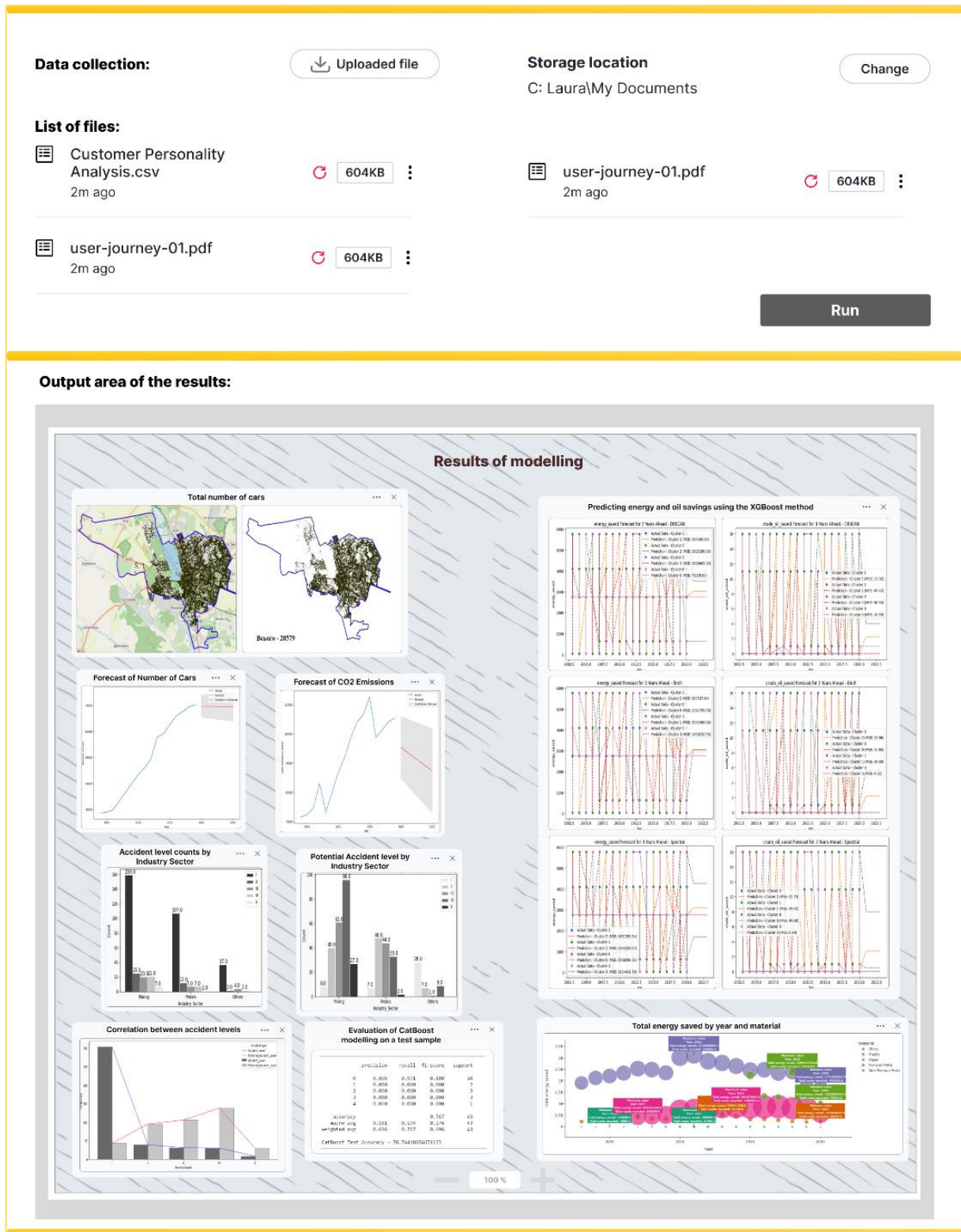


Рис. 7.15. Інтелектуальний модуль екологічного та техногенного моніторингу

Для адміністрації громади цей модуль є надзвичайно цінним, оскільки він допомагає у визначенні пріоритетних напрямів діяльності в екологічній сфері, а також в плануванні заходів зі зниження негативного впливу на довкілля. Крім того, інструментарій моніторингу сприяє своєчасному виявленню та ефективному реагуванню на екологічні та техногенні виклики, забезпечуючи безпеку та здоров'я мешканців громади.

Прогнозування обсягів відходів продемонструвало високу ефективність роботи системи (див. додаток Г), яка дозволила значно скоротити час аналізу та підвищити точність прогнозів. Час з використанням системи варіювався від 1 до 2 годин, тоді як без неї експерти витрачали від 5 до 10 днів на виконання аналогічних завдань. Загальна ефективність процесу становила ~99,72%, що свідчить про суттєве покращення в управлінні даними про відходи та їх прогнозуванні завдяки впровадженню інтелектуальної технології.

## **7.5 Висновки**

У розділі представлено метод гібридного аналізу соціально–економічних даних ТГ, який інтегрує класифікацію, кластеризацію та прогнозування для вирішення складних управлінських задач. Розглянуто підходи для аналізу різнорідних даних, що дозволяють ефективно підтримувати стратегічні рішення в умовах постійної зміни соціально–економічних показників.

Розроблено метод гібридного аналізу, який використовує багаторівневу обробку даних: кластеризацію та класифікацію на першому рівні, прогнозування на другому. Це дозволяє комплексно аналізувати різні типи даних (структуровані, напівструктуровані, неструктуровані) і створювати багатофакторні прогнози. Метод дозволяє виявляти кластерні особливості регіонів, оцінювати рівень стабільності громад та визначати оптимальні управлінські рішення. Застосування цього підходу забезпечує точність прогнозів і гнучкість у використанні аналітичних інструментів.

Розроблено інтелектуальний метод класифікації рівня техногенних катастроф, який враховує текстові та кількісні дані. Модель інтегрує методи машинного навчання (CatBoost, XGBoost) та використовує адаптивне перенавчання, що забезпечує оперативну реакцію на кризові ситуації. Проведений аналіз показав високу точність (до 95,83%) у класифікації рівня небезпеки, що є критично важливим для ефективного управління надзвичайними ситуаціями. Використання цього підходу дозволяє значно оптимізувати процес прийняття рішень у сфері безпеки громад.

Розроблено метод прогнозування обсягів відходів громади, який включає кластеризацію місць утворення відходів (K-Means, DBSCAN) і прогнозування обсягів за допомогою ARIMA, DNN, XGBoost. Метод забезпечив точність прогнозів до 98%, дозволяючи оптимізувати управління ресурсами громад. Такий підхід сприяє зменшенню екологічного навантаження, більш ефективному плануванню та зниженню витрат на управління відходами.

Експериментальні результати підтвердили ефективність запропонованих методів, зокрема в задачах класифікації рівня техногенних катастроф і прогнозування обсягів відходів. Висока точність моделей (до 99,72%) демонструє практичну цінність для сталого розвитку громад. Розроблені інструменти також дозволяють зменшити часові витрати на аналіз даних з 5–10 днів до 1–2 годин.

Запропонований метод гібридного аналізу забезпечує ефективну інтеграцію різноманітних даних для вирішення ключових соціально–економічних задач ТГ. Його застосування дозволяє підвищити точність прогнозів, скоротити час прийняття рішень та забезпечити екологічну і соціально–економічну стабільність. Отримані результати підтверджують доцільність впровадження таких підходів у практику управління громадами.

## ВИСНОВКИ

У цьому дослідженні розроблено інноваційну інформаційну технологію інтелектуального аналізу соціально–економічних даних ТГ, яка забезпечує підвищення соціальної стійкості та ефективності управління ресурсами. Запропоновані методи та моделі дозволяють інтегрувати різноманітні дані, враховувати динамічність соціально–економічних процесів та приймати обґрунтовані управлінські рішення в режимі реального часу.

При цьому отримано такі основні наукові й практичні результати:

1. Проведено аналіз сучасного стану інтелектуального аналізу соціально–економічних даних для ТГ, яке виявило суттєві протиріччя та обмеженість існуючих методів щодо забезпечення точного прогнозування, адаптивності та інтеграції даних з різних джерел у режимі реального часу. Це підтвердило актуальність розроблення інтелектуальних технологій для ефективного управління та прийняття рішень у динамічних умовах.

2. Вперше запропоновано узагальнені принципи синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ, які забезпечують інтеграцію структурованих, неструктурованих і напівструктурованих даних та враховують динамічність соціально–економічних процесів. Це дозволило підвищити соціальну стійкість інфраструктури громад за рахунок адаптивного вибору методів аналізу та формування управлінських рішень, скоротивши час на обробку даних і прийняття рішень з днів до годин, що підтверджено експертною ефективністю ~99,72%.

3. Розроблено методологію аналізу соціально–економічних даних ТГ, що включає сукупність методів класифікаційного, кластерного та гібридного аналізу, а також прогнозування. Особливістю методології є адаптивність до нестационарних даних і можливість роботи в режимі реального часу, що підвищило точність прогнозування, класифікації та прозорість управлінських рішень.

4. Розроблено новий метод формування навчальної вибірки для нестационарних даних на основі RFM–аналізу та кластерного аналізу. Це дозволило більш точно відбирати інформативні зразки, порівнюючи їх з аналогами, що підтверджено експериментальними результатами з точністю класифікації до 99,1%.

5. Вперше розроблено метод ансамблевих адаптивних прогнозних моделей для багатовимірного аналізу, особливістю якого є оцінювання на «ковзному вікні» та використання метамоделей першого і другого рівнів на основі оптимізації функції Лагранжа. Це забезпечило підвищення точності прогнозів багатовимірних нестационарних процесів до 98%, що критично для прийняття рішень у реальному часі.

6. Удосконалено метод класифікаційного аналізу кількісних соціально–економічних даних, який забезпечує інтеграцію різнорідних даних. Це дозволило підвищити точність класифікації кількісних показників для реалізації управлінських рішень та скоротити час аналізу з днів до годин. Експертні оцінки показали ефективність методу ~99,72% у задачі виявлення фіктивних підприємств, що підтверджує значне зменшення часу на ідентифікацію ризиків.

7. Удосконалено метод класифікаційного аналізу текстових соціально–економічних даних з інтеграцією текстових джерел різної структури та застосуванням технологій обробки природної мови. Це дозволило зменшити час виявлення закономірностей з 1–5 днів до 1–2 годин, підвищивши ефективність виявлення закономірностей для реалізації управлінських рішень, що підтверджено експертними оцінками ефективності ~99,17%.

8. Удосконалено метод кластерного аналізу соціально–економічних даних, який забезпечує ідентифікацію груп об'єктів з подібними характеристиками. Це дозволило підвищити точність розподілу ресурсів для реалізації управлінських рішень, з ефективністю методу ~99,61% за експертними оцінками.

9. Удосконалено метод прогнозування соціально–економічних даних, який інтегрує різні типи даних та застосовує адаптивні методи інтелектуального аналізу. Це скоротило час прогнозування з днів до годин, підвищивши ефективність процесу

на ~98,75%, що дозволяє швидко реагувати на зміни попиту та інших важливих показників.

10. Удосконалено метод гібридного аналізу соціально–економічних даних, який враховує як кількісні, так і якісні показники через інтеграцію різнорідних даних та багаторівневий підхід до аналізу. Це зменшило час прийняття комплексних рішень з тижнів до годин, підвищивши гнучкість і адаптивність процесу на ~99,72% за експертними оцінками.

11. Реалізовано розроблену інформаційну технологію інтелектуального аналізу соціально–економічних даних ТГ у вигляді модулів інформаційної системи та проведено розрахунок її ефективності. Впровадження технології підтвердило теоретичні результати на практиці, забезпечивши підвищення точності прогнозів до 98%, а ефективність кластеризації та класифікації даних до 95%. Розрахунок ефективності продемонстрував скорочення часових витрат на аналіз даних з 5–30 днів до 1–120 годин. Ці кількісні показники підтверджують практичну цінність розробленої технології для управління в ТГ, забезпечуючи більш швидке та точне прийняття управлінських рішень та оптимізацію використання ресурсів.

Результати дисертаційної роботи впроваджено в освітній процес Західноукраїнського національного університету, зокрема для підготовки здобувачів спеціальності «Комп'ютерні науки» за навчальними курсами «Інтелектуальний аналіз даних», «Машинне навчання», «Методи та засоби штучного інтелекту» та іншими. Також результати використано у науково–дослідних проєктах Erasmus+: «WORK4CE» і «My Farm», та Creative Europe «AURA», а також у державних наукових роботах, зокрема, TruScanAI, «Методи та засоби структурно–статистичної ідентифікації», «Інтелектуальні методи соціально–економічного розвитку ТГ» і в рамках перспективного плану розвитку наукового напрямку «Технічні науки» ЗУНУ. Крім того, результати дисертаційної роботи апробовано в ІТ–компанії «МагнетікВан» (м. Тернопіль), Тернопільській обласній державній адміністрації та Великоберезовицькій територіальній громаді.

**СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ**

1. Abdullah–All–Tanvir, Mahir, E. M., Akhter, S., & Huq, M. R. (2019). Detecting fake news using machine learning and deep learning algorithms. *2019 7th International Conference on Smart Computing & Communications (ICSCC)*. IEEE. <https://doi.org/10.1109/icsc.2019.8843612>
2. Achmad, K. A., Nugroho, L. E., Djunaedi, A., & Widyawan. (2018). Smart city for development: Towards a conceptual framework. *2018 4th International Conference on Science and Technology (ICST)*. IEEE. <https://doi.org/10.1109/icst.2018.8528677>
3. Adel, H., Dahou, A., Mabrouk, A., Abd Elaziz, M., Kayed, M., El–Henawy, I. M., & Amin Ali, A. (2022). Improving crisis events detection using distilbert with hunger games search algorithm. *Mathematics*, *10*(3), 447. <https://doi.org/10.3390/math10030447>
4. Aguilar, J., & Garcia, G. (2018). An adaptive intelligent management system of advertising for social networks: A case study of Facebook. *IEEE Transactions on Computational Social Systems*, *5*(1), 20–32. <https://doi.org/10.1109/tcss.2017.2759188>
5. Ahmad, Z., Thaheem, M. J., & Maqsoom, A. (2018). Building information modeling as a risk transformer: An evolutionary insight into the project uncertainty. *Automation in Construction*, *92*, 103–119. <https://doi.org/10.1016/j.autcon.2018.03.032>
6. Ahmed, M., Irfan, M., Meero, A., Tariq, M., Comite, U., Abdul Rahman, A. A., & Gunnlaugsson, S. B. (2021). Bubble identification in the emerging economy fuel price series: Evidence from generalized sup Augmented Dickey–Fuller Test. *Processes*, *10*(1), 65.
7. Ahmetoğlu Taşdemir, F. (2022). Machine learning sales forecasting for food supplements in pandemic era. *Journal of risk analysis and crisis response*, *12*(2). <https://doi.org/10.54560/jracr.v12i2.326>
8. Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social network analysis and mining*, *13*(1). <https://doi.org/10.1007/s13278-023-01028-5>

9. Akhtar, P., Ghouri, A. M., Khan, H. U. R., Amin ul Haq, M., Awan, U., Zahoor, N., Ashraf, A. (2022). Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions. *Annals of operations research*. <https://doi.org/10.1007/s10479-022-05015-5>
10. Alajali, W., Wen, S., & Zhou, W. (2017). On-Street car parking prediction in smart city: a multi-source data analysis in sensor-cloud environment. In *Security, privacy, and anonymity in computation, communication, and storage* (pp. 641–652). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-72395-2\\_58](https://doi.org/10.1007/978-3-319-72395-2_58)
11. Alam, S., Sonbhadra, S. K., Agarwal, S., & Nagabhushan, P. (2020). One-class support vector classifiers: a survey. *Knowledge-Based systems*, 196, 105754. <https://doi.org/10.1016/j.knosys.2020.105754>
12. Alberts, I. L., Mercolli, L., Pyka, T., Prenosil, G., Shi, K., Rominger, A., & Afshar-Oromieh, A. (2023). Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be? *European journal of nuclear medicine and molecular imaging*. <https://doi.org/10.1007/s00259-023-06172-w>
13. Al-Fedaghi, S., & Makdessi, M. (2020). Modeling business process and events. In *Intelligent algorithms in software engineering* (pp. 366–379). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-51965-0\\_32](https://doi.org/10.1007/978-3-030-51965-0_32)
14. Alhendi, A. A., Al-Sumaiti, A. S., Elmay, F. K., Wescaot, J., Kavousi-Fard, A., Heydarian-Forushani, E., & Alhelou, H. H. (2022). Artificial intelligence for water-energy nexus demand forecasting: a review. *International journal of low-carbon technologies*. <https://doi.org/10.1093/ijlct/ctac043>
15. Ali, G., Le, H.-Q., Kim, J., Hwang, S.-W., & Hwang, J.-I. (2019). Design of seamless multi-modal interaction framework for intelligent virtual agents in wearable mixed reality environment. In *CASA '19: computer animation and social agents*. New York, NY, USA: ACM. <https://doi.org/10.1145/3328756.3328758>
16. Aliprantis, J., & Caridakis, G. (2019). A survey of augmented reality applications in cultural heritage. *International journal of computational methods in heritage science*, 3(2), 118–147. <https://doi.org/10.4018/ijcmhs.2019070107>



17. Al-Kharusi, H., & Al-Bahadly, I. (2014). Intelligent parking management system based on image processing. *World journal of engineering and technology*, 02(02), 55–67. <https://doi.org/10.4236/wjet.2014.22006>
18. Alsmadi, I., Rice, N. M., & O'Brien, M. J. (2022). Fake or not? Automated detection of COVID-19 misinformation and disinformation in social networks and digital media. *Computational and mathematical organization theory*. <https://doi.org/10.1007/s10588-022-09369-w>
19. Al-Zahrani, A., & Al-Hebbi, M. (2022). Big data major security issues. *Tehnički glasnik*, 16(2), 197–204. <https://doi.org/10.31803/tg-20220124135330>
20. AlZu'bi, S., Hawashin, B., EIBes, M., & Al-Ayyoub, M. (2018). A novel recommender system based on apriori algorithm for requirements engineering. In *2018 fifth international conference on social networks analysis, management and security (SNAMS)*. IEEE. <https://doi.org/10.1109/snams.2018.8554909>
21. Amato, A., Venticinque, S., & Di Martino, B. (2013). Image recognition and augmented reality in cultural heritage using opencv. In *International conference*. New York, New York, USA: ACM Press. <https://doi.org/10.1145/2536853.2536878>
22. Amato, F., Fonisto, M., Giacalone, M., & Sansone, C. (2023). An intelligent conversational agent for the legal domain. *Information*, 14(6), 307. <https://doi.org/10.3390/info14060307>
23. Amindoust, A. (2018). A resilient-sustainable based supplier selection model using a hybrid intelligent method. *Computers & industrial engineering*, 126, 122–135. <https://doi.org/10.1016/j.cie.2018.09.031>
24. Amri, S., Sallami, D., & Aïmeur, E. (2022). EXMULF: an explainable multimodal content-based fake news detection system. In *Foundations and practice of security* (pp. 177–187). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-08147-7\\_12](https://doi.org/10.1007/978-3-031-08147-7_12)
25. Andeobu, L., Wibowo, S., & Grandhi, S. (2022). Artificial intelligence applications for sustainable solid waste management practices in Australia: a systematic review. *Science of the total environment*, 155389. <https://doi.org/10.1016/j.scitotenv.2022.155389>

26. Andersson, J. (2021). Statistical methods in recommender systems. In *Statistical analysis with swift* (pp. 167–198). Berkeley, CA: Apress. [https://doi.org/10.1007/978-1-4842-7765-2\\_7](https://doi.org/10.1007/978-1-4842-7765-2_7)
27. Anitha, P., & Patil, M. M. (2019). RFM model for customer purchase behavior using K–Means algorithm. *Journal of king saud university – computer and information sciences*. <https://doi.org/10.1016/j.jksuci.2019.12.011>
28. AnyLogic cloud: online simulation tools. (n.d.). Джерело: <https://cloud.anylogic.com/model/>
29. Aranda, J. R., Campos, I., Cosculluela, C., San Martin, J., & De Pablos, C. (2023). Continuous vocational training in response to the challenge of industry 4.0: required skills and business results. *Journal of industrial engineering and management*, 16(2), 319. <https://doi.org/10.3926/jiem.4665>
30. Arasu, B. S., Seelan, B. J. B., & Thamaraiselvan, N. (2020). A machine learning–based approach to enhancing social media marketing. *Computers & electrical engineering*, 86, 106723. <https://doi.org/10.1016/j.compeleceng.2020.106723>
31. Arkes, J. (2022). Regression analysis basics. In *Regression analysis* (2nd ed., pp. 12–50). London: Routledge. <https://doi.org/10.4324/9781003285007-2>
32. Arunachalam, D., & Kumar, N. (2018). Benefit–based consumer segmentation and performance evaluation of clustering approaches: an evidence of data–driven decision–making. *Expert systems with applications*, 111, 11–34. <https://doi.org/10.1016/j.eswa.2018.03.007>
33. Ashfaq, M., Yun, J., Yu, S., & Loureiro, S. M. C. (2020). I, Chatbot: Modeling the determinants of users’ satisfaction and continuance intention of AI–powered service agents. *Telematics and informatics*, 54, 101473. <https://doi.org/10.1016/j.tele.2020.101473>
34. Avvenuti, M., Cimino, M. G. C. A., Cresci, S., Marchetti, A., & Tesconi, M. (2016). A framework for detecting unfolding emergencies using humans as sensors. *SpringerPlus*, 5(1). <https://doi.org/10.1186/s40064-016-1674-y>
35. Baden–Fuller, C., & Haefliger, S. (2013). Business models and technological innovation. *Long range planning*, 46(6), 419–426. <https://doi.org/10.1016/j.lrp.2013.08.023>

36. Bahad, P., Saxena, P., & Kamal, R. (2019). Fake news detection using bi-directional lstm-recurrent neural network. *Procedia computer science*, 165, 74–82. <https://doi.org/10.1016/j.procs.2020.01.072>
37. Bai, L., Qi, M., & Liang, J. (2023). Spectral clustering with robust self-learning constraints. *Artificial intelligence*, 103924. <https://doi.org/10.1016/j.artint.2023.103924>
38. Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert systems with applications*, 42(20), 7046–7056. <https://doi.org/10.1016/j.eswa.2015.05.013>
39. Balzano, W., & Vitale, F. (2017). DiG-Park: A smart parking availability searching method using V2V/V2I and dgp-class problem. In *2017 31st international conference on advanced information networking and applications workshops (WAINA)*. IEEE. <https://doi.org/10.1109/waina.2017.104>
40. Bandyopadhyay, M., & Singh, V. (2016). Development of agent based model for predicting emergency response time. *Perspectives in science*, 8, 138–141. <https://doi.org/10.1016/j.pisc.2016.04.017>
41. Barabash, O., Shevchenko, G., Dakhno, N., Neshcheret, O., & Musienko, A. (2017). Information technology of targeting: optimization of decision making process in a competitive environment. *International journal of intelligent systems and applications*, 9(12), 1–9. <https://doi.org/10.5815/ijisa.2017.12.01>
42. Barenfanger, R., & Otto, B. (2015). Proposing a capability perspective on digital business models. In *2015 IEEE 17th conference on business informatics (CBI)*. IEEE. <https://doi.org/10.1109/cbi.2015.18>
43. Bartie, P., Mackaness, W., Lemon, O., Dalmas, T., Janarthanam, S., Hill, R. L., Liu, X. (2018). A dialogue based mobile virtual assistant for tourists: The SpaceBook Project. *Computers, environment and urban systems*, 67, 110–123. <https://doi.org/10.1016/j.compenvurbsys.2017.09.010>
44. Basu, M., Bit, S. D., & Ghosh, S. (2021). Utilizing microblogs for optimized real-time resource allocation in post-disaster scenarios. *Social network analysis and mining*, 12(1). <https://doi.org/10.1007/s13278-021-00841-0>

45. Behera, D. K. (2023). Promoting sustainable development through environmental policy, green technologies, and effective waste management: a comprehensive review. *Journal of multidisciplinary science: mIKAILALSYS*, 1(2), 179–198. <https://doi.org/10.58578/mikailalsys.v1i2.1675>
46. Behnassi, M., & El Haiba, M. (2022). Implications of the Russia–Ukraine war for global food security. *Nature human behaviour*. <https://doi.org/10.1038/s41562-022-01391-x>
47. Bekele, M. K., Pierdicca, R., Frontoni, E., Malinverni, E. S., & Gain, J. (2018). A survey of augmented, virtual, and mixed reality for cultural heritage. *Journal on computing and cultural heritage*, 11(2), 1–36. <https://doi.org/10.1145/3145534>
48. Bento, F. R. d. S. R. (2018). *Predicting start-up success with machine learning* (Master's thesis). Джерело: <http://hdl.handle.net/10362/33785>
49. Bezverhny, E., Dadteev, K., Barykin, L., Nemshaev, S., & Klimov, V. (2020). Use of chat bots in learning management systems. *Procedia computer science*, 169, 652–655. <https://doi.org/10.1016/j.procs.2020.02.195>
50. Bhattacharya, S., Scott, E., & Arthur, M. (2006). The phoenix rises from the ashes: advertising and content monetization in a digital world. *Journal of digital asset management*, 2(6), 269–278. <https://doi.org/10.1057/palgrave.dam.3650046>
51. Bhoi, A., Pujari, S. P., & Balabantaray, R. C. (2020). A deep learning–based social media text analysis framework for disaster resource management. *Social network analysis and mining*, 10(1). <https://doi.org/10.1007/s13278-020-00692-1>
52. Bishop, C. M., & Bishop, H. (2023). Deep neural networks. In *Deep learning* (pp. 171–207). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-45468-4\\_6](https://doi.org/10.1007/978-3-031-45468-4_6)
53. Bisong, E. (2019). Google automl: cloud vision. In *Building machine learning and deep learning models on google cloud platform* (pp. 581–598). Berkeley, CA: Apress. [https://doi.org/10.1007/978-1-4842-4470-8\\_42](https://doi.org/10.1007/978-1-4842-4470-8_42)
54. Black, J. S., & van Esch, P. (2020). AI–enabled recruiting: what is it and how should a manager use it? *Business horizons*, 63(2), 215–226. <https://doi.org/10.1016/j.bushor.2019.12.001>

55. Bocken, N. M. P., Short, S. W., Rana, P., & Evans, S. (2014). A literature and practice review to develop sustainable business model archetypes. *Journal of cleaner production*, 65, 42–56. <https://doi.org/10.1016/j.jclepro.2013.11.039>
56. Bodyanskiy, Y. V., Lipianina–Honcharenko, K. V., & Sachenko, A. O. (2022). Ensemble of adaptive predictors for multivariate nonstationary sequences and its online learning. *Radio Electronics, Computer Science, Control*, (4(67)), 91–97. <https://doi.org/10.15588/1607–3274–2023–4–9>
57. Bodyanskiy, Y., & Popov, S. (2004). Fuzzy selection mechanism for multimodel prediction. In *Lecture notes in computer science* (pp. 772–778). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978–3–540–30133–2\\_101](https://doi.org/10.1007/978–3–540–30133–2_101)
58. Bodyanskiy, Y., Otto, P., Pliss, I., & Popov, S. (2003). An optimal algorithm for combining multivariate forecasts in hybrid systems. In *Lecture notes in computer science* (pp. 967–972). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978–3–540–45226–3\\_132](https://doi.org/10.1007/978–3–540–45226–3_132)
59. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5–е вид.). Hoboken, NJ: Wiley.
60. Boz, E., Çizmecioğlu, S., & Çalık, A. (2022). A novel MDCM approach for sustainable supplier selection in healthcare system in the era of logistics 4.0. *Sustainability*, 14(21), 13839. <https://doi.org/10.3390/su142113839>
61. Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of retailing*, 93(1), 79–95. <https://doi.org/10.1016/j.jretai.2016.12.004>
62. Brožová, H., & Růžička, M. (2020). The prediction of parking space availability. *Transport*, 35(5), 462–473. <https://doi.org/10.3846/transport.2020.14016>
63. Brych, V., Manzhula, V., Borysiak, O., & Bondarchuk, M. (2021). Factor analysis of financial and economic activities of energy enterprises of Ukraine. *2021 11th international conference on advanced computer information technologies (ACIT)*. <https://doi.org/10.1109/ACIT52158.2021.9548358>

64. Brzeziński, Ł., & Wyrwicka, M. K. (2022). Fundamental directions of the development of the smart cities concept and solutions in Poland. *Energies*, *15*(21), 8213. <https://doi.org/10.3390/en15218213>
65. Bühlmann, P., & Yu, B. (2009). Boosting. *WIREs computational statistics*, *2*(1), 69–74. <https://doi.org/10.1002/wics.55>
66. Cabral, J. S. (2017). *Project risk management strategies for IT project managers* (ScholarWorks). Джерело: <https://scholarworks.waldenu.edu/dissertations/3663>
67. Cabrelles, M., Blanco–Pons, S., Carrión–Ruiz, B., & Lerma, J. L. (2017). From multispectral 3D recording and documentation to development of mobile apps for dissemination of cultural heritage. In *Cyber–Archaeology and grand narratives* (pp. 67–90). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-65693-9\\_5](https://doi.org/10.1007/978-3-319-65693-9_5)
68. Cai, Y., & Omay, T. (2021). Using double frequency in Fourier Dickey–Fuller unit root test. *Computational Economics*, 1–26.
69. Camero, A., Toutouh, J., Stolfi, D. H., & Alba, E. (2018). Evolutionary deep learning for car park occupancy prediction in smart cities. In *Lecture notes in computer science* (pp. 386–401). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-05348-2\\_32](https://doi.org/10.1007/978-3-030-05348-2_32)
70. Campbell, C., Plangger, K., Sands, S., & Kietzmann, J. (2021). Preparing for an era of deepfakes and AI–generated ads: a framework for understanding responses to manipulated advertising. *Journal of advertising*, 1–17. <https://doi.org/10.1080/00913367.2021.1909515>
71. Cao, J., Lam, K.–Y., Lee, L.–H., Liu, X., Hui, P., & Su, X. (2022). Mobile augmented reality: user interfaces, frameworks, and intelligence. *ACM computing surveys*. <https://doi.org/10.1145/3557999>
72. Casillo, M., Clarizia, F., D'Aniello, G., De Santo, M., Lombardi, M., & Santaniello, D. (2020). CHAT–Bot: a cultural heritage aware teller–bot for supporting touristic experiences. *Pattern recognition letters*, *131*, 234–243. <https://doi.org/10.1016/j.patrec.2020.01.003>

73. Cecil, R. R., & Soares, J. (2019). IBM watson studio: a platform to transform data to intelligence. In *Pharmaceutical supply chains – medicines shortages* (pp. 183–192). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-15398-4\\_13](https://doi.org/10.1007/978-3-030-15398-4_13)
74. Celotto, E., Ellero, A., & Ferretti, P. (2012). Short–medium term tourist services demand forecasting with rough set theory. *Procedia economics and finance*, 3, 62–67. [https://doi.org/10.1016/s2212-5671\(12\)00121-9](https://doi.org/10.1016/s2212-5671(12)00121-9)
75. Chakraborty, U., & Bhat, S. (2018). Credibility of online reviews and its impact on brand image. *Management research review*, 41(1), 148–164. <https://doi.org/10.1108/mrr-06-2017-0173>
76. Chepurnenko, A., Kondratieva, T., Deberdeev, T., Akopyan, V., Avakov, A., & Chepurnenko, V. (2023). Prediction of rheological parameters of polymers, using gradient boosting algorithm CatBoost. *All the materials. encyclopedic reference book.*, (6), 21–29. <https://doi.org/10.31044/1994-6260-2023-0-6-21-29>
77. Chiang, L.–L. & Yang, C.–S. (2018). Does country–of–origin brand personality generate retail customer lifetime value? A Big Data analytics approach. *Technological forecasting and social change*, 130, 177–187. <https://doi.org/10.1016/j.techfore.2017.06.034>
78. Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1). <https://doi.org/10.1186/s12864-019-6413-7>
79. Chiu, C.–C., Wei, W.–J., Lee, L.–C., & Lu, J.–C. (2019). Augmented reality system for tourism using image–based recognition. *Microsystem technologies*. <https://doi.org/10.1007/s00542-019-04600-2>
80. Christensson, P. (2006). Array. Джерело: <https://techterms.com/definition/array>
81. Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: an experimental study of human–chatbot interaction. *Future generation computer systems*, 92, 539–548. <https://doi.org/10.1016/j.future.2018.01.055>

82. Civera, M., Sibille, L., Zanotti Fragonara, L., & Ceravolo, R. (2023). A dbscan–based automated operational modal analysis algorithm for bridge monitoring. *Measurement*, 112451. <https://doi.org/10.1016/j.measurement.2023.112451>
83. Client cabinet. (n.d.). Джерело: <https://seller.rozetka.com.ua/newseller/ua>
84. Comendador, B. E. V., Francisco, B. M. B., Medenilla, J. S., Nacion, S. M. T., & Serac, T. B. E. (2015). Pharmabot: a pediatric generic medicine consultant chatbot. *Journal of automation and control engineering*, 3(2), 137–140. <https://doi.org/10.12720/joace.3.2.137-140>
85. Contributors to Wikimedia projects. (2001). Array (data structure) – Wikipedia. Джерело: [https://en.wikipedia.org/wiki/Array\\_\(data\\_structure\)](https://en.wikipedia.org/wiki/Array_(data_structure))
86. Cooper, V., V. (2020). Algorithm inspection for chatbot performance evaluation. *Procedia computer science*, 171, 2267–2274. <https://doi.org/10.1016/j.procs.2020.04.245>
87. Čopič Pucihar, K., & Kljun, M. (2018). ART for art: augmented reality taxonomy for art and cultural heritage. In *Springer series on cultural computing* (pp. 73–94). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-69932-5\\_3](https://doi.org/10.1007/978-3-319-69932-5_3)
88. Costa, N. L., Llobodanin, L. A. G., Castro, I. A., & Barbosa, R. (2019). Using support vector machines and neural networks to classify merlot wines from south america. *Information processing in agriculture*, 6(2), 265–278. <https://doi.org/10.1016/j.inpa.2018.10.003>
89. Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). SuperAgent: a customer service chatbot for e–commerce websites. In *Proceedings of ACL 2017, system demonstrations*. Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/p17-4017>
90. Customer personality analysis. (n.d.). Джерело: <https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis>
91. Data base management systems (DBMS): the beginner’s guide. (2023, May 11). Джерело: [https://www.splunk.com/en\\_us/blog/learn/dbms-database-management-systems.html](https://www.splunk.com/en_us/blog/learn/dbms-database-management-systems.html)



92. Data visualization tools used for decision making: a review. (2023). *International research journal of modernization in engineering technology and science*. <https://doi.org/10.56726/irjmets33640>
93. Dataiku | everyday AI, extraordinary people. (n.d.). Джерело: <https://www.dataiku.com/>
94. DataRobot homepage | DataRobot AI Platform. (n.d.). Джерело: <https://datarobot.com>
95. De Capitani di Vimercati, S., Foresti, S., Livraga, G., & Samarati, P. (2022). Digital infrastructure policies for data security and privacy in smart cities. In *Smart cities policies and financing* (pp. 249–261). Elsevier. <https://doi.org/10.1016/b978-0-12-819130-9.00007-3>
96. Dehghani, Z. (2022). *Data mesh: delivering data-driven value at scale*. O'Reilly Media, Incorporated.
97. Demil, B., & Lecocq, X. (2010). Business model evolution: in search of dynamic consistency. *Long range planning*, 43(2–3), 227–246. <https://doi.org/10.1016/j.lrp.2010.02.004>
98. Descriptive analyses. (2022). In *Statistics and data analysis literacy for nurses*. New York, NY: Springer Publishing Company. <https://doi.org/10.1891/9780826165824.0003>
99. Devagiri, J. S., Paheding, S., Niyaz, Q., Yang, X., & Smith, S. (2022). Augmented Reality and Artificial Intelligence in industry: trends, tools, and future challenges. *Expert systems with applications*, 207, 118002. <https://doi.org/10.1016/j.eswa.2022.118002>
100. Dialogflow documentation | google cloud. (n.d.). Джерело: <https://dialogflow.com/>
101. Discovery Walking Tours TV. (2019). Shenzhen electronics market (huaqiangbei) china shopping tour [Video file]. Джерело: <https://www.youtube.com/watch?v=jnWDR9XsfR4>

102. Do, H.–N., Shih, W., & Ha, Q.–A. (2020). Effects of mobile augmented reality apps on impulse buying behavior: an investigation in the tourism field. *Heliyon*, 6(8), Article e04667. <https://doi.org/10.1016/j.heliyon.2020.e04667>
103. Dong, J.–Y., & Wan, S.–P. (2016). Virtual enterprise partner selection integrating LINMAP and TOPSIS. *Journal of the operational research society*, 67(10), 1288–1308. <https://doi.org/10.1057/jors.2016.22>
104. Dong, Q., Gong, S., & Zhu, X. (2017). Multi–task curriculum transfer deep learning of clothing attributes. In *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE. <https://doi.org/10.1109/wacv.2017.64>
105. Dulishkovych, O., Lipyana, H., Sachenko, O., Lendyuk, T., Osolinskiy, O., & Kit, I. (2021). Business process management of virtual enterprise based on augmented reality. *CEUR proceedings of the 2nd international workshop IT project management (ITPM 2021)*, (2851), 422–431. Джерело: <https://ceur-ws.org/Vol-2851/paper39.pdf>
106. Egorov, E. E., Lebedeva, T. E., Prokhorova, M. P., Tsapina, T. N., & Shkunova, A. A. (2020). Opportunities and prospects of using chatbots in HR. In *Scientific and technical revolution: yesterday, today and tomorrow* (pp. 782–791). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-47945-9\\_83](https://doi.org/10.1007/978-3-030-47945-9_83)
107. Eicher, A. (2019). Big business with big data big business mit big data. *GIS Business*, 12(3), 20–25. <https://doi.org/10.26643/gis.v12i3.5173>
108. El Morr, C., Jammal, M., Ali–Hassan, H., & El–Hallak, W. (2022). Logistic regression. In *International series in operations research & management science* (pp. 231–249). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-16990-8\\_7](https://doi.org/10.1007/978-3-031-16990-8_7)
109. El Ouadi, J., Malhene, N., Benhadou, S., & Medromi, H. (2022). Towards a machine–learning based approach for splitting cities in freight logistics context: benchmarks of clustering and prediction models. *Computers & industrial engineering*, 166, 107975. <https://doi.org/10.1016/j.cie.2022.107975>
110. Enders, W. (2014). *Applied Econometric Time Series* (4–е вид.). Hoboken, NJ: Wiley.

111. Errouso, H., Malhene, N., Benhadou, S., & Medromi, H. (2020). Predicting car park availability for a better delivery bay management. *Procedia computer science*, 170, 203–210. <https://doi.org/10.1016/j.procs.2020.03.026>
112. Esposito, E., & Evangelista, P. (2014). Investigating virtual enterprise models: literature review and empirical findings. *International journal of production economics*, 148, 145–157. <https://doi.org/10.1016/j.ijpe.2013.10.003>
113. Etaati, L. (2019). Azure machine learning studio. In *Machine learning with microsoft technologies* (pp. 201–223). Berkeley, CA: Apress. [https://doi.org/10.1007/978-1-4842-3658-1\\_12](https://doi.org/10.1007/978-1-4842-3658-1_12)
114. Fanini, B., Pagano, A., Pietroni, E., Ferdani, D., Demetrescu, E., & Palombini, A. (2021). Augmented reality for cultural heritage. In *Springer handbooks* (pp. 391–411). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-67822-7\\_16](https://doi.org/10.1007/978-3-030-67822-7_16)
115. Favaro, P., & Vedaldi, A. (2020). AdaBoost. In *Computer vision* (pp. 1–4). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-03243-2\\_663-1](https://doi.org/10.1007/978-3-030-03243-2_663-1)
116. Ferreira, A. J., & Figueiredo, M. A. T. (2012). Boosting algorithms: a review of methods, theory, and applications. In *Ensemble machine learning* (pp. 35–85). Boston, MA: Springer US. [https://doi.org/10.1007/978-1-4419-9326-7\\_2](https://doi.org/10.1007/978-1-4419-9326-7_2)
117. Ferreira, L., Lopes, N., Ávila, P. S., Castro, H., Varela, M. L. R., Putnik, G. D., Cruz-Cunha, M. M. (2017). Virtual Enterprise integration management based on a Meta-enterprise – a PMBoK approach. *Procedia computer science*, 121, 1112–1118. <https://doi.org/10.1016/j.procs.2017.12.120>
118. Fowler, B. (2023). Cloud services and technologies. In *AWS for public and private sectors* (pp. 1–21). Berkeley, CA: Apress. [https://doi.org/10.1007/978-1-4842-9048-4\\_1](https://doi.org/10.1007/978-1-4842-9048-4_1)
119. Fränti, P., & Sieranoja, S. (2018). K-means properties on six clustering benchmark datasets. *Applied intelligence*, 48(12), 4743–4759. <https://doi.org/10.1007/s10489-018-1238-7>

120. Freund, Y., & Schapire, R. E. (1997). A decision–theoretic generalization of on–line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
121. Fu, D., Ban, Y., Tong, H., Maciejewski, R., & He, J. (2022). DISCO: comprehensive and explainable disinformation detection. In *CIKM '22: the 31st ACM international conference on information and knowledge management*. New York, NY, USA: ACM. <https://doi.org/10.1145/3511808.3557202>
122. Fustos, J., Morris, G., & Haga, W. (2016). Web developer knowledge and skills 2014. *International journal of information and education technology*, 6(7), 538–542. <https://doi.org/10.7763/ijiet.2016.v6.747>
123. Gacutan, J., & Selvadurai, N. (2020). A statutory right to explanation for decisions generated using artificial intelligence. *International journal of law and information technology*, 28(3), 193–216. <https://doi.org/10.1093/ijlit/ehaa016>
124. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2012). A review on ensembles for the class imbalance problem: bagging–, boosting–, and hybrid–based approaches. *IEEE transactions on systems, man, and cybernetics, part C (applications and reviews)*, 42(4), 463–484. <https://doi.org/10.1109/tsmcc.2011.2161285>
125. Gammeter, S., Gassmann, A., Bossard, L., Quack, T., & Van Gool, L. (2010). Server–side object recognition and client–side object tracking for mobile augmented reality. In *2010 IEEE computer society conference on computer vision and pattern recognition workshops (CVPR workshops)*. IEEE. <https://doi.org/10.1109/cvprw.2010.5543248>
126. Ganguly, D., Conrad, J. G., Ghosh, K., Ghosh, S., Goyal, P., Bhattacharya, P., Paul, S. (2023). Legal IR and NLP: the history, challenges, and state–of–the–art. In *Lecture notes in computer science* (pp. 331–340). Cham: Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-28241-6\\_34](https://doi.org/10.1007/978-3-031-28241-6_34)
127. Garcia–Dias, R., Vieira, S., Lopez Pinaya, W. H., & Mechelli, A. (2020). Clustering analysis. In *Machine learning* (pp. 227–247). Elsevier. <https://doi.org/10.1016/b978-0-12-815739-8.00013-4>

128. García-Sánchez, F., Colomo-Palacios, R., & Valencia-García, R. (2020). A social-semantic recommender system for advertisements. *Information processing & management*, 57(2), 102153. <https://doi.org/10.1016/j.ipm.2019.102153>
129. Gassmann, O. (2014). *The business model navigator: 55 models that will revolutionise your business*.
130. Gedikli, A., Taş, C. Y., & Taş, N. B. (2022). Redefining smart cities, urban energy, and green technologies for sustainable development. In *Practice, progress, and proficiency in sustainability* (pp. 216–232). IGI Global. <https://doi.org/10.4018/978-1-7998-8482-8.ch014>.
131. Geetha, T. V., & Sendhilkumar, S. (2023). Classification algorithms. In *Machine learning* (pp. 127–151). Boca Raton: Chapman and Hall/CRC. <https://doi.org/10.1201/9781003290100-6>.
132. Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial intelligence in healthcare* (pp. 295–336). Elsevier. <https://doi.org/10.1016/b978-0-12-818438-7.00012-5>
133. Ghose, A., Li, B., & Liu, S. (2019). Mobile targeting using customer trajectory patterns. *Management science*, 65(11), 5027–5049. <https://doi.org/10.1287/mnsc.2018.3188>
134. Gillespie, A. A., & Magor, S. (2019). Tackling online fraud. *ERA forum*, 20(3), 439–454. <https://doi.org/10.1007/s12027-019-00580-y>
135. Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. *Applied soft computing*, 101, 106991. <https://doi.org/10.1016/j.asoc.2020.106991>
136. Gomes, M., Oliveira, B., & Sousa, C. (2022). Enriching legal knowledge through intelligent information retrieval techniques: a review. In *Progress in artificial intelligence* (pp. 119–130). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-16474-3\\_11](https://doi.org/10.1007/978-3-031-16474-3_11)
137. Gopnarayan, A., & Deshpande, S. (2020). Tweets analysis for disaster management: preparedness, emergency response, impact, and recovery. In *Innovative data*

*communication technologies and application* (pp. 760–764). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-38040-3\\_87](https://doi.org/10.1007/978-3-030-38040-3_87)

138. Govindasamy, R. (2018). Cluster analysis of wine market segmentation – A consumer based study in the mid-atlantic USA. *Economic affairs*, 63(1). <https://doi.org/10.30954/0424-2513.2018.00150.19>

139. Gramyak, R., Lipyanina–Goncharenko, H., Sachenko, A., Lendyuk, T., & Zahorodnia, D. (2022). Intelligent method of a competitive product choosing based on the emotional feedbacks coloring. *Proceedings of the 2nd international workshop on intelligent information technologies & systems of information security with CEUR–WS*, 2853, 346–357. Джерело: <https://ceur-ws.org/Vol-2853/paper31.pdf>

140. Granger, C. W. J., & Newbold, P. (1986). *Forecasting Economic Time Series* (2–е вид.). Orlando, FL: Academic Press.

141. Griva, A., Bardaki, C., Pramadari, K., & Papakiriakopoulos, D. (2018). Retail business analytics: customer visit segmentation using market basket data. *Expert systems with applications*, 100, 1–16. <https://doi.org/10.1016/j.eswa.2018.01.029>

142. Gu, X., Gao, F., Tan, M., & Peng, P. (2020). Fashion analysis and understanding with artificial intelligence. *Information processing & management*, 57(5), 102276. <https://doi.org/10.1016/j.ipm.2020.102276>

143. Gue, I. H. V., Lopez, N. S. A., Chiu, A. S. F., Ubando, A. T., & Tan, R. R. (2022). Predicting waste management systems performance from city and country attributes. *Journal of cleaner production*, 132951. <https://doi.org/10.1016/j.jclepro.2022.132951>

144. Guerzoni, M., Nava, C. R., & Nuccio, M. (2020). Start-ups survival through a crisis. Combining machine learning with econometrics to measure innovation. *Economics of innovation and new technology*, 1–26. <https://doi.org/10.1080/10438599.2020.1769810>

145. Guich, J. A., & Osorio, V. L. A. (2021). *Design Thinking Vs. Lean Startup, un análisis desde el enfoque educativo* (Bachelor's thesis, Universidad Peruana de Ciencias Aplicadas (UPC)). Джерело: <http://hdl.handle.net/10757/655548>

146. Guo, S., Jin, Z., Sun, F., Li, J., Li, Z., Shi, Y., & Cao, N. (2021). Vinci: an intelligent graphic design system for generating advertising posters. In *CHI '21: CHI*

*conference on human factors in computing systems*. New York, NY, USA: ACM.  
<https://doi.org/10.1145/3411764.3445117>

147. Guo, S., Liu, Y., Chen, R., Sun, X., & Wang, X. (2018). Improved SMOTE algorithm to deal with imbalanced activity classes in smart homes. *Neural processing letters*, 50(2), 1503–1526. <https://doi.org/10.1007/s11063-018-9940-3>

148. Gupta, Y. P., Mukul, & Gupta, N. (2022). Deep learning model based multimedia retrieval and its optimization in augmented reality applications. *Multimedia tools and applications*. <https://doi.org/10.1007/s11042-022-13555-y>

149. Gurcan, F., Ayaz, A., Menekse Dalveren, G. G., & Derawi, M. (2023). Business intelligence strategies, best practices, and latest trends: analysis of scientometric data from 2003 to 2023 using machine learning. *Sustainability*, 15(13), 9854. <https://doi.org/10.3390/su15139854>

150. Ha, T. M., & Duong, V. H. (2018). Using system dynamics modelling and communication strategies for a resilient and smart city in vietnam. *International journal of environmental science & sustainable development*, 3(1), 10. <https://doi.org/10.21625/essd.v3iss1.280>

151. Hailiang, Z., Khokhar, M., Islam, T., & Sharma, A. (2023). A model for green–resilient supplier selection: fuzzy best–worst multi–criteria decision–making method and its applications. *Environmental science and pollution research*. <https://doi.org/10.1007/s11356-023-25749-4>

152. Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

153. Han, D.–I. D., Tom Dieck, M. C., & Jung, T. (2019). Augmented Reality Smart Glasses (ARSG) visitor adoption in cultural tourism. *Leisure studies*, 38(5), 618–633. <https://doi.org/10.1080/02614367.2019.1604790>

154. Han, J.–G., Park, K.–W., Ban, K.–J., & Kim, E.–K. (2013). Cultural heritage sites visualization system based on outdoor augmented reality. *AASRI procedia*, 4, 64–71. <https://doi.org/10.1016/j.aasri.2013.10.011>

155. Hancock, B., Bordes, A., Mazare, P.–E., & Weston, J. (2019). Learning from dialogue after deployment: feed yourself, chatbot! In *Proceedings of the 57th annual*

*meeting of the association for computational linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/p19-1358>

156. Haouel, C. (2023). Assessment of the impact of russia's war on ukraine on EU and UK oil and gas imports and their energy supply security. In *CEEeGov 2023: central and eastern european edem and egov days 2023*. New York, NY, USA: ACM. <https://doi.org/10.1145/3603304.3603344>

157. Harika, A., Sunil Kumar, M., Anantha Natarajan, V., & Kallam, S. (2021). Business process reengineering: issues and challenges. In *Algorithms for intelligent systems* (pp. 363–382). Singapore: Springer Singapore. [https://doi.org/10.1007/978-981-15-6707-0\\_35](https://doi.org/10.1007/978-981-15-6707-0_35)

158. He, W., & Liu, Y. (2018). To regularize or not: Revisiting SGD with simple algorithms and experimental studies. *Expert systems with applications*, *112*, 1–14. <https://doi.org/10.1016/j.eswa.2018.06.026>

159. Henz, P. (2021). Ethical and legal responsibility for Artificial Intelligence. *Discover artificial intelligence*, *1*(1). <https://doi.org/10.1007/s44163-021-00002-4>

160. Hidalgo-Crespo, J., Álvarez-Mendoza, C. I., Soto, M., & Amaya-Rivas, J. L. (2022). Quantification and mapping of domestic plastic waste using GIS/GPS approach at the city of Guayaquil. *Procedia CIRP*, *105*, 86–91. <https://doi.org/10.1016/j.procir.2022.02.015>

161. Hoang, T. A.-D., & Nguyen, T. B. (2009). State of the art and emerging rule-driven perspectives towards service-based business process interoperability. In *2009 IEEE-RIVF international conference on computing and communication technologies*. IEEE. <https://doi.org/10.1109/rivf.2009.5174659>

162. Hoffmann, J. P. (2021). Review of elementary statistical concepts. In *Linear regression models* (pp. 7–36). Boca Raton: Chapman and Hall/CRC. <https://doi.org/10.1201/9781003162230-2>

163. Howlader, A. M., Patel, D., & Gammariello, R. (2023). Data-driven approach for instantaneous vehicle emission predicting using integrated deep neural network. *Transportation research part D: transport and environment*, *116*, 103654. <https://doi.org/10.1016/j.trd.2023.103654>



164. Hrybinenko, O. H., Bulatova, O., & Zakharova, O. (2020). Financial indicators in the system of economic security of the world countries. In *11th international scientific conference „business and management 2020“*. VGTU Technika. <https://doi.org/10.3846/bm.2020.672>
165. Indans, R., Hauthal, E., & Burghardt, D. (2019). Towards an audio–locative mobile application for immersive storytelling. *KN – journal of cartography and geographic information*, 69(1), 41–50. <https://doi.org/10.1007/s42489-019-00007-1>
166. Industrial safety and health analytics database. (n.d.). Джерело: <https://www.kaggle.com/datasets/ihmstefanini/industrial-safety-and-health-analytics-database>
167. Inoue, N., Simo–Serra, E., Yamasaki, T., & Ishikawa, H. (2017). Multi–label fashion image classification with minimal human supervision. In *2017 IEEE international conference on computer vision workshop (ICCVW)*. IEEE. <https://doi.org/10.1109/iccvw.2017.265>
168. Ismail, N., & Yusof, U. K. (2022). Recent trends of machine learning predictions using open data: a systematic review. *Journal of information and communication technology*, 21(No.3), 337–381. <https://doi.org/10.32890/jict2022.21.3.3>
169. Izonin, I., Hovorushchenko, T., & Shandilya, S. K. (2024). Quality and Security of Critical Infrastructure Systems. *Big Data and Cognitive Computing*, 8(1), 10.. <https://doi.org/10.3390/bdcc8010010>
170. Izquierdo–Horna, L., Damazo, M., & Yanayaco, D. (2022). Identification of urban sectors prone to solid waste accumulation: a machine learning approach based on social indicators. *Computers, environment and urban systems*, (96), Article 101834. <https://doi.org/10.1016/j.compenvurbsys.2022.101834>
171. Jadhav, S. S., & Thepade, S. D. (2019). Fake news identification and classification using DSSM and improved recurrent neural network classifier. *Applied artificial intelligence*, 33(12), 1058–1068. <https://doi.org/10.1080/08839514.2019.1661579>
172. Jan, M. T., & Ammari, D. (2016). Advertising online by educational institutions and students' reaction: a study of Malaysian Universities. *Journal of marketing for higher education*, 26(2), 168–180. <https://doi.org/10.1080/08841241.2016.1245232>

173. Janczyk, M., & Pfister, R. (2023). Decision errors, effect sizes, and power. In *Understanding inferential statistics* (pp. 81–96). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-662-66786-6\\_7](https://doi.org/10.1007/978-3-662-66786-6_7)

174. Janczyk, M., & Pfister, R. (2023). Factorial Analysis of Variance (ANOVA). In *Understanding Inferential Statistics* (pp. 127–144). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-662-66786-6\\_9](https://doi.org/10.1007/978-3-662-66786-6_9)

175. Janczyk, M., & Pfister, R. (2023). Hypothesis testing and significance. In *Understanding inferential statistics* (pp. 35–44). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-662-66786-6\\_4](https://doi.org/10.1007/978-3-662-66786-6_4)

176. Janczyk, M., & Pfister, R. (2023). Introduction to inferential statistics 2: population and parameter estimation. In *Understanding inferential statistics* (pp. 23–33). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-662-66786-6\\_3](https://doi.org/10.1007/978-3-662-66786-6_3)

177. Javani, B., & Rwelamila, P. M. D. (2016). Risk management in IT projects – a case of the South African public sector. *International journal of managing projects in business*, 9(2), 389–413. <https://doi.org/10.1108/ijmpb-07-2015-0055>

178. Javed, A., & Yasir, M. (2019). Virtual social enterprise: modeling sustainability of an enterprise by digital intervention. *World journal of entrepreneurship, management and sustainable development*, 15(2), 182–196. <https://doi.org/10.1108/wjemsd-03-2018-0032>

179. Jiang, R., Bouridane, A., Li, C.–T., Crookes, D., Boussakta, S., Hao, F., & A. Edirisinghe, E. (Eds.). (2022). *Big data privacy and security in smart cities*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-031-04424-3>

180. Jin, S., Chen, S., & Xie, X. (2021). Property-based test for part-of-speech tagging tool. In *2021 36th IEEE/ACM international conference on automated software engineering (ASE)*. IEEE. <https://doi.org/10.1109/ase51524.2021.9678807>

181. Jomaa, H. S., Grabocka, J., Schmidt–Thieme, L., & Borek, A. (2019). A hybrid convolutional approach for parking availability prediction. In *2019 international joint conference on neural networks (IJCNN)*. IEEE. <https://doi.org/10.1109/ijcnn.2019.8852400>

182. Joshi, A. V. (2019). Amazon's machine learning toolkit: sagemaker. In *Machine learning and artificial intelligence* (pp. 233–243). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-26622-6\\_24](https://doi.org/10.1007/978-3-030-26622-6_24)
183. Jung, D., Tran Tuan, V., Quoc Tran, D., Park, M., & Park, S. (2020). Conceptual framework of an intelligent decision support system for smart city disaster management. *Applied sciences*, *10*(2), 666. <https://doi.org/10.3390/app10020666>
184. Juvenile Ehwi, R., Holmes, H., Maslova, S., & Burgess, G. (2022). The ethical underpinnings of smart city governance: decision-making in the smart cambridge programme, UK. *Urban studies*, 004209802110649. <https://doi.org/10.1177/00420980211064983>
185. Kaczmarek, K., Narebski, J., Piotrowski, S., & Przymus, P. (2022). Fast JSON parser using metaprogramming on GPU. In *2022 IEEE 9th international conference on data science and advanced analytics (DSAA)*. IEEE. <https://doi.org/10.1109/dsaa54385.2022.10032381>
186. Kalepalli, Y., Tasneem, S., Phani Teja, P. D., & Manne, S. (2020). Effective comparison of LDA with LSA for topic modelling. In *2020 4th international conference on intelligent computing and control systems (ICICCS)*. IEEE. <https://doi.org/10.1109/iciccs48265.2020.9120888>
187. Kamilov, M., Hudayberdiev, M., & Khamroev, A. (2019). Algorithm for the development of a training set that best describes the objects of recognition. *Procedia computer science*, *150*, 116–122. <https://doi.org/10.1016/j.procs.2019.02.024>
188. Kanavos, A., Iakovou, S., Sioutas, S., & Tampakas, V. (2018). Large scale product recommendation of supermarket ware based on customer behaviour analysis. *Big data and cognitive computing*, *2*(2), 11. <https://doi.org/10.3390/bdcc2020011>
189. Kanezaki, A. (2018). Unsupervised image segmentation by backpropagation. In *ICASSP 2018 – 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. <https://doi.org/10.1109/icassp.2018.8462533>
190. Kapała, A. (2022). Legal instruments to support short food supply chains and local food systems in france. *Laws*, *11*(2), 21. <https://doi.org/10.3390/laws11020021>

191. Kaplenko, H. K., Sytnyk, N. S., & Kołeczka, M. K. (2021). Implementation and development of the smart city system. *Socio world–social research & behavioral sciences*, 06(04(01)), 60–71. [https://doi.org/10.36962/swd0604\(01\)2021–60](https://doi.org/10.36962/swd0604(01)2021–60)
192. Karam, E., Hussein, W., & Gharib, T. F. (2021). Integrating location and textual information for detecting affected people in a crisis. *Social network analysis and mining*, 11(1). <https://doi.org/10.1007/s13278–020–00715–x>
193. Kateri, M. (2014). Association models. In *Contingency table analysis* (pp. 153–196). New York, NY: Springer New York. [https://doi.org/10.1007/978–0–8176–4811–4\\_6](https://doi.org/10.1007/978–0–8176–4811–4_6)
194. Kaushik, K., Mishra, R., Rana, N. P., & Dwivedi, Y. K. (2018). Exploring reviews and review sequences on e-commerce platform: a study of helpful reviews on amazon.in. *Journal of retailing and consumer services*, 45, 21–32. <https://doi.org/10.1016/j.jretconser.2018.08.002>
195. Key, Z., Parrish, A., Snavely, C., & Shafiee–Jood, M. (2022). Emergency management and underserved communities: using big data to improve emergency management preparedness, response and resilience. In *2022 systems and information engineering design symposium (SIEDS)*. IEEE. <https://doi.org/10.1109/sieds55548.2022.9799307>
196. Khajouei, R., & Farahani, F. (2019). The evaluation of users' satisfaction with the social security electronic system in iran. *Health and technology*, 9(5), 797–804. <https://doi.org/10.1007/s12553–019–00347–y>
197. Khan, M. A., Israr, S., S Almogren, A., Din, I. U., Almogren, A., & Rodrigues, J. J. P. C. (2020). Using augmented reality and deep learning to enhance Taxila Museum experience. *Journal of real–time image processing*. <https://doi.org/10.1007/s11554–020–01038–y>
198. Khan, M. M., Bashar, I., Minhaj, G. M., Wasi, A. I., & Hossain, N. U. I. (2023). Resilient and sustainable supplier selection: an integration of SCOR 4.0 and machine learning approach. *Sustainable and resilient infrastructure*, 1–17. <https://doi.org/10.1080/23789689.2023.2165782>

199. Khedkar, S., & Shinde, S. (2020). Deep learning–based approach to classify praises or complaints from customer reviews. In *Proceeding of international conference on computational science and applications* (pp. 391–402). Singapore: Springer Singapore. [https://doi.org/10.1007/978-981-15-0790-8\\_38](https://doi.org/10.1007/978-981-15-0790-8_38)
200. Kherwa, P., & Bansal, P. (2020). A comparative empirical evaluation of topic modeling techniques. In *Advances in intelligent systems and computing* (pp. 289–297). Singapore: Springer Singapore. [https://doi.org/10.1007/978-981-15-5148-2\\_26](https://doi.org/10.1007/978-981-15-5148-2_26)
201. Kianisarkaleh, A., & Ghassemian, H. (2016). Nonparametric feature extraction for classification of hyperspectral images with limited training samples. *ISPRS journal of photogrammetry and remote sensing*, *119*, 64–78. <https://doi.org/10.1016/j.isprsjprs.2016.05.009>
202. Kim, J., & Irizarry, J. (2020). Evaluating the use of augmented reality technology to improve construction management student’s spatial skills. *International journal of construction education and research*, 1–18. <https://doi.org/10.1080/15578771.2020.1717680>
203. Kit, I., Lipyanina–Goncharenko, H., Lendyuk, T., Sachenko, A., & Komar, M. (2022). Neural network method of items catalog forming for online store. *Lecture Notes on Data Engineering and Communications Technologies*, *135*, 157–169. [https://doi.org/10.1007/978-3-031-04809-8\\_14](https://doi.org/10.1007/978-3-031-04809-8_14)
204. Klappenecker, A., Lee, H., & Welch, J. L. (2014). Finding available parking spaces made easy. *Ad hoc networks*, *12*, 243–249. <https://doi.org/10.1016/j.adhoc.2012.03.002>
205. Kleinbaum, D. G., & Klein, M. (2010). Introduction to logistic regression. In *Statistics for biology and health* (pp. 1–39). New York, NY: Springer New York. [https://doi.org/10.1007/978-1-4419-1742-3\\_1](https://doi.org/10.1007/978-1-4419-1742-3_1)
206. Kolassa, S. (2022). Do we want coherent hierarchical forecasts, or minimal MAPEs or MAEs? (We won’t get both!). *International journal of forecasting*. <https://doi.org/10.1016/j.ijforecast.2022.11.006>
207. Komar, M., Lipianina–Honcharenko, K., Domanskyi, V., & Melnyk, N. (2024). Regression–based method for real–time solar power plant efficiency forecasting.

Proceedings of the Modern Machine Learning Technologies Workshop (MoMLeT 2024) Lviv, Ukraine, May 31 – June 1, 2024. (pp. 235–245) <https://ceur-ws.org/Vol-3711/paper14.pdf>

208. Komar, M., Savchyshyn, R., Lipianina–Honcharenko, K., & Osolinskyi, O. (2023). Intelligent method for counting cars from satellite images. *III international scientific symposium “intelligent solutions” (intsol-2023)*, (3538), Article 295–303. Джерело: [https://ceur-ws.org/Vol-3538/Short\\_1.pdf](https://ceur-ws.org/Vol-3538/Short_1.pdf)

209. Komorowski, M., Marshall, D. C., Saliccioli, J. D., & Crutain, Y. (2016). Exploratory data analysis. In *Secondary analysis of electronic health records* (pp. 185–203). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-43742-2\\_15](https://doi.org/10.1007/978-3-319-43742-2_15)

210. Koniaris, M., Galanis, D., Giannini, E., & Tsanakas, P. (2023). Evaluation of automatic legal text summarization techniques for greek case law. *Information*, 14(4), 250. <https://doi.org/10.3390/info14040250>

211. Koziuk, V., & Lipyanina–Goncharenko, H. (2021). Intelligent Method of Predicting the Discount Rate Trend. In *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. IEEE. <https://doi.org/10.1109/idaacs53288.2021.9660835>

212. Krawczyk, B., Minku, L. L., Gama, J., Stefanowski, J., & Woźniak, M. (2017). Ensemble learning for data stream analysis: A survey. *Information Fusion*, 37, 132–156.

213. Kreutzer, R. T., & Sirrenberg, M. (2019). Fields of application of artificial intelligence–security sector and military sector. In *Understanding artificial intelligence* (pp. 225–233). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-25271-7\\_9](https://doi.org/10.1007/978-3-030-25271-7_9)

214. Krylov, V., Sachenko, A., Strubytskyi, P., Lendiuk, D., Lipyanina, H., Zahorodnia, D., Lendyuk, T. (2019). Multiple regression method for analyzing the tourist demand considering the influence factors. In *2019 10th IEEE international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS)*. IEEE. <https://doi.org/10.1109/idaacs.2019.8924461>

215. Krysovaty, A., Lipianina–Honcharenko, H., Sachenko, S., Desyatnyuk, O., Banasik, A., & Lukasevych–Krutnyk, I. (2022). Recognizing the fictitious business entity on logistic regression base. *Proceedings of the 3rd international workshop on intelligent information technologies & systems of information security Khmelnytskyi*, 3156. Джерело: <https://ceur-ws.org/Vol-3156/paper15.pdf>

216. Krysovaty, A., Lipyanina–Goncharenko, H., Desyatnyuk, O., & Sachenko, S. (2021). Classification method of fictitious enterprises based on gaussian naive bayes. *2021 IEEE 16th international conference on computer sciences and information technologies (CSIT), LVIV, ukraine, 22–25 september 2021*. <https://doi.org/10.1109/csit52700.2021.9648584>

217. Krysovaty, A., Lipyanina–Goncharenko, H., Sachenko, S., & Desyatnyuk, O. (2021). Economic crime detection using support vector machine classification. *Modern Machine Learning Technologies and Data Science Workshop. Proc. 3rd International Workshop (Momlet&ds 2021)*, (2917), 830–840. Джерело: <https://ceur-ws.org/Vol-2917/paper46.pdf>

218. Kurze, M., & Roselius, A. (2011). Smart glasses linking real live and social network's contacts by face recognition. In *The 2nd augmented human international conference*. New York, New York, USA: ACM Press. <https://doi.org/10.1145/1959826.1959857>

219. Kutty, A. A., Wakjira, T. G., Kucukvar, M., Abdella, G. M., & Onat, N. C. (2022). Urban resilience and livability performance of European smart cities: a novel machine learning approach. *Journal of cleaner production*, 378, 134203. <https://doi.org/10.1016/j.jclepro.2022.134203>

220. Kwiatek, C., Sharif, M., Li, S., Haas, C., & Walbridge, S. (2019). Impact of augmented reality and spatial cognition on assembly in construction. *Automation in construction*, 108, 102935. <https://doi.org/10.1016/j.autcon.2019.102935>

221. Lambach, D., Bayer, M., Bethke, F. S., Dressler, M., & Dudouet, V. (2020). Statistical analysis. In *Nonviolent resistance and democratic consolidation* (pp. 41–61). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-39371-7\\_3](https://doi.org/10.1007/978-3-030-39371-7_3)

222. Lampropoulos, G., Keramopoulos, E., & Diamantaras, K. (2020). Enhancing the functionality of augmented reality using deep learning, semantic web and knowledge graphs: a review. *Visual informatics*, 4(1), 32–42. <https://doi.org/10.1016/j.visinf.2020.01.001>
223. Lange, S., Altroch, C.–M., Gossmann, E., Fegert, J. M., & Jud, A. (2022). COVID–19–What price do children pay? An analysis of economic and social policy factors. *International journal of environmental research and public health*, 19(13), 7604. <https://doi.org/10.3390/ijerph19137604>
224. Le, H. H., Nguyen, C.–T., Ngo, T. P., Vinh, P. V., Nguyen, B. T., Huynh, A. T., & Nguyen, H. D. (2023). Intelligent retrieval system on legal information. In *Intelligent information and database systems* (pp. 97–108). Singapore: Springer Nature Singapore. [https://doi.org/10.1007/978–981–99–5834–4\\_8](https://doi.org/10.1007/978–981–99–5834–4_8)
225. Lee, J., Babcock, J., Pham, T. S., Bui, T. H., & Kang, M. (2022). Smart city as a social transition towards inclusive development through technology: a tale of four smart cities. *International journal of urban sciences*, 1–26. <https://doi.org/10.1080/12265934.2022.2074076>
226. Lehmann, E. L., & Romano, J. P. (2022). *Testing statistical hypotheses*. Cham: Springer International Publishing. <https://doi.org/10.1007/978–3–030–70578–7>
227. Lelyk, L., Olikhovskiy, V., Mahas, N., & Olikhovska, M. (2022). An integrated analysis of enterprise economy security. *Decision science letters*, 299–310. <https://doi.org/10.5267/j.dsl.2022.2.003>
228. Lessmann, S., Haupt, J., Coussement, K., & De Bock, K. W. (2019). Targeting customers for profit: an ensemble learning framework to support marketing decision–making. *Information sciences*. <https://doi.org/10.1016/j.ins.2019.05.027>
229. Li, G. (2013). Application of improved k–means clustering algorithm in customer segmentation. *Applied mechanics and materials*, 411–414, 1081–1084. <https://doi.org/10.4028/www.scientific.net/amm.411–414.1081>
230. Li, N., & Chen, Y. (2023). Municipal solid waste classification and real–time detection using deep learning methods. *Urban climate*, 49, 101462. <https://doi.org/10.1016/j.uclim.2023.101462>



231. Li, Q., & Mao, Y. (2014). A review of boosting methods for imbalanced data classification. *Pattern analysis and applications*, 17(4), 679–693. <https://doi.org/10.1007/s10044-014-0392-8>
232. Li, V., Amponis, G., Nebel, J.–C., Argyriou, V., Lagkas, T., & Sarigiannidis, P. (2021). Object recognition for augmented reality applications. *Azerbaijan journal of high performance computing*, 4(1), 15–28. <https://doi.org/10.32010/26166127.2021.4.1.15.28>
233. Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism management*, 59, 57–66. <https://doi.org/10.1016/j.tourman.2016.07.005>
234. Li, Y., Hu, C., Huang, C., & Duan, L. (2017). The concept of smart tourism in the context of tourism information services. *Tourism management*, 58, 293–300. <https://doi.org/10.1016/j.tourman.2016.03.014>
235. Lima, M., & Baudier, P. (2017). Business model canvas acceptance among french entrepreneurship students: principles for enhancing innovation artefacts in business education. *Journal of innovation economics*, 23(2), 159. <https://doi.org/10.3917/jie.pr1.0008>
236. Lin, K., Zhao, Y., Kuo, J.–H., Deng, H., Cui, F., Zhang, Z., Wang, T. (2022). Toward smarter management and recovery of municipal solid waste: a critical review on deep learning approaches. *Journal of cleaner production*, 346, 130943. <https://doi.org/10.1016/j.jclepro.2022.130943>
237. Lin, T., Rivano, H., & Le Mouel, F. (2017). A survey of smart parking solutions. *IEEE transactions on intelligent transportation systems*, 18(12), 3229–3253. <https://doi.org/10.1109/tits.2017.2685143>
238. Lin, Y., Ge, X.–h., Liu, P., Zhang, J., & Jiang, L.–p. (2023). A retrospective analysis of coping competence among community health centers during the COVID–19 pandemic in Shanghai, China: coping strategies for future public health emergency events. *Global health & medicine*. <https://doi.org/10.35772/ghm.2023.01042>
239. Lipianina–Honcharenko K., Sachenko A., Wolff C. and Bodyanskiy Y. (2023). Simulation Model for Determining Quality of Life in Ukrainian Cities During the War. 2023

IEEE European Technology and Engineering Management Summit (E-TEMS), Kaunas, Lithuania, pp. 97–101, <https://doi.org/10.1109/E-TEMS57541.2023.10424587>.

240. Lipianina–Honcharenko, K., Bodyanskiy Y., Sachenko A., Kit I., Podchasova T. and Lendiuk T. (2023). An Intelligent Method of Prediction the Demand for Goods/Services in Crisis Conditions, Proceedings of the 3rd International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2023) 2023 Waterloo, Canada, November 20–22, (3641), 224–232 <https://ceur-ws.org/Vol-3641/short2.pdf>

241. Lipianina–Honcharenko, K., Bodyanskiy, Y., & Sachenko, A. (2023). Intelligent information system of the city's socio-economic infrastructure. *System research and information technologies*, (3), 108–120. <https://doi.org/10.20535/srit.2308-8893.2023.3.08>

242. Lipianina–Honcharenko, K., Bodyanskiy, Y., Kustra, N., & Ivasechko, A. OLTW–TEC: Online Learning with Sliding Windows for Text Classifier Ensembles. *Frontiers in Artificial Intelligence*, 7, 1401126. <https://doi.org/10.3389/frai.2024.1401126>

243. Lipianina–Honcharenko, K., Kit, I., Zahorodnia, D., & Osolinskyi, O. (2023). Optimization Method of Advertising Texts Based on Generative Models. Proceedings of the 3rd International Workshop on Information Technologies: Theoretical and Applied Problems 2023, Vol–3628, 126–135. Джерело: <https://ceur-ws.org/Vol-3628/paper10.pdf>

244. Lipianina–Honcharenko, K., Komar, M., Lendyuk, T., & Gramyak, R. (2021). Method of choosing a competitive product based on the emotional color of the calls. *Herald of khmelnytskyi national university*, 303(6), 86–88. <https://doi.org/10.31891/2307-5732-2021-303-6-86-88>

245. Lipianina–Honcharenko, K., Komar, M., Melnyk, N. & Komarnytsky, R. (2024). Sustainable Information System for Enhancing Virtual Company Resilience Through Machine Learning in Smart City Socio-Economic Scenarios. *ECONOMICS*, 12(2). <https://doi.org/10.2478/eoik-2024-0022>

246. Lipianina–Honcharenko, K., Komar, M., Osolinskyi, O., Shymanskyi, V., Havryliuk, M., & Semaniuk, V. (2023). Intelligent waste-volume management method in the smart city concept. *Smart Cities*, 7(1), 78–98. <https://doi.org/10.3390/smartcities7010004>

247. Lipianina–Honcharenko, K., Lendiuk, T., Sachenko, A., & Wołoszyn, J. (2022). Method of forming the context of advertising and target audience based on associative rules learning. *Proceedings of the Fifth International Workshop on Computer Modeling and Intelligent Systems (CMIS–2022) Zaporizhzhia, Ukraine, May 12, 2022. CEUR Workshop Proceedings*, (3137), 98–107. Джерело: <https://ceur-ws.org/Vol-3137/paper9.pdf>

248. Lipianina–Honcharenko, K., Lendiuk, T., Sachenko, A., Osolinskyi, O., Zahorodnia, D., & Komar, M. (2022). An intelligent method for forming the advertising content of higher education institutions based on semantic analysis. In *Communications in computer and information science* (pp. 169–182). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-14841-5\\_11](https://doi.org/10.1007/978-3-031-14841-5_11)

249. Lipianina–Honcharenko, K., Lukasevych–Krutnyk, I., Butryn–Boka, N., & Sachenko, A. (2022). Intelligent method for identifying the fraudulent online stores. *2021 IEEE 8th international conference on problems of infocommunications, science and technology (PIC s&t)*. <https://doi.org/10.1109/PICST54195.2021.9772195>

250. Lipianina–Honcharenko, K., Maika, N., Sachenko, S., Kopania, L., & Soia, M. A Cyclical Approach to Legal Document Analysis: Leveraging AI for Strategic Policy Evaluation. *CEUR Workshop Proceedings*, 2024, 3736, pp. 201–211 Джерело: <https://ceur-ws.org/Vol-3736/paper15.pdf>

251. Lipianina–Honcharenko, K., Melnychuk, A., Yurkiv, K., Hladiy, G., & Telka, M. (2024). Integrated Approach to the International Aspects of Online Dispute Resolution Formation. *Proceedings of the First International Workshop of Young Scientists on Artificial Intelligence for Sustainable Development Ternopil, Ukraine, May 10–11, 2024*. (pp. 88–98)

252. Lipianina–Honcharenko, K., Sachenko, A., Kulyk, V., Savchyshyn, R., Provozin, O., Shchur, S., & Kurpita, L. (2022). Simulation model structure of business processes for a product based on auralization technology. *Computer systems and information technologies*, (4), 114–120. <https://doi.org/10.31891/csit-2022-4-15>

253. Lipianina–Honcharenko, K., Sachenko, A., Semaniuk, V., Badasian, A., & Kopania, Ł. (2023). Intelligent method for selecting business location in smart city. In *2023*

*IEEE 12th international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS)*. IEEE.  
<https://doi.org/10.1109/idaacs58523.2023.10348823>

254. Lipianina–Honcharenko, K., Soia M, Yurkiv K., Ivasechko A. Evaluation of the Effectiveness of Machine Learning Methods for Detecting Disinformation in Ukrainian Text Data. Proceedings of The Seventh International Workshop on Computer Modeling and Intelligent Systems (CMIS–2024), Zaporizhzhia, Ukraine, May 3, 2024. P. 97–109  
 Джерело: <https://ceur-ws.org/Vol-3702/paper9.pdf>

255. Lipianina–Honcharenko, K., Wolff, C., Chyzhovska, Z., Sachenko, A., Lendiuk, T., & Grodskyi, S. (2022). Intelligent method for forming the consumer basket. In *Communications in computer and information science* (pp. 221–231). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-16302-9\\_17](https://doi.org/10.1007/978-3-031-16302-9_17)

256. Lipianina–Honcharenko, K., Wolff, C., Sachenko, A., Desyatnyuk, O., Sachenko, S., & Kit, I. (2023). Intelligent information system for product promotion in internet market. *Applied sciences*, 13(17), 9585. <https://doi.org/10.3390/app13179585>

257. Lipianina–Honcharenko, K., Wolff, C., Sachenko, A., Kit, I., & Zahorodnia, D. (2023). Intelligent method for classifying the level of anthropogenic disasters. *Big data and cognitive computing*, 7(3), 157. <https://doi.org/10.3390/bdcc7030157>

258. Lipianina–Honcharenko, K., Yarych, V., Ivasechko, A., Filinyuk, A., Yurkiv, K., Lebid, T., & Soia11, M. (2024). Evaluating the Effectiveness of Attention–Gated–CNN–BGRU Models for Historical Manuscript Recognition in Ukraine. Proceedings of the First International Workshop of Young Scientists on Artificial Intelligence for Sustainable Development Ternopil, Ukraine, May 10–11, 2024. (pp. 99–108)

259. Lipyanina, H., Maksymovych, V., Sachenko, A., Lendyuk, T., Fomenko, A., & Kit, I. (2020). Assessing the investment risk of virtual IT company based on machine learning. In *Communications in computer and information science* (pp. 167–187). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-61656-4\\_11](https://doi.org/10.1007/978-3-030-61656-4_11)

260. Lipyanina, H., Sachenko, A., Lendyuk, T., & Nadvynychny, S. (n.d.). Decision tree based targeting model of customer interaction with business page. *Proceedings of the third international workshop on computer modeling and intelligent systems (CMIS–2020)*,

*CEUR workshop proceedings, 2608*, 1001–1012. Джерело: <http://ceur-ws.org/Vol-2608/paper75.pdf>

261. Lipyana, H., Sachenko, O., Lendyuk, T., Sachenko, A., & Vasylykiv, N. (2020). Intelligent method of forming the HR management short-term project. In *Advances in intelligent systems and computing* (pp. 1045–1055). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-63270-0\\_71](https://doi.org/10.1007/978-3-030-63270-0_71)

262. Lipyana, H., Sachenko, S., Lendyuk, T., & Sachenko, A. (n.d.). Targeting model of HEI video marketing based on classification tree. *16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2020, CEUR*, (2732), 487–498. Джерело: <https://ceur-ws.org/Vol-2732/20200487.pdf>

263. Lipyana, H., Sachenko, S., Lendyuk, T., Brych, V., Yatskiv, V., & Osolinskiy, O. (2021). Method of detecting a fictitious company on the machine learning base. In *Advances in computer science for engineering and education IV* (pp. 138–146). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-80472-5\\_12](https://doi.org/10.1007/978-3-030-80472-5_12)

264. Lipyana–Goncharenko, H., Brych, V., Sachenko, S., Lendyuk, T., Bykovyya, P., & Zahorodnia, D. (2021). Method of forming a training sample for segmentation of tender organizers on machine learning basis. *Proceedings of the 5th international conference on computational linguistics and intelligent systems (COLINS 2021)*., 2870, 1843–1852. Джерело: <https://ceur-ws.org/Vol-2870/paper134.pdf>

265. Listiyoko, L., & Purno, M. (2020). Sample identification approach by k-means clustering in thinner retail market segmentation. *Journal of physics: conference series, 1477*, 022004. <https://doi.org/10.1088/1742-6596/1477/2/022004>

266. Liu, L., Zhang, H., Ji, Y., & Jonathan Wu, Q. M. (2019). Toward AI fashion design: An Attribute-GAN model for clothing match. *Neurocomputing*, 341, 156–167. <https://doi.org/10.1016/j.neucom.2019.03.011>

267. Liu, L., Zhang, H., Xu, X., Zhang, Z., & Yan, S. (2020). Collocating clothes with generative adversarial networks cosupervised by categories and attributes: a multidiscriminator framework. *IEEE transactions on neural networks and learning systems*, 31(9), 3540–3554. <https://doi.org/10.1109/tnnls.2019.2944979>

268. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, *234*, 11–26. <https://doi.org/10.1016/j.neucom.2016.12.038>
269. Liu, Z. (2021). Development of advertising art design based on information technology. In *Lecture notes on data engineering and communications technologies* (pp. 3–10). Singapore: Springer Singapore. [https://doi.org/10.1007/978-981-16-5854-9\\_1](https://doi.org/10.1007/978-981-16-5854-9_1)
270. Liu, Z. H., Hammerschmidt, B., McMahon, D., Chang, H., Lu, Y., Spiegel, J., Arora, V. (2020). Native JSON datatype support. *Proceedings of the VLDB Endowment*, *13*(12), 3059–3071. <https://doi.org/10.14778/3415478.3415534>
271. Liu, Z., Chen, H., Sun, X., & Chen, H. (2020). Data-Driven real-time online taxi-hailing demand forecasting based on machine learning method. *Applied sciences*, *10*(19), 6681. <https://doi.org/10.3390/app10196681>
272. Lo, Y. C., & Fang, C.-Y. (2018). Facebook marketing campaign benchmarking for a franchised hotel. *International journal of contemporary hospitality management*, *30*(3), 1705–1723. <https://doi.org/10.1108/ijchm-04-2017-0206>
273. Long, Y., Liu, J., Fang, M., Wang, T., & Jiang, W. (2018). Prediction of employee promotion based on personal basic features and post features. In *The international conference*. New York, New York, USA: ACM Press. <https://doi.org/10.1145/3224207.3224210>
274. Lopik, K., Sinclair, M., Sharpe, R., Conway, P., & West, A. (2020). Developing augmented reality capabilities for industry 4.0 small enterprises: lessons learnt from a content authoring case study. *Computers in industry*, *117*, 103208. <https://doi.org/10.1016/j.compind.2020.103208>
275. Lorbeer, B., Kosareva, A., Deva, B., Softić, D., Ruppel, P., & Küpper, A. (2018). Variations on the clustering algorithm BIRCH. *Big data research*, *11*, 44–53. <https://doi.org/10.1016/j.bdr.2017.09.002>
276. Lunsford, D. L. (2019). An output model for human resource development analytics. *Performance improvement quarterly*, *32*(1), 13–35. <https://doi.org/10.1002/piq.21284>

277. Lv, S., Liu, G., & Bai, X. (2021). Multifeature pool importance fusion based GBDT (MPIF-GBDT) for short-term electricity load prediction. *IOP conference series: earth and environmental science*, 702(1), 012012. <https://doi.org/10.1088/1755-1315/702/1/012012>
278. Lyneis, J. M., & Ford, D. N. (2007). System dynamics applied to project management: a survey, assessment, and directions for future research. *System dynamics review*, 23(2-3), 157-189. <https://doi.org/10.1002/sdr.377>
279. Lytvyn, V., Vysotska, V., Veres, O., Rishnyak, I., & Rishnyak, H. (2017). The risk management modelling in multi project environment. In *2017 12th international scientific and technical conference on computer sciences and information technologies (CSIT)*. IEEE. <https://doi.org/10.1109/stc-csit.2017.8098730>
280. Machine learning applications in supply chain management: A deep learning model using an optimized LSTM network for demand forecasting. (2022). *International journal of intelligent engineering and systems*, 15(2), 464-478. <https://doi.org/10.22266/ijies2022.0430.42>
281. Machine learning. (2022). In *Research methods* (2nd ed., pp. 223-247). WORLD SCIENTIFIC. [https://doi.org/10.1142/9789811256943\\_0017](https://doi.org/10.1142/9789811256943_0017)
282. Madichetty, S., & M, S. (2021). A neural-based approach for detecting the situational information from twitter during disaster. *IEEE transactions on computational social systems*, 8(4), 870-880. <https://doi.org/10.1109/tcss.2021.3064299>
283. Maillo, J., Ramírez, S., Triguero, I., & Herrera, F. (2017). KNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors classifier for big data. *Knowledge-Based systems*, 117, 3-15. <https://doi.org/10.1016/j.knosys.2016.06.012>
284. Malik, M., & Hussain, A. (2018). An analysis of review content and reviewer variables that contribute to review helpfulness. *Information processing & management*, 54(1), 88-104. <https://doi.org/10.1016/j.ipm.2017.09.004>
285. Mansouri, M., & Khansari, N. (2018). A conceptual model for intelligent urban governance: influencing energy behaviour in cognitive cities. In *Designing cognitive cities* (pp. 185-202). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-00317-3\\_8](https://doi.org/10.1007/978-3-030-00317-3_8)

286. Marienko, M. V., Nosenko, Y. H., & Shyshkina, M. P. (2020). *Personalization of learning using adaptive technologies and augmented reality*. [б. б.]. <https://doi.org/10.31812/123456789/4418>
287. Mathew, R. B., Varghese, S., Joy, S. E., & Alex, S. S. (2019). Chatbot for disease prediction and treatment recommendation using machine learning. In *2019 3rd international conference on trends in electronics and informatics (ICOEI)*. IEEE. <https://doi.org/10.1109/icoei.2019.8862707>
288. Meng, F., Wang, N., & Xu, Y. (2019). Interval neutrosophic preference relations and their application in virtual enterprise partner selection. *Journal of ambient intelligence and humanized computing*, 10(12), 5007–5036. <https://doi.org/10.1007/s12652-019-01178-5>
289. Mirtalaie, M. A., & Hussain, O. K. (2020). Sentiment aggregation of targeted features by capturing their dependencies: making sense from customer reviews. *International journal of information management*, 53, 102097. <https://doi.org/10.1016/j.ijinfomgt.2020.102097>
290. Mishra, A., Jat, D. S., & Mishra, D. K. (2021). Machine intelligence for predicting new start-ups success. In *DSMLAI '21': international conference on data science, machine learning and artificial intelligence*. New York, NY, USA: ACM. <https://doi.org/10.1145/3484824.3484919>
291. Mishra, H. R. (2022). Big data security challenges. *International journal of research publication and reviews*, 693–696. <https://doi.org/10.55248/gengpi.2022.3.10.31>
292. Mohite, M. T. H. (2021). Smart City using IOT. *International journal for research in applied science and engineering technology*, 9(4), 382–389. <https://doi.org/10.22214/ijraset.2021.33627>
293. Mrazovic, P., Eravci, B., Larriba-Pey, J. L., Ferhatosmanoglu, H., & Matskin, M. (2017). Understanding and predicting trends in urban freight transport. In *2017 18th IEEE international conference on mobile data management (MDM)*. IEEE. <https://doi.org/10.1109/mdm.2017.26>
294. Mrs. J. Sirisha, V. Lakshmi Prathyusha, P. Naga Anupriya, M. Suma Sri, & P. Naga Hema. (2022). Customer Segmentation using K-Means Clustering. *International*



*journal of advanced research in science, communication and technology*, 170–175.  
<https://doi.org/10.48175/ijarsct-7618>

295. Munusamy, S., & Murugesan, P. (2020). Modified dynamic fuzzy c-means clustering algorithm – Application in dynamic customer segmentation. *Applied intelligence*, 50(6), 1922–1942. <https://doi.org/10.1007/s10489-019-01626-x>

296. Navío-Marco, J., Ruiz-Gómez, L. M., & Sevilla-Sevilla, C. (2018). Progress in information technology and tourism management: 30 years on and 20 years after the internet – Revisiting Buhalis & Law's landmark study about eTourism. *Tourism management*, 69, 460–470. <https://doi.org/10.1016/j.tourman.2018.06.002>

297. Nazer, T. H., Xue, G., Ji, Y., & Liu, H. (2017). Intelligent disaster response via social media analysis A survey. *ACM SIGKDD explorations newsletter*, 19(1), 46–59. <https://doi.org/10.1145/3137597.3137602>

298. Ndukwe, I. G., Daniel, B. K., & Amadi, C. E. (2019). A machine learning grading system using chatbots. In *Lecture notes in computer science* (pp. 365–368). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-23207-8\\_67](https://doi.org/10.1007/978-3-030-23207-8_67)

299. Necula, S.-C., & Strîmbei, C. (2019). People analytics of semantic web human resource résumés for sustainable talent acquisition. *Sustainability*, 11(13), 3520. <https://doi.org/10.3390/su11133520>

300. Nezhyva, M., & Mysiuk, V. (2023). Global economic consequences of the military-political crisis in Ukraine. *Foreign trade: economics, finance, law*, 126(1), 4–12. [https://doi.org/10.31617/3.2023\(126\)01](https://doi.org/10.31617/3.2023(126)01)

301. Nie, Y., Yang, W., Chen, Z., Lu, N., Huang, L., & Huang, H. (2021). Public curb parking demand estimation with POI distribution. *IEEE transactions on intelligent transportation systems*, 1–11. <https://doi.org/10.1109/tits.2020.3046841>

302. Niesen, T., Houy, C., Fettke, P., & Loos, P. (2016). Towards an integrative big data analysis framework for data-driven risk management in industry 4.0. In *2016 49th hawaii international conference on system sciences (HICSS)*. IEEE. <https://doi.org/10.1109/hicss.2016.627>

303. Nowotko, P. M. (2021). III in judicial application of law and the right to a court. *Procedia computer science*, 192, 2220–2228. <https://doi.org/10.1016/j.procs.2021.08.235>
304. Nunes, S. A. S., Ferreira, F. A. F., Govindan, K., & Pereira, L. F. (2021). “Cities go smart!”: a system dynamics–based approach to smart city conceptualization. *Journal of cleaner production*, 313, 127683. <https://doi.org/10.1016/j.jclepro.2021.127683>
305. Nuruzzaman, M., & Hussain, O. K. (2018). A survey on chatbot implementation in customer service industry through deep neural networks. In *2018 IEEE 15th international conference on e-business engineering (ICEBE)*. IEEE. <https://doi.org/10.1109/icebe.2018.00019>
306. Nuruzzaman, M., & Hussain, O. K. (2020). IntelliBot: A Dialogue–based chatbot for the insurance industry. *Knowledge–Based systems*, 196, 105810. <https://doi.org/10.1016/j.knosys.2020.105810>
307. Obelets, T. (2022). Data mining tools for complex socio–economic processes and systems. *System research and information technologies*, (4), 68–78. <https://doi.org/10.20535/srit.2308–8893.2022.4.06>
308. Olowu, K., & Gabasa, P. (2020). Financial crime, ICT & e–governance: libraries role. *Advances in social sciences research journal*, 7(1), 609–612. <https://doi.org/10.14738/assrj.71.7476>
309. Osolinskyi, O., Lipianina–Honcharenko, K., Kochan, V., Sachenko, A., & Zahorodnia, D. (2023). Energy Consumption of Methods for Pattern Recognition using Microcontrollers. *International Journal of Computing*, 22(4), 502–508. <https://doi.org/10.47839/ijc.22.4.3358>
310. Pan, B., Chenguang Wu, D., & Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of hospitality and tourism technology*, 3(3), 196–210. <https://doi.org/10.1108/17579881211264486>
311. Panasiuk, O., Akbar, Z., Şimşek, U., & Fensel, D. (2018). Enabling conversational tourism assistants through schema.org mapping. *Lecture notes in computer science*, 11155, Article 137–141. [https://doi.org/10.1007/978–3–319–98192–5\\_26](https://doi.org/10.1007/978–3–319–98192–5_26)

312. Panou, C., Ragia, L., Dimelli, D., & Mania, K. (2018). An architecture for mobile outdoors augmented reality for cultural heritage. *ISPRS international journal of geo-information*, 7(12), 463. <https://doi.org/10.3390/ijgi7120463>
313. Pantano, E. (2019). The role of smart technologies in decision making: developing, supporting and training smart consumers. *Journal of marketing management*, 35(15–16), 1367–1369. <https://doi.org/10.1080/0267257x.2019.1688927>
314. Pantano, E., & Pizzi, G. (2020). Forecasting artificial intelligence on online customer assistance: evidence from chatbot patents analysis. *Journal of retailing and consumer services*, 55, 102096. <https://doi.org/10.1016/j.jretconser.2020.102096>
315. Pariy, V., Lipianina–Honcharenko, K., Brukhanskyi, R., Sachenko, A., Tkachyk, F., & Lendiuk, D. (2023, November). Intelligent Verbal Interaction Methods with Non–Player Characters in Metaverse Applications. In 2023 IEEE 5th International Conference on Advanced Information and Communication Technologies (AICT) (pp. 67–71). IEEE.
316. Parveen, N., Santhi, M. V. B. T., Ramani Burra, L., Pellakuri, V., & Pellakuri, H. (2021). Women’s e–commerce clothing sentiment analysis by probabilistic model LDA using R–SPARK. *Materials today: proceedings*. <https://doi.org/10.1016/j.matpr.2020.10.064>
317. Pasayat, A. K., Bhowmick, B., & Roy, R. (2020). Factors responsible for the success of a start–up: a meta–analytic approach. *IEEE transactions on engineering management*, 1–11. <https://doi.org/10.1109/tem.2020.3016613>
318. Pathak, A., Srihari, R. K., & Natu, N. (2021). Disinformation: analysis and identification. *Computational and mathematical organization theory*, 27(3), 357–375. <https://doi.org/10.1007/s10588–021–09336–x>
319. Paul, S. K. (2015). Supplier selection for managing supply risks in supply chain: a fuzzy approach. *The international journal of advanced manufacturing technology*, 79(1–4), 657–664. <https://doi.org/10.1007/s00170–015–6867–y>
320. Petersen, S. A., Pourzolfaghar, Z., Alloush, I., Ahlers, D., Krogstie, J., & Helfert, M. (2019). Value–Added services, virtual enterprises and data spaces inspired enterprise architecture for smart cities. In *Collaborative networks and digital transformation*

(pp. 393–402). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-28464-0\\_34](https://doi.org/10.1007/978-3-030-28464-0_34)

321. Pigneur, Y., & Osterwalder, A. (2013). *Business model generation: a handbook for visionaries, game changers, and challengers*. Wiley & Sons, Incorporated, John.

322. Piza, D. M., & Moroz, G. V. (2018). Methods of forming classified training sample for adaptation of weight coefficient of automatic interference compensator. *Radioelectronics and communications systems*, 61(1), 32–37. <https://doi.org/10.3103/s0735272718010041>

323. Piza, D. M., Bugrova, T. I., Lavrentiev, V. N., & Semenov, D. S. (2018). Method of forming classified training sample in case of spacial signal processing under influence of combined interference. *Radioelectronics and communications systems*, 61(7), 325–331. <https://doi.org/10.3103/s0735272718070051>

324. PortalÑas, C., Lerma, J. L., & PÑarez, C. (2009). Photogrammetry and augmented reality for cultural heritage applications. *The photogrammetric record*, 24(128), 316–331. <https://doi.org/10.1111/j.1477-9730.2009.00549.x>

325. Prasad, S. (2022). Basic concepts in research and data analysis. In *Elementary statistical methods* (pp. 1–36). Singapore: Springer Nature Singapore. [https://doi.org/10.1007/978-981-19-0596-4\\_1](https://doi.org/10.1007/978-981-19-0596-4_1)

326. Praveena, M., & Jaiganesh, V. (2017). A literature review on supervised machine learning algorithms and boosting process. *International journal of computer applications*, 169(8), 32–35. <https://doi.org/10.5120/ijca2017914816>

327. Premalatha, M., Viswanathan, V., Suganya, G., Kaviya, M., & Vijaya, A. (2018). Educational data mining and recommender systems survey. *International journal of web portals*, 10(1), 39–53. <https://doi.org/10.4018/ijwp.2018010104>

328. Prozorro. (n.d.). Джерело: <https://bi.prozorro.sale/#/participantsCard>

329. Qiangwei Wang, Boyang Li, & Jinglu Hu. (2009). Feature selection for Human resource selection based on Affinity Propagation and SVM sensitivity analysis. In *2009 world congress on nature & biologically inspired computing (nabic)*. IEEE. <https://doi.org/10.1109/nabic.2009.5393596>

330. Qisman, M., Rosadi, R., & Abdullah, A. S. (2021). Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores). *Journal of physics: conference series*, 1722, 012020. <https://doi.org/10.1088/1742-6596/1722/1/012020>
331. Rafew, S. M., & Rafizul, I. M. (2021). Application of system dynamics model for municipal solid waste management in Khulna city of Bangladesh. *Waste management*, 129, 1–19. <https://doi.org/10.1016/j.wasman.2021.04.059>
332. Rajender Kumar Surana, C. S., Shriya, Gupta, D. B., & Shankar, S. P. (2019). Intelligent chatbot for requirements elicitation and classification. In *2019 4th international conference on recent trends on electronics, information, communication & technology (RTEICT)*. IEEE. <https://doi.org/10.1109/rteict46194.2019.9016907>
333. Ramtohil, A., & Khedo, K. K. (2019). A prototype mobile augmented reality systems for cultural heritage sites. In *Advances in intelligent systems and computing* (pp. 175–185). Singapore: Springer Singapore. [https://doi.org/10.1007/978-981-13-3338-5\\_17](https://doi.org/10.1007/978-981-13-3338-5_17)
334. Ramzan, A., Cisneros-Cabrera, S., Sampaio, P., Mehandjiev, N., & Kazantsev, N. (2020). Digital services for industry 4.0: assessing collaborative technology readiness. In *Information systems* (pp. 609–622). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-44322-1\\_45](https://doi.org/10.1007/978-3-030-44322-1_45)
335. Rani, S., Bhambri, P., Kataria, A., & Khang, A. (2022). Smart city ecosystem. In *AI-Centric smart city ecosystems* (pp. 1–20). Boca Raton: CRC Press. <https://doi.org/10.1201/9781003252542-1>
336. Rathnayaka, S. D., Selvanathan, E. A., & Selvanathan, S. (2022). Modelling the consumption patterns in the Asian countries. *Economic analysis and policy*, 74, 277–296. <https://doi.org/10.1016/j.eap.2022.02.004>
337. Remane, G., Hanelt, A., Tesch, J. F., & Kolbe, L. M. (2017). The business model pattern database – a tool for systematic business model innovation. *International journal of innovation management*, 21(01), 1750004. <https://doi.org/10.1142/s1363919617500049>

338. Results / AURA – Auralisation of acoustic heritage sites using Augmented and Virtual Reality. (n.d.). Джерело: <https://aura-project.eu/en/results/>

339. Rinaldi, A. M., Russo, C., & Tommasino, C. (2022). An approach based on linked open data and augmented reality for cultural heritage content-based information retrieval. In *Computational science and its applications – ICCSA 2022* (pp. 99–112). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-10450-3\\_8](https://doi.org/10.1007/978-3-031-10450-3_8)

340. Rochdane, H., & Echatibi, S. (2022). Applicable smart city strategies for a smart sustainable city to ensure energy efficiency and renewable energy integration: casablanca case study. *International business & economics studies*, 4(3), p107. <https://doi.org/10.22158/ibes.v4n3p107>

341. Ronaghi, M. H., & Ronaghi, M. (2022). A contextualized study of the usage of the augmented reality technology in the tourism industry. *Decision analytics journal*, 100136. <https://doi.org/10.1016/j.dajour.2022.100136>

342. Roy, S. K., Singh, G., Hope, M., Nguyen, B., & Harrigan, P. (2019). The rise of smart consumers: role of smart servicescape and smart consumer experience co-creation. *Journal of marketing management*, 35(15–16), 1480–1513. <https://doi.org/10.1080/0267257x.2019.1680569>

343. Russo, M. (2021). AR in the architecture domain: state of the art. *Applied sciences*, 11(15), 6800. <https://doi.org/10.3390/app11156800>

344. Ryzhanskyi, O., Manziuk, E., Barmak, O., Krak, I., & Bacanin, N. (2024). An Approach to Optimizing CO2 Emissions in Traffic Control via Reinforcement Learning. In *IntellITSIS* (pp. 137–155).

345. Saavedra, S., & Romero, M. (2021). Local incentives and national tax evasion: The response of illegal mining to a tax reform in Colombia. *European economic review*, 138, 103843. <https://doi.org/10.1016/j.euroecorev.2021.103843>

346. Sachenko, A., Lendiuk, T., Lipianina-Honcharenko, K., Dobrowolski, M., Boguta, G., & Bytsyura, L. (2024). Method of Determining the Text Sentiment by Thematic Rubrics. In *COLINS* (3) (pp. 404–414). Джерело: <https://ceur-ws.org/Vol-3688/paper26.pdf>

347. Sachenko, A., Lendiuk, T., Lipianina–Honcharenko, K., Koval, V., Hladiy, G., & Halias, Y. (2024). Evaluation of ensemble machine learning models for movie recommendation systems. Proceedings of the Modern Machine Learning Technologies Workshop (MoMLeT 2024) Lviv, Ukraine, May 31 – June 1, 2024. (pp. 273–286)  
Джерело: <https://ceur-ws.org/Vol-3711/paper17.pdf>

348. Sadigh, B. L., Nikghadam, S., Ozbayoglu, A. M., Unver, H. O., Dogdu, E., & Kilic, S. E. (2017). An ontology–based multi–agent virtual enterprise system (OMAVE). *International journal of computer integrated manufacturing*, 30(10), 1072–1092. <https://doi.org/10.1080/0951192x.2017.1285424>

349. Saeed, T., Kiong Loo, C., & Shahreeza Safiruz Kassim, M. (2022). Artificial intelligence based sentiment analysis for health crisis management in smart cities. *Computers, materials & continua*, 71(1), 143–157. <https://doi.org/10.32604/cmc.2022.021502>

350. Saeed, U., Sarim, M., Usmani, A., & Mukhtar, A. (2015). Application of machine learning algorithms in crime classification and classification rule mining. *Research journal of applied sciences*, 4(3), 106–114. Джерело: <https://www.isca.me/rjrs/archive/v4/i3/15.ISCA-RJRS-2013-1005.pdf>

351. Saha, S., & Chaki, R. (2023). IoT based smart waste management system in aspect of COVID–19. *Journal of open innovation: technology, market, and complexity*, 100048. <https://doi.org/10.1016/j.joitmc.2023.100048>

352. Şahin, C., Rokne, J., & Alhajj, R. (2018). Emergency detection and evacuation planning using social media. In *Lecture notes in social networks* (pp. 149–164). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-78256-0\\_9](https://doi.org/10.1007/978-3-319-78256-0_9)

353. Saito, R., Otake, K., & Namatame, T. (2020). Analysis of fashion market trend using advertising data of shopping information site. In *Lecture notes in computer science* (pp. 389–400). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-49576-3\\_28](https://doi.org/10.1007/978-3-030-49576-3_28)

354. Samara, D., Magnisalis, I., & Peristeras, V. (2020). Artificial intelligence and big data in tourism: a systematic literature review. *Journal of hospitality and tourism technology*, *11*(2), 343–367. <https://doi.org/10.1108/jhtt-12-2018-0118>
355. Sansone, C., & Sperlí, G. (2022). Legal Information Retrieval systems: state-of-the-art and open issues. *Information systems*, *106*, 101967. <https://doi.org/10.1016/j.is.2021.101967>
356. Santos, F. C. C. (2023). Artificial intelligence in automated detection of disinformation: a thematic analysis. *Journalism and media*, *4*(2), 679–687. <https://doi.org/10.3390/journalmedia4020043>
357. Santoso, T., Darmawan, A., Sari, N., Syadza, M. A. F., Himawan, E. C. B., & Rahman, W. A. (2023). Clusterization of agroforestry farmers using k-means cluster algorithm and elbow method. *Jurnal sylvia lestari*, *11*(1), 107–122. <https://doi.org/10.23960/jsl.v11i1.646>
358. Sarantitis, G. A., Papadimitriou, T., & Gogas, P. (2016). A network analysis of the united kingdom's consumer price index. *Computational economics*, *51*(2), 173–193. <https://doi.org/10.1007/s10614-016-9625-9>
359. Schardosin, F. Z., De Rolt, C. R., Batista, A. M. L., Bier, C. A., & Lentez, A. A. (2020). Green virtual enterprise breeding environment: a proposal of web platform model for a circular economy. In *Boosting collaborative networks 4.0* (pp. 71–80). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-62412-5\\_6](https://doi.org/10.1007/978-3-030-62412-5_6)
360. Schiffeler, N., Stehling, V., Haberstroh, M., & Isenhardt, I. (2019). Collaborative augmented reality in engineering education. In *Lecture notes in networks and systems* (pp. 719–732). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-23162-0\\_65](https://doi.org/10.1007/978-3-030-23162-0_65)
361. Schmelzer, R. (2023, January 5). Top trends in big data for 2023 and beyond | techtarget. Джерело: <https://www.techtargget.com/searchdatamanagement/feature/Top-trends-in-big-data-for-2021-and-beyond>
362. Schweidel, D. A., Reisenbichler, M., Reutterer, T., & Zhang, K. (2023). Leveraging AI for content generation: a customer equity perspective. In *Review of marketing*



*research* (pp. 125–145). Emerald Publishing Limited. <https://doi.org/10.1108/s1548-643520230000020006>

363. Selwood, K. E., Antos, M., Bramwell, M., Lee, A., Lynch, M., Magrath, M. J. L., Clarke, R. H. (2021). Emergency conservation interventions during times of crisis: A case study for a threatened bird species in the Australian Black Summer bushfires. *Conservation science and practice*, 4(2). <https://doi.org/10.1111/csp2.606>

364. Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. In *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)*. IEEE. <https://doi.org/10.1109/dsaa49011.2020.00096>

365. Shao, W., Zhang, Y., Guo, B., Qin, K., Chan, J., & Salim, F. D. (2019). Parking availability prediction with long short term memory model. In *Green, pervasive, and cloud computing* (pp. 124–137). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-15093-8\\_9](https://doi.org/10.1007/978-3-030-15093-8_9)

366. Sharma, S., Gamoura, S., Prasad, D., & Aneja, A. (2021). Emerging legal informatics towards legal innovation: current status and future challenges and opportunities. *Legal information management*, 21(3–4), 218–235. <https://doi.org/10.1017/s1472669621000384>

367. Sharma, V. (2020). Machine learning algorithms via detection of fake news. *International journal for research in applied science and engineering technology*, 8(6), 780–784. <https://doi.org/10.22214/ijraset.2020.6125>

368. Shazad, B., Ullah khan, H., Rehman, Z.-u., Farooq, M., Mahmood, A., Mehmood, I., Nam, Y. (2019). Finding temporal influential users in social media using association rule learning. *Intelligent automation and soft computing*, 1–1. <https://doi.org/10.31209/2019.100000130>

369. Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: anthropomorphism and adoption. *Journal of business research*, 115, 14–24. <https://doi.org/10.1016/j.jbusres.2020.04.030>

370. Sheikh, S. A., Tiwari, V., & Singhal, S. (2019). Generative model chatbot for human resource using deep learning. In *2019 international conference on data science and engineering (ICDSE)*. IEEE. <https://doi.org/10.1109/icdse47409.2019.8971795>

371. Shiralkar, K., Bongale, A., Kumar, S., & Bongale, A. M. (2023). An intelligent method for supply chain finance selection using supplier segmentation: a payment risk portfolio approach. *Cleaner logistics and supply chain*, 100115. <https://doi.org/10.1016/j.clscn.2023.100115>
372. Shupta, A., Barmak, O., Wierzbicki, A., & Skrypnyk, T. (2023). An Adaptive Approach to Detecting Fake News Based on Generalized Text Features. In COLINS (1) (pp. 300–310).
373. Singapore waste management. (n.d.). Джерело: <https://www.kaggle.com/datasets/kingabzpro/singapore-waste-management>
374. Singh, M. K., Sárközy, H., Singh, S. K., & Zéman, Z. (2022). Impact of Ukraine–Russia war on global trade and development: an empirical study. *Acta academiae beregsasiensis. economics*, (1), 80–92. <https://doi.org/10.58423/2786-6742/2022-1-80-92>
375. Sisodia, D., & Sisodia, D. S. (2021). Quad division prototype selection–based k–nearest neighbor classifier for click fraud detection from highly skewed user click dataset. *Engineering science and technology, an international journal*. <https://doi.org/10.1016/j.jestch.2021.05.015>
376. Smart city transformation strategies. (2022). In *Advances in electronic government, digital divide, and regional development* (pp. 168–179). IGI Global. <https://doi.org/10.4018/978-1-6684-3509-0.ch007>
377. Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & education*, 151, 103862. <https://doi.org/10.1016/j.compedu.2020.103862>
378. Song, H., Liang, H., Li, H., Dai, Z., & Yun, X. (2019). Vision–based vehicle detection and counting system using deep learning in highway scenes. *European transport research review*, 11(1). <https://doi.org/10.1186/s12544-019-0390-4>
379. Starke, P., Kaasch, A., & Van Hooren, F. (2014). Political parties and social policy responses to global economic crises: constrained partisanship in mature welfare states. *Journal of social policy*, 43(2), 225–246. <https://doi.org/10.1017/s0047279413000986>

380. Ștefan, L., Gheorghiu, D., Hodea, M., & Moțăianu, M. (2023). The history of furniture objects: an intelligent augmented reality application. In *Augmented reality and artificial intelligence* (pp. 171–190). Cham: Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-27166-3\\_10](https://doi.org/10.1007/978-3-031-27166-3_10)

381. Stolfi, D. H., Alba, E., & Yao, X. (2017). Predicting car park occupancy rates in smart cities. In *Smart cities* (pp. 107–117). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-59513-9\\_11](https://doi.org/10.1007/978-3-319-59513-9_11)

382. Stolfi, D. H., Alba, E., & Yao, X. (2019). Can I park in the city center? Predicting car park occupancy rates in smart cities. *Journal of urban technology*, 1–15. <https://doi.org/10.1080/10630732.2019.1586223>

383. Structured arrays – numpy v1.26 manual. (n.d.). Джерело: <https://numpy.org/doc/stable/user/basics.rec.html>

384. Structured data: examples, sources, and how it works. (n.d.). Джерело: <https://www.datamation.com/big-data/structured-data/>

385. Su, Y. M. (2014). The research of urban development system model based on system dynamics. *Advanced materials research*, 926–930, 4394–4397. <https://doi.org/10.4028/www.scientific.net/amr.926-930.4394>

386. Sukarsa, I. M., Pandika Pinata, N. N., Dwi Rusjayanthi, N. K., & Wisswani, N. W. (2021). Estimation of gourami supplies using gradient boosting decision tree method of xgboost. *TEM journal*, 144–151. <https://doi.org/10.18421/tem101-17>

387. Sun, G.-L., Cheng, Z.-Q., Wu, X., & Peng, Q. (2017). Personalized clothing recommendation combining user social circle and fashion style consistency. *Multimedia tools and applications*, 77(14), 17731–17754. <https://doi.org/10.1007/s11042-017-5245-1>

388. Sun, Q., Gao, X., Wang, Z., Liu, S., Guo, S., & Li, Y. (2019). Quantifying the risk of price fluctuations based on weighted Granger causality networks of consumer price indices: evidence from G7 countries. *Journal of economic interaction and coordination*, 15(4), 821–844. <https://doi.org/10.1007/s11403-019-00273-2>

389. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *2016 IEEE conference on*

*computer vision and pattern recognition (CVPR)*. IEEE. <https://doi.org/10.1109/cvpr.2016.308>

390. Tahir, H., Shahbaz Khan, M., & Owais Tariq, M. (2021). Performance analysis and comparison of faster R–CNN, mask R–CNN and resnet50 for the detection and counting of vehicles. In *2021 international conference on computing, communication, and intelligent systems (ICCCIS)*. IEEE. <https://doi.org/10.1109/icccis51004.2021.9397079>

391. Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., & Asadpour, M. (2020). Boosting methods for multi–class imbalanced data classification: an experimental review. *Journal of big data*, 7(1). <https://doi.org/10.1186/s40537-020-00349-y>

392. Tao, T., Yan, K., & Yang, S. (2019). Classification of mutual fund investment types with advanced machine learning models. In *2019 IEEE international conference on big data, cloud computing, data science & engineering (BCD)*. IEEE. <https://doi.org/10.1109/bcd.2019.8885073>

393. Tao, Z. (2019). Research on the degree of coupling between the urban public infrastructure system and the urban economic, social, and environmental system: A case study in Beijing, China. *Mathematical Problems in Engineering*, 2019(1), 8206902.

394. Ten trends in artificial intelligence law. (2020). In *Artificial intelligence* (pp. 225–239). Singapore: Springer Singapore. [https://doi.org/10.1007/978-981-15-6548-9\\_22](https://doi.org/10.1007/978-981-15-6548-9_22)

395. Tereshchenko, T., Khytra, O., Hnatchuk, Y., Hnatchuk, A., & El Bouhissi, H. (2024). Decision support system for assessing the economic development potential of a territorial community. In *IntelITSIS* (pp. 100–117).

396. The future of data engineering: trends and predictions. (n.d.). Джерело: <https://iabac.org/blog/the-future-of-data-engineering-trends-and-predictions>

397. Thejas, G. S., Soni, J., Boroojeni, K. G., Iyengar, S. S., Srivastava, K., Badrinath, P., Upadhyay, H. (2019). A multi–time–scale time series analysis for click fraud forecasting using binary labeled imbalanced dataset. In *2019 4th international conference on computational systems and information technology for sustainable solution (CSITSS)*. IEEE. <https://doi.org/10.1109/csitss47250.2019.9031036>

398. Tokuda, E. K., Comin, C. H., & Costa, L. d. F. (2021). Revisiting agglomerative clustering. *Physica A: statistical mechanics and its applications*, 126433. <https://doi.org/10.1016/j.physa.2021.126433>
399. Topsakal, O., & Akinci, T. C. (2023). Creating large language model applications utilizing langchain: A primer on developing LLM apps fast. *International conference on applied engineering and natural sciences*, 1(1), 1050–1056. <https://doi.org/10.59287/icaens.1127>
400. Trofymchuk, O., Stenin, A., Soldatova, M., & Drozdovich, I. (2022). Intelligent decision support systems in the development of megalopolis infrastructure. *System research and information technologies*, (2), 61–74. <https://doi.org/10.20535/srit.2308–8893.2022.2.04>
401. Tyler, N. (2023). Smart city 2023. *New electronics*, 56(06), 14–16. [https://doi.org/10.12968/s0047–9624\(23\)60769–2](https://doi.org/10.12968/s0047–9624(23)60769–2)
402. Ukrainian news. (n.d.). Kaggle: Your Machine Learning and Data Science Community. Джерело: <https://www.kaggle.com/datasets/zepopo/ukrainian-fake-and-true-news>
403. Vafoopoulos, M. (2011). The web economy: goods, users, models, and policies. *Foundations and trends® in web science*, 3(1–2), 1–136. <https://doi.org/10.1561/18000000015>
404. Vanaja, S., & Belwal, M. (2018). Aspect–Level sentiment analysis on e-commerce data. In *2018 international conference on inventive research in computing applications (ICIRCA)*. IEEE. <https://doi.org/10.1109/icirca.2018.8597286>
405. Wan, J., Ling, C., & Liu, M. (2021). Research on success factors of internet product project management with grounded theory. *International journal of information systems and change management*, 12(3), 259. <https://doi.org/10.1504/ijiscm.2021.120330>
406. Wang, D., Zhang, Y., & Zhao, Y. (2017). LightGBM. In *The 2017 international conference*. New York, New York, USA: ACM Press. <https://doi.org/10.1145/3155077.3155079>
407. Wang, F., Li, K., Duić, N., Mi, Z., Hodge, B.–M., Shafie–khah, M., & Catalão, J. P. S. (2018). Association rule mining based quantitative analysis approach of household

characteristics impacts on residential electricity consumption patterns. *Energy conversion and management*, 171, 839–854. <https://doi.org/10.1016/j.enconman.2018.06.017>

408. Wang, G., Xiao, C., Qi, Z., Meng, F., & Liang, X. (2021). Development tendency analysis for the water resource carrying capacity based on system dynamics model and the improved fuzzy comprehensive evaluation method in the Changchun city, China. *Ecological indicators*, 122, 107232. <https://doi.org/10.1016/j.ecolind.2020.107232>

409. Wang, H., & Hong, M. (2019). Online ad effectiveness evaluation with a two-stage method using a Gaussian filter and decision tree approach. *Electronic commerce research and applications*, 35, 100852. <https://doi.org/10.1016/j.elerap.2019.100852>

410. Wang, J., Nguyen, D. Q., Bonkalo, T., & Grebennikov, O. (2021). Smart governance of urban data. *E3S web of conferences*, 301, 05005. <https://doi.org/10.1051/e3sconf/202130105005>

411. Wang, N., Chen, J., & Wang, H. (2023). Resilient supply chain optimization considering alternative supplier selection and temporary distribution center location. *Mathematics*, 11(18), 3955. <https://doi.org/10.3390/math11183955>

412. Wang, X., Han, Y., Leung, V. C. M., Niyato, D., Yan, X., & Chen, X. (2020). Artificial intelligence applications on edge. In *Edge III* (pp. 51–63). Singapore: Springer Singapore. [https://doi.org/10.1007/978-981-15-6186-3\\_4](https://doi.org/10.1007/978-981-15-6186-3_4)

413. Wang, Y., Song, C., Cheng, C., Wang, H., Wang, X., & Gao, P. (2023). Modelling and evaluating the economy–resource–ecological environment system of a third-polar city using system dynamics and ranked weights–based coupling coordination degree model. *Cities*, 133, 104151. <https://doi.org/10.1016/j.cities.2022.104151>

414. Weber, E. (2019). Digital social security: outline of a concept for the 21st century. Джерело: <https://www.econstor.eu/bitstream/10419/216058/1/hbs-fofoe-wp-138-2019.pdf>

415. What constitutes a machine–learning–driven business model? A. (n.d.). Джерело: <https://ideas.repec.org/p/dar/wpaper/132193.html>

416. What is volume of data? (n.d.). Джерело: <https://www.indicative.com/resource/volume-of-data/>

417. Wu, C., Chen, Y., Dong, Y., Zhou, F., Zhao, Y., & Liang, C. J. (2023). VizOPTICS: Getting insights into OPTICS via interactive visual analysis. *Computers and electrical engineering*, *107*, 108624. <https://doi.org/10.1016/j.compeleceng.2023.108624>
418. Wu, C., Zou, H., & Barnes, D. (2023). A supply risk perspective integrated sustainable supplier selection model in the intuitionistic fuzzy environment. *Soft computing*. <https://doi.org/10.1007/s00500-023-08336-0>
419. Wu, P., Li, X., Shen, S., & He, D. (2020). Social media opinion summarization using emotion cognition and convolutional neural networks. *International journal of information management*, *51*, 101978. <https://doi.org/10.1016/j.ijinfomgt.2019.07.004>
420. Xiao, Z., Sen, L., Yunfei, F., Bin, L., Boyuan, Z., & Bang, L. (2017). Tourism route decision support based on neural net buffer analysis. *Procedia computer science*, *107*, 243–247. <https://doi.org/10.1016/j.procs.2017.03.086>
421. Xie, J., Fang, W., Cai, Y., & Lin, Z. (2021). Comparison question generation based on potential compared attributes extraction. In *Web and big data* (pp. 237–252). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-85899-5\\_18](https://doi.org/10.1007/978-3-030-85899-5_18)
422. Xu, H., Zhou, W., Zhu, J., Huang, X., & Wang, W. (2016). Vehicle counting based on double virtual lines. *Signal, image and video processing*, *11*(5), 905–912. <https://doi.org/10.1007/s11760-016-1038-7>
423. Xu, M., & Liu, S. (2019). Semantic-Enhanced and context-aware hybrid collaborative filtering for event recommendation in event-based social networks. *IEEE access*, *7*, 17493–17502. <https://doi.org/10.1109/access.2019.2895824>
424. Xu, Z., & Song, B. (2006). A machine learning application for human resource data mining problem. In *Advances in knowledge discovery and data mining* (pp. 847–856). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/11731139\\_99](https://doi.org/10.1007/11731139_99)
425. Yadav, D., Singh, R., Vatsa, M., & Noore, A. (2014). Recognizing age-separated face images: humans and machines. *PLoS ONE*, *9*(12), Article e112234. <https://doi.org/10.1371/journal.pone.0112234>
426. Yakovlev, A., Moghadam, H. F., Moharrer, A., Cai, J., Chavoshi, N., Varadarajan, V., Agarwal, N. (2020). Oracle AutoML. *Proceedings of the VLDB Endowment*, *13*(12), 3166–3180. <https://doi.org/10.14778/3415478.3415542>

427. Yang, T., Liang, R., & Huang, L. (2021). Vehicle counting method based on attention mechanism SSD and state detection. *The visual computer*. <https://doi.org/10.1007/s00371-021-02161-y>
428. Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism management*, 46, 386–397. <https://doi.org/10.1016/j.tourman.2014.07.019>
429. Yaning, G., & Qianwen, W. (2021). Analysis of collaborative co-governance path of public crisis emergency management in an all-media environment: –theoretical research based on multi-agent. In *2021 international conference on management science and software engineering (ICMSSE)*. IEEE. <https://doi.org/10.1109/icmsse53595.2021.00057>
430. Yigitcanlar, T., Corchado, J. M., Mehmood, R., Li, R. Y. M., Mossberger, K., & Desouza, K. (2021). Responsible urban innovation with local government artificial intelligence (AI): a conceptual framework and research agenda. *Journal of open innovation: technology, market, and complexity*, 7(1), 71. <https://doi.org/10.3390/joitmc7010071>
431. Yildirim, H. (2021). Testing bubbles formation at real-time commodity prices. *Journal of Public Affairs*, 21(3), e2243.
432. Yongmei, X., & Juli, Z. (2020). Network public opinion crisis management mechanism based on big data technology. In *2020 international conference on robots & intelligent system (ICRIS)*. IEEE. <https://doi.org/10.1109/icris52159.2020.00055>
433. Yu, H., Mihai, I. C., & Srivastava, A. (2021). Study and research on iot and big data analysis for smart city development. *Scalable computing: practice and experience*, 22(2). <https://doi.org/10.12694/scpe.v22i2.1898>
434. Zehtabian, S., Khodadadeh, S., Bölöni, L., & Turgut, D. (2021). Modeling an intelligent controller for predictive caching in AR/VR-enabled home scenarios. *Pervasive and mobile computing*, 71, 101334. <https://doi.org/10.1016/j.pmcj.2021.101334>
435. Zhang, C., & Ma, Y. (Eds.). (2012). *Ensemble machine learning*. Boston, MA: Springer US. <https://doi.org/10.1007/978-1-4419-9326-7>
436. Zhang, C., Dong, H., Geng, Y., Liang, H., & Liu, X. (2022). Machine learning based prediction for China's municipal solid waste under the shared socioeconomic



pathways. *Journal of environmental management*, 312, 114918. <https://doi.org/10.1016/j.jenvman.2022.114918>

437. Zhang, H., Veltri, A., Calvo–Amodio, J., & Haapala, K. R. (2020). Making the business case for sustainable manufacturing in small and medium–sized manufacturing enterprises: a systems decision making approach. *Journal of cleaner production*, 125038. <https://doi.org/10.1016/j.jclepro.2020.125038>

438. Zhang, Q., Zhang, X., Mu, X., Wang, Z., Tian, R., Wang, X., & Liu, X. (2021). Recyclable waste image recognition based on deep learning. *Resources, conservation and recycling*, 171, 105636. <https://doi.org/10.1016/j.resconrec.2021.105636>

439. Zhang, S., Liu, S., Cao, X., Song, Z., & Zhou, J. (2018). Watch fashion shows to tell clothing attributes. *Neurocomputing*, 282, 98–110. <https://doi.org/10.1016/j.neucom.2017.12.027>

440. Zhang, S., Song, Z., Cao, X., Zhang, H., & Zhou, J. (2020). Task–Aware attention model for clothing attribute prediction. *IEEE transactions on circuits and systems for video technology*, 30(4), 1051–1064. <https://doi.org/10.1109/tcsvt.2019.2902268>

441. Zhang, T., Wang, D., Chen, H., Zeng, Z., Guo, W., Miao, C., & Cui, L. (2020). BDANN: bert–based domain adaptation neural network for multi–modal fake news detection. In *2020 international joint conference on neural networks (IJCNN)*. IEEE. <https://doi.org/10.1109/ijcnn48605.2020.9206973>

442. Zhang, X., Liu, X., & Guo, H. (2018). A click fraud detection scheme based on cost sensitive BPNN and ABC in mobile advertising. In *2018 IEEE 4th international conference on computer and communications (ICCC)*. IEEE. <https://doi.org/10.1109/compcomm.2018.8780941>

443. Zhang, Y., Zhang, P., Yuan, C., & Wang, Z. (2020). Texture and shape biased two–stream networks for clothing classification and attribute recognition. In *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. IEEE. <https://doi.org/10.1109/cvpr42600.2020.01355>

444. Zhang, Z., Yang, W., & Wushour, S. (2020). Traffic accident prediction based on LSTM–GBRT model. *Journal of control science and engineering*, 2020, 1–10. <https://doi.org/10.1155/2020/4206919>

445. Zheng, Q., Han, Y., Li, S., Dong, J., Yan, L., & Qin, J. (2009). Computer system integration and e-commerce. In *Introduction to e-commerce* (pp. 336–372). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-49645-8\\_10](https://doi.org/10.1007/978-3-540-49645-8_10)
446. Zheng, X., Qiao, X., Cao, Y., & Lau, R. W. H. (2019). Content-aware generative modeling of graphic design layouts. *ACM transactions on graphics*, 38(4), 1–15. <https://doi.org/10.1145/3306346.3322971>
447. Zheng, Z., Zhong, Y., Wang, J., Ma, A., & Zhang, L. (2021). Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: from natural disasters to man-made disasters. *Remote sensing of environment*, 265, 112636. <https://doi.org/10.1016/j.rse.2021.112636>
448. Zhou, L., Gao, J., Li, D., & Shum, H.-Y. (2020). The design and implementation of xiaoice, an empathetic social chatbot. *Computational linguistics*, 46(1), 53–93. [https://doi.org/10.1162/coli\\_a\\_00368](https://doi.org/10.1162/coli_a_00368)
449. Zhou, S., Ng, S. T., Yang, Y., & Xu, J. F. (2021). Integrating computer vision and traffic modeling for near-real-time signal timing optimization of multiple intersections. *Sustainable cities and society*, 68, 102775. <https://doi.org/10.1016/j.scs.2021.102775>
450. Zhou, Z.-H. (2021). Ensemble learning. In *Machine learning* (pp. 181–210). Singapore: Springer Singapore. [https://doi.org/10.1007/978-981-15-1967-3\\_8](https://doi.org/10.1007/978-981-15-1967-3_8)
451. Барановська І. В., Ковалів І. Р., Куценко В. І., Ольшевський В. Й., Савченко О. В., Стегній О. Г., Тетіор О. Н., “Інфраструктура”, *Енциклопедія Сучасної України*, т. 11, редкол. І. М. Дзюба, А. І. Жуковський, М. Г. Железняк та ін. Київ: Інститут енциклопедичних досліджень НАН України, 2011. Джерело: <https://esu.com.ua/article-12489>
452. Вимоги до структури та змісту XML-схеми архівних електронних документів, Наказ Міністерства юстиції України No. 1886/5 (2014) (Україна). Джерело: <https://zakon.rada.gov.ua/laws/show/z1423-14#Text>
453. Держстат України. (n.d.). Джерело: <https://ukrstat.gov.ua/>
454. Закон України від 21 травня 1997 року № 280/97-ВР «Про місцеве самоврядування в Україні» Джерело: <https://zakon.rada.gov.ua/laws/show/280/97-%D0%92%D0%A0#Text>

455. Комар, М., Ліп'яніна–Гончаренко, Х., Кіт, І., Мадараш, Р., & Юрків, Х. (2023). Інтелектуальний метод виявлення джерел мультилінгвальної дезінформації. *Measuring and Computing Devices in Technological Processes*, (2), 221–230. <https://doi.org/10.31891/2219-9365-2023-74-31>
456. Ліп'яніна–Гончаренко, Х. (2023). Інтелектуальний метод вибору локації для старту бізнесу в ТГ. *Системні технології*, 4(147), 132–140. <https://doi.org/10.34185/1562-9945-4-147-2023-12>
457. Ліп'яніна–Гончаренко, Х. (2024). Методи аналізу соціально–економічних даних територіальних громад для адаптивного управління ресурсами. *Computer Systems and Information Technologies*, (3), 92–97. <https://doi.org/10.31891/csit-2024-3-12>
458. Ліп'яніна–Гончаренко, Х. (2024). Методологія аналізу соціально–економічних даних територіальних громад. *Herald of Khmelnytskyi National University. Technical Sciences*, 337(3(2)), 446–450. <https://doi.org/10.31891/2307-5732-2024-337-3-67>
459. Ліп'яніна–Гончаренко, Х. (2024). Узагальнений Принцип Синтезу Інформаційної Технології Інтелектуального Аналізу Соціально–Економічних Даних Тг. *Measuring And Computing Devices In Technological Processes*, (1), 359–367. <https://doi.org/10.31891/2219-9365-2024-77-48>
460. Ліп'яніна–Гончаренко, Х. В. (2022). Інтелектуальний метод формування людських ресурсів на короткостроковий проект. *Вісник Черкаського державного технологічного університету*, (3), 49–58. <https://doi.org/10.24025/2306-4412.3.2022.259775>
461. Ліп'яніна–Гончаренко, Х. В. (2023). Метод генерування рекламного зображення на основі ключових слів. *Телекомунікаційні та інформаційні технології*, 81(4). <https://doi.org/10.31673/2412-4338.2023.043745>
462. Ліп'яніна–Гончаренко, Х., & Кіт, І. (2023). Метод удосконалення рекламних текстів на основі генеративних моделей. *Вісник Приазовського Державного Технічного Університету. Серія: Технічні науки*, (46), 6–13. <https://doi.org/10.31498/2225-6733.46.2023.288087>

463. Лип'яніна–Гончаренко, Х., & Юрків, Х. (2023). Методи бустингового машинного навчання для нестационарних часових рядів. *Measuring and Computing Devices in Technological Processes*, (3), 19–30. <https://doi.org/10.31891/2219-9365-2023-75-2>

464. Лип'яніна–Гончаренко, Х., Комар, М., Саченко, А., & Лендюк, Т. (2022). Метод формування контексту реклами та цільової аудиторії на основі навчання асоціативних правил. *Вісник Хмельницького національного університету. Серія: Технічні науки*, 313(5), 279–287. <https://doi.org/10.31891/2307-5732-2022-313-5-279-287>

465. Лип'яніна–Гончаренко, Х., Комар, М., Саченко, А., & Лендюк, Т. (2022). Оцінка інвестиційних ризиків віртуальної іт-компанії на основі МН. *Measuring and Computing Devices in Technological Processes*, (3), 45–60. <https://doi.org/10.31891/2219-9365-2022-71-3-6>

466. Лип'яніна–Гончаренко, Х. В. (2023). Метод формування навчальної вибірки для масивів даних на основі МН. *Наукові записки НаУКМА. Комп'ютерні науки*. – том 6, 2023 р 30–35. [HTTPS://DOI.ORG/ 10.18523/2617-3808.2023.6.30-35](https://doi.org/10.18523/2617-3808.2023.6.30-35)

467. Лип'яніна–Гончаренко, Х. В. (2023). Метод генерування рекламного зображення на основі відео потоку. *Науковий вісник Ужгородського університету. Серія: Математика і інформатика*, 43(2), 130–135. [https://doi.org/10.24144/2616-7700.2023.43\(2\).130-135](https://doi.org/10.24144/2616-7700.2023.43(2).130-135)

468. Лип'яніна–Гончаренко, Х. В., Комар, М. П., Саченко, А. О., & Лендюк, Т. В. (2022). Метод виявлення фіктивних підприємств на підставі Гаусового наївного класифікатора Байєса. *Scientific Bulletin of UNFU*, 32(5), 92–96. <https://doi.org/10.36930/40320513>

469. Лип'яніна–Гончаренко, Х., Комар, М., Юрків Х., Лукянчук В. (2023). Концептуальна модель інтелектуальної оцінки наслідків техногенних катастроф. *Herald of Khmelnytskyi National University. Technical Sciences*, 329(6), 230–237. <https://doi.org/10.31891/2307-5732-2023-329-6-230-237>

470. Мельник, А., & Лип'яніна–Гончаренко, Х. В. (2024). Інтерактивний мобільний додаток для підтримки осіб із порушеннями мовлення. *Measuring and*

computing devices in technological processes, (2), 26–32. <https://doi.org/10.31891/2219-9365-2024-78-3>

471. Одяг та взуття в Україні порівняльні ціни на prom.ua. (n.d.). Джерело: <https://prom.ua/ua/>

472. Попова, О. Ю. (2014). Аналіз моделей та методів стимулювання попиту на туристичні послуги в сучасних умовах. *Держава та регіони. Серія : Економіка та підприємництво*, 1, 69–74. Джерело: [http://nbuv.gov.ua/UJRN/drep\\_2015\\_1\\_14](http://nbuv.gov.ua/UJRN/drep_2015_1_14)

473. Про затвердження Методики виявлення ризиків здійснення державно–приватного партнерства, їх оцінки та визначення форми управління ними, Постанова Кабінету Міністрів України No. 232 (2015) (Україна). Джерело: <https://zakon.rada.gov.ua/laws/show/232-2011-п#Text>

474. Твір наукового характеру. «Метод виявлення фіктивних підприємств на основі МН» : Авторське право на твір 111509 Україна / А. І. Крисоватий, Х.В. Ліп'яніна–Гончаренко [та ін.]. – № 111509; опубл. 31.03.2022, Бюл. № 69.

475. Федорова, Ю., Мірющенко, М., & Івченко, В. (2021). Цифрові технології в управлінні персоналом. *Adaptive Management Theory and Practice Economics*, 12(24). [https://doi.org/10.33296/2707-0654-12\(24\)-11](https://doi.org/10.33296/2707-0654-12(24)-11)

ДОДАТОК А. СТРУКТУРНІ СХЕМИ МЕТОДІВ

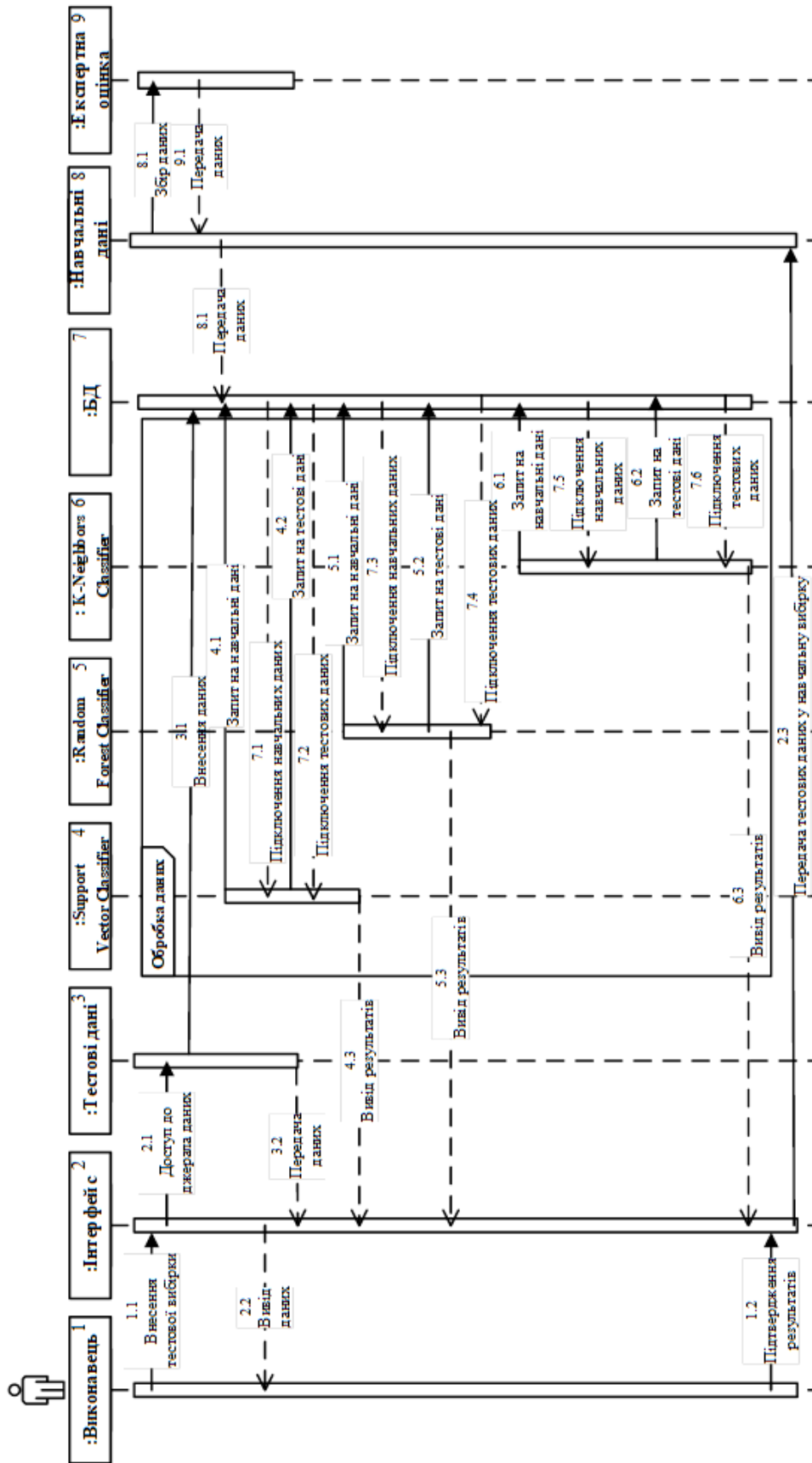


Рис. А.1. Діаграма послідовності процесу аналізу інвестиційних ризиків

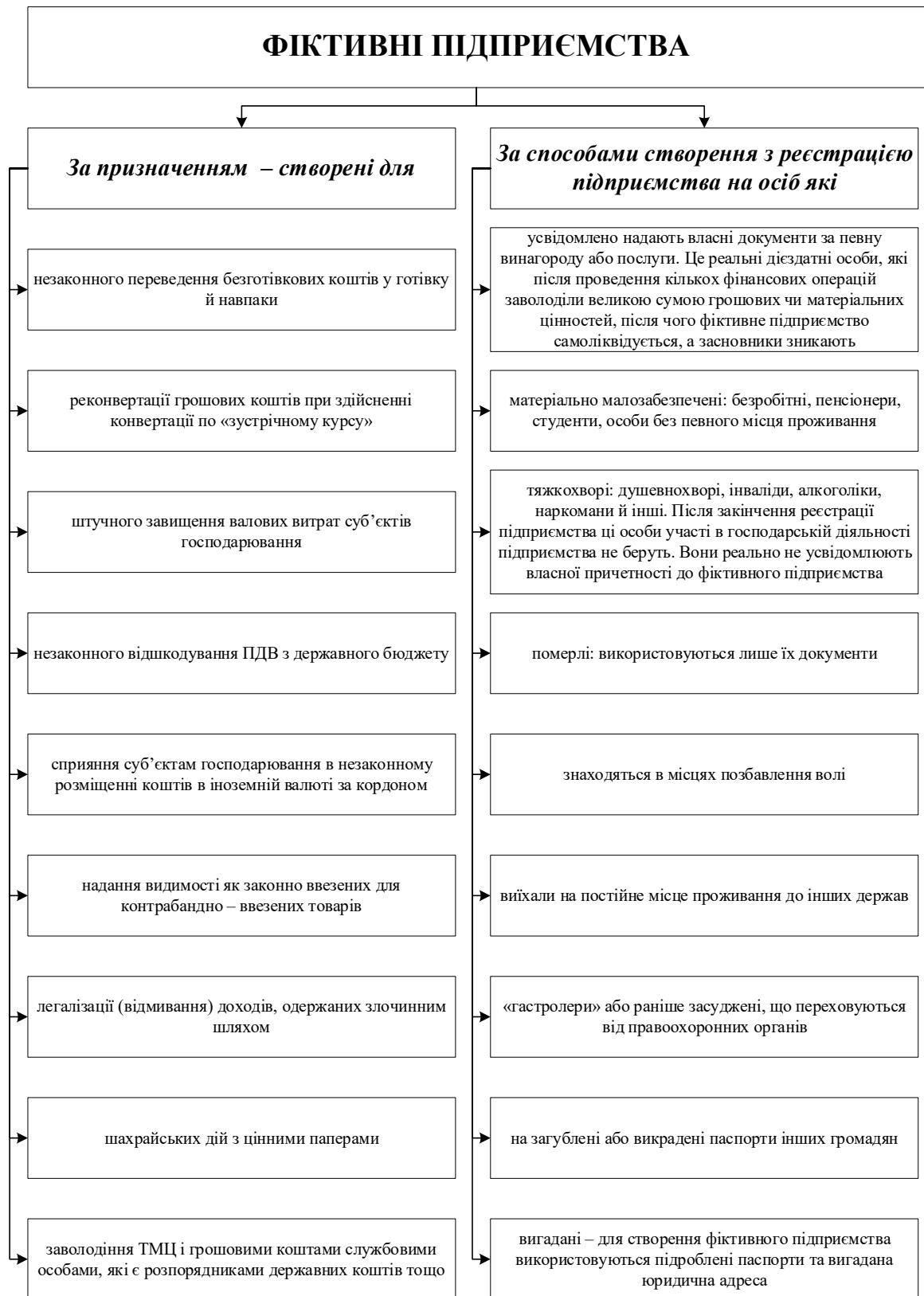


Рис. А.2. Класифікація фіктивних підприємств

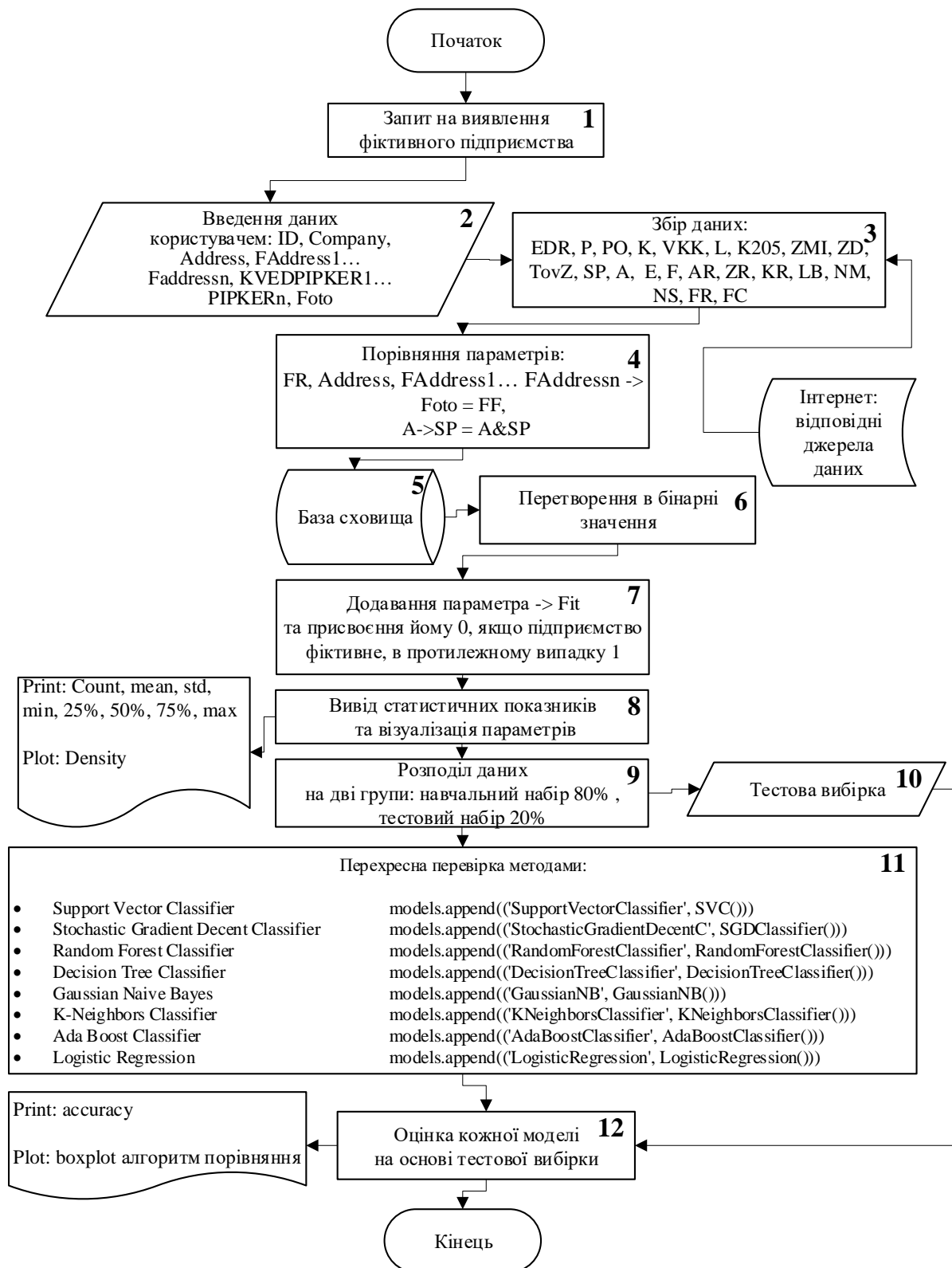


Рис. А.3. Структура методу виявлення фіктивного підприємства на основі МН



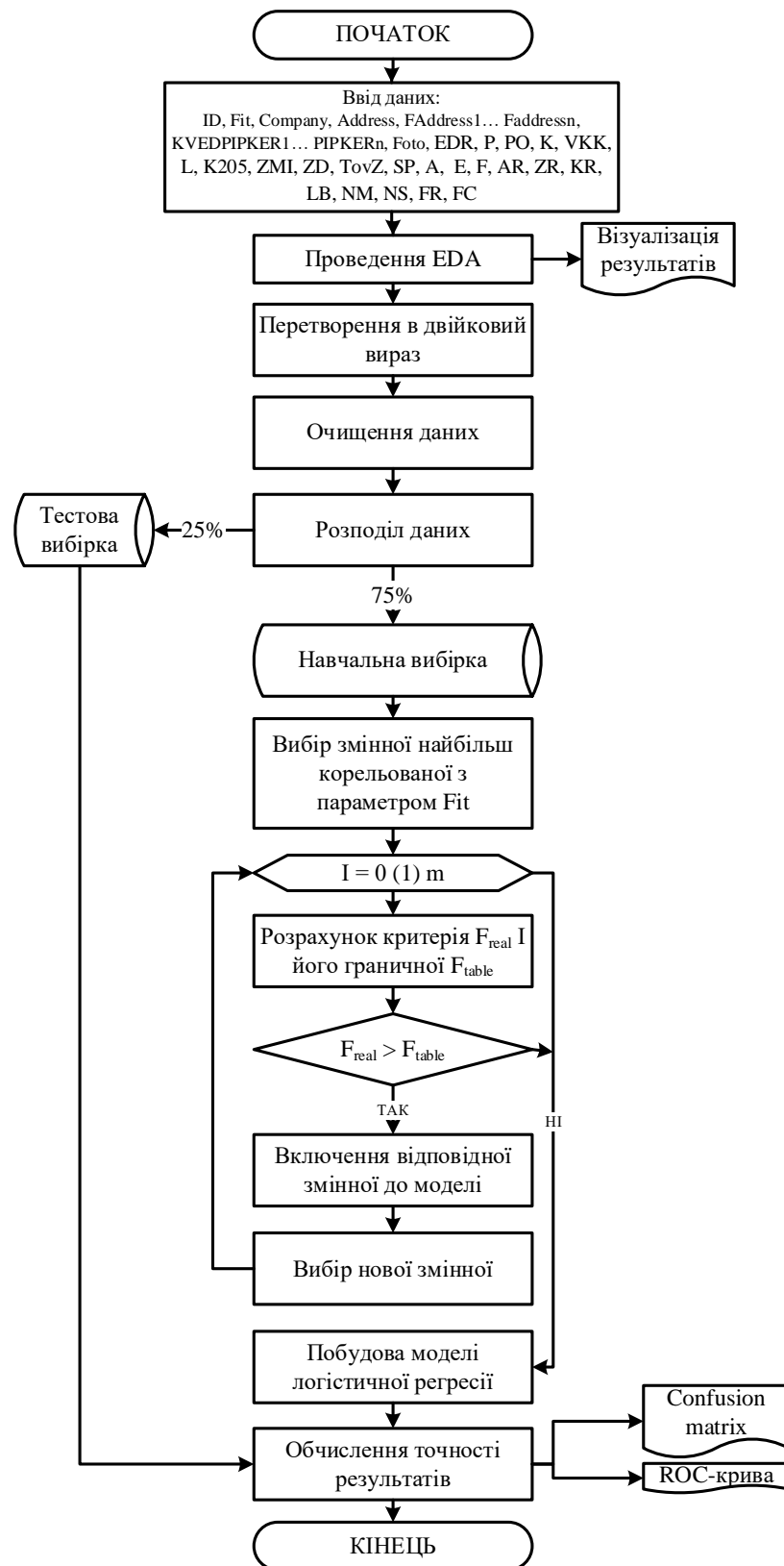


Рис. А.4. Структура методу класифікації фіктивних підприємств на основі MN методом Logistic Regression

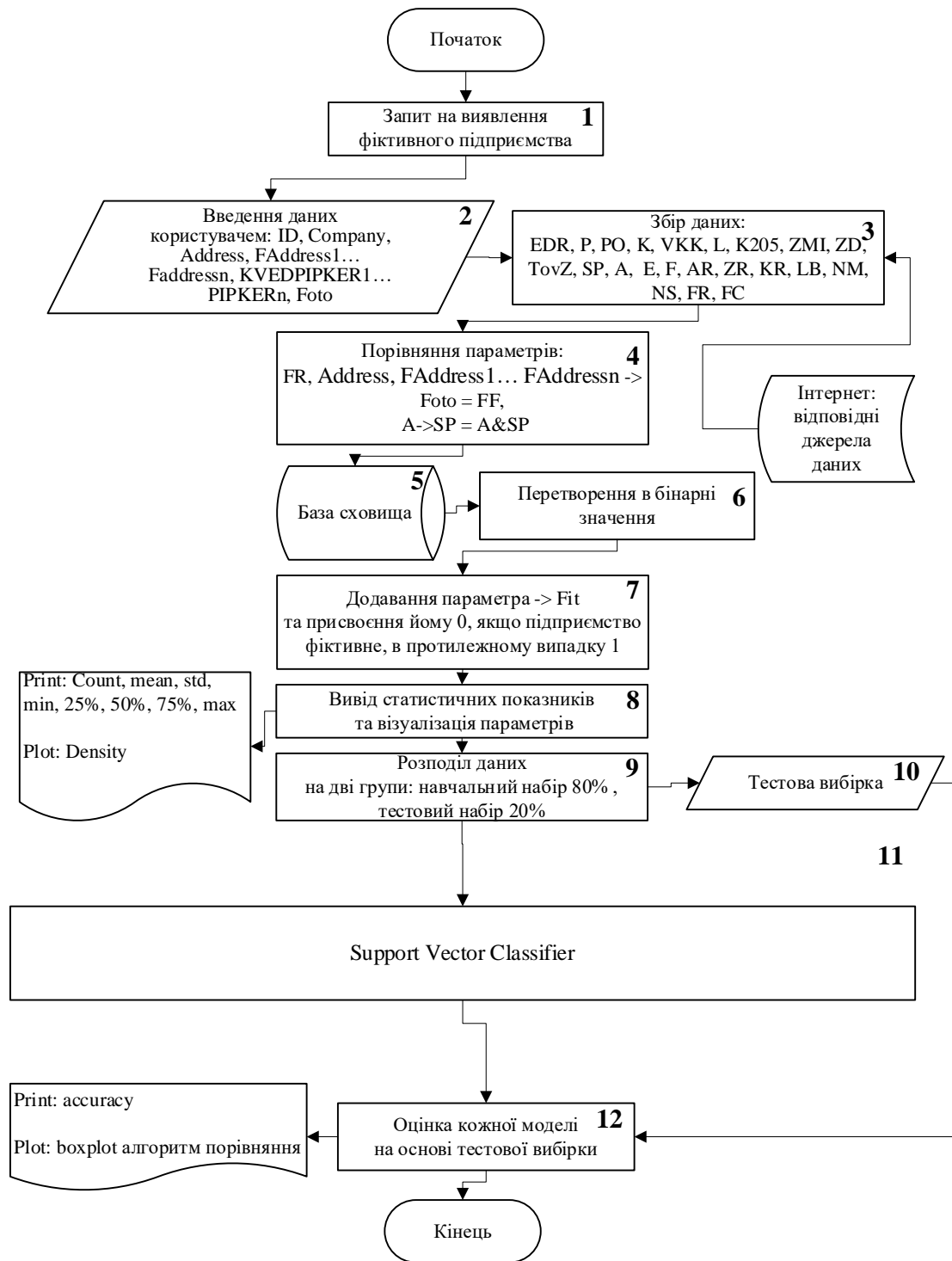


Рис. А.5. Структура методу виявлення фіктивного підприємства на основі Support Vector Machine Classification

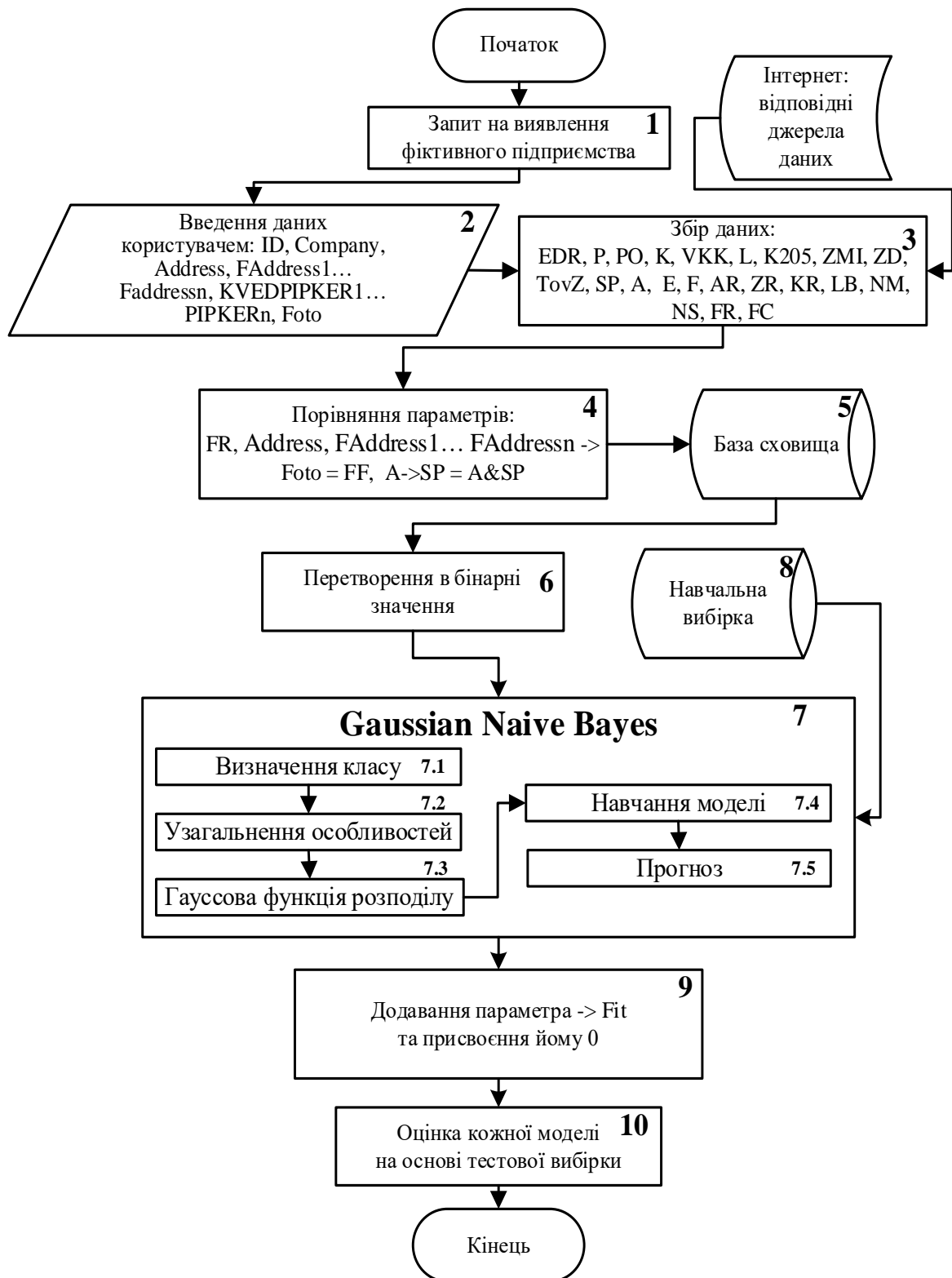


Рис. А.6. Структура методу виявлення фіктивного підприємства на основі Gaussian Naive Bayes

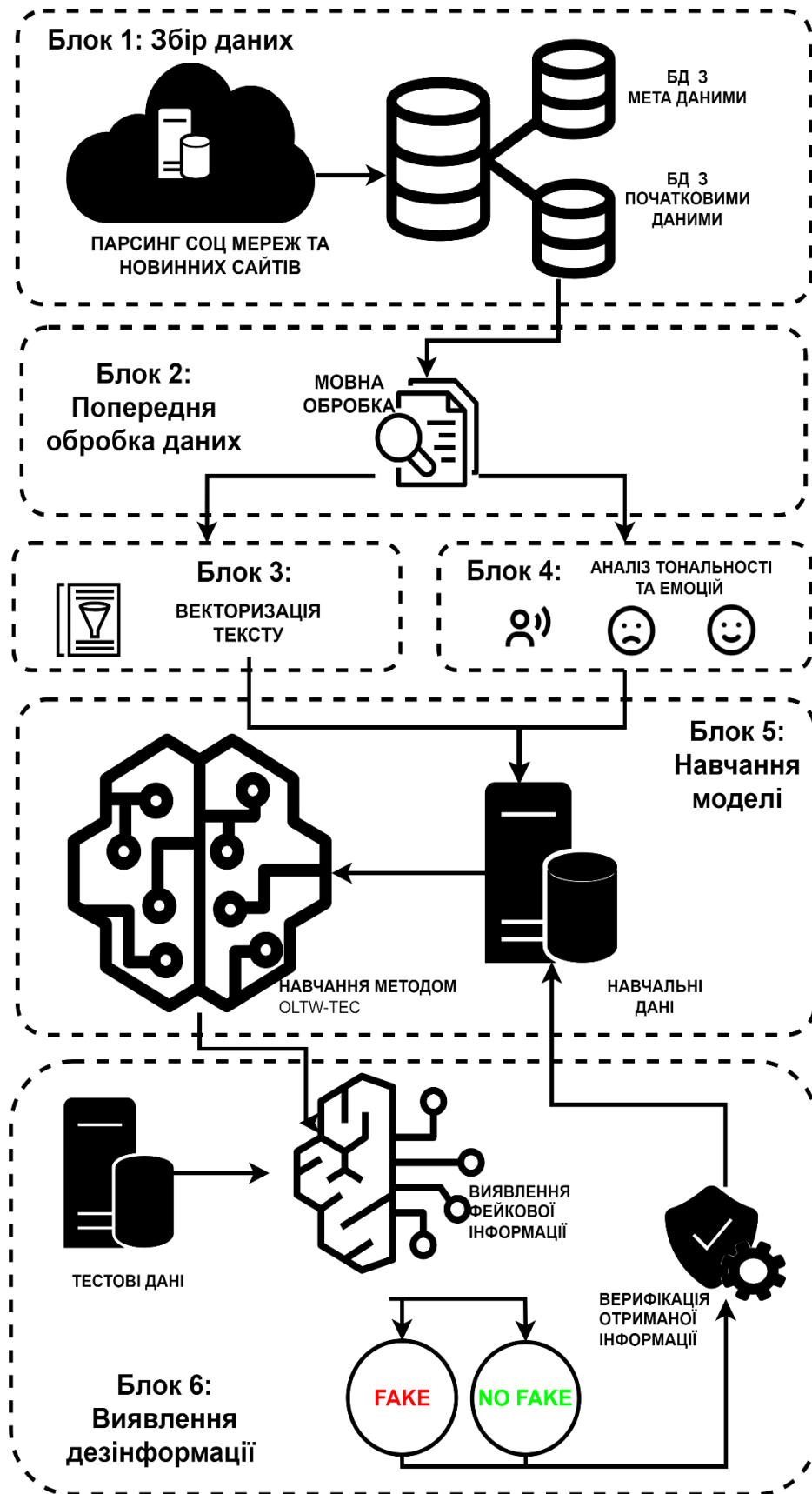


Рис. А.7. Структура інтелектуального методу виявлення дезінформації в онлайн-режимі

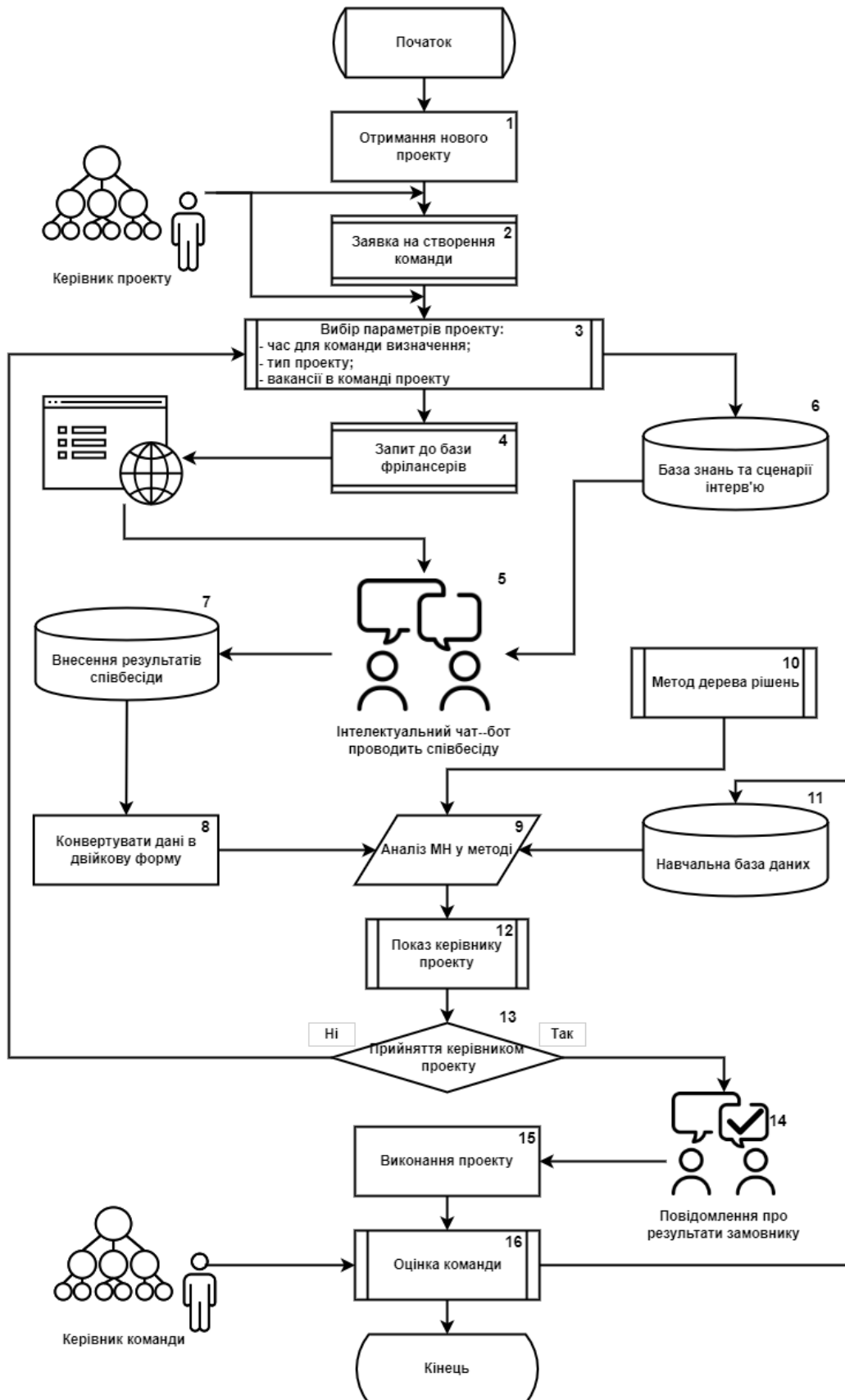


Рис. А.8. Структура методу формування короткострокового HR-проекту на основі МН

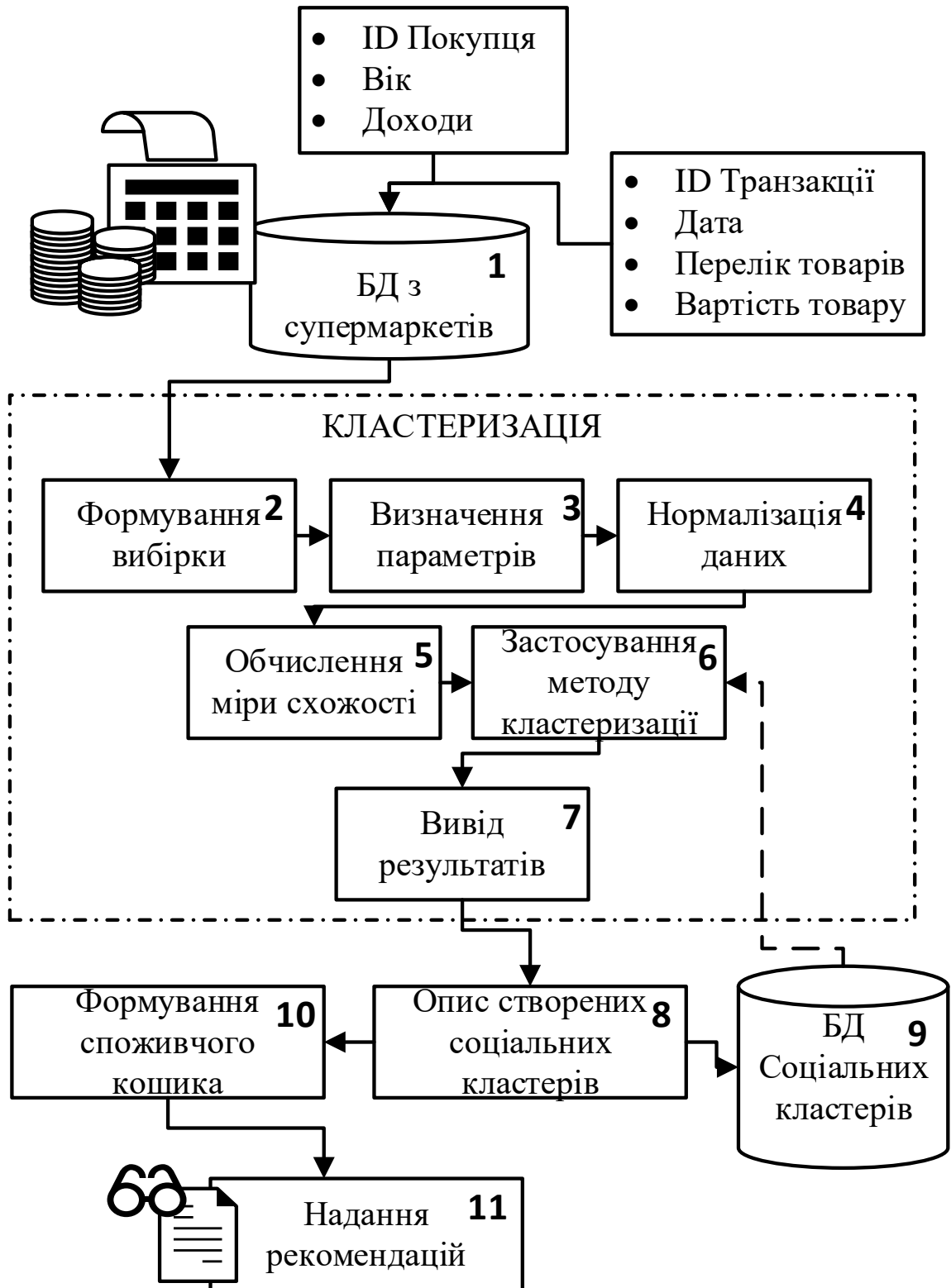


Рис. А.9. Структура інтелектуального методу формування споживчого кошика

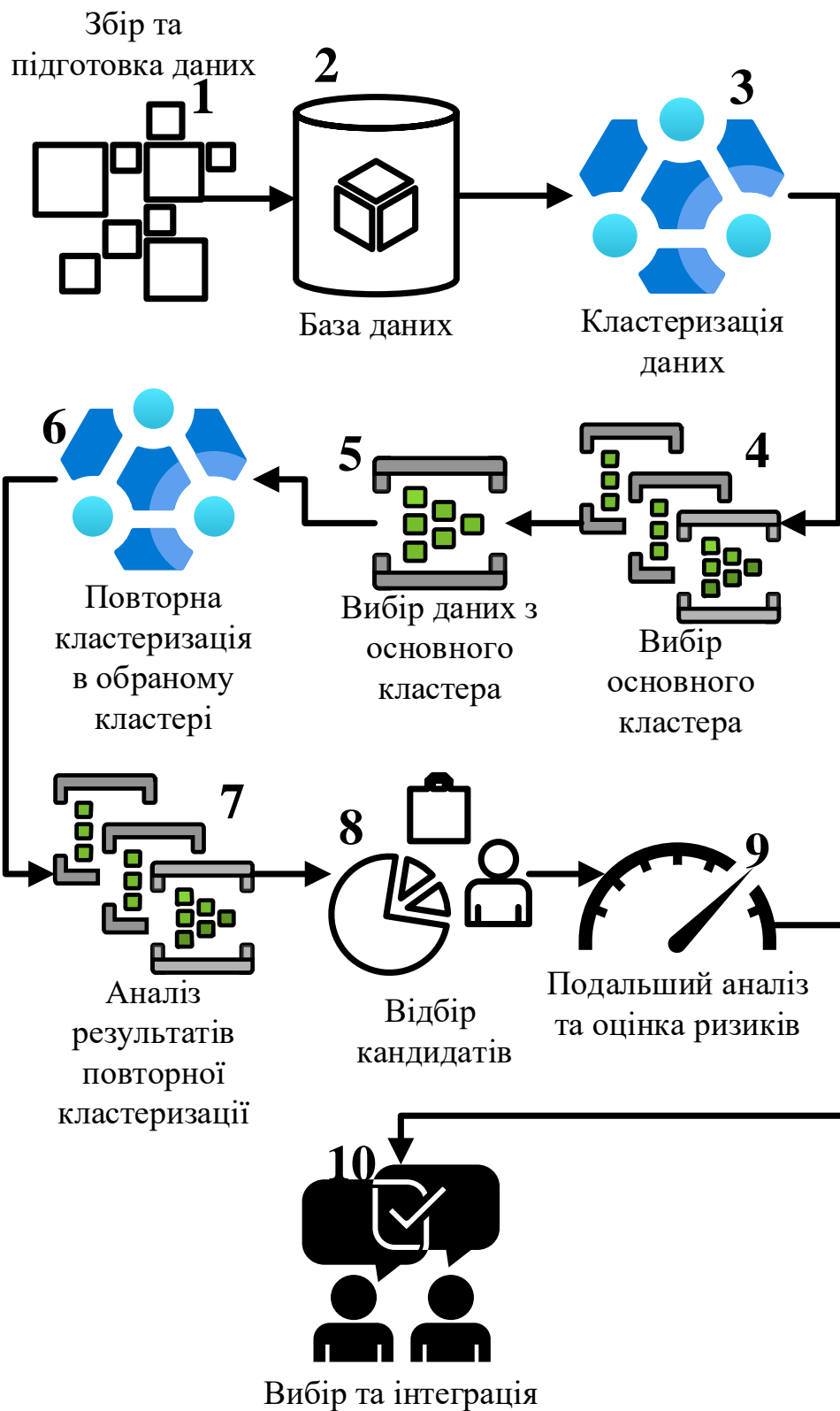


Рис. А.10. Структура інтелектуального методу пошуку альтернативних постачальників

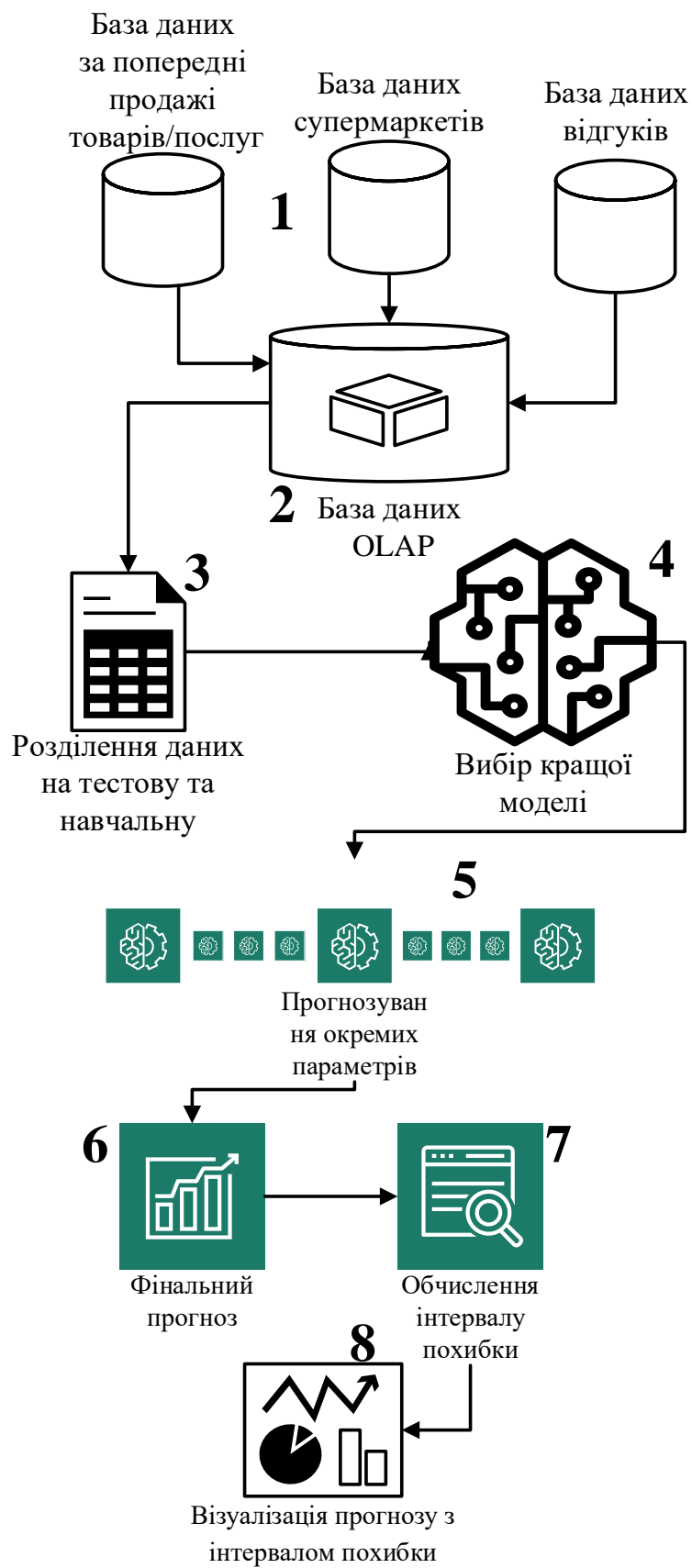


Рис. А.11. Структура інтелектуального методу прогнозування попиту товарів/послуг в умовах кризи



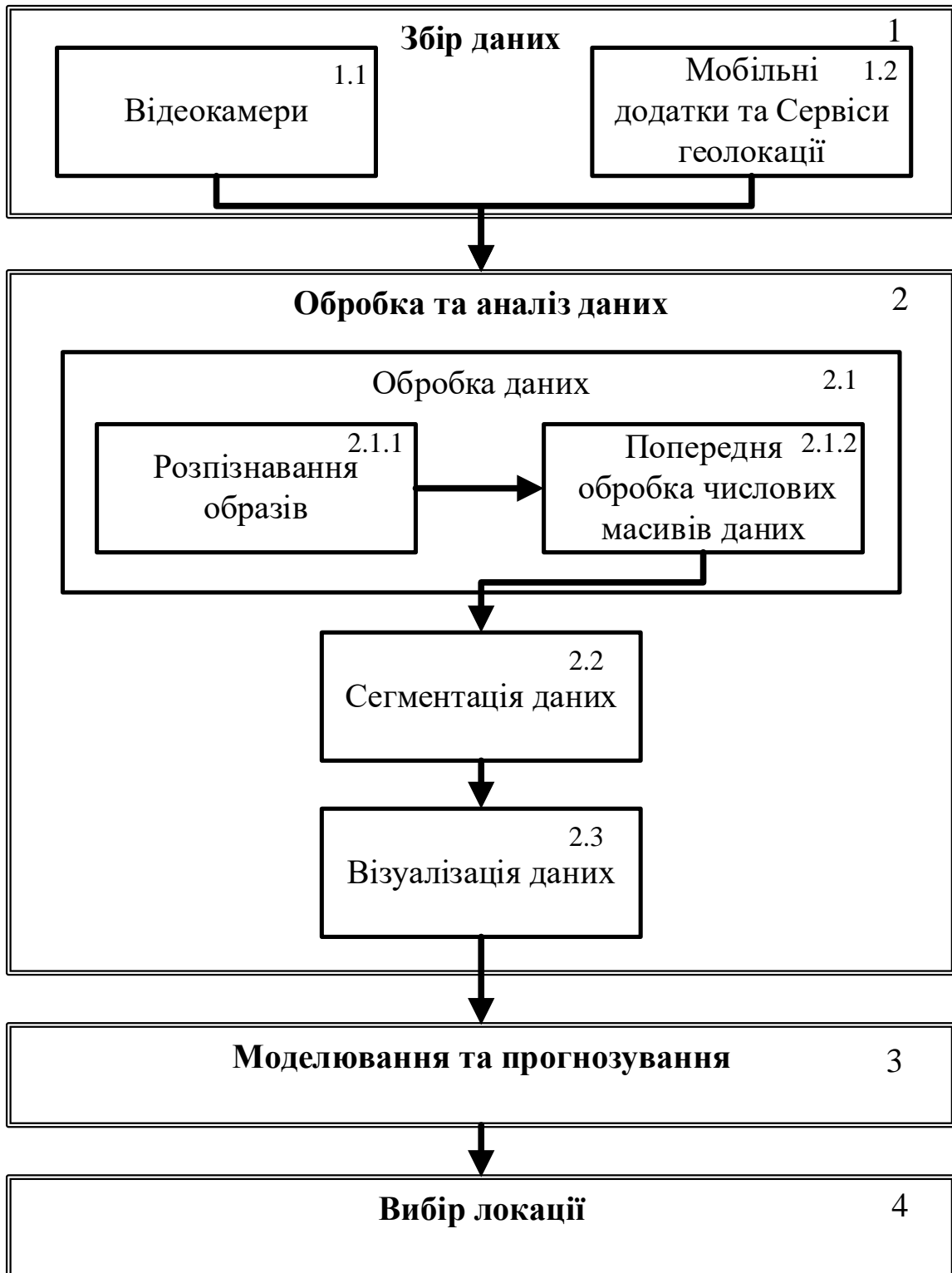


Рис. А.12. Схематичне представлення інтелектуального методу вибору локації для старту бізнесу в ТГ

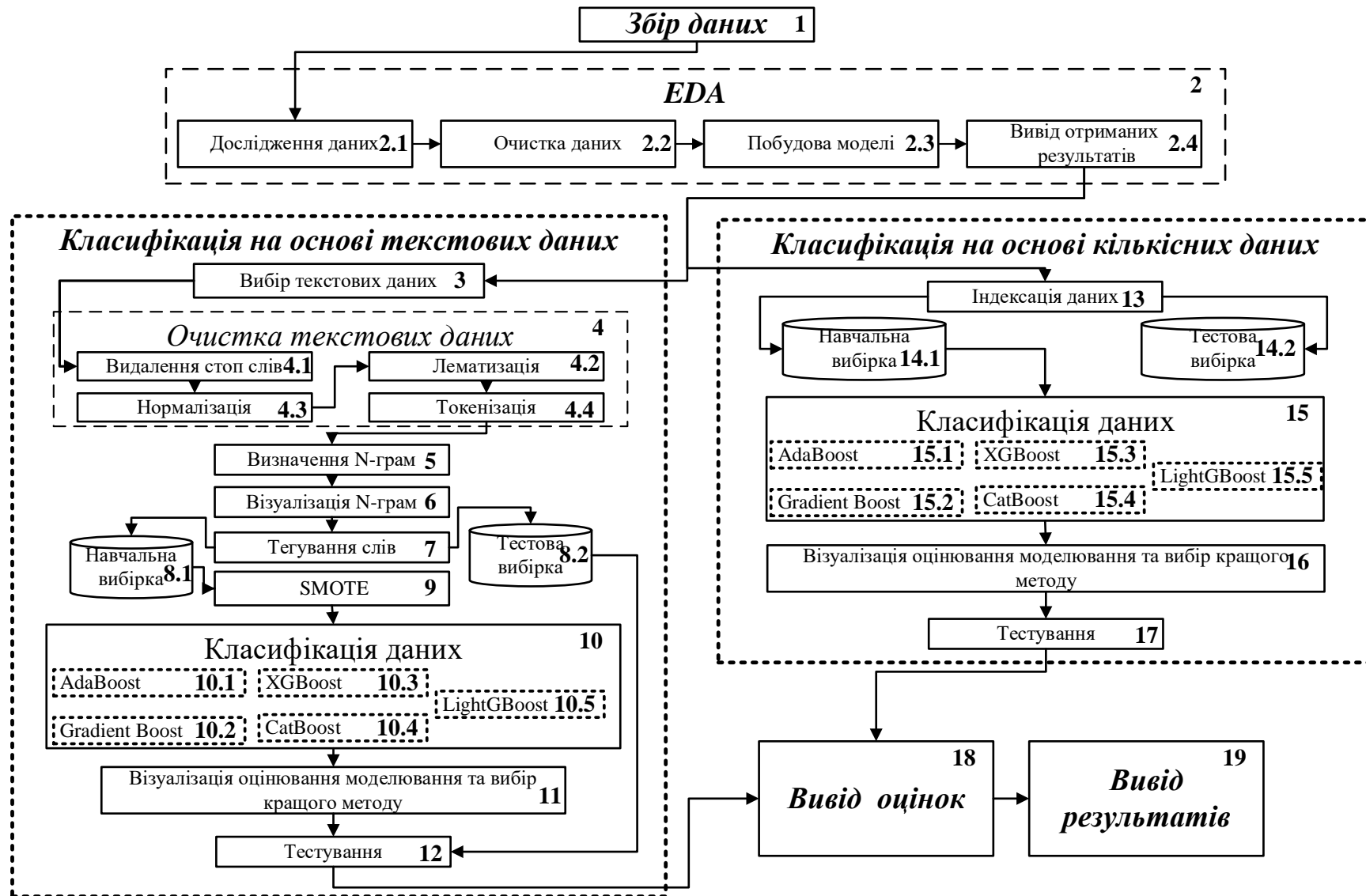


Рис. А.13. Структура інтелектуального методу класифікації техногенних катастроф

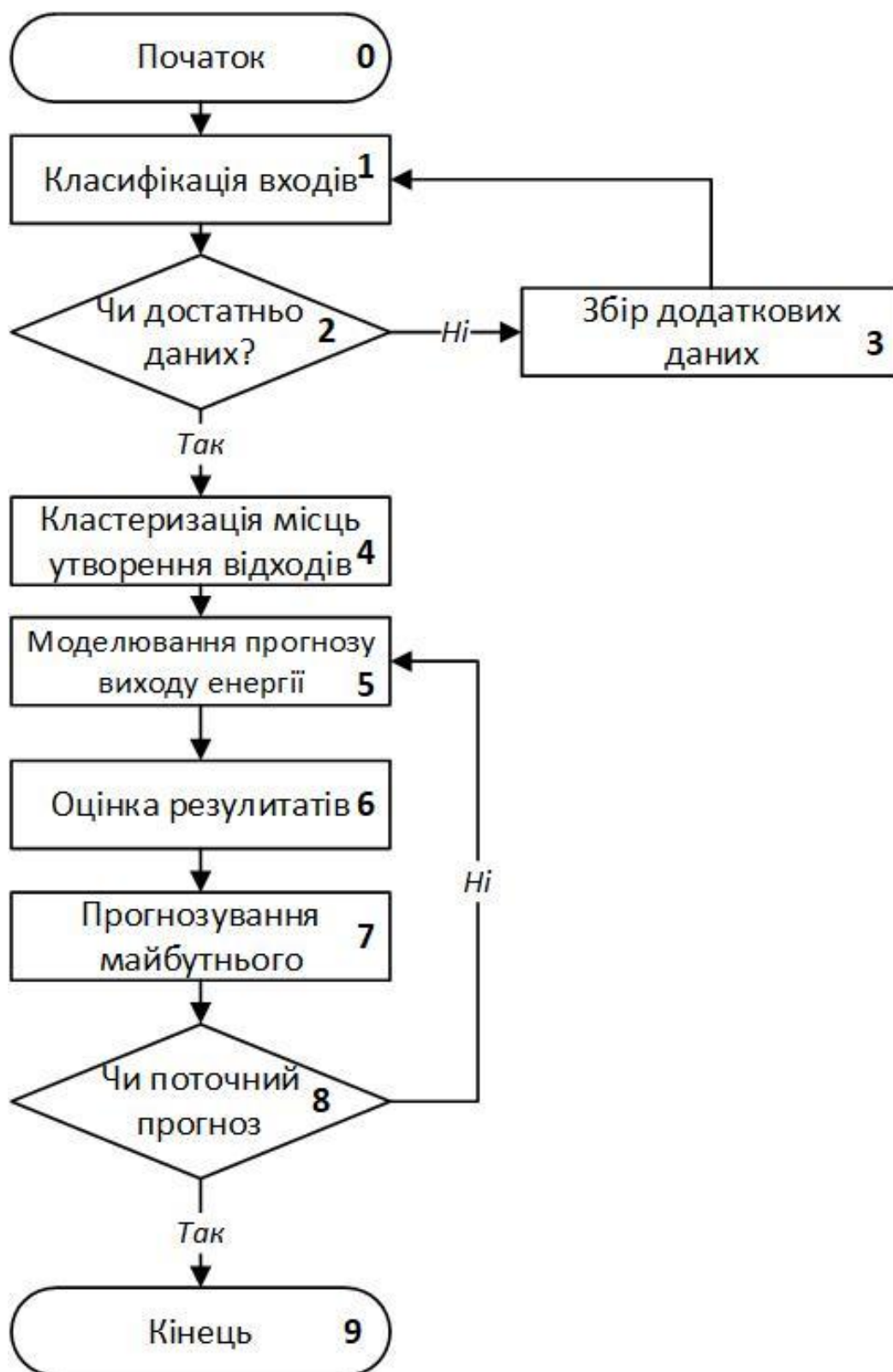


Рис. А.14. Алгоритм інтелектуального управління обсягами відходів міста

## ДОДАТОК Б. ОПИС ВХІДНИХ ПАРАМЕТРІВ ДЛЯ МЕТОДІВ

Таблиця Б.1

Характеристика даних для оцінки інвестиційних ризиків віртуального ІТ-  
підприємства

№	Параметр	Позначення	Тип даних
1.	Код компанії	ID_p	object
2.	Код експерта	ID_exp	object
3.	Попит та ринок збуту	DM	int64
4.	Постачання та вхідні матеріали	SIM	int64
5.	Недооцінка витрат	UC	int64
6.	Недоліки технології	ST	int64
7.	Нормативно-правові ризики	RegR	int64
8.	Наявність фінансування	AF	int64
9.	Фінансова здатність приватного партнера	FCPP	int64
10.	Відсоткова ставка	IR	int64
11.	Інфляція	inflation	int64
12.	Курс гривні до іноземної валюти	ERHFC	int64
13.	Форс-мажорні обставини	FM	int64
14.	Ризик упущеної фінансової вигоди	RLFG	int64
15.	Ризик впровадження	RSup	int64
16.	Не вдасться підібрати персонал необхідної кваліфікації	PS	int64
17.	Невиконання зобов'язань компанією	Failure	int64
18.	Помилки у проектуванні	DE	int64
19.	Зміна платоспроможності клієнтів	CSC	int64
20.	Коливання цін на вироблену продукцію	FPMP	int64
21.	Репутація	Reputational	int64
22.	Робота персоналу	IS	int64
23.	Втрата актуальності продукту	LR	int64
24.	Акумуляування та зберігання інформації	ISe	int64
25.	Ведення бухгалтерського обліку	accounting	int64

Таблиця Б.2

Збір даних за відкритими даними України

Параметр	Пояснення	Тип даних	Джерело даних
ID	Код підприємства	int64	
EDR	Наявність в єдиній базі реєстру юридичних та фізичних осіб	bool	<a href="https://usr.minjust.gov.ua/">https://usr.minjust.gov.ua/</a>
P	Наявність в базі платників ПДВ, ЄСВ та єдиного податку	bool	<a href="https://cabinet.sfs.gov.ua/cabinet/faces/public/reestr.jsp">https://cabinet.sfs.gov.ua/cabinet/faces/public/reestr.jsp</a> x

## Продовження таблиці Б.2

Параметр	Пояснення	Тип даних	Джерело даних
PO	Проведення своєчасної оплати податків	bool	<a href="https://cabinet.sfs.gov.ua/cabinet/faces/public/reestr.jspx">https://cabinet.sfs.gov.ua/cabinet/faces/public/reestr.jspx</a>
K	Наявність розрахунків з контрагентами	bool	<a href="https://youcontrol.com.ua/landing_002/">https://youcontrol.com.ua/landing_002/</a>
VKK	Відомості про наявність керівників компанії в державному реєстрі декларацій	bool	<a href="https://public.nazk.gov.ua/">https://public.nazk.gov.ua/</a>
L	Наявність ліцензій відповідно до КВЕД	bool	<a href="http://irc.gov.ua/ua/Poshuk-v-YeLR.html">http://irc.gov.ua/ua/Poshuk-v-YeLR.html</a>
K205	Наявність кримінальних справ за ст. 205 КК України	bool	<a href="http://www.reyestr.court.gov.ua/">http://www.reyestr.court.gov.ua/</a>
ZMI	Наявність згадувань керівників компаній з ключовими словами: кримінальна справа, корупція, офшорні рахунки та ін.	bool	<a href="http://dzmi.informjust.ua/">http://dzmi.informjust.ua/</a>
ZD	Наявність земельної ділянки за юридичною чи фізичною адресами	bool	<a href="http://map.land.gov.ua/kadastrova-karta">http://map.land.gov.ua/kadastrova-karta</a>
TovZ	Наявність зареєстрованих товарних знаків та послуг, база даних промислових знаків, база даних винаходів та інші бази даних Інституту промислової власності України	bool	<a href="http://www.uipv.org/ua/bases2.html">http://www.uipv.org/ua/bases2.html</a>
SP	Наявність виданих страхових полісів автоцивілки, перевірка поліса МТСБУ, база даних автоцивілки, пошук за державним номером автомобіля, перевірка статусу поліса «Зеленої картки», для авто, що належать компанії	bool	<a href="https://mail.mtibu.kiev.ua/Login.aspx?ReturnUrl=/Cbd/MTSBU_Pages/Tree.aspx">https://mail.mtibu.kiev.ua/Login.aspx?ReturnUrl=/Cbd/MTSBU_Pages/Tree.aspx</a>
A	Наявність автомобілів та їхніх власників, оформлених на компанію.	bool	<a href="https://igov.org.ua/service/1397/general">https://igov.org.ua/service/1397/general</a>
A&SP	Співпадіння зареєстрованих авто зі страховими полісами	bool	
E	Наявність в базі даних експортерів	bool	<a href="http://ukrexport.gov.ua/rus/ukr_export_exporters/?country=ukr">http://ukrexport.gov.ua/rus/ukr_export_exporters/?country=ukr</a>
F	Наявність в базі даних фондового ринку	bool	<a href="http://www.nssmc.gov.ua/fund/registers">http://www.nssmc.gov.ua/fund/registers</a>
AR	Наявність автомобілів та їхніх власників, оформлених на компанію в розшуку	bool	<a href="http://wanted.mvs.gov.ua/searchtransport/">http://wanted.mvs.gov.ua/searchtransport/</a>
ZR	Наявність зброї власників компанії в розшуку	bool	<a href="http://wanted.mvs.gov.ua/searchorj/">http://wanted.mvs.gov.ua/searchorj/</a>
KR	Наявність культурних цінностей власників компанії в розшуку	bool	<a href="http://wanted.mvs.gov.ua/searchart/">http://wanted.mvs.gov.ua/searchart/</a>

## Продовження таблиці Б.2

Параметр	Пояснення	Тип даних	Джерело даних
LB	Наявність ліцензій на будівництво в компанії	bool	<a href="http://dabi.gov.ua/license/list.php">http://dabi.gov.ua/license/list.php</a>
NM	Наявність нерухомого майна в компанії	bool	<a href="https://kap.minjust.gov.ua/services?keywords=&amp;product_id=1&amp;usertype">https://kap.minjust.gov.ua/services?keywords=&amp;product_id=1&amp;usertype</a>
NS	Наявність сайту компанії	bool	www.google.com
FR	Наявність обладнання, розпізнавання обладнання за наявним фото та визначення відповідності геолокації з виробничою адресою	bool	Фото обладнання з геолокацією
FC	Наявність в соціальних мережах компанії та приєднаних працівників	bool	Facebook.com

Таблиця Б.3

## Приклади типових запитань для розробників на співбесіді

Ідентифікатор запитання	Запитання	Ідеальна відповідь	Тип даних
1_1	Ім'я, адреса електронної пошти	Ім'я	об'єкт
1_2	Вік	27	int64
1_3	На яку посаду він претендує?	розробник	об'єкт
1_4	Досвід роботи у відповідних проєктах	Так	булеве значення
1_5	Скільки років	3	int64
1_6	В якому проєкті ви працювали минулого разу?	Проєкт	об'єкт
1_7	Освіта за напрямом проєкту	Так	булеве значення
1_8	Сімейний стан – Одружений?	Ні	булеве значення
1_9	Наявність основної вакансії	Ні	булеве значення
2_1	Виконання робіт за проєктом в узгодженні термінів	Так	булеве значення
2_2	Досвід управління проєктами	Ні	булеве значення
2_3	Робота за стандартами Agile	Так	булеве значення
	Робота за стандартами Scrum	Так	булеве значення
2_4	Ставлення до трудової дисципліни	Так	булеве значення
2_5	Надання клієнтської бази з попередніх проєктів	Ні	булеве значення

## Продовження таблиці Б.3

Ідентифікатор запитання	Запитання	Ідеальні відповіді	Тип даних
2_7	Перевага зеленому кольору	Так	булеве значення
2_8	Перевага червоному кольору	Ні	булеве значення
2_9	Перевага сірому кольору	Так	булеве значення
2_10	Ставлення до критики	Так	булеве значення
3_1	Володіння англійською мовою на рівні B2	Так	булеве значення
3_2	Досвід програмування від 3–х років	Так	булеве значення
3_3	Досвід програмування від 1 року	Так	булеве значення
3_4	Досвід програмування до 1 року	Так	булеве значення
3_5	Наявність облікового запису на GitHub	Так	булеве значення
3_6	Досвід роботи з HTML, JavaScript	Так	булеве значення
3_7	Досвід роботи з Unity	Так	булеве значення
3_8	Досвід C#	Так	булеве значення
3_9	Досвід роботи з Wordpress	Так	булеве значення
3_10	Досвід роботи з MySQL	Так	булеве значення
3_11	Досвід роботи з Python, R	Так	булеве значення
3_12	Чи влаштовує зарплата?	Так	булеве значення

## ДОДАТОК В. ЕКСПЕРТНИЙ ВИСНОВОК

### ВИСНОВОК

експертної комісії громадської організації "Прогресивні Люди" щодо оцінки ефективності інформаційної системи інтелектуального аналізу соціально-економічних даних територіальних громад (ТГ)

Експертна комісія створена на базі громадської організації "Прогресивні Люди" (ЄДРПОУ 42108496, зареєстрована 04.05.2018) у складі:

**Голова комісії — експерт:** Щур Світлана — Голова Організації, Співзасновниця, проєктний менеджер.

**Члени комісії — експерти:** Турський Ігор — Заступник Голови Організації, Голова наглядової ради, Співзасновник, професор, доктор економічних наук; Корендій Світлана — Комунікаційниця, SMM-менеджер; Созанська Тетяна — Аналітикиня; Мар'янченко Тетяна — Юристка.

Експертна комісія провела експертизу ефективності інтелектуальних систем для оптимізації соціально-економічної інфраструктури в ТГ. Метою експертизи було визначення рівня впливу впровадження інтелектуальних систем на підвищення продуктивності та якості управління в громадах.

За результатами проведеної експертизи встановлено, що з метою уніфікації експертних оцінок ефективності інтелектуальних систем у сфері соціально-економічної інфраструктури ТГ було розроблено підхід, що включає розрахунок ефективності на основі часу виконання завдань. Формула для розрахунку середнього часу виконання кожного завдання виглядає так:

$$k_c = \frac{1}{n} \sum_{i=1}^n P_i$$

де  $P_i$  — показник часу виконання завдання, визначений експертом;  $n$  — кількість експертів (у нашому випадку  $n=5$ ). Експерти оцінювали час виконання кожної задачі з використанням системи та без неї. На основі цих оцінок визначалися межі часу виконання завдань — від мінімального до максимального значення.

Ефективність системи розраховується як відносне скорочення часу виконання завдань при використанні системи порівняно з виконанням без неї. Формула ефективності для кожного завдання виглядає так:

$$\text{Ефективність (\%)} = \left( 1 - \frac{\text{Час з системою}}{\text{Час без системи}} \right) \times 100\%$$

де "Час з системою" та "Час без системи" представлені у відповідних одиницях виміру (годинах або днях). Ця формула дозволяє отримати відсотковий показник, який відображає, наскільки використання інтелектуальної системи скорочує час виконання завдання. Для кожного завдання ефективність обчислюється на основі усереднених значень часу, визначених експертами.

Числові узагальнені значення експертних оцінок ефективності інтелектуальних систем для соціально-економічної інфраструктури ТГ, відповідно до оцінок експертів, вважаються наступними (таблиця 1):



Таблиця 1. Експертні оцінки

№ експерта	Оцінка інвестиційних ризиків	Виявлення фактивних підприємств	Виявлення дезінформації	Формування HR	Вибір конкурентного товару	Формування складового кошика	Пошук альтернативних клієнтів та постачальників	Прогнозування попиту в умовах кризи	Вибір локації для старту бізнесу та прогнозування потоку відвідувачів	Виявлення технологічних катастроф	Прогнозування обсягів відходів
1	2	2	2	1	1	1	1	2	1	2	1
2	1	2	2	10	1	1	4	4	1	1	2
3	3	3	1	25	2	2	2	1	2	2	1
4	1	6	2	120	2	3	5	3	3	4	2
5	2	2	1	100	1	2	4	5	2	4	1
Час з системою в годинах	~1-3	~2-6	~1-2	~1-120	~1-2	~1-3	~1-5	~1-5	~1-3	~1-4	~1-2
1	15	8	3	5	14	2	8	1	2	3	5
2	40	12	2	10	25	10	7	2	10	1	9
3	3	3	1	25	35	15	36	1	15	3	5
4	25	6	2	20	90	60	60	5	60	5	8
5	60	20	5	30	80	5	10	5	5	4	10
Час без системи в днях	~2-60	~3-20	~1-5	~5-30	~14-90	~2-60	~7-60	~1-5	~2-10	~1-5	~5-10
Ефективність, %	~99,72	~99,58	~99,17	~80,00	~99,84	~98,75	~99,61	~95,83	~98,75	~95,83	~99,72

**Рекомендації експертної комісії:** Результати експертних оцінок ефективності інтелектуальних систем, представлені в таблиці 1, показали значне скорочення часу виконання ключових завдань, що, у свою чергу, суттєво підвищує ефективність управлінських процесів у ТГ. Загальний показник ефективності інтелектуальної системи розраховувався як середнє значення ефективності для всіх завдань, що дозволяє оцінити її вплив на оптимізацію процесів у ТГ.

Експертна комісія підтверджує високий рівень ефективності інтелектуальних систем, встановлений на основі середніх значень. У зв'язку з цим комісія рекомендує подальшу інтеграцію цих систем у інфраструктуру ТГ для підвищення продуктивності та якості управління. Такий підхід сприятиме більш ефективному та швидкому виконанню управлінських завдань, що є важливим аспектом розвитку сучасних громад.

Числові значення експертних оцінок ефективності можуть бути доповнені додатковими показниками, важливими для конкретних умов застосування. При цьому кількісні оцінки можуть коригуватися залежно від результатів новітніх досліджень та впровадження інноваційних підходів у сфері управління ТГ. Допустиме відхилення оцінок — у межах  $\pm 10-20\%$  — є допустимим і може бути зумовлене розвитком нових технологій та методів у цій галузі.

#### Експерти:

Голова експертної комісії:

Голова організації, співзасновниця,

Світлана ЩУР

Член експертної комісії:

Заступник Голови Організації,  
Голова наглядової ради, співзасновник,  
професор, доктор економічних наук

Ігор ТУРСЬКИЙ

Комунікаційниця, SMM-менеджер

Світлана КОРЕНДІЙ

Аналітикиня

Тетяна СОЗАНСЬКА

Юристка

Тетяна МАР'ЯНЧЕНКО

З висновками експертизи ознайоmlена:

к.т.н., доцент

Христина ЛП'ЯНІНА-ГОНЧАРЕНКО



## ДОДАТОК Г. СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ТА ВІДОМОСТІ ПРО АПРОБАЦІЮ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЇ

*Статті у наукових виданнях, включених до Переліку наукових фахових видань України:*

1. Ліп'яніна–Гончаренко, Х. (2024). Узагальнений принцип синтезу інформаційної технології інтелектуального аналізу соціально–економічних даних ТГ. *Measuring and Computing Devices in Technological Processes*, (1), 359–367. <https://doi.org/10.31891/2219-9365-2024-77-48>
2. Ліп'яніна–Гончаренко, Х. (2024). Методологія аналізу соціально–економічних даних територіальних громад. *Herald of Khmelnytskyi National University. Technical Sciences*, 337(3(2)), 446–450. <https://doi.org/10.31891/2307-5732-2024-337-3-67>
3. Ліп'яніна–Гончаренко, Х. (2024). Методи аналізу соціально–економічних даних територіальних громад для адаптивного управління ресурсами. *Computer Systems and Information Technologies*, (3), 92–97. <https://doi.org/10.31891/csit-2024-3-12>
4. Ліп'яніна–Гончаренко, Х. В. (2023). Метод формування навчальної вибірки для масивів даних на основі машинного навчання. *Наукові записки НаУКМА. Комп'ютерні науки*, (6), 30–35. <https://doi.org/10.18523/2617-3808.2023.6.30-35>
5. Ліп'яніна–Гончаренко, Х., & Юрків, Х. (2023). Методи бустингового машинного навчання для нестационарних часових рядів. *Measuring and Computing Devices in Technological Processes*, (3), 19–30. <https://doi.org/10.31891/2219-9365-2023-75-2>
6. Ліп'яніна–Гончаренко, Х., Комар, М., Юрків, Х., & Лук'янчук, В. (2023). Концептуальна модель інтелектуальної оцінки наслідків техногенних катастроф. *Вісник Хмельницького національного університету. Серія: Технічні науки*, 329(6), 230–237. <https://doi.org/10.31891/2307-5732-2023-329-6-230-237>
7. Комар, М., Ліп'яніна–Гончаренко, Х., Кіт, І., Мадараш, Р., & Юрків, Х. (2023). Інтелектуальний метод виявлення джерел мультилінгвальної дезінформації.

*Measuring and Computing Devices in Technological Processes*, (2), 221–230.  
<https://doi.org/10.31891/2219-9365-2023-74-31>

8. Ліп'яніна–Гончаренко, Х. В. (2022). Інтелектуальний метод формування людських ресурсів на короткостроковий проєкт. *Вісник Черкаського державного технологічного університету*, (3), 49–58. <https://doi.org/10.24025/2306-4412.3.2022.259775>

9. Ліп'яніна–Гончаренко, Х., Комар, М., Саченко, А., & Лендюк, Т. (2022). Оцінка інвестиційних ризиків віртуальної ІТ–компанії на основі машинного навчання. *Measuring and Computing Devices in Technological Processes*, (3), 45–60. <https://doi.org/10.31891/2219-9365-2022-71-3-6>

10. Ліп'яніна–Гончаренко, Х. В., Комар, М. П., Саченко, А. О., & Лендюк, Т. В. (2022). Метод виявлення фіктивних підприємств на підставі Гаусового наївного класифікатора Байєса. *Scientific Bulletin of UNFU*, 32(5), 92–96. <https://doi.org/10.36930/40320513>

11. Ліп'яніна–Гончаренко, Х. (2023). Інтелектуальний метод вибору локації для старту бізнесу в розумному місті. *Системні технології*, 4(147), 132–140. <https://doi.org/10.34185/1562-9945-4-147-2023-12>

12. Ліп'яніна–Гончаренко, Х., Комар, М., Саченко, А., & Лендюк, Т. (2022). Метод формування контексту реклами та цільової аудиторії на основі навчання асоціативних правил. *Вісник Хмельницького національного університету. Серія: Технічні науки*, 313(5), 279–287. <https://doi.org/10.31891/2307-5732-2022-313-5-279-287>

13. Ліп'яніна–Гончаренко, Х. В. (2023). Метод генерування рекламного зображення на основі ключових слів. *Телекомунікаційні та інформаційні технології*, 81(4). <https://doi.org/10.31673/2412-4338.2023.043745>

14. Ліп'яніна–Гончаренко, Х., & Кіт, І. (2023). Метод удосконалення рекламних текстів на основі генеративних моделей. *Вісник Приазовського державного технічного університету. Серія: Технічні науки*, (46), 6–13. <https://doi.org/10.31498/2225-6733.46.2023.288087>

15. Ліп'яніна–Гончаренко, Х. В. (2023). Метод генерування рекламного зображення на основі відеопотоку. *Науковий вісник Ужгородського університету*.

Серія: Математика і інформатика, 43(2), 130–135. [https://doi.org/10.24144/2616-7700.2023.43\(2\).130-135](https://doi.org/10.24144/2616-7700.2023.43(2).130-135)

Статті у періодичних виданнях, включених до категорії «А» Переліку наукових фахових видань України, або у закордонних виданнях, проіндексованих у базах даних *Web of Science Core Collection* та/або *Scopus*:

16. Lipianina–Honcharenko, K., Wolff, C., Sachenko, A., Kit, I., & Zahorodnia, D. (2023). Intelligent method for classifying the level of anthropogenic disasters. *Big Data and Cognitive Computing*, 7(3), 157. <https://doi.org/10.3390/bdcc7030157> (Індексована в наукометричній базі *Web of Science* та *Scopus*, що відноситься до Q2 відповідно до *SCImago Journal & Country Rank*)

17. Lipianina–Honcharenko, K., Wolff, C., Sachenko, A., Desyatnyuk, O., Sachenko, S., & Kit, I. (2023). Intelligent information system for product promotion in internet market. *Applied Sciences*, 13(17), 9585. <https://doi.org/10.3390/app13179585> (Індексована в наукометричній базі *Web of Science* та *Scopus*, що відноситься до Q2 відповідно до *SCImago Journal & Country Rank*)

18. Lipianina–Honcharenko, K., Komar, M., Osolinskyi, O., Shymanskyi, V., Havryliuk, M., & Semaniuk, V. (2024). Intelligent waste–volume management method in the smart city concept. *Smart Cities*, 7(1), 78–98. <https://doi.org/10.3390/smartcities7010004> (Індексована в наукометричній базі *Web of Science* та *Scopus*, що відноситься до Q1 відповідно до *SCImago Journal & Country Rank*)

19. Lipianina–Honcharenko, K., Komar, M., Melnyk, N., & Komarnytsky, R. (2024). Sustainable information system for enhancing virtual company resilience through machine learning in smart city socio–economic scenarios. *Economics*, 12(2). <https://doi.org/10.2478/eoik-2024-0022> (Індексована в наукометричній базі *Web of Science* та *Scopus*, що відноситься до Q2 відповідно до *SCImago Journal & Country Rank*)

20. Lipianina–Honcharenko, K., Bodyanskiy, Y., Kustra, N., & Ivasechko, A. (2024). OLTW–TEC: Online learning with sliding windows for text classifier ensembles. *Frontiers in Artificial Intelligence*, 7, 1401126. <https://doi.org/10.3389/frai.2024.1401126>

(Індексована в наукометричній базі Scopus та відноситься до Q2 відповідно до SCImago Journal & Country Rank)

21. Bodyanskiy, Y. V., Lipianina–Honcharenko, K. V., & Sachenko, A. O. (2022). Ensemble of adaptive predictors for multivariate nonstationary sequences and its online learning. *Radio Electronics, Computer Science, Control*, (4(67)), 91–97, <https://doi.org/10.15588/1607-3274-2023-4-9> (Індексована в наукометричній базі Web of Science)

22. Lipianina–Honcharenko, K., Bodyanskiy, Y., & Sachenko, A. (2023). Intelligent information system of the city's socio–economic infrastructure. *System Research and Information Technologies*, (3), 108–120. <https://doi.org/10.20535/srit.2308-8893.2023.3.08> (Індексована в наукометричній базі Scopus та відноситься до Q4 відповідно до SCImago Journal & Country Rank)

*Монографії (розділи у колективних монографіях):*

23. Lipyanina, H., Sachenko, O., Lendyuk, T., Sachenko, A., & Vasylykiv, N. (2020). Intelligent method of forming the HR management short–term project. In *Advances in Intelligent Systems and Computing* (pp. 1045–1055). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-63270-0\\_71](https://doi.org/10.1007/978-3-030-63270-0_71)

24. Lipianina–Honcharenko, K., Wolff, C., Chyzhovska, Z., Sachenko, A., Lendiuk, T., & Grodskyi, S. (2022). Intelligent method for forming the consumer basket. In *Communications in Computer and Information Science* (pp. 221–231). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-16302-9\\_17](https://doi.org/10.1007/978-3-031-16302-9_17) (Індексована в наукометричній базі Scopus)

25. Lipyanina, H., Maksymovych, V., Sachenko, A., Lendyuk, T., Fomenko, A., & Kit, I. (2020). Assessing the investment risk of virtual IT company based on machine learning. In *Communications in Computer and Information Science* (pp. 167–187). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-61656-4\\_11](https://doi.org/10.1007/978-3-030-61656-4_11) (Індексована в наукометричній базі Scopus)

26. Lipianina–Honcharenko, K., Lendiuk, T., Sachenko, A., Osolinskyi, O., Zahorodnia, D., & Komar, M. (2022). An intelligent method for forming the advertising content of higher education institutions based on semantic analysis. In *Communications in*

*Computer and Information Science* (pp. 169–182). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-14841-5\\_11](https://doi.org/10.1007/978-3-031-14841-5_11) (Індексована в наукометричній базі Scopus)

*Статті у матеріалах міжнародних конференцій, які індексуються у наукометричних базах Scopus та Web of Science:*

27. Lipyanina–Goncharenko, H., Brych, V., Sachenko, S., Lendyuk, T., Bykovyy, P., & Zahorodnia, D. (2021). Method of forming a training sample for segmentation of tender organizers on a machine learning basis. In *COLINS* (pp. 1843–1852). <https://ceur-ws.org/Vol-2870/paper134.pdf> (Індексована в наукометричній базі Scopus та Web of Science)

28. Lipianina–Honcharenko, K., Kit, I., Zahorodnia, D., & Osolinskyi, O. (2023). Optimization method of advertising texts based on generative models. In *Proceedings of the 3rd International Workshop on Information Technologies: Theoretical and Applied Problems 2023* (Vol. 3628, pp. 126–135). <https://ceur-ws.org/Vol-3628/paper10.pdf> (Індексована в наукометричній базі Scopus)

29. Komar, M., Savchyshyn, R., Lipianina–Honcharenko, K., & Osolinskyi, O. (2023). Intelligent method for counting cars from satellite images. In *IntSol* (pp. 295–303). [https://ceur-ws.org/Vol-3538/Short\\_1.pdf](https://ceur-ws.org/Vol-3538/Short_1.pdf) (Індексована в наукометричній базі Scopus)

30. Lipianina–Honcharenko, K., Bodyanskiy, Y., Sachenko, A., Kit, I., Podchasova, T., & Lendiuk, T. (2023). An intelligent method of predicting the demand for goods/services in crisis conditions. In *Proceedings of the 3rd International Workshop of IT–Professionals on Artificial Intelligence (ProfIT AI 2023)* (Vol. 3641, pp. 224–232). <https://ceur-ws.org/Vol-3641/short2.pdf> (Індексована в наукометричній базі Scopus)

31. Lipyanina, H., Sachenko, S., Lendyuk, T., Brych, V., Yatskiv, V., & Osolinskyi, O. (2021). Method of detecting a fictitious company on a machine learning basis. In *Advances in Computer Science for Engineering and Education IV* (pp. 138–146). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-80472-5\\_12](https://doi.org/10.1007/978-3-030-80472-5_12) (Індексована в наукометричній базі Scopus)

32. Krysovaty, A., Lipianina–Honcharenko, H., Sachenko, S., Desyatnyuk, O., Banasik, A., & Lukasevych–Krutnyk, I. (2022). Recognizing the fictitious business entity

on a logistic regression basis. In *Proceedings of the 3rd International Workshop on Intelligent Information Technologies & Systems of Information Security* (Vol. 3156, pp. 315–327). <https://ceur-ws.org/Vol-3156/paper15.pdf> (Індексована в наукометричній базі Scopus )

33. Krysovatyu, A., Lipyanina–Goncharenko, H., Desyatnyuk, O., & Sachenko, S. (2021). Classification method of fictitious enterprises based on Gaussian Naive Bayes. In *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)* (pp. 210–216). <https://doi.org/10.1109/CSIT52700.2021.9648584> (Індексована в наукометричній базі Scopus )

34. Krysovatyu, A., Lipyanina–Goncharenko, H., Sachenko, S., & Desyatnyuk, O. (2021). Economic crime detection using support vector machine classification. In *Modern Machine Learning Technologies and Data Science Workshop (MoMLLeT&DS 2021)* (Vol. 2917, pp. 830–840). <https://ceur-ws.org/Vol-2917/paper46.pdf> (Індексована в наукометричній базі Scopus )

35. Lipianina–Honcharenko, K., Sachenko, A., Semaniuk, V., Badasian, A., & Kopania, Ł. (2023). Intelligent method for selecting a business location in a smart city. In *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. (pp. 1184–1188). <https://doi.org/10.1109/IDAACS58523.2023.10348823> (Індексована в наукометричній базі Scopus)

36. Gramyak, R., Lipyanina–Goncharenko, H., Sachenko, A., Lendyuk, T., & Zahorodnia, D. (2022). Intelligent method for choosing a competitive product based on emotional feedback coloring. In *Proceedings of the 2nd International Workshop on Intelligent Information Technologies & Systems of Information Security* (Vol. 2853, pp. 346–357). <https://ceur-ws.org/Vol-2853/paper31.pdf> (Індексована в наукометричній базі Scopus)

37. Lipianina–Honcharenko, K., Lendiuk, T., Sachenko, A., & Wołoszyn, J. (2022). Method of forming the context of advertising and target audience based on associative rule learning. In *Proceedings of the Fifth International Workshop on Computer Modeling and*



*Intelligent Systems (CMIS–2022)* (Vol. 3137, pp. 98–107). <https://ceur-ws.org/Vol-3137/paper9.pdf> (Індексована в наукометричній базі Scopus)

38. Lipyanina, H., Sachenko, S., Lendyuk, T., & Sachenko, A. (n.d.). Targeting model of HEI video marketing based on a classification tree. In *16th International Conference on ICT in Education, Research, and Industrial Applications. Integration, Harmonization, and Knowledge Transfer (ICTERI 2020)* (Vol. 2732, pp. 487–498). <https://ceur-ws.org/Vol-2732/20200487.pdf> (Індексована в наукометричній базі Scopus)

39. Lipianina–Honcharenko, K., Lukasevych–Krutnyk, I., Butryn–Boka, N., & Sachenko, A. (2022). Intelligent method for identifying fraudulent online stores. In *2021 IEEE 8th International Conference on Problems of Infocommunications, Science, and Technology (PIC S&T)*. <https://doi.org/10.1109/PICST54195.2021.9772195> (Індексована в наукометричній базі Scopus)

40. Koziuk, V., & Lipyanina–Goncharenko, H. (2021). Intelligent method of predicting the discount rate trend. In *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. IEEE. <https://doi.org/10.1109/IDAACS53288.2021.9660835> (Індексована в наукометричній базі Scopus)

41. Lipianina–Honcharenko, K., Sachenko, A., Wolff, C., & Bodyanskiy, Y. (2023). Simulation model for determining the quality of life in Ukrainian cities during the war. In *2023 IEEE European Technology and Engineering Management Summit (E–TEMS)* (pp. 97–101). <https://doi.org/10.1109/E–TEMS57541.2023.10424587> (Індексована в наукометричній базі Scopus)

42. Pariy, V., Lipianina–Honcharenko, K., Brukhanskyi, R., Sachenko, A., Tkachyk, F., & Lendiuk, D. (2023, November). Intelligent verbal interaction methods with non–player characters in metaverse applications. In *2023 IEEE 5th International Conference on Advanced Information and Communication Technologies (AICT)* (pp. 67–71). <https://doi.org/10.1109/AICT61584.2023.10452688> (Індексована в наукометричній базі Scopus)

43. Lipianina–Honcharenko, Kh., Soia, M., Yurkiv, Kh., & Ivasechko, A. (2024). Evaluation of the effectiveness of machine learning methods for detecting disinformation in

Ukrainian text data. In *Proceedings of the Seventh International Workshop on Computer Modeling and Intelligent Systems (CMIS-2024)* (pp. 97–109). <https://ceur-ws.org/Vol-3702/paper9.pdf> (Індексована в наукометричній базі Scopus)

44. Sachenko, A., Lendiuk, T., Lipianina–Honcharenko, K., Dobrowolski, M., Boguta, G., & Bytsyura, L. (2024). Method of determining the text sentiment by thematic rubrics. In *COLINS* (pp. 404–414). <https://ceur-ws.org/Vol-3688/paper26.pdf> (Індексована в наукометричній базі Scopus)

45. Lipianina–Honcharenko, K., Melnychuk, A., Yurkiv, K., Hladiy, G., & Telka, M. (2024). Integrated approach to the international aspects of online dispute resolution formation. In *Proceedings of the First International Workshop of Young Scientists on Artificial Intelligence for Sustainable Development* (pp. 88–98). <https://ceur-ws.org/Vol-3716/paper7.pdf> (Індексована в наукометричній базі Scopus)

46. Komar, M., Lipianina–Honcharenko, K., Domanskyi, V., & Melnyk, N. (2024). Regression–based method for real–time solar power plant efficiency forecasting. In *Proceedings of the Modern Machine Learning Technologies Workshop (MoMLeT 2024)* (pp. 235–245). <https://ceur-ws.org/Vol-3711/paper14.pdf> (Індексована в наукометричній базі Scopus)

47. Sachenko, A., Lendiuk, T., Lipianina–Honcharenko, K., Koval, V., Hladiy, G., & Halias, Y. (2024). Evaluation of ensemble machine learning models for movie recommendation systems. In *Proceedings of the Modern Machine Learning Technologies Workshop (MoMLeT 2024)* (pp. 273–286). <https://ceur-ws.org/Vol-3711/paper17.pdf> (Індексована в наукометричній базі Scopus)

48. Lipianina–Honcharenko, K., Maika, N., Sachenko, S., Kopania, L., & Soia, M. (2024). A cyclical approach to legal document analysis: Leveraging AI for strategic policy evaluation. In *CEUR Workshop Proceedings* (Vol. 3736, pp. 201–211). <https://ceur-ws.org/Vol-3736/paper15.pdf> (Індексована в наукометричній базі Scopus)

*Публікації, які додатково відображають наукові результати дисертації:*

49. Крисоватий, А. І., Лип’яніна–Гончаренко, Х. В., та ін. (2022). *Метод виявлення фіктивних підприємств на основі машинного навчання*. Авторське право на твір № 111509, Україна. Опубл. 31.03.2022, Бюл. № 69.

## ДОДАТОК Д. АКТИ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ РОБОТИ



ТЕРНОПІЛЬСЬКА ОБЛАСНА ДЕРЖАВНА АДМІНІСТРАЦІЯ  
ТЕРНОПІЛЬСЬКА ОБЛАСНА ВІЙСЬКОВА АДМІНІСТРАЦІЯ  
ДЕПАРТАМЕНТ ЦИФРОВОЇ ТРАНСФОРМАЦІЇ

вул. Грушевського, 8, м.Тернопіль, 46021, тел./факс (0352) 51-70-10  
E-mail: [digital@te.gov.ua](mailto:digital@te.gov.ua), Web: [digital.te.gov.ua](http://digital.te.gov.ua) Код згідно ЄДРПОУ 44253982

від 569 № 20.10.2024

### АКТ

про впровадження результатів  
дисертаційної роботи Ліп'яніної-Гончаренко Христини Володимирівни

Результати дисертаційної роботи Ліп'яніної-Гончаренко Христини Володимирівни на тему «Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад», поданої на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.06 «Інформаційні технології», апробовано в Департаменті цифрової трансформації Тернопільської обласної державної адміністрації, зокрема:

1. Інтелектуальний метод виявлення дезінформації, який дозволяє з точністю 93% ефективно ідентифікувати фейкові новини. Метод підвищує правильну класифікацію фейкових новин на 95% та забезпечує відновлення на рівні 72%, що сприяє захисту інформаційного простору та підвищенню довіри громадськості.

2. Інтелектуальний метод класифікації рівня техногенних катастроф, що забезпечує точність класифікації на рівні 77%, дозволяючи зменшити час реакції на надзвичайні ситуації. Метод досягає точності 98% у визначенні критичних рівнів небезпеки, підтримуючи стабільність інфраструктури в умовах криз.

3. Модель оцінки інвестиційного ризику віртуальної ІТ-компанії на основі машинного навчання, яка досягає точності 98% з середньою абсолютною

похибкою 5%. Це дозволяє приймати інвестиційні рішення з високою достовірністю та мінімізувати ризики, пов'язані з інвестуванням у віртуальні ІТ-проекти.

4. Інтелектуальний метод прогнозування попиту на продукцію, що забезпечує середню абсолютну похибку 8,97% та точність прогнозування попиту на рівні 90%. Метод дозволяє ефективно управляти ресурсами та задовольняти потреби громади, особливо в умовах кризових ситуацій.

Зазначені результати дозволили підвищити ефективність управління інформаційними та техногенними ризиками, оптимізувати економічне планування та забезпечити стабільний розвиток територіальних громад.

Отримані результати також можуть бути використані Департаментом цифрової трансформації Тернопільської обласної державної адміністрації для подальшого вдосконалення цифрової інфраструктури та підвищення інформаційної безпеки в області.

**Виконувач обов'язків директора  
департаменту цифрової трансформації**



**Сергій ВЕРБОВИЙ**

№ 20 від 29.11.2023

АКТ ВПРОВАДЖЕННЯ  
щодо впровадження результатів дисертаційної роботи  
Ліп'яніної-Гончаренко Х.В.  
на тему «Теоретичні та прикладні засади інформаційної технології  
інтелектуального аналізу соціально-економічних даних територіальних громад»  
у дослідницької роботи в рамках проекту  
AURA – Auralisation of Acoustic Heritage Sites using Augmented and Virtual  
Reality (№ 101008547), фінансованого фондом "Creative Europe",  
на базі компанії MagneticOne

Ми, нижчепідписані, представники компанії MagneticOne:

Директор компанії MagneticOne: Савчишин Руслан Володимирович

Керівник дослідницького відділу: Щур Світлана Володимирівна

склали цей акт про те, що результати дослідницької роботи, виконані в рамках проекту AURA – Auralisation of Acoustic Heritage Sites using Augmented and Virtual Reality (№ 101008547), були успішно впроваджені в діяльність нашої компанії та проекту.

Опис впроваджених результатів:

1. Метод класифікаційного аналізу текстових соціально-економічних даних територіальних громад

— Мета: Використання методу для аналізу текстових даних з метою ефективної обробки інформації про туристичні показники у територіальних громадах.

— Практична реалізація: Метод інтегрований у процес розробки чат-бота для інтелектуального обслуговування громадян, що дозволяє автоматизувати надання консультаційних послуг на основі аналізу текстових даних.

2. Метод прогнозування соціально-економічних даних територіальних громад

— Мета: Прогнозування ключових показників соціально-економічного розвитку для оптимізації управлінських рішень в туристичній галузі.

— Практична реалізація: Метод застосований для формування імітаційної моделі бізнес-процесів, що дозволяє аналізувати ефективність процесів у інфраструктурі територіальних громад.

Впроваджені результати забезпечують:

- Підвищення ефективності туристичних сервісів.

- Збереження та популяризацію культурної спадщини за допомогою віртуальної реставрації.
- Оптимізацію бізнес-процесів та збільшення доходів від впровадження новітніх технологій у сфері культури та туризму.

Даний акт складений в двох примірниках, по одному для кожної зі сторін.

Підписи:

Директор компанії MagneticOne:  Савчишин Р.В./

Керівник дослідницького відділу:  /Щур С.В./

Дата: 29.01.2023 року





МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

вул. Львівська, 11, м. Тернопіль, 46009; тел./факс +380 (352) 51-75-75;  
www.wunu.edu.ua; rektor@wunu.edu.ua; ідентифікаційний код за ЄДРПОУ 33680120

27 листопада 2024

ДОВІДКА

Даний документ складений з метою підтвердження того, що результати дисертаційної роботи Ліп'яніної-Гончаренко Христини Володимирівни «Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад» апробовано та впроваджено при виконанні наступних міжнародних проектів:

1. «Мультифункціональне Фермерство / Multifunctionality Farming» (MyFarm) 101140288 - ERASMUS-EDU-2023-PI-ALL-INNO-EDU-ENTERP

2. «Міждоменні компетенції для забезпечення здорової та безпечної роботи у 21 столітті (Cross-domain competences for health and safe work in the 21st century)» (WORK4CE) 619034-EPP-1-2020-1-UA-EPPKA2-CBHE-JP у межах програми Еразмус+.

Проректор з науково-педагогічної роботи



Уляна КОРУЦ

ПРОГРЕСИВНІ ЛЮДИ  
ГРОМАДСЬКА ОРГАНІЗАЦІЯ



PROGRESSIVE PEOPLE  
PUBLIC ORGANIZATION

№ 23 від 18.10.2024

### АКТ

про впровадження результатів дисертаційної роботи Ліп'яніної-Гончаренко  
Христини Володимирівни

Результати дисертаційної роботи Ліп'яніної-Гончаренко Х. В. на тему «Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад», поданої на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.06 «Інформаційні технології», апробовано у Громадській Організації «Прогресивні Люди», зокрема:

1. Інтелектуальний метод формування короткострокового проєкту управління персоналом, що дає можливість ефективніше управляти людськими ресурсами та підвищувати продуктивність роботи підприємств у громаді. Застосування цього методу сприяє залученню та утриманню кваліфікованих кадрів, що є важливим для сталого розвитку бізнесу.
2. Модель оцінки інвестиційного ризику для віртуальної ІТ-компанії на основі машинного навчання, яка використовується для зменшення ризиків при запуску нових проєктів та залученні інвестицій. Така модель допомагає підприємствам приймати більш обґрунтовані рішення в умовах невизначеності, що підвищує їхню фінансову стійкість.
3. Інтелектуальний метод прогнозування попиту на продукцію, який дозволив громаді краще розуміти ринкові тенденції та адаптувати виробничі процеси відповідно до потреб клієнтів. Це забезпечує ефективне планування бізнес-діяльності та покращення якості обслуговування.
4. Інтелектуальний метод інтерпретації правової інформації, застосований для аналізу законодавчих актів та нормативних документів. Це сприяло здійсненню реформ на всіх рівнях влади та підвищенню прозорості в суспільстві.

Зазначені результати дозволили ГО «Прогресивні Люди» підвищити ефективність своєї діяльності в напрямках розвитку громадянського суспільства, зміцнення міжнародного співробітництва та підтримки розвитку громад через реалізацію проєктів у сфері освіти, культури та економіки.



Отримані результати також можуть бути використані в майбутніх проєктах організації для подальшого розвитку інфраструктури, залучення додаткових ресурсів та підтримки реформ на всіх рівнях влади та в суспільстві.

Засновник ГО «Прогресивні Люди»



Ігор ТУРСЬКИЙ

№ 15 від 26.09.2024

### АКТ

про впровадження результатів дисертаційної роботи Ліп'яніної-Гончаренко Христини Володимирівни

Результати дисертаційної роботи Ліп'яніної-Гончаренко Христини Володимирівни на тему «Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад», поданої на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.06 «Інформаційні технології», апробовано у ТОВ «МагнетікВан Муніципальні Технології» для підвищення ефективності роботи муніципалітетів через впровадження геоінформаційних систем, зокрема:

1. Гібридний метод аналізу соціально-економічних даних територіальних громад для прогнозування обсягів відходів. Цей метод дозволяє з точністю 99,72% прогнозувати обсяги відходів на основі класифікації за типом та кластеризації місць їх утворення. Запропонований підхід забезпечує ефективне управління відходами у громадах, сприяє оптимізації ресурсів та покращенню екологічної ситуації завдяки швидкому та точному аналізу соціально-економічних даних.

2. Інтелектуальний метод вибору локації для старту бізнесу в територіальних громадах. Цей метод дозволяє з точністю 98,75% визначати оптимальні місця для відкриття нових бізнес-проектів на основі прогнозування соціально-економічних даних та технологій машинного навчання. Метод враховує географічні особливості та ринкові чинники територіальної громади, що забезпечує ефективне планування розвитку економічної інфраструктури та підвищує ймовірність успішного запуску бізнесу.

3. Метод прогнозування соціально-економічних даних для визначення динаміки кількості автомобілів на період 2020–2030 років. Використовуючи даний метод, прогноз показує зростання кількості автомобілів у місті Тернополі до 15% у порівнянні з попередніми роками, що дозволяє муніципалітетам вчасно планувати стратегії для скорочення екологічного навантаження.

Зазначені результати дали змогу ТОВ «МагнетікВан Муніципальні Технології» розробити сучасне програмне рішення для моніторингу сталого

розвитку інфраструктури та управління екологічними показниками в межах муніципалітетів.

Даний акт не є підставою для фінансових взаєморозрахунків.

Директор

ТОВ «МагнетікВан Муніципальні Технології» Подібівський





УКРАЇНА

ВЕЛИКОБЕРЕЗОВИЦЬКА СЕЛИЩНА РАДА

ПЕТРИКІВСЬКИЙ СТАРОСТИНСЬКИЙ ОКРУГ с.ПЕТРИКІВ

Вул. Шептицького, буд.118 а, с. Петриків Тернопільського району, Тернопільської області.

☎ 47720, 📠 код: (0352) факс: 27-42-31; 27-41-42; 29-10-84, [ybsr.gov.ua@ukr.net](mailto:ybsr.gov.ua@ukr.net)

код ЄДРПОУ: 04393462

Від 28.10. 2024 р. № 336

## АКТ

впровадження результатів наукового дослідження Ліп'яніної-Гончаренко Х.В.  
«Теоретичні та прикладні засади інформаційної технології інтелектуального  
аналізу соціально-економічних даних територіальних громад»

Великобerezовицька селищна громада Тернопільського району Тернопільської області цим актом підтверджує, що розроблені у процесі підготовки наукового дослідження Ліп'яніною-Гончаренко Христиною Володимирівною наукові положення і висновки за темою «Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад» були детально вивчені та впроваджені у практичну діяльність громади.

Зокрема, дослідником було надано рукопис наукового дослідження та перелік наукових праць за темою. За результатами вивчення матеріалів, відзначено, що окремі наукові й практичні результати, одержані особисто здобувачем наукового ступеня Ліп'яніною-Гончаренко Х.В., використані для підвищення ефективності управління соціально-економічним розвитком територіальної громади. Зокрема, застосовано такі положення:

1. Проведено аналіз стану і перспектив розвитку аналізу соціально-економічних даних в інфраструктурі територіальних громад.
  - Виконано оцінку наявного стану збору, обробки та використання соціально-економічних даних у територіальних громадах.
  - Визначено ключові виклики, зокрема відсутність інтегрованих систем та низьку ефективність традиційних методів обробки даних.

- Обґрунтовано необхідність впровадження сучасних аналітичних інструментів для підвищення точності моніторингу та прогнозування розвитку громади.
2. Реалізовано узагальнений принцип синтезу інформаційної технології для інтелектуального аналізу соціально-економічних даних ТГ.
- Запропоновано комплексний підхід до розробки інформаційної технології, що базується на інтеграції методів класифікації, кластеризації та прогнозування соціально-економічних показників.
  - Створено архітектуру системи, що включає модулі збору даних, їх інтелектуальної обробки та візуалізації результатів для ефективного прийняття рішень.
  - Узагальнено принципи адаптивності технології для її застосування в різних громадах з урахуванням їхніх специфічних потреб.

Вказані результати наукового дослідження, викладені у пропозиціях Лип'яніної-Гончаренко Х.В., можуть бути надалі враховані та використані у повсякденній діяльності органів місцевого самоврядування для вдосконалення управлінських процесів та забезпечення стійкого розвитку громади.

**Голова Великоберезовицької селищної громади**



Затверджую

Проректор з наукової роботи

Західноукраїнського

національного університету

Д.т.н., професор Дивак М.П.



«29»  
2024р.

*М.П. Дивак*

### АКТ

про впровадження результатів дисертаційної роботи «Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад» здобувачки наукового ступеня Ліп'яніної-Гончаренко Христини Володимирівни при виконанні держбюджетних та науково-дослідних робіт Західноукраїнського національного університету

Даний акт складений з метою підтвердження того, що результати дисертаційної роботи Ліп'яніної-Гончаренко Х. В. «Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад» апробовано та впроваджено при виконанні наступних держбюджетних і науково-дослідних робіт:

1. Держбюджетна наукова робота молодих учених № 0119U100755 «Методи та засоби структурно-статистичної ідентифікації ієрархічних об'єктів за характерними точками їх контурів» (термін виконання: 01.01.2019 – 31.12.2020 рр.). Впроваджені результати сприяли розробці методів аналізу та ідентифікації об'єктів з урахуванням їхньої структурної особливості, що підвищило точність аналізу.

2. Держбюджетна наукова робота молодих учених № 0124U00002 «TruScanAI: інструмент виявлення фейкової інформації на основі технологій ШІ для боротьби з дезінформацією» (термін виконання: 01.01.2024 – 31.12.2025 рр.). Розроблені методи та моделі аналізу даних для ідентифікації неправдивої інформації дозволяють ефективно відокремлювати дезінформаційні елементи в соціальних мережах та інших інформаційних ресурсах.

3. Науково-дослідна робота ІОСУ-2023 «К» «Інтелектуальні методи, моделі та технології соціально-економічного розвитку територіальних громад в умовах

сьогодення» (протокол №4 від 09.11.2022 р., термін виконання: 01.2023 12.2027 рр.). Впровадження методів і моделей аналізу сприяє розробці інноваційних технологій для покращення соціально-економічного розвитку територіальних громад.

4. Договір БФ/4-2021 від 01.06.2021 р. «Виконання завдань перспективного плану розвитку наукового напрямку “Технічні науки” Західноукраїнського національного університету» (номер державної реєстрації 0121U114705, термін виконання: 06.2021 12.2025 рр.). Результати дисертаційної роботи забезпечили підвищення ефективності розробок у сфері автоматизації та інтелектуального аналізу великих обсягів даних.

Впроваджені теоретичні основи, моделі, методи та засоби аналізу даних у сукупності сприяють підвищенню ефективності й достовірності інтелектуального аналізу соціально-економічних даних територіальних громад.

Начальник науково-дослідної частини  
д.е.н., професор



Віта СЕМАНЮК



Затверджую  
 проректор з науково-педагогічної роботи  
 Західноукраїнського національного університету  
 к.е.н., доцент

Островерхов В. М.  
 2024р

Акт

про впровадження в освітній процес Західноукраїнського національного університету (ЗУНУ) результатів дисертаційної роботи здобувача наукового ступеня доктора технічних наук Ліп'яніної-Гончаренко Христини Володимирівни на тему «Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад»

Ми, комісія у складі: декана факультету комп'ютерних інформаційних технологій, доцента, к.т.н. Якименка І.З, в.о. завідувача кафедри інформаційно-обчислювальних систем і управління, доцента, к.т.н. Васильків Н.М, гаранта освітньо-професійної програми «Штучний інтелект» першого (освітнього) рівня, к.т.н., доцента, Ковалю В.С., гаранта освітньо-професійної програми «Комп'ютерні науки» другого (освітнього) рівня, к.т.н., доцента, Загородньої Д.І., гаранта освітньо-наукової програми «Комп'ютерні науки» третього (освітньо-наукового) рівня, д.т.н., професора Комара М.П. склали цей акт про те, що результати дисертаційної роботи Ліп'яніної-Гончаренко Х.В. на тему «Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу економічних даних територіальних громад» впроваджено в освітньому процесі ЗУНУ для підготовки здобувачів вищої освіти на першому, другому (освітніх) та третьому (освітньо-науковому) рівнях за спеціальністю «Комп'ютерні науки».

Результати впроваджені у навчальні курси, зокрема:

- «Інтелектуальний аналіз даних» — використано методи машинного навчання, що забезпечують якісний аналіз даних, а також алгоритми класифікації, аналізу та прогнозування.
- «Машинне навчання» — інтеграція ансамблевих методів навчання, що підвищують точність обробки соціально-економічних даних.
- «Методи та засоби штучного інтелекту» — реалізовано технології інтелектуальної обробки даних, що дозволяють використовувати сучасні алгоритми обробки та класифікації текстів.
- «Інтелектуальна обробка тексту та природної мови» — використано методи обробки природної мови для аналізу контексту та визначення тональності текстів для боротьби з дезінформацією.



— «Методи та засоби обробки зображень» — впроваджено методи інтелектуальної обробки зображень, що застосовуються у задачах класифікації та розпізнавання об'єктів на зображеннях культурної спадщини.

— «Інтелектуальна обробка даних у складних системах» — використано підходи до обробки багатовимірних даних з різних джерел, що підвищують ефективність управління територіальними громадами.

— «Методи обробки й аналізу експериментальних даних» — реалізовано адаптивні методи прогнозування та аналізу експериментальних даних для соціально-економічних досліджень територіальних громад.

Декан факультету комп'ютерних  
інформаційних технологій,  
к.т.н., доцент

Якименко І.І.

В.о. завідувача кафедри  
інформаційно-обчислювальних систем і управління,  
к.т.н.

Васильків Н.М.

Гарант освітньо-професійної програми  
«Штучний інтелект» першого (освітнього) рівня,  
к.т.н. доцент

Коваль В.С.

Гарант освітньо-професійної програми  
«Комп'ютерні науки» другого (освітнього) рівня,  
к.т.н., доцент

Загородня Д.І.

Гарант освітньо-наукової програми  
«Комп'ютерні науки» третього (освітньо-наукового) рівня,  
д.т.н., професор

Комар М.П.