

ВІДГУК

офіційного опонента, доктора технічних наук, професора,
доцента кафедри комп'ютерних наук

Чернівецького національного університету імені Юрія Федьковича,

Угрин Дмитра Ілліча

на дисертаційну роботу

Ліп'яніної-Гончаренко Христини Володимирівни

«Теоретичні та прикладні засади інформаційної технології інтелектуального
аналізу соціально-економічних даних територіальних громад»,
подану на здобуття наукового ступеня доктора технічних наук
за спеціальністю 05.13.06 – інформаційні технології

1. Актуальність теми дисертації

Сучасні інформаційні технології відіграють ключову роль у вдосконаленні процесів управління соціально-економічною інфраструктурою територіальних громад (ТГ). Вони значно підвищують ефективність прийняття рішень через інтегрований аналіз великих масивів структурованих, неструктурованих і напівструктурованих даних, знижують часові витрати на їх обробку, забезпечують прозорість використання ресурсів, зміцнюють соціальну та економічну стійкість громад і створюють умови для швидкої адаптації до кризових викликів. Упровадження таких технологій дозволяє оперативно реагувати на зміни в економічній, екологічній і соціальній сферах, сприяє точному прогнозуванню та моделюванню сценаріїв розвитку, залученню інвестицій і формуванню комфортного середовища для мешканців.

У процесі управління ТГ виникає нагальна потреба в оперативній обробці великих і гетерогенних масивів соціально-економічних даних, динамічному прийнятті рішень та точному прогнозуванні. Умови постійної невизначеності економічних, демографічних та екологічних показників ускладнюють процес управління, що зумовлює потребу в інтелектуальних технологіях підтримки рішень. Саме ці аспекти визначають актуальність створення адаптивних систем, здатних інтегрувати різноманітні джерела даних у реальному часі та ефективно працювати в умовах нестаціонарності процесів.

Управлінські завдання ТГ нерідко носять міждисциплінарний характер і потребують застосування сучасних підходів, таких як машинне навчання, обробка природної мови, комп'ютерне бачення та адаптивні алгоритми

прогнозування. Тому розроблення комплексної інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад набуває особливої наукової та прикладної значущості.

Актуальність обраної проблематики підтверджується її впровадженням у науково-дослідні роботи, освітні програми, а також практичним застосуванням в органах місцевого самоврядування та бізнес-структурах. Тематика дисертації Ліп'яніної-Гончаренко Х.В. гармонійно поєднується з пріоритетними напрямками розвитку науки і техніки, визначеними постановою Кабінету Міністрів України № 942 від 7 вересня 2011 р. (у редакції від 9 травня 2023 р. № 463), зокрема у сферах «Інтелектуальні інтерактивні інформаційно-аналітичні системи» та «Методи й засоби інформаційно-аналітичного забезпечення процесів прийняття рішень».

Дослідження, наведені в дисертації, здійснювалися в рамках міжнародних та національних науково-дослідних програм і проєктів. Серед них — програми Erasmus+: KA2 CBHE «Міждоменні компетенції для забезпечення здорової та безпечної роботи у 21 столітті» (WORK4CE), Creative Europe: AURA, ERASMUS-EDU-2023-PI-ALL-INNO «My Farm». Крім того, робота базується на результатах держбюджетних науково-дослідних тем, таких як «TruScanAI: інструмент виявлення фейкової інформації на основі технологій ШІ для боротьби з дезінформацією» та «Інтелектуальні методи, моделі та технології соціально-економічного розвитку ТГ в умовах сьогодення».

2. Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих в дисертаційній роботі

Наукові положення, висновки та рекомендації, наведені в дисертаційній роботі, базуються на ґрунтовному застосуванні сучасних теоретичних і практичних методів, що забезпечують високий рівень обґрунтованості отриманих результатів. Для досягнення поставленої мети було використано широкий спектр інструментів, які відповідають вимогам міждисциплінарного підходу до вирішення складних задач управління територіальними громадами.

Теоретичний аналіз літературних джерел та існуючих підходів став основою для визначення ключових проблем і формування концептуальних засад розробки нових інформаційних технологій для обробки соціально-економічних даних. Математичне моделювання було використано для

створення інтелектуальних моделей і методів, таких як ансамблеві адаптивні предиктори, що забезпечують високу точність прогнозування та класифікації навіть у складних умовах нестационарності даних. Особливу увагу приділено розробці методів формування навчальних вибірок, що підвищують ефективність роботи алгоритмів машинного навчання.

Інструменти машинного навчання, включно з сучасними алгоритмами глибокого навчання та адаптивної оптимізації, були інтегровані для вирішення задач прогнозування, класифікації та оптимізації управлінських рішень у соціальній, екологічній і бізнес-сферах територіальних громад. Для перевірки працездатності й ефективності запропонованих методів проводилося експериментальне моделювання та комп'ютерні симуляції з використанням як реальних, так і синтетичних даних. Це дозволило оцінити адаптивність і масштабованість запропонованих підходів.

Верифікація та підтвердження практичної значущості результатів були здійснені за допомогою статистичних методів аналізу, які забезпечили об'єктивність і точність порівняння запропонованих підходів з існуючими. Зокрема, проводилося порівняння точності прогнозування, швидкості роботи алгоритмів та ефективності використання обчислювальних ресурсів. Дослідження продемонстрували, що запропоновані методи мають суттєві переваги в контексті обробки великих обсягів даних і прийняття управлінських рішень у режимі реального часу.

Завдяки комплексному підходу до вирішення поставлених завдань, використані в роботі методи забезпечили високий рівень достовірності отриманих результатів і їх відповідність сучасним науковим стандартам. Це підкреслює наукову та практичну значущість розроблених моделей, методів і технологій, що мають потенціал для впровадження в практику управління територіальними громадами та подальшого розвитку в суміжних наукових напрямках.

3. Наукова новизна отриманих в дисертації результатів

В дисертаційній роботі авторкою отримано результати, наукова новизна яких полягає у наступному:

1) вперше запропоновано узагальнені принципи синтезу інформаційної технології інтелектуального аналізу соціально-економічних даних ТГ, особливістю якої є здатність інтегрувати різноманітні 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 нестаціонарними 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 нестаціонарних процесів на основі RFM 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 нестаціонарних 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 нестаціонарних 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.

4. Зв'язок матеріалів кандидатської дисертації з докторською дисертацією

Наукові положення, результати, висновки, рекомендації та інші матеріали наукових досліджень, представлені та захищені здобувачкою у кандидатській дисертації, не стали предметом розгляду та не були використані в дослідженнях, представлених у поданій до захисту докторській дисертації Ліп'яніної-Гончаренко Х.В.

5. Повнота викладення результатів досліджень у наукових публікаціях за темою докторської дисертації

Наукові положення, результати, висновки та рекомендації, представлені у дисертаційному дослідженні та подані до захисту, опубліковані в необхідному обсязі.

За темою дисертації з викладенням основних її результатів опубліковано 49 наукових праць, з них: 15 статей у наукових виданнях Переліку наукових фахових видань України; 7 статей у періодичних виданнях категорії «А» Переліку наукових фахових видань України або у закордонних виданнях, проіндексованих у базах даних Web of Science Core Collection та/або Scopus; 4 монографії (розділи у колективних монографіях); 22 статті у матеріалах міжнародних конференцій, індексованих у наукометричних базах Scopus та

Web of Science; а також 1 публікацію, яка додатково відображає наукові результати дисертації.

Аналіз внеску авторки в публікаціях з питань, висвітлених в дисертації, показав, що внесок Ліп'яніної-Гончаренко Х.В. є вирішальним, усі основні наукові результати та основні матеріали докторської дисертації опубліковані в наукових публікаціях здобувачки.

Зазначений перелік публікацій відповідає вимогам МОН України до опублікування результатів дисертації на здобуття наукового ступеня доктора наук.

6. Практичне значення результатів дисертаційної роботи

Розроблена в дисертаційній роботі інформаційна технологія інтелектуального аналізу соціально-економічних даних територіальних громад (ТГ) усуває обмеження традиційних підходів до обробки даних, забезпечуючи адаптивність до динамічних змін у соціально-економічних процесах. Використання сучасних методів машинного навчання уможливорює оперативне реагування на зміни в показниках, враховуючи специфіку різномірних джерел даних, таких як статистичні звіти, відкриті дані й результати соціологічних опитувань. Це дозволяє підвищити точність прогнозування ризиків, що є критично важливим для прийняття ефективних управлінських рішень в умовах сучасних викликів.

Інтеграція різномірних джерел даних у рамках запропонованої технології забезпечує виконання завдань комплексного аналізу та моделювання стану територіальних громад. Такий підхід сприяє виявленню тенденцій, прогнозуванню змін і оцінюванню впливу прийнятих рішень на соціально-економічний розвиток громад. Розробка модулів для обробки як структурованих, так і неструктурованих даних спрощує інтеграцію інформації з різних джерел, усуваючи технічні бар'єри. Завдяки прозорості та пояснюваності результатів аналізу, розроблена система підвищує обґрунтованість і ефективність управлінських рішень.

Практична реалізація запропонованої технології не лише забезпечує високу точність і швидкість аналізу даних, але й сприяє покращенню комунікації між зацікавленими сторонами завдяки доступності результатів для кінцевих користувачів. Впровадження технології сприяє сталому розвитку

громад, підвищенню соціальної стійкості та адаптивності до кризових ситуацій.

Результати дисертаційного дослідження впроваджено в освітній процес Західноукраїнського національного університету (ЗУНУ) для підготовки здобувачів вищої освіти за спеціальністю «Комп'ютерні науки». Зокрема, вони використані для вдосконалення навчальних курсів: «Інтелектуальний аналіз даних», «Машинне навчання», «Методи та засоби штучного інтелекту», «Інтелектуальна обробка тексту та природної мови», «Методи та засоби обробки зображень». На третьому (освітньо-науковому) рівні результати дослідження застосовуються у дисциплінах «Інтелектуальна обробка даних у складних системах» та «Методи обробки й аналізу експериментальних даних».

Окрім освітніх ініціатив, результати дисертації успішно впроваджені у виконання науково-дослідних робіт та проєктів, таких як:

- 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» (№ 101008547);
- ERASMUS-EDU-2023-PI-ALL-INNO: «My Farm» (Proposal № 101140288);
- Державний договір БФ/4-2021 «Виконання завдань перспективного плану розвитку наукового напрямку “Технічні науки” ЗУНУ» (реєстрація № 0121U114705);
- Проєкт «TruScanAI» – інструмент виявлення фейкової інформації на основі технологій штучного інтелекту для боротьби з дезінформацією (реєстрація № 0117U003871);
- Держбюджетна наукова робота ІОСУ2023 «К»: «Інтелектуальні методи, моделі та технології соціально-економічного розвитку ТГ в умовах сьогодення».

Отримані результати є важливим внеском у розвиток інтелектуальних методів аналізу даних, створюючи нові можливості для наукових досліджень, освітніх програм та практичних впроваджень у сфері управління територіальними громадами.

7. Оцінка змісту дисертаційної роботи

Дисертаційна робота складається з анотації, вступу, семи розділів, висновків, списку використаних джерел та чотирьох додатків. Загальний обсяг роботи становить 299 сторінок основного тексту. Список використаних джерел включає 475 найменувань, що свідчить про глибокий аналіз сучасних досягнень у досліджуваній галузі. Зміст дисертації повністю відповідає визначеній меті та завданням, відзначається структурованістю, логічністю та завершеністю викладу.

У *вступі* автор обґрунтовує актуальність обраної теми, висвітлює зв'язок роботи з науковими програмами, планами та темами, формулює мету та завдання дослідження, визначає об'єкт і предмет роботи. Крім того, описано наукову новизну та практичну цінність отриманих результатів, наведено інформацію про апробацію роботи та публікації за темою дисертації, а також викладено структуру роботи. Авторка ретельно аналізує актуальність дослідження в умовах сучасних викликів, зокрема кризових ситуацій, що виникають у соціально-економічних процесах територіальних громад.

У *першому* розділі проведено комплексний аналіз літературних джерел щодо сучасного стану і перспектив розвитку інтелектуального аналізу соціально-економічних даних. Виявлено ключові проблеми, такі як гетерогенність і динамічність даних, нестаціонарність часових рядів і недостатня інтеграція структурованих, напівструктурованих і неструктурованих даних. Особливу увагу приділено питанням оперативного реагування на глобальні виклики, включаючи економічні кризи, пандемії та військові конфлікти. Постановка наукової проблеми підкріплюється визначенням гіпотези дослідження та обґрунтуванням необхідності створення інтегрованої інформаційної технології для аналізу даних у режимі реального часу.

Другий розділ присвячений розробленню узагальненої інформаційної технології для інтелектуального аналізу соціально-економічних даних територіальних громад. Авторка формулює базові методологічні засади, що охоплюють інтеграцію різних типів джерел даних (структурованих, напівструктурованих, неструктурованих) та адаптивне моделювання нестаціонарних процесів. Запропоновано методологію, яка включає адаптивні методи класифікації, кластеризації та прогнозування, а також інноваційні

підходи до формування навчальних вибірок і ансамблевого прогнозування багатовимірних даних.

Третій розділ детально аналізує методи класифікаційного аналізу кількісних даних. Розглянуто можливості таких алгоритмів, як Logistic Regression, SVM, Random Forest і Gradient Boosting. Особливу увагу приділено оцінці інвестиційних ризиків і виявленню фіктивних підприємств. Експериментальні результати показують високу точність (до 99,7%) і значне скорочення часу прийняття управлінських рішень.

Четвертий розділ зосереджено на класифікації текстових даних. Авторка інтегрує технології обробки природної мови (NLP) із сучасними алгоритмами машинного навчання, такими як LSTM і трансформери. Запропоновані підходи використовуються для виявлення дезінформації та розробки HR-проектів. Досягнута точність у задачах класифікації фейкових новин становить близько 93%, що дозволяє значно скоротити час на пошук персоналу.

У *п'ятому* розділі представлено метод кластерного аналізу, орієнтований на групування об'єктів за схожими характеристиками. Розглянуто алгоритми k-means, DBSCAN і GMM для формування споживчого кошика та визначення альтернативних постачальників. Дослідження підтверджують ефективність методу (до 99,61%) для розподілу ресурсів і прийняття управлінських рішень.

Шостий розділ присвячений прогнозуванню соціально-економічних даних за допомогою алгоритмів ARIMA, LSTM і XGBoost. Описано практичні кейси, які демонструють скорочення часу реагування на зміни з декількох днів до годин, що підвищує ефективність управління ресурсами.

Сьомий розділ пропонує гібридний підхід, що поєднує класифікацію, кластеризацію та прогнозування в багаторівневій системі. Запропоновані алгоритми забезпечують точність прогнозування (97–98%) і скорочують час прийняття рішень до 1–2 годин.

Загальні висновки дисертаційної роботи належним чином узагальнюють проведені дослідження та отримані результати.

Список використаних джерел відображає сучасний рівень наукових досліджень в предметній області дисертаційної роботи, є достатньо повним та змістовним.

Представлені в роботі *додатки* містять інформацію, необхідну для забезпечення повноти наукових положень, висновків, рекомендацій, результатів експериментальних досліджень. Додатки мають необхідні посилання в тексті для розкриття викладеного змісту та підтвердження проведених досліджень.

Дисертаційна робота відповідає вимогам «Порядку присудження та позбавлення наукового ступеня доктора наук», затвердженого Постановою КМУ №1197 від 17 липня 2021 р., та вимогам наказу МОН України від 12 січня 2017 р. № 40 «Про затвердження вимог до оформлення дисертації».

8. Зауваження та дискусійні питання

Даючи загальну позитивну оцінку науковим положенням, результатам, рекомендаціям, наведеним у дисертаційній роботі, та рівню їх обґрунтування, слід відзначити наступні зауваження та дискусійні питання:

1. У першому розділі, зокрема в п. 1.3 для опрацювання нестационарних даних використовуються бустингові методи, по яких проведено широкий аналіз, але не порівняно з іншими сучасними методами.

2. На мою думку, автор не завжди коректно оперує окремими термінами, наприклад використовує «екстракція знань» (с. 55) замість загально прийнятого «видобування знань», «комп'ютерне бачення» замість «комп'ютерний зір» (с. 92) та ін.

3. На с. 68 зазначено «Крім того, важливо приділяти увагу інтеграції різних джерел даних та забезпеченню етичності і безпеки використання інтелектуальних систем у маркетингових процесах», але в подальшому в роботі питання «етичності і безпеки» не розглядаються.

4. Результати моделювання демонструють 100% точність групування за методами RFM-оцінки та K-Means, а також високі показники точності на тренувальній та тестовій вибірках для алгоритмів Logistic Regression та Naive Bayes. Такий рівень точності є надзвичайно рідкісним у реальних умовах і може свідчити про переоснащення моделі (overfitting) або можливе витік даних (data leakage).

5. На с. 101 приведена таблиця 2.2 «Тестування на стаціонарність», яку через великий розмір слід було перенести до додатків. Крім того, до таблиці слід було внести ознаку сезонності, що впливає на коефіцієнти параметрів.

6. Опис методів класифікаційного аналізу є досить загальним і не завжди містить детальні пояснення щодо конкретних кроків, алгоритмів та їх інтеграції. Наприклад, як саме методи обробляють різні типи даних (структуровані, неструктуровані, напівструктуровані) та, як вони взаємодіють у рамках загальної системи.

7. Опис процесу інтеграції даних ($D_{\text{інтегровані}} = D_1 \oplus D_2 \oplus D_3$) є досить загальним і не надає інформації про конкретні методи чи інструменти, які використовуються для злиття різнорідних даних. Також відсутні деталі щодо обробки відсутніх даних, аномалій чи шуму в даних.

8. У дослідженні під різними номерами повторюються формули, зокрема: (3.1) на с. 130, (4.1) на с. 158, (5.4) на с. 184 та (6.4) на с. 208, а також (2.2) на с. 91, (5.2) на с. 183 та (6.2) на с. 208.

9. У дисертації по запропонованій методології, що включає класифікаційний, кластерний і гібридний аналіз даних та прогностичні моделі, не достатньо розкрито опис та реалізацію гібридного аналізу.

10. У роботі зустрічаються граматичні та стилістичні неточності, зокрема на с. 42 «у забезпеченні сталого та функціонування середовища громади», на с. 44 «інтелектуального управління соціально–економічними для ТГ» та ін.

Зазначені зауваження не здійснюють визначального впливу на зміст та результати дисертаційного дослідження та є предметом наукової дискусії.

7. Загальні висновки

Дисертаційна робота Ліп'яніної-Гончаренко Христини Володимирівни «Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад» є завершеним науковим дослідженням, в якому вирішено актуальну науково-прикладну проблему щодо розроблення методологічних основ інформаційної технології інтелектуального аналізу соціально-економічних даних ТГ.

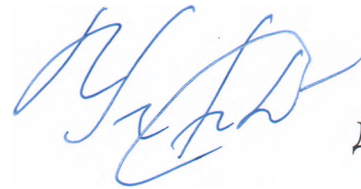
Представлені в дисертаційній роботі наукові положення мають відповідне теоретичне обґрунтування та належним чином підтверджені на практиці.

05.13.06 -

, , . 6, 7, 8, 9 «
»,
2021 ., , -

1197 17

05.13.06 -



Д. УГРИН

Підпис Угрин Д засвідчую
Учений секретар Чернівецького національного
університету імені Юрія Федьковича
Сироговерка Н. С.
"13" січня 2021

