

ВІДГУК

офіційного опонента,

доктора технічних наук, професора, професора кафедри інформатики та програмної інженерії Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського"

Стеценко Інни Вячеславівни

на дисертацію 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

Актуальність теми дослідження та зв'язок з науковими програмами, планами та темами

В управлінні територіальною громадою спостерігається потреба в оперативній обробці великих та гетерогенних масивів соціально-економічних даних, що відповідала б сучасним вимогам щодо динамічності процесів прийняття рішень та точності прогнозування. Спонукає до вдосконалення процесів обробки даних також збільшення доступних для обробки джерел даних та обсягів даних, що характеризують життєдіяльність громад. Різноманітність даних, що використовують для аналізу соціально-економічного стану територіальної громади, та різноманітність джерел даних, з яких можуть бути почерпнуті дані з урахуванням сучасних інформаційних технологій, з одного боку, відкривають нові можливості для інтелектуального аналізу стану, з іншого боку – висувають нові виклики щодо спільного та узгодженого використання великої кількості різноманітних джерел для досягнення поставлених цілей. Тому актуальною є розробка інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад для забезпечення комплексної комп'ютерної підтримки прийняття управлінських рішень.

Інформаційні технології суттєво впливають на вдосконалення процесів управління соціально-економічною інфраструктурою територіальних громад та підвищують ефективність прийняття рішень завдяки комплексному аналізу великих обсягів структурованих, неструктурованих та напівструктурованих даних, а також зменшують витрати зусиль та часу на обробку інформації, забезпечують прозорість використання ресурсів, підсилюють соціальну та економічну стійкість громад, створюють умови для швидкої адаптації в умовах кризових викликів. Впровадження нових інтелектуальних технологій обробки даних дає змогу оперативно відстежувати зміни в економіці та соціальній сфері, прогнозувати подальший розвиток подій та моделювати сценарії розвитку з урахування прогнозованих значень.

Дисертаційне дослідження Ліп'яніної-Гончаренко Х. В. містить обґрунтування, концепцію, методологію та методи, що утворюють інформаційну технологію інтелектуального аналізу соціально-економічних

даних територіальної громади, є актуальною та важливою для комплексного аналізу життєдіяльності громади.

Тематика дисертаційного дослідження відповідає пріоритетному напрямку розвитку науки і техніки «Інтелектуальні інтерактивні інформаційно-аналітичні системи» та, частково, напрямку «Методи та засоби інформаційно-аналітичного та нормативно-методичного забезпечення процесів прийняття рішень у сфері національної безпеки і оборони», визначеним постановою Кабінету Міністрів України від 7 вересня 2011 р. № 942 (в редакції постанови Кабінету Міністрів України від 9 травня 2023 р. № 463).

Дисертаційне дослідження виконувалось у межах низки міжнародних та державних науково-дослідних проєктів і програм, зокрема Erasmus+: KA2 CBHE «Міждоменні компетенції для забезпечення здорової та безпечної роботи у 21 столітті» (WORK4CE), Creative Europe: AURA та ERASMUS-EDU-2023-PI-ALL-INNO «My Farm», а також у держбюджетних науково-дослідних темах, зокрема «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».

Ступінь обґрунтованості наукових положень, висновків і рекомендацій

Ступінь обґрунтованості наукових методів, положень, висновків і рекомендацій дисертаційної роботи визначається коректністю виконаних теоретичних та експериментальних досліджень, практичним застосуванням одержаних результатів у діяльність ІТ-компанії «МагнетікВан Муніципальні Технології», Великоберезовицької територіальної громади, громадська організація «Прогресивні люди» та Департаменту цифрової трансформації Тернопільської ОДА.

Наукові результати, висновки та практичні рекомендації дисертаційного дослідження Ліп'яніної-Гончаренко Х. В. отримані авторкою в рамках 5 міжнародних та державних науково-дослідних проєктів, що засвідчує опрацювання основних концепцій у тісній співпраці з науковою спільнотою.

Сформульовані в дисертації положення, висновки й рекомендації ґрунтуються на теорії систем та системного аналізу (зокрема принципах ієрархічності та декомпозиції), теоретико-множинному описі систем, методах моделювання, прогнозування, інтелектуального аналізу даних, а також методах експериментального дослідження.

Наукові положення та висновки логічно структуровані на основі комплексного аналізу та узагальнення відомих і нових одержаних результатів, а також підтверджені експериментальним дослідженням запропонованих методів та модулів розробленої інформаційної технології.

Наукова новизна результатів дослідження

Дисертаційне дослідження спрямовано на вирішення важливої наукової проблеми автоматизації обробки соціально-економічних даних, отриманих з різнорідних джерел, в реальному часі з метою їх комплексного використання для прийняття обґрунтованих рішень в управлінні територіальною громадою, що забезпечують сталий розвиток громади.

Отримано такі нові наукові результати:

1) вперше запропоновано узагальнені принципи синтезу інформаційної технології інтелектуального аналізу соціально-економічних даних територіальної громади, що забезпечує підвищення соціальної стійкості інфраструктури громад за рахунок комплексного використання гетерогенних даних, адаптивного вибору методів аналізу та формування управлінських рішень;

2) вперше розроблено методологію інтелектуального аналізу соціально-економічних даних територіальної громади як сукупність методів класифікаційного, кластерного та гібридного аналізу, а також методів прогнозування, особливістю якої є адаптивність до обробки нестационарних даних, отриманих з різнорідних джерел даних, в реальному часі для оперативного прийняття управлінських рішень, що дає змогу підвищити точність прогнозування та класифікації;

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, для підвищення точності опрацювання 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 дає змогу підвищити точність прогнозів багатовимірних нестационарних процесів;

5) удосконалено метод класифікаційного аналізу кількісних соціально-економічних даних, який, на відміну від відомих підходів, забезпечує інтеграцію структурованих, неструктурованих і напівструктурованих даних, що дає змогу 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 їх аналіз за допомогою технологій обробки природної мови, що дає змогу 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 зі схожими характеристиками, що дає змогу підвищити точність розподілу ресурсів для реалізації управлінських рішень;

8) удосконалено метод прогнозування соціально-економічних даних, який, на відміну від відомих підходів, забезпечує достатньо точне прогнозування динамічних процесів шляхом інтеграції структурованих, неструктурованих і напівструктурованих даних із застосуванням адаптивних методів інтелектуального аналізу, що дає змогу 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 багаторівневному підходу до аналізу даних, що дає змогу підвищити гнучкість та адаптивність процесу прийняття управлінських рішень.

Практичне значення результатів дисертаційної роботи

Практичне значення дисертаційної роботи полягає у розробці інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад, що забезпечує автоматизацію збору гетерогенних даних з різних джерел, моделювання нестаціонарних соціально-економічних процесів; прогнозування соціально-економічних показників у реальному часі; застосування сучасних методів інтелектуального аналізу даних. Це дозволяє оперативно реагувати на зміни у соціально-економічних показниках життєдіяльності громади та приймати обґрунтовані управлінські рішення в умовах економічної нестабільності, екологічних ризиків чи військової агресії, забезпечуючи стабільний розвиток та якість життя громадян.

Розроблені методи та інформаційна технологія впроваджено у науково-дослідні проекти Erasmus+: «WORK4CE», «My Farm», та Creative Europe «AURA». Крім того, результати використано у державних наукових роботах, зокрема TruScanAI, «Методи та засоби структурно-статистичної ідентифікації», «Інтелектуальні методи соціально-економічного розвитку ТГ». Практична апробація проведена в ІТ-компанії «МагнетікВан» (м. Тернопіль), Тернопільській обласній державній адміністрації та Великоберезовицькій територіальній громаді. Запропоновані методи дозволили значно скоротити час аналізу даних та прийняття управлінських рішень, підвищивши точність прогнозів та ефективність управління ресурсами. Використання розроблених технологій сприяло оптимізації процесів управління, підвищенню соціальної стійкості та забезпеченню сталого розвитку територіальних громад, що підтверджено позитивними відгуками та експертними оцінками впроваджених рішень.

Повнота викладу в наукових публікаціях

Усі основні положення та найбільш важливі результати дисертації, подані до захисту, опубліковані в необхідному обсязі у фахових наукових

виданнях України та закордонних виданнях, пройшли відповідну апробацію на міжнародних науково-технічних конференціях. 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.

Загальна кількість публікацій та їх зміст в достатній мірі висвітлюють результати дисертаційної роботи. Результати аналізу публікацій здобувача за темою дисертаційної роботи вказують на повноту викладу основних наукових положень та дотримання здобувачем принципів академічної доброчесності у процесі підготовки докторської дисертації.

Зміст дисертації та відповідність її встановленим вимогам

Дисертація за змістом та отриманими результатами повністю відповідає паспорту спеціальності 05.13.06 – інформаційні технології.

У вступі наведено обґрунтування актуальності вибору теми дослідження, зазначено зв'язок з науковими програмами, планами, темами, сформульовано мету і завдання дослідження, визначено об'єкт та предмет дослідження, описано наукову новизну та практичну цінність одержаних результатів дослідження, наведено результати впровадження дисертаційної роботи, а також наведено відомості про апробацію роботи та публікації за темою дисертації із виділенням особистого внеску здобувача та описано структуру дисертаційної роботи.

Перший розділ дисертації присвячено аналізу сучасного стану інтелектуального аналізу соціально-економічних даних територіальних громад, наведено огляд джерел літератури та виявлено основні проблеми, пов'язані з гетерогенністю та динамічністю великих обсягів даних, нестаціонарністю часових рядів, а також із необхідністю інтеграції структурованих, напівструктурованих і неструктурованих даних. Розглянуто питання реагування на кризові виклики та необхідність оперативної обробки даних. Обґрунтовано доцільність розроблення комплексної інформаційної технології, здатної обробляти соціально-економічні дані в режимі реального часу, для формування обґрунтованих управлінських рішень.

Другий розділ роботи містить опис узагальненої інформаційної технології інтелектуального аналізу соціально-економічних даних ТГ та базових методологічних засад. У ньому обґрунтовано загальний принцип синтезу, що охоплює інтеграцію даних із різних джерел (структурованих, напівструктурованих і неструктурованих). Запропоновано методологію інтелектуального аналізу даних, яка містить методи класифікації,

кластеризації та прогнозування. Представлено оригінальний підхід до ансамблевого прогнозування багатовимірних даних у режимі реального часу.

У третьому розділі описано розроблений авторкою метод класифікаційного аналізу кількісних соціально-економічних даних територіальних громад. Проаналізовано можливості та обмеження алгоритмів машинного навчання Logistic Regression, SVM, Random Forest, Gradient Boosting тощо. Наведено приклади застосування розробленого методу для оцінювання інвестиційних ризиків та виявлення фіктивних підприємств. Наведено експериментальні результати, які підтверджують високу точність та істотне скорочення часу прийняття управлінських рішень, що свідчить про практичну цінність запропонованого підходу.

У наступному розділі описано розроблений авторкою метод класифікаційного аналізу текстових соціально-економічних даних, який дає змогу застосувати для обробки даних сучасні технології обробки природної мови (NLP) та методи машинного навчання. Наведено приклад виокремлення методу для виявлення дезінформації та формування HR-проектів із застосуванням глибоких нейронних мереж (LSTM, трансформери).

П'ятий розділ роботи містить опис розробленого авторкою методу кластерного аналізу соціально-економічних даних ТГ, який реалізує підхід до групування об'єктів (бізнес, соціальна, логістична та інформаційна інфраструктури) за схожими характеристиками. Розглянуто застосування таких алгоритмів DBSCAN, GMM для формування споживчого кошика та пошуку альтернативних постачальників та наведені результати експериментів, що підтверджують дієвість методу для оперативного розподілу ресурсів та прийняття управлінських рішень у громадах.

У наступному розділі наведено розроблений метод прогнозування соціально-економічних даних територіальних громад із використанням сучасних алгоритмів ARIMA, LSTM, XGBoost та інших. Описано інтелектуальні системи, що надають можливість достатньо точно прогнозувати попит на товари / послуги та визначати локації для старту бізнесу на основі відеоданих. На прикладі кейсу «СмартМед» продемонстровано скорочення часу реагування на ринкові зміни, що забезпечує гнучкість управління ресурсами в громаді.

В останньому розділі зосереджено увагу на методі гібридного аналізу соціально-економічних даних ТГ, який поєднує класифікацію / кластеризацію з прогнозуванням у багаторівневій системі. Запропоновано інтелектуальні методи для класифікації рівня техногенних катастроф та прогнозування обсягів відходів у громадах. Експериментально доведено, що розроблений метод дає змогу достатньо швидко виконувати аналіз даних для прийняття рішень в управлінні територіальною громадою.

Висновки по роботі відповідають отриманим в роботі результатам. За своїм рівнем висновки відповідають вимогам, які висуваються до результатів докторської дисертації.

Автореферат відповідає змісту дисертації, написаний українською мовою з використанням сучасної української наукової термінології. Опубліковані в авторефераті положення співпадають з основними положеннями дисертаційної роботи.

Оцінка мови, стилю та оформлення дисертації й автореферату

Дисертаційна робота складається з анотації, вступу, шести розділів, висновків, списку використаних джерел та чотирьох додатків. Робота містить 299 сторінок основного тексту. Список використаних літературних джерел містить 475 найменувань.

Обсяг, структура, оформлення матеріалів дослідження в цілому відповідають 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».

Зауваження та дискусійні питання:

1. Показники соціальної стійкості територіальної громади, на підвищення яких спрямоване наукове дослідження, не визначені в тексті дисертації. За результатами дослідження не зазначено умови, за яких забезпечується стійкість соціально-економічної інфраструктури територіальної громади.

2. Опис процесів інтелектуальної обробки даних, які розроблені для ІТ, не формалізований. Є опис у дисертації множин даних, але теоретико-множинний опис далі не розвинений. Оператор, який утворює з різних типів даних єдину множину даних, не розкритий і в тексті дисертації називається по різному - як об'єднання у розділі 4, як злиття у розділі 5. Саме цей оператор, схоже, відповідає за інтеграцію даних, яка важлива в контексті всього наукового дослідження. Тому визначення цього оператора та дослідження його властивостей додало б значимості дисертаційному дослідженню.

3. У тексті використовується опис функціональних залежностей без вказівки на те, який вигляд має функція та яким чином вона може бути визначена, тому залишається незрозумілим як в практичних умовах відшукати зазначену функціональну залежність. Наприклад, у розділі 2 так визначена функція застосовності f , у розділі 4 функція f для формування команди проекту, у розділах 3, 4, 5 та 6 функція g визначення оптимального управлінського рішення.

4. Авторка констатує підвищення ефективності за рахунок використання методів на 80-99%. На мою думку, така оцінка не є справедливою, оскільки ефективність переважної більшості розроблених методів авторка порівнює з виконанням аналізу даних без програмних засобів. У разі порівняння з існуючими на поточний момент технологіями та програмними системами, які використовуються для аналогічного аналізу даних, такого високого підвищення ефективності точно не спостерігалось би. Таке порівняння з

провідними технологіями сучасності значно підсилило б обґрунтування наукових результатів здобувачки.

5. У таблиці 4.1 наведено порівняння з класичними підходами, але у науковій дисертації потрібно порівнювати з існуючими сучасними підходами.

6. У підрозділі 2.3 (стор.98) авторка зазначає, що «Методологія забезпечує системний підхід до аналізу соціально–економічних даних, дозволяючи адаптувати процес аналізу до специфічних умов ТГ, підвищуючи прийняття рішень та оптимізуючи управління ресурсами» і обґрунтовує цей висновок посиланням на свої публікації, в яких наведені приклади застосування методів для обробки соціально-економічних даних, але не методології в цілому.

7. У назвах підрозділів 3.2 вказано на опис методу, проте в них описано приклади застосування викладеного вище методу класифікаційного аналізу кількісних соціально–економічних даних територіальної громади. Кожне практичне застосування методу не може вважатись окремим новим методом. Підрозділ 4.3 – аналогічне зауваження.

8. Прогнозна найякісніша модель обирається у дисертаційному дослідженні за критерієм найменших квадратів, розрахованого на тестових даних. Проте відомі дослідження академіка Івахненка О.Г. та його учнів, які розробили МГУА-методи для пошуку найліпшої моделі, в яких, зокрема, доведено, що критерій мінімуму зсуву ліпше визначає істинну модель у порівнянні з критерієм найменших квадратів. По-друге, були проведені дослідження щодо найліпшого поділу на тренувальну та тестову частини та експериментально доведено, що для методу найменших квадратів оптимальним є поділ на частини 1/1. Авторка дисертації використовує поділ 4/1 без будь-якого обґрунтування.

9. Для багатьох процесів, що характеризують територіальну громаду, притаманною є періодичність, пов'язана з сезонними змінами. Наявність такої періодичності авторка не досліджує і не враховує в побудованих моделях, що, на мою думку, зменшує точність прогнозу.

10. Формулювання проблеми, на вирішення якої мало бути наведено у підрозділі 1.4, який названий “Висновки, спрямоване дисертаційне дослідження, постановка проблеми та задач дослідження”, та у висновках до роботи.

11. Зауваження до оформлення та термінології:

- на стор. 144, 155, 175, 180, 200, 205, 221, 252, 258 міститься вказівка на Додаток Г (список публікацій авторки) замість Додатку В (експертний висновок), водночас посилання на додаток В відсутнє в тексті дисертації;

- у тексті наявні такі одруки: на стор. 114 замість λ вказаний коефіцієнт α , на стор. 151 замість трансформер – трансформатор;

- зауваження до термінології: термін «матриця плутанини» ліпше замінити на «матриця невідповідностей (або матриця помилок)», термін «перенавчання» (стор.165) - на «повторне навчання» (бо перенавчання має інше тлумачення), терміни «F1-бал», «F1-оцінка» – на «F1-міра»;

- на стор. 169 є формулювання «метод реалізовано в експериментальних дослідженнях...», яке, оскільки метод не може бути реалізованим в експериментальних дослідженнях, ліпше замінити на «ефективність методу експериментально досліджено...».

Вказані зауваження не мають принципового значення та не зменшують наукової цінності дисертаційної роботи.

Загальні висновки

З урахуванням вищезазначеного можна стверджувати, що дисертаційна робота Ліп'яніної-Гончаренко Христини Володимирівни «Теоретичні та прикладні засади інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад» є завершеним науковим дослідженням, яке вирішує актуальну прикладну проблему автоматизації обробки соціально-економічних даних, отриманих з різнорідних джерел, в реальному часі з метою їх комплексного використання для прийняття обґрунтованих рішень в управлінні територіальною громадою, має вагомe теоретичне та практичне значення. Тема та зміст дисертаційної роботи повністю відповідають паспорту спеціальності 05.13.06 – інформаційні технології.

За обсягом досліджень, науковим рівнем, практичною цінністю отриманих результатів **дисертаційна робота відповідає** чинним вимогам до дисертації, зокрема, пп. 6, 7, 8, 9 «Порядку присудження та позбавлення наукового ступеня доктора наук», затвердженого Постановою КМУ №1197 від 17 липня 2021 р., а її авторка, **Ліп'яніна-Гончаренко Христина Володимирівна, заслуговує** на присудження їй наукового ступеня доктора технічних наук за спеціальністю 05.13.06 – інформаційні технології.

Офіційний опонент:

доктор технічних наук, професор,
професор кафедри інформатики та програмної
інженерії

Національного технічного університету України
"Київський політехнічний інститут
імені Ігоря Сікорського"



20 січня 2025 р.