

<https://doi.org/10.31891/2219-9365-2024-77-48>

УДК 004.9

ЛІП'ЯНИНА-ГОНЧАРЕНКО Христина

Західноукраїнський національний університет

<https://orcid.org/0000-0002-2441-6292>

e-mail: [kh.lipianina@wunu.edu.ua](mailto:kh.lipianina@wunu.edu.ua)

## УЗАГАЛЬНЕНИЙ ПРИНЦИП СИНТЕЗУ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ДАНИХ ТГ

Наразі актуальною проблемою є забезпечення ефективного аналізу соціально-економічних даних територіальних громад (ТГ), які характеризуються гетерогенністю, динамічністю та великим обсягом. Проведене дослідження існуючих підходів показало, що: наявні методи зазвичай орієнтовані на аналіз лише одного типу даних (структурованих, неструктурованих або напівструктурованих), а також не враховують динамічності соціально-економічних процесів, що знижує ефективність прийняття управлінських рішень. Тому метою загального дослідження є розробка інформаційної технології, яка інтегрує різноманітні дані, адаптується до змін у соціально-економічних процесах та забезпечує підтримку стратегічних рішень у ТГ. Метою даного дослідження є запропонування узагальнених принципів синтезу такої технології та удосконалення методів класифікаційного аналізу кількісних і текстових даних. Розроблений у статті підхід дозволяє інтегрувати структуровані, неструктуровані та напівструктуровані дані, адаптувати методи аналізу до специфіки соціально-економічних процесів і скоротити час прийняття управлінських рішень з днів до годин. Крім того, удосконалені методи класифікаційного аналізу кількісних і текстових даних забезпечують підвищення точності аналізу показників і виявлення закономірностей для підтримки управлінських рішень.

Ключові слова: соціально-економічні дані, територіальні громади, інформаційна технологія, класифікаційний аналіз, динамічність процесів.

LIPIANINA-HONCHARENKO Khrystyna

West Ukrainian National University

## GENERALIZED PRINCIPLE FOR SYNTHESIZING INFORMATION TECHNOLOGY FOR INTELLIGENT ANALYSIS OF SOCIO-ECONOMIC DATA OF TERRITORIAL COMMUNITIES

In conditions of rapid socio-economic changes and crisis situations, such as a pandemic or war, there is a need to develop effective information technologies for analyzing data of territorial communities (TC). This is due to the need to increase the resilience of the socio-economic infrastructure of communities and ensure its adaptation to the challenges of the time. The generalized principle of information technology synthesis of intelligent analysis of socio-economic data of TC is an important step towards the integration of heterogeneous data, the formation of adaptive approaches to their analysis and the adoption of informed management decisions.

Currently, an urgent problem is ensuring the effective analysis of socio-economic data of territorial communities (TCs), which are characterized by heterogeneity, dynamism, and large volumes. A review of existing approaches has shown that current methods are typically focused on analyzing only one type of data (structured, unstructured, or semi-structured) and do not consider the dynamic nature of socio-economic processes, reducing the efficiency of decision-making. Therefore, the overall aim of this research is to develop an information technology that integrates heterogeneous data, adapts to changes in socio-economic processes, and supports strategic decision-making in TCs. The aim of this study is to propose generalized principles for synthesizing such technology and to improve classification analysis methods for quantitative and textual data. The approach developed in this article enables the integration of structured, unstructured, and semi-structured data, adapts analytical methods to the specifics of socio-economic processes, and reduces decision-making time from days to hours. Additionally, the improved methods for classification analysis of quantitative and textual data enhance the accuracy of indicator analysis and the identification of patterns to support managerial decisions.

Keywords: socio-economic data, territorial communities, information technology, classification analysis, dynamic processes.

### Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

В умовах стрімких соціально-економічних змін та кризових ситуацій, таких як пандемія чи війна, постає необхідність розробки ефективних інформаційних технологій для аналізу даних територіальних громад (ТГ). Це обумовлено необхідністю підвищення стійкості соціально-економічної інфраструктури громад та забезпечення її адаптації до викликів часу. Узагальнений принцип синтезу інформаційної технології інтелектуального аналізу соціально-економічних даних ТГ є важливим кроком у напрямку інтеграції гетерогенних даних, формування адаптивних підходів до їх аналізу та прийняття обґрунтованих управлінських рішень.

У зв'язку з цим у даній статті представлено наступні розроблені пункти новизни:

1. Вперше запропоновані 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. 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.

3. 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.

Розроблені підходи спрямовані для аналізу соціально-економічних даних, що дозволяє не лише підвищити точність обробки даних, але й забезпечити можливість адаптації моделей аналізу до умов реального часу. Це створює перспективи для ефективного управління ресурсами та реалізації стратегій сталого розвитку територіальних громад.

### Узагальнений принцип синтезу інформаційної технології інтелектуального аналізу соціально-економічних даних ТГ

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 рисунку 1 представлено узагальнену структуру, що ілюструє ключові компоненти інформаційної технології, її архітектуру та 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.

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 та усувати 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, дозволяючи швидко шукати альтернативні постачальники 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.

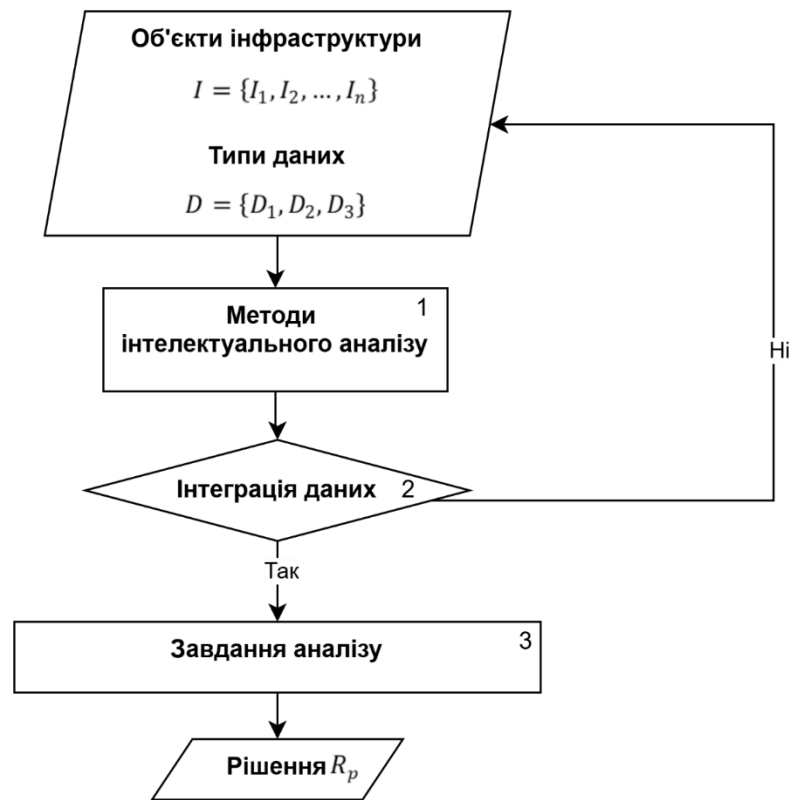


Рис. 1. Узагальнена структура інформаційної технології інтелектуального аналізу соціально-економічних даних ТГ

Отже, нехай  $I$  — множина об'єктів інфраструктури ТГ, що включає наступні компоненти:

$$I = \{I_1, I_2, \dots, I_n\} \quad 1$$

де кожен  $I_j$  (для  $j = 1, 2, \dots, n$ ) представляє окремий елемент інфраструктури, наприклад:

$I_1$  — Бізнес-інфраструктура, включаючи підприємства, ІТ-компанії, фіктивні підприємства та ринки.

$I_2$  — Освітня інфраструктура, що враховує рівень кваліфікації персоналу, освітні програми та можливості для короткострокових проектів 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.

– та інші.

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 та 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:

$$D = \{D^1, D^2, D^3\} \quad 2$$

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:

$$A = \{ \text{Машинне навчання, Обробка природної мови,} \\ \text{Комп'ютерне бачення, ... } \} \quad 3$$

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

$$f: D \times A \rightarrow \{0,1\}, \quad 4$$

де  $f(D_i, A_k) = 1$  означає, що метод  $A_k$  застосовний 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 забезпечує взаємозв'язок між об'єктами інфраструктури, типами даних та відповідними підходами. Така

модель дозволяє не лише структурувати інформацію, але й визначити 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:

$$D_{\text{інтегровані}} = D_1 \oplus D_2 \oplus \dots \oplus D_n \quad 5$$

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  $A_{I_j} \subseteq A$ :

$$A_{I_j} = \{A_k \in A \mid f(D_i, A_k) = 1, \forall D_i \in D_{I_j}\} \quad 6$$

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  $f(D_i, A_k) = 1$ :

$$Z_l = (I_j, D_i, A_k) \quad 7$$

Таким чином, множина всіх можливих завдань аналізу:

$$Z = \{Z_l \mid Z_l = (I_j, D_i, A_k), \forall I_j \in I, D_i \in D, A_k \in A, f(D_i, A_k) = 1\}, \quad 8$$

де функція  $f(D_i, A_k) = 1$  означає, що метод аналізу  $A_k$  застосовний 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 бути 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:

$$R_p = g(Z_l) \forall Z_l \in Z \quad 9$$

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.

#### Метод класифікаційного аналізу кількісних соціально-економічних даних територіальних громад

На основі узагальненого принципу синтезу інформаційної технології інтелектуального аналізу соціально-економічних даних територіальних громад (див. Рисунок 1 та 1-9) запропоновано метод класифікаційного аналізу кількісних соціально-економічних даних ТГ. Розроблений метод враховує

специфіку завдань, таких як оцінка інвестиційних ризиків віртуальних ІТ-компаній та виявлення фіктивних підприємств, шляхом інтеграції різних типів даних (2) з відповідними об'єктами інфраструктури (а 1) та застосуванням методів інтелектуального аналізу (3).

Метод передбачає роботу з трьома типами даних: структурованими ( $D_1$ ), неструктурованими ( $D_2$ ) та напівструктурованими ( $D_3$ ) (2). Для забезпечення точності аналізу використовується оператор інтеграції даних (5), який об'єднує дані з різних джерел у єдину базу. Кожне завдання аналізу визначається трійкою  $(I_j, D_i, A_k)$  відповідно до (7), де  $I_j \in \{I^1, I^6\}$  представляє об'єкти інфраструктури,  $D \in \{D^1, D^2, D^3\}$  — типи даних, а  $A_k \in \{LR, SVM, RF, kNN, GNB\}$  — методи аналізу (4).

Застосування методів машинного навчання, таких як Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (kNN) та Gaussian Naive Bayes (GNB). Це забезпечує ефективну обробку та аналіз кількісних даних, сприяючи прийняттю обґрунтованих управлінських рішень у територіальних громадах.

На основі запропонованого методу класифікаційного аналізу кількісних соціально-економічних даних ТГ розроблено методи до оцінки інвестиційного ризику віртуальної ІТ-компанії на основі методів машинного навчання та метод виявлення фіктивного підприємства навчання, які реалізовані відповідно у [1] та [2].

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

Метод оцінки інвестиційного ризику для віртуальних ІТ-компаній базується на аналізі соціально-економічних даних територіальних громад (ТГ) із використанням алгоритмів машинного навчання. Відповідно до джерел [5, 6], потенційні збитки можуть виникати через політичні чинники, економічну та соціальну нестабільність, швидкість науково-технічного прогресу, застарівання товарів та послуг, неефективність управлінських методів, некомпетентність персоналу, непорядність клієнтів, а також через недосконалу кредитно-грошову та правову політику. Основні кроки реалізації методу наступні:

Крок 1. Збір даних. Акумуляція релевантних даних про економічну діяльність та ринкову ситуацію, включаючи ключові фінансові, соціальні та операційні показники.

Крок 2. Формування вхідних даних. Підготовка набору даних, включаючи кількісні показники ризиків та експертні оцінки, для забезпечення повного розуміння контексту.

Крок 3. Попередня обробка даних. Стандартизація, нормалізація та фільтрація даних для підготовки до моделювання.

Крок 4. Обробка ризиків. Аналіз ступеня впливу та ймовірності виникнення ризиків із використанням формули  $O(R) = Q(R) \times I(R)$ , де  $O(R)$  — загальний рівень ризику.

Крок 5. Навчання моделей МН. Розподіл даних на навчальну (80%) та тестову (20%) вибірки. Навчання моделей SVC, Random Forest та kNN для оцінки ризиків.

Крок 6. Крос-валідація. Проведення перехресної перевірки для об'єктивного порівняння точності моделей та вибору оптимального алгоритму.

Крок 7. Визначення ризиків. Виявлення проєктів з високим інвестиційним ризиком на основі прогнозів моделей.

Крок 8. Адаптація моделей. Перевірка та оновлення моделей на основі нових даних для підвищення точності прогнозування ризиків.

Крок 9. Інтеграція результатів. Виведення результатів оцінки ризиків для ухвалення обґрунтованих управлінських рішень.

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

Метод виявлення фіктивного підприємства базується на аналізі кількісних соціально-економічних даних ТГ із застосуванням сучасних алгоритмів машинного навчання. У аналогічному дослідженні [7] використовуються методи МН 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 МН для ідентифікації фіктивних підприємств, а не загальних злочинів, що забезпечує більшу точність та цілеспрямованість у виявленні економічних злочинів. Основні кроки реалізації наступні:

Крок 1. Збір даних. Формування бази даних з інформацією про підприємства, включаючи юридичну адресу, фізичні адреси, КВЕД, фото обладнання з геолокацією та інші параметри.

Крок 2. Попередня обробка даних. Перевірка даних на повноту, точність та консистентність, їх фільтрація та нормалізація.

Крок 3. Структуризація даних. Перетворення даних у формат, зручний для аналізу, та підготовка до застосування алгоритмів МН.

Крок 4. Аналіз вхідних параметрів. Порівняння отриманих даних між собою (геолокація, відповідність обладнання, страхові поліси) для виявлення невідповідностей.

Крок 5. Бінаризація даних. Перетворення ознак у бінарний формат для підвищення ефективності обробки.

Крок 6. Навчання моделей МН. Розподіл даних на навчальну (80%) та тестову (20%) вибірки. Навчання моделей SVC, Random Forest та Logistic Regression.

Крок 7. Перехресна перевірка. Оцінка моделей із використанням крос-валідації для забезпечення об'єктивності результатів.

Крок 8. Оцінка фіктивності. Прогнозування фіктивності підприємства на основі результатів аналізу моделей.

Крок 9. Візуалізація результатів. Створення графіків для ілюстрації тенденцій та залежностей у даних.

Крок 10. Моніторинг та оновлення. Оновлення моделей на основі нових даних для підвищення їхньої адаптивності та точності.

### Метод класифікаційного аналізу текстових соціально-економічних даних територіальних громад

Відповідно до узагальненого принципу синтезу інформаційної технології інтелектуального аналізу соціально-економічних даних ТГ (див. Рисунок 1 та 1-9) розроблено метод класифікаційного аналізу текстових соціально-економічних даних ТГ. Цей метод враховує специфіку завдань, таких як виявлення дезінформації, формування короткострокових HR-проектів та вибір конкурентоспроможних товарів на основі аналізу емоційного забарвлення відгуків, використовуючи методи класифікації текстових даних.

Метод передбачає інтеграцію гетерогенних даних (5) з різних об'єктів інфраструктури ( $I = \{I_1, I_2, I_3\}$ ) та застосування відповідних методів інтелектуального аналізу ( $A = \{A_1, A_2, A_3, A_4\}$ ) (3). Використання методів логістичної регресії ( $A_1$ ) для класифікації текстів за рівнем достовірності, методів опорних векторів ( $A_2$ ) для аналізу емоційного забарвлення текстів, алгоритмів випадкових лісів ( $A_3$ ) для визначення важливості факторів та глибоких нейронних мереж, таких як LSTM і трансформери ( $A_4$ ), забезпечує високий рівень точності аналізу текстових даних.

Кожне завдання аналізу визначається трійкою  $(I_j, D_i, A_k)$  відповідно до (7), де  $I_j \in \{I^1, I^2, I^3\}$  представляє об'єкти інфраструктури,  $D_i \in \{D^1, D^2, D^3\}$  — типи даних, а  $A_k \in \{A^1, A^2, A^3, A^4\}$  — методи аналізу. Це дозволяє ефективно поєднувати об'єкти інфраструктури з відповідними типами даних та методами аналізу для досягнення оптимальних результатів.

На основі методу класифікаційного аналізу текстових соціально-економічних даних ТГ розроблено методи до виявлення дезінформації та формування короткострокового HR-проекту, які реалізовані відповідно у [3] та [4].

#### Інтелектуальний метод виявлення дезінформації

Метод виявлення дезінформації, спрямований на глибокий аналіз текстових даних соціально-економічного характеру, деталізується наступними кроками:

Крок 1. Збір даних. Збір текстових даних з різноманітних джерел, таких як новинні портали, соціальні мережі, блоги, для формування навчального набору даних.

Крок 2. Попередня обробка даних. Виконання токенизації, стемінгу, розпізнавання іменованих сутностей та інших технік, що забезпечують якісну підготовку текстових даних.

Крок 3. Векторизація тексту. Перетворення текстових даних у векторні представлення за допомогою таких методів, як Word2Vec, GloVe або BERT.

Крок 4. Навчання ансамблю моделей. Створення ансамблю класифікаційних моделей, оптимізація їх ваг та формування інтегрованої метамоделі.

Крок 5. Використання методу "ковзного вікна". Адаптація моделей до нових даних у реальному часі шляхом регулярного оновлення навчального набору.

Крок 6. Виявлення фейкової інформації. Застосування розробленої метамоделі для прогнозування автентичності текстових даних.

Крок 7. Оцінка точності. Регулярна перевірка ефективності моделі на тестових даних з урахуванням змін у характері дезінформації.

Крок 8. Адаптація моделі. Застосування результатів моніторингу для перенавчання моделі та удосконалення алгоритмів.

#### Інтелектуальний метод формування короткострокового проекту управління персоналом

Метод формування короткострокового HR-проекту базується на використанні інтелектуальних технологій для підбору команди та оцінки її ефективності. Метод деталізується наступними кроками:

Крок 1. З'являється новий проєкт. Характеризація проєкту ключовими параметрами, такими як тип, складність, обсяг роботи.

Крок 2. Ініціювання запиту. Формування запиту менеджером проєкту для відбору команди відповідно до вимог.

Крок 3. Встановлення критеріїв відбору. Визначення набору параметрів, таких як необхідні навички, досвід, бюджет.

Крок 4. Запит до бази даних. Автоматичний відбір кандидатів з бази даних фрилансерів на основі заданих критеріїв.

Крок 5. Активізація чат-бота. Здійснення первинної співбесіди з кандидатами за допомогою чат-бота, який оцінює їх відповідність вимогам.

Крок 6. Аналіз результатів. Використання алгоритмів машинного навчання для класифікації кандидатів.

Крок 7. Формування команди. Застосування моделі дерев рішень для остаточного вибору кандидатів, які відповідають вимогам.

Крок 8. Завершення проєкту. Підведення підсумків виконання завдань, оцінка роботи членів команди та оновлення профілів кандидатів у базі даних.

### Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

На основі проведених досліджень сформульовано низку нових наукових результатів, які мають практичне та теоретичне значення для підвищення ефективності аналізу соціально-економічних даних територіальних громад:

1. 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 та прийняття рішень з днів до годин.

2. 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 для них.

3. 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.

### Література

1. Lipyana, 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. doi:10.1007/978-3-030-61656-4\_11

2. Krysovaty, A., Lipyana-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. doi:10.1109/csit52700.2021.9648584

3. Lipyana-Goncharenko, K., Bodyanskiy, Y., Kustra, N., & Ivasechko, A. OLTW-TEC: Online Learning with Sliding Windows for Text Classifier Ensembles. *Frontiers in Artificial Intelligence*, 7, 1401126. doi: 10.3389/frai.2024.1401126

4. Lipyana, H., Sachenko, O., Lendyuk, T., Sachenko, A., & Vasylyuk, 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. doi:10.1007/978-3-030-63270-0\_71

5. Cabral, J. S. (2017). Project risk management strategies for IT project managers (ScholarWorks). Retrieved from <https://scholarworks.waldenu.edu/dissertations/3663>

6. 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. doi:10.1108/ijmpb-07-2015-0055

7. 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. Retrieved from <https://www.isca.me/rjrs/archive/v4/i3/15.ISCA-RJRS-2013-1005.pdf>

### References

1. Lipyana, 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. doi:10.1007/978-3-030-61656-4\_11

2. Krysovaty, A., Lipyana-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. doi:10.1109/csit52700.2021.9648584



3. 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. doi: 10.3389/frai.2024.1401126
4. Lipianina, 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. doi:10.1007/978-3-030-63270-0\_71
5. Cabral, J. S. (2017). Project risk management strategies for IT project managers (ScholarWorks). Retrieved from <https://scholarworks.waldenu.edu/dissertations/3663>
6. 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. doi:10.1108/ijmpb-07-2015-0055
7. 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. Retrieved from <https://www.isca.me/rjrs/archive/v4/i3/15.ISCA-RJRS-2013-1005.pdf>