

МЕТОДОЛОГІЯ АНАЛІЗУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ДАНИХ ТЕРИТОРІАЛЬНИХ ГРОМАД

У сучасних умовах цифрової трансформації та стрімкого зростання обсягів даних територіальні громади стикаються з викликами, які потребують нових підходів до аналізу соціально-економічної інформації. Соціально-економічні дані територіальних громад характеризуються високим ступенем гетерогенності, нестаціонарності та динамічності, що значно ускладнює їх обробку традиційними методами. Актуальною задачею є забезпечення можливості оперативного прийняття рішень на основі аналізу таких даних, що сприятиме підвищенню ефективності управління ресурсами громади та покращенню якості життя її мешканців.

У статті розглянуто нову методологію аналізу соціально-економічних даних територіальних громад, яка поєднує класифікаційний, кластерний та гібридний аналіз із прогнозуванням. Методологія забезпечує адаптивність до роботи з нестаціонарними даними, інтеграцію різнорідних джерел інформації та можливість роботи в режимі реального часу.

Запропоновано новий метод формування навчальної вибірки для задач класифікації та прогнозування нестаціонарних даних, який базується на RFM-аналізі та кластеризації.

Також вперше розроблено метод ансамблевих адаптивних прогнозних моделей, який забезпечив точність прогнозування багатомірних процесів, що є критично важливим для оперативного прийняття управлінських рішень.

Результати дослідження можуть бути використані для ефективного управління соціально-економічними процесами територіальних громад, адаптивного прогнозування та розробки інтелектуальних систем підтримки прийняття рішень.

Ключові слова: територіальні громади, соціально-економічні дані, класифікація, кластеризація, прогнозування, адаптивні моделі, нестаціонарні дані.

LIPIANINA-HONCHARENKO KHRYSTYNA

West Ukrainian National University

METHODOLOGY FOR ANALYZING SOCIO-ECONOMIC DATA OF TERRITORIAL COMMUNITIES

In the current era of digital transformation and rapid data growth, territorial communities face challenges that require innovative approaches to analyzing socio-economic information. Socio-economic data of territorial communities is characterized by high heterogeneity, non-stationarity, and dynamism, which significantly complicate its processing using traditional methods. A pressing task is to enable operational decision-making based on the analysis of such data, contributing to improved resource management efficiency and the quality of life of community residents.

This article presents a novel methodology for analyzing socio-economic data of territorial communities, combining classification, clustering, and hybrid analysis with forecasting. The methodology ensures adaptability to non-stationary data, integration of heterogeneous information sources, and real-time processing capability.

A new method for constructing training datasets for classification and forecasting of non-stationary data is proposed, based on RFM analysis and clustering. Additionally, an ensemble method of adaptive predictive models is developed for the first time, ensuring high accuracy in forecasting multidimensional processes, which is critically important for operational decision-making.

The methodology provides a systematic approach to the analysis of socio-economic data, allowing to adapt the analysis process to the specific conditions of the TG, increasing decision-making and optimizing resource management. The research outcomes can be utilized for efficient management of socio-economic processes in territorial communities, adaptive forecasting, and the development of intelligent decision-support systems.

Keywords: territorial communities, socio-economic data, classification, clustering, forecasting, adaptive models, non-stationary data.

Вступ

У сучасному світі цифрова трансформація та динамічні зміни в соціально-економічних процесах створюють значні виклики для територіальних громад (ТГ). Зростаючі обсяги даних, їхня гетерогенність, нестаціонарність та необхідність оперативного прийняття рішень вимагають нових підходів до аналізу й управління. Запропонована методологія спрямована на інтеграцію сучасних інтелектуальних технологій для підвищення ефективності управління ресурсами, підтримки соціальної стійкості, розвитку підприємницької діяльності та забезпечення прозорості управлінських рішень.

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

1. 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.

2. Розроблено новий метод формування навчальної вибірки для нестаціонарних даних на основі RFM-аналізу та кластерного аналізу, що на відміну від існуючих підходів забезпечує сегментацію та кластеризацію, що дозволило підвищити точність опрацювання нових вхідних наборів даних у задачах прогнозування, класифікації та аналізу нестаціонарних соціально-економічних даних.

3. 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.

Запропонована методологія базується на принципах адаптивності, аналізу в реальному часі, інтеграції різномірних джерел даних, точності прогнозування та прозорості результатів. Вона включає такі етапи, як формування навчальної вибірки, використання інтелектуальних методів, створення ансамблю моделей та інтерактивного інструментарію для управлінських рішень. Завдяки цьому методологія може бути ефективно використана для управління соціально-економічними процесами в ТГ, забезпечуючи адаптивність і точність в умовах динамічного середовища.

Методологія 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.) 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 1), що базуються на визначених протиріччях, 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.

Таблиця 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 різномірних джерел даних для створення повної картини соціально-економічної ситуації.	Принцип інтеграції	Передбачає об'єднання різномірних 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.

Далі представлено розроблену методологію (див. Рис.) 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:

- 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.

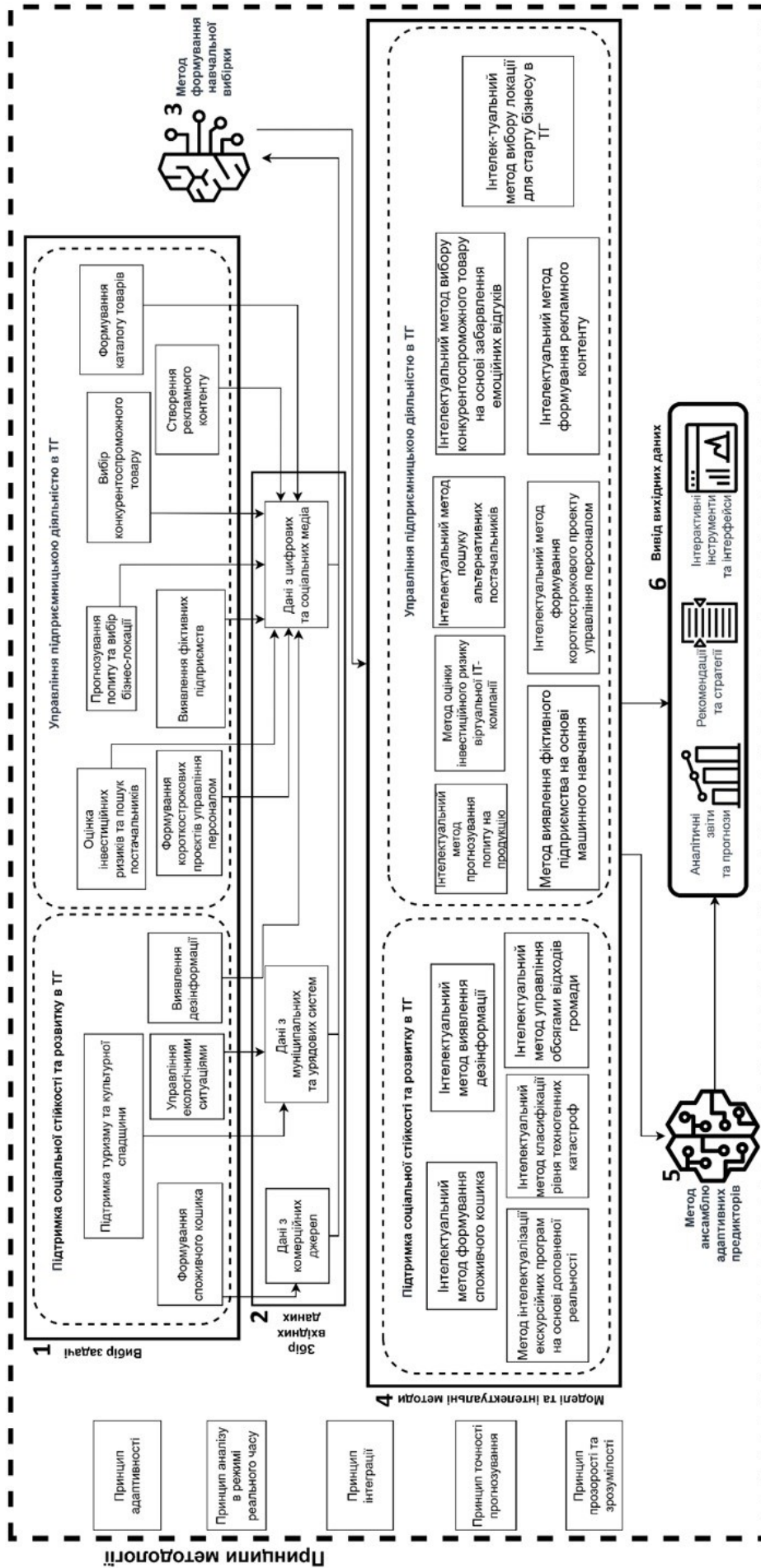


Рис. 1. Концепція методології аналізу соціально-економічних даних IT

• Управління підприємницькою діяльністю в ТГ охоплює такі завдання, як оцінка інвестиційних ризиків і пошук постачальників, прогнозування попиту та вибір оптимальних бізнес-локацій, формування короткострокових проектів 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:

- Дані з комерційних джерел, наприклад, для формування споживчого кошика;
- Дані з муніципальних та урядових систем для задач управління підприємницькою діяльністю;
- Дані з цифрових та соціальних медіа, що особливо важливо для задач просування товарів та виявлення дезінформації.

Етап 3. Метод формування навчальної вибірки для нестационарних даних на основі RFM-аналізу [7] та кластерного аналізу дозволяє більш точно відбирати інформативні зразки, порівнюючи їх з аналогами, що було доведено в [3, 4, 5, 6]. Запропонований метод складається з наступних кроків:

Крок 1. Введення вхідних даних.

Крок 2. Аналіз RFM. RFM аналіз [1].

Крок 3. Формування квантилів.

Крок 4. Групування за сумою квантилів.

Крок 5. Зменшення розмірності даних.

Крок 6. Кластеризація [2].

Крок 7. Інтеграція даних аналізу RFM та кластерного аналізу в базу даних.

Крок 8. Тестування навчального набору даних на основі алгоритмів класифікації машинного навчання.

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:

— Для підтримки соціальної стійкості та розвитку використовуються методи оцінки якості життя, управління екологічними ситуаціями, підтримки культурної спадщини та інші.

— 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 ансамблю адаптивних передбачувальних моделей для аналізу в реальному часі багатовимірних даних [8]. Побудова ансамблю метамodelей, які використовують багатовимірні сигнали для виведення прогнозів, може бути деталізована наступними кроками:

Крок 1. Збір даних.

Крок 2. Формування вхідних даних.

Крок 3. Обробка даних.

Крок 4. Оцінка на "повзучому вікні" випадкових значень.

Крок 5. Синтез метамodelей.

Крок 6. Формування бази даних результатів.

Крок 7. Синтез метамodelі другого рівня.

Крок 8. Вибір найкращої метамodelі.

Крок 9. Прогнозування: фінальна метамodelь використовується для виведення прогнозів на основі вхідних даних.

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.

Висновки

У даній роботі вирішено низку важливих задач, спрямованих на вдосконалення аналізу соціально-економічних даних територіальних громад. Основні результати дослідження включають:

1. 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.
2. Розроблено новий метод формування навчальної вибірки для нестационарних даних на основі RFM-аналізу та кластерного аналізу.
3. Вперше розроблено метод ансамблевих адаптивних прогнозних моделей для багатовимірного аналізу, особливістю якого є оцінювання на "ковзному вікні" та використання метамodelей першого і другого рівнів на основі оптимізації функції Лагранжа.

Література

1. 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*. doi:10.1016/j.jksuci.2019.12.011.
2. Fränti, P., & Sieranoja, S. (2018). K-means properties on six clustering benchmark datasets. *Applied intelligence*, 48(12), 4743–4759. doi:10.1007/s10489-018-1238-7
3. 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. doi:10.3103/s0735272718070051.
4. 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. doi:10.3103/s0735272718010041
5. 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. doi:10.1016/j.isprsjprs.2016.05.009
6. 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. doi:10.1016/j.procs.2019.02.024
7. Lipyanina-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. Retrieved from <https://ceur-ws.org/Vol-2870/paper134.pdf>
8. 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. doi:10.15588/1607-3274-2023-4-9

References

1. 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*. doi:10.1016/j.jksuci.2019.12.011.
2. Fränti, P., & Sieranoja, S. (2018). K-means properties on six clustering benchmark datasets. *Applied intelligence*, 48(12), 4743–4759. doi:10.1007/s10489-018-1238-7
3. 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. doi:10.3103/s0735272718070051.
4. 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. doi:10.3103/s0735272718010041
5. 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. doi:10.1016/j.isprsjprs.2016.05.009
6. 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. doi:10.1016/j.procs.2019.02.024
7. Lipyanina-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. Retrieved from <https://ceur-ws.org/Vol-2870/paper134.pdf>
8. 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. doi:10.15588/1607-3274-2023-4-9