https://doi.org/10.31891/csit-2024-3-12
UDC 004.9

Khrystyna LIPIANINA-HONCHARENKO
West Ukrainian National University

METHODS FOR ANALYZING SOCIO-ECONOMIC DATA OF TERRITORIAL COMMUNITIES FOR ADAPTIVE RESOURCE MANAGEMENT

The socio-economic development of territorial communities in modern conditions requires adaptive approaches to resource management based on the intelligent analysis of large volumes of data of various types. Effective decision-making depends on the ability to integrate structured, semi-structured, and unstructured data, enabling the prediction of dynamic processes, identification of cluster groups of objects, and evaluation of key development indicators. The proposed information technology integrates modern methods of machine learning, natural language processing, and computer vision for socio-economic data analysis, ensuring accuracy, speed, and flexibility in decision-making.

Based on the proposed approach, methods for cluster analysis, forecasting, and hybrid analysis have been improved, allowing consideration of the specifics of territorial communities and adaptation to crisis conditions. The obtained results lay the foundation for creating an innovative decision-support system that promotes sustainable community development, efficient resource management, and improved quality of life for the population.

Keywords: socio-economic data, territorial communities, adaptive management, cluster analysis, forecasting, hybrid analysis, intelligent technologies.

Христина ЛІП'ЯНІНА-ГОНЧАРЕНКО
Західноукраїнський національний університет

МЕТОДИ АНАЛІЗУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ДАНИХ ТЕРИТОРІАЛЬНИХ ГРОМАД ДЛЯ АДАПТИВНОГО УПРАВЛІННЯ РЕСУРСАМИ

Соціально-економічний розвиток територіальних громад у сучасних умовах вимагає адаптивних підходів до управління ресурсами, які базуються на інтелектуальному аналізі великих обсягів даних різної природи. Ефективне прийняття управлінських рішень залежить від здатності інтегрувати структуровані, напівструктуровані та неструктуровані дані, що дозволяє прогнозувати динамічні процеси, ідентифікувати кластерні групи об'єктів та оцінювати ключові показники розвитку. Запропонована інформаційна технологія інтегрує сучасні методи машинного навчання, обробки природної мови та комп'ютерного бачення для аналізу соціально-економічних даних, забезпечуючи точність, швидкість та гнучкість прийняття рішень.

На основі запропонованого підходу удосконалено методи кластерного аналізу, прогнозування та гібридного аналізу, які дозволяють враховувати специфіку територіальних громад та адаптуватися до кризових умов. Отримані результати формують основу для створення інноваційної системи підтримки прийняття рішень, яка сприяє стійкому розвитку громад, ефективному управлінню ресурсами та підвищенню якості життя населення.

Ключові слова: соціально-економічні дані, територіальні громади, адаптивне управління, кластерний аналіз, прогнозування, гібридний аналіз, інтелектуальні технології.

Introduction

Modern territorial communities (TCs) face numerous challenges associated with the rapid dynamics of socio-economic processes, particularly in crisis situations such as pandemics or wars. Effective resource management, community resilience, and adaptation to change depend on the ability to analyze large volumes of diverse data: structured, unstructured, and semi-structured. In this context, developing information technologies that enable the integration of data from various sources, facilitate their analysis, and support informed managerial decision-making is of paramount importance.

One of the primary directions for addressing this challenge is the creation of a generalized principle for synthesizing an information technology framework for intelligent analysis of socio-economic data in TCs. This principle is founded on the formalization of interconnections between the key infrastructure objects within communities, the data types that characterize them, and the methods of intelligent analysis. The integration of these components ensures adaptability and flexibility in the technology, which are critical for fostering the sustainable development of communities.

The proposed information technology has several key features. First, it accounts for the heterogeneity of data, which may be presented in the form of statistical tables, textual documents, images, or other formats. Second, it leverages adaptive analysis methods, such as machine learning, natural language processing, and computer vision, to ensure the accuracy and relevance of information processing. Third, the technology aims to establish a decision-support system that formalizes the processes of modeling, analyzing, and forecasting socio-economic processes within TCs.

Thus, the developed generalized principle for synthesizing an information technology framework for the intelligent analysis of socio-economic data in territorial communities forms the foundation for innovative approaches to managing socio-economic development. It facilitates data structuring, integration of heterogeneous information, and application of adaptive methods to address managerial tasks.

In this regard, the scientific novelty of this article is as follows:

1. The method of cluster analysis of socio-economic data has been improved, which, unlike the known

approaches, ensures the identification of groups of objects with similar characteristics, which has allowed to increase the accuracy of resource allocation for the implementation of management decisions.

2. The method of forecasting socio-economic data has been improved, which, unlike the known approaches, provides highly accurate forecasting of dynamic processes by integrating structured, unstructured and semi-structured data using adaptive methods of intellectual analysis, which has increased the accuracy of forecasting for the implementation of management decisions.

3. The method of hybrid analysis of socio-economic data has been improved, which, unlike the known approaches, ensures the consideration of quantitative and qualitative indicators through the integration of heterogeneous data and a multi-level approach to analysis, which has increased the flexibility and adaptability of the decision-making process for the implementation of management decisions.

Method of Cluster Analysis of Socio-Economic Data in Territorial Communities

Based on the proposed generalized principle for synthesizing an information technology framework for the intelligent analysis of socio-economic data in territorial communities (TCs) and the described objectives, a method for cluster analysis of socio-economic data in TCs has been developed to support effective decision-making. The proposed method focuses on identifying groups of objects with similar characteristics to improve resource management and support socio-economic stability.

Let the set of infrastructure objects in TCs be denoted as:

$$I = \{I_1, I_2, I_3, I_4\} \tag{1}$$

where::

I_1 — business infrastructure, including enterprises and IT companies;

I_2 — social infrastructure responsible for analyzing consumer needs;

I_3 — logistics infrastructure that ensures supply and resource management;

I_4 — information infrastructure, including sources of data on population sentiment and media content.

The socio-economic data used in the analysis are heterogeneous in nature and formalized as a set of data types:

$$D = \{D_1, D_2, D_3\} \tag{2}$$

where:

D_1 — structured data, such as financial reports and demographic indicators;

D_2 — unstructured data, including text documents, social media, and videos;

D_3 — semi-structured data in JSON and XML formats obtained via APIs.

To perform cluster analysis, a set of intelligent analysis methods is utilized:

$$A = \{k - means, DBSCAN, GMM, PCA\} \tag{3}$$

where each method A_k has its application depending on the data characteristics:

— k-means — a clustering method that divides data into k groups by minimizing the sum of squared distances between each point and its cluster center. It is used to identify similar groups in the data.

— DBSCAN (Density-Based Spatial Clustering of Applications with Noise) — a clustering method based on data density. It detects clusters of arbitrary shapes and isolates noise (outliers).

— GMM (Gaussian Mixture Model) — a clustering method using Gaussian distribution models to represent clusters. It is suitable for identifying groups in data with overlapping clusters.

— PCA (Principal Component Analysis) — a dimensionality reduction method that preserves as much information as possible by projecting data onto new coordinates that maximize variance. It is used for visualization and data preparation before clustering.

The data integration process, a key stage, is formalized as:

$$D_{\text{інтегровані}} = D_1 \oplus D_2 \oplus D_3 \tag{4}$$

where the operator $\oplus$ performs the merging of data from various sources to form a unified database.

Each cluster analysis task is formulated as a triplet:

$$Z_l = \left(I_j, D_i, A_k\right) \tag{5}$$

where $I_j \in I, D_i \in D, A_k \in A$. The task involves selecting an infrastructure object, a data type, and an appropriate clustering method.

The solution derived from the analysis is denoted as:

$$R_p = g(Z_l) \tag{6}$$

where $g(Z_l)$ is a function that determines the optimal managerial decision, for example:

- Selection of social groups for targeted resource allocation;
- Identification of enterprise clusters with high investment potential;
- Development of logistics plans to meet community needs.

Based on the cluster analysis method for socio-economic data in TCs, methods for forming a consumer basket and identifying alternative suppliers have been developed, as implemented in [1] and [2].

Intelligent Method for Forming a Consumer Basket

The method for forming a consumer basket is based on the analysis of socio-economic data and the clustering of consumers, enabling consideration of the needs of various social groups and adaptation to crisis conditions. Unlike existing approaches [3, 4], the method ensures the rapid adaptation of the set of goods and determination of the minimum cost of living. The main implementation stages of the method include:

1. Data Collection and Preprocessing. Accumulation of information on purchases, demographic data of consumers, and transaction records. Normalization is performed to ensure comparability of features.
2. Data Clustering. Application of clustering methods (e.g., K-means) to group consumers based on behavioral and demographic features.
3. Cluster Analysis. Description of cluster characteristics, including average expenditures, purchase frequency, and the variety of acquired goods.
4. Consumer Basket Formation. Determination of a typical set of goods for each cluster based on the probability of product purchases and expected expenditures.
5. Recommendations for Basket Optimization. Development of suggestions for improving the structure of the consumer basket according to the needs of each cluster.
6. Visualization and Implementation of Results. Creation of graphs and tables for the visual presentation of the obtained data and recommendations.

The developed method enables the formation of consumer baskets that align with the current needs of various social groups.

Intelligent Method for Finding Alternative Suppliers

The method for finding alternative suppliers relies on the analysis of socio-economic data and clustering of potential partners. Compared to existing approaches [5, 6], this method evaluates suppliers based on critical parameters such as price, quality, and reputation, which is particularly significant in crisis conditions. The main stages of the method's implementation include:

1. Data Collection and Preparation. Accumulation of data on suppliers, including key parameters (price, quality, production capacity), and normalization of values to ensure comparability.
2. Data Clustering. Application of clustering methods (e.g., K-means, DBSCAN) to identify groups of suppliers with similar characteristics.
3. Selection of the Primary Cluster. Evaluation of clusters based on business criteria and selection of the optimal cluster for further analysis.
4. Analysis of the Primary Cluster.Detailed analysis of the internal structure of the selected cluster through repeated clustering to identify subgroups of suppliers.
5. Supplier Ranking.Evaluation of suppliers within the primary cluster using weighted coefficients that consider the importance of each parameter.
6. Selection of the Best Supplier. Identification of the supplier that best meets the specified criteria using ratings and parameter analysis.
7. Visualization and Presentation of Results. Creation of graphs and tables to represent cluster characteristics and perform a comparative analysis of suppliers.

The developed method enables the identification of reliable suppliers, ensuring the resilience of supply chains under crisis conditions.

Method for Forecasting Socio-Economic Data of Territorial Communities

The forecasting method for socio-economic data in TCs is based on the generalized principle of synthesizing information technology and aims to generate accurate forecasts for strategic resource management. Unlike the cluster analysis method, this approach utilizes adaptive models to work with time series and structured data, enabling the consideration of socio-economic process dynamics.

Forecasting employs a set of intelligent analysis methods (A), including LSTM, ARIMA, and XGBoost. These methods facilitate modeling and forecasting dependencies in structured (D_1), semi-structured (D_3), and unstructured (D_2) data, which are pre-integrated into a unified database using the operator $\oplus$ (4).

Forecasting tasks are formalized as triplets:

$$Zl = \left(I_j, D_i, A_k\right) \quad (5),$$

where I_j represents an infrastructure object, D_i is the data type, and A_k is the respective forecasting method. Forecast results serve as the foundation for managerial decisions (R_p), described by the function $g(Z_l)$ (6).

Based on the forecasting method for socio-economic data in TCs, methods for predicting product demand and selecting optimal locations for business startups in TCs have been developed, as implemented in [7] and [8]..

Intelligent Method for Predicting Product Demand

This method for predicting product demand is designed using modern machine learning technologies to ensure adaptability and accuracy in forecasting. The main implementation stages include:

1. Data Collection.Gathering information from customer reviews, retail point-of-sale data, and historical time series.
2. Data Preprocessing. Integration and normalization of data into a structured format for analysis; anomaly detection and removal, and transformation using scaling methods.
3. Data Splitting. Creating training and test datasets to ensure the objectivity of model evaluations.
4. Model Selection. Evaluating machine learning models, including ensemble methods and specialized algorithms for time series; selecting the best-performing model based on accuracy metrics (e.g., RMSE, MAE).
5. Model Training. Training the selected model on the training data and forecasting key parameters.
6. Result Integration. Combining individual forecasts to form the final product demand prediction.
7. Visualization. Presenting forecasts with error intervals in the form of graphs and charts to support decision-making.

The developed method has been tested in experimental studies, confirming its effectiveness under crisis conditions to support strategic planning in territorial communities.

Intelligent Method for Selecting Business Startup Locations in TCs

This method is based on the analysis of geographic, market, and demographic data using machine learning algorithms. The main implementation stages include:

1. Data Collection. Accumulating data from surveillance cameras, geolocation services, and mobile applications.
2. Data Preprocessing. Analyzing images to determine gender and age, cleaning data, and clustering it using machine learning algorithms.
3. Data Analysis. Using algorithms to identify key trends and factors influencing business success in various locations.
4. Modeling and Forecasting. Building regression models to forecast sales volumes and assess the competitiveness of each location.
5. Location Selection. Calculating the optimal location based on forecasted profits, competition levels, and operational costs.

The developed method enables *TCs* to effectively plan their economic infrastructure, ensuring a high probability of success for new business projects. Experimental results have validated the practical value and effectiveness of this method.

Hybrid Method for Analyzing Socio-Economic Data of TCs

The hybrid method combines classification, clustering, and forecasting to provide a comprehensive approach to solving managerial tasks. Its key distinction from the cluster analysis method is a multi-level structure that involves preliminary grouping at the first level (A^{I}) and detailed forecasting or classification at the second level (A^{II}).

At the first level (A^{I}), clustering methods such as k-means and DBSCAN can be applied, similar to the cluster analysis method. After initial grouping, the data are passed to the second level (A^{II}), where forecasting methods such as LSTM, ARIMA, and XGBoost are employed to analyze dynamic processes within clusters.

Tasks for hybrid analysis are formulated as triplets (I_j, D_i, A_k) (4), where $A_k \in A^{\mathrm{I}} \cup A^{\mathrm{II}}$, reflecting combined classification and forecasting methods distinct from clustering approaches. Analysis results are expressed as managerial decisions $R_p = g(Z_l)$ (6), including classification of regions by economic stability, forecasting community needs for key resources, and identifying clusters with high investment potential.

The hybrid method combines classification, clustering, and forecasting, offering a comprehensive approach to analysis and decision-making. The multi-level approach effectively handles heterogeneous data, considers dynamic changes, and formulates strategic solutions aimed at socio-economic development, distinguishing it significantly from standalone clustering or forecasting methods.

Based on this method, approaches for classifying the level of man-made disasters and forecasting community waste volumes have been developed and implemented in [9] and [10], respectively.

Intelligent Method for Classifying the Level of Man-Made Disasters

This method leverages hybrid analysis of textual and quantitative data to ensure rapid and accurate identification of risks in a region. Despite significant progress in developing tools for analyzing and automating responses to man-made disasters, decision-making in this field remains complex [11, 12]. Key implementation stages include:

1. Data Collection. Accumulation of textual descriptions of accidents and quantitative characteristics of man-made events (e.g., location, time, impact intensity), marked by hazard levels.
2. Data Preprocessing:
   — Textual Data: Normalization, lemmatization, tokenization, and removal of stop words to prepare texts. Text vectorization is performed using TF-IDF or word embeddings.
   — Quantitative Data: Normalization of numerical characteristics, anomaly detection, handling missing values, and scaling for compatibility with classification models.
3. Training and Test Data Formation. Splitting data into training (80%) and test (20%) sets. The SMOTE method is used for balancing classes in textual and quantitative data.
4. Model Training:
   — Textual Data: Ensemble methods (AdaBoost, XGBoost, CatBoost) model dependencies between text features and risk classes.
   — Quantitative Data: Models based on decision trees and boosting approaches classify risks using numerical characteristics.
   — Model Integration: A hybrid model combines predictions from textual and quantitative classifiers through a weighted ensemble mechanism for improved accuracy.
5. Evaluation of Modeling Results. Results are assessed using metrics such as Accuracy, Recall, and F1-Score for individual textual and quantitative models and the hybrid approach.
6. Visualization of Results. Graphs, such as ROC curves, are used to analyze classification accuracy and compare the performance of individual and combined models.
7. Risk Classification. Applying the hybrid model to test datasets to determine disaster levels and evaluate potential consequences.

The developed method ensures effective and rapid hazard level determination, enabling strategic decisions to prevent man-made disasters.

Intelligent Method for Forecasting Community Waste Volumes

This method integrates data on types and locations of waste generation using clustering and machine learning models. Compared to existing approaches [13, 14, 15, 16], it offers more accurate forecasting and efficient waste management. Key implementation stages include:

1. Waste Classification. Classification of waste by composition, quantity, and size using machine learning methods (logistic regression, decision trees).
2. Clustering of Waste Generation Locations. Grouping waste generation sites into clusters using algorithms such as K-means, DBSCAN, and Spectral Clustering.
3. Model Development for Forecasting. Application of ARIMA, XGBoost, and DNN models to forecast waste volumes in each cluster.
4. Result Evaluation. Using MAE and MSE metrics to select the best forecasting model for each cluster.
5. Forecasting Based on the Best Model. Generating forecasts for future periods for each cluster, considering the unique features of waste generation sites.
6. Forecast Visualization. Presenting results as graphs for assessing forecast accuracy and supporting managerial decisions.

The proposed method optimizes waste management, improving environmental conditions in communities.

Conclusions

As part of the study, methods for analyzing and forecasting socio-economic data were improved, significantly enhancing the accuracy, speed, and efficiency of managerial decision-making. Specifically:

1. The method of cluster analysis of socio-economic data was improved, enabling the identification of groups of objects with similar characteristics.
2. The method of forecasting socio-economic data was enhanced by integrating various data types and employing adaptive methods of intelligent analysis.
3. The method of hybrid analysis of socio-economic data was refined to incorporate both quantitative and qualitative indicators through the integration of heterogeneous data and a multi-level analytical approach.

References

1. Lipianina-Honcharenko, K., Wolff, C., Chyzhovska, Z., Sachenko, A., Lendiuk, T., & Grodskyi, S. (2022). Intelligent method for forming the consumer basket. In Communications in computer and information science (pp. 221–231). Cham: Springer International Publishing. doi:10.1007/978-3-031-16302-9_17
2. Lipianina-Honcharenko, K., Komar, M., Melnyk, N. & Komarnytsky, R. (2024). Sustainable Information System for Enhancing Virtual Company Resilience Through Machine Learning in Smart City Socio-Economic Scenarios. ECONOMICS, 12(2). https://doi.org/10.2478/eoik-2024-0022

3. Listiyoko, L., & Purno, M. (2020). Sample identification approach by k-means clustering in thinner retail market segmentation. Journal of physics: conference series, 1477, 022004. doi:10.1088/1742-6596/1477/2/022004

4. Munusamy, S., & Murugesan, P. (2020). Modified dynamic fuzzy c-means clustering algorithm – Application in dynamic customer segmentation. Applied intelligence, 50(6), 1922–1942. doi:10.1007/s10489-019-01626-x

5. Khan, M. M., Bashar, I., Minhaj, G. M., Wasi, A. I., & Hossain, N. U. I. (2023). Resilient and sustainable supplier selection: an integration of SCOR 4.0 and machine learning approach. Sustainable and resilient infrastructure, 1–17. doi:10.1080/23789689.2023.2165782

6. Boz, E., Çizmecioğlu, S., & Çalık, A. (2022). A novel MDCM approach for sustainable supplier selection in healthcare system in the era of logistics 4.0. Sustainability, 14(21), 13839. doi:10.3390/su142113839

7. Lipianina-Honcharenko, K., Bodyanskiy Y., Sachenko A., Kit I., Podchasova T. and Lendiuk T. (2023). An Intelligent Method of Prediction the Demand for Goods/Services in Crisis Conditions, Proceedings of the 3rd International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2023) 2023 Waterloo, Canada, November 20-22, (3641), 224-232 https://ceur-ws.org/Vol-3641/short2.pdf

8. Lipianina-Honcharenko, K., Sachenko, A., Semaniuk, V., Badasian, A., & Kopania, Ł. (2023). Intelligent method for selecting business location in smart city. In 2023 IEEE 12th international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS). IEEE. doi:10.1109/idaacs58523.2023.10348823

9. Lipianina-Honcharenko, K., Wolff, C., Sachenko, A., Kit, I., & Zahorodnia, D. (2023). Intelligent method for classifying the level of anthropogenic disasters. Big data and cognitive computing, 7(3), 157. doi:10.3390/bdcc7030157

10. Lipianina-Honcharenko, K., Komar, M., Osolinskyi, O., Shymanskyi, V., Havryliuk, M., & Semaniuk, V. (2023). Intelligent waste-volume management method in the smart city concept. Smart Cities, 7(1), 78–98. doi:10.3390/smartcities7010004

11. Karam, E., Hussein, W., & Gharib, T. F. (2021). Integrating location and textual information for detecting affected people in a crisis. Social network analysis and mining, 11(1). doi:10.1007/s13278-020-00715-x

12. Adel, H., Dahou, A., Mabrouk, A., Abd Elaziz, M., Kayed, M., El-Henawy, I. M., Amin Ali, A. (2022). Improving crisis events detection using distilbert with hunger games search algorithm. Mathematics, 10(3), 447. doi:10.3390/math10030447

13. El Ouadi, J., Malhene, N., Benhadou, S., & Medromi, H. (2022). Towards a machine-learning based approach for splitting cities in freight logistics context: benchmarks of clustering and prediction models. Computers & industrial engineering, 166, 107975. doi:10.1016/j.cie.2022.107975

14. Gue, I. H. V., Lopez, N. S. A., Chiu, A. S. F., Ubando, A. T., & Tan, R. R. (2022). Predicting waste management systems performance from city and country attributes. Journal of cleaner production, 132951. doi:10.1016/j.jclepro.2022.132951

15. Izquierdo-Horna, L., Damazo, M., & Yanayaco, D. (2022). Identification of urban sectors prone to solid waste accumulation: a machine learning approach based on social indicators. Computers, environment and urban systems, (96), Article 101834. doi:10.1016/j.compenvurbsys.2022.101834

16. Zhang, C., Dong, H., Geng, Y., Liang, H., & Liu, X. (2022). Machine learning based prediction for China's municipal solid waste under the shared socioeconomic pathways. Journal of environmental management, 312, 114918. doi:10.1016/j.jenvman.2022.114918

Khrystyna Lipianina-Honcharenko **Христина Ліп'яніна-Гончаренко**	Associate professor, Ph.D. in information technologies Department for Information Computer Systems and Control, West Ukrainian National University https://orcid.org/0000-0002-2441-6292 e-mail: kh.lipianina@wunu.edu.ua	Кандидат технічних наук, доцент, доцент кафедри інформаційно-обчислювальних систем і управління, Західноукраїнський національний університет