УДК 004.9:681.3
DOI https://doi.org/10.36994/2788-5518-2025-01-09-16

APPLICATION OF MONTE CARLO METHODS FOR MODELING AND PREDICTING USER BEHAVIOR ON SOCIAL MEDIA

Mysiuk I. V., *PhD student, Ivan Franko National University of Lviv, Lviv, Ukraine. https://orcid.org/0000-0002-3641-4518, iruna.musyk8a@gmail.com*

Mysiuk R. V., *Ph.D, Ivan Franko National University of Lviv, Lviv, Ukraine. https://orcid.org/0000-0002-7843-7646, mysyukr@ukr.net*

Abstract. *Understanding and predicting user behavior on social media is challenging due to the dynamic nature of users' online interactions. Monte Carlo methods are one of the methods for modeling user engagement, content virality, sentiment evolution and personalized recommendations in dynamic environments.*

The article investigates the application of Monte Carlo methods for modeling and predicting user behavior in social networks. Several variants of this approach are considered, including the quasi-Monte Carlo, the classical Monte Carlo method, and the Metropolis-Hastings methods, which demonstrated better efficiency for the selected data set due to more uniform coverage of user behavior scenarios, taking into account the selected parameters of user activities. Convergences by Monte Carlo methods with a sample of 10000 values are also displayed, and based on the main selected parameters for modeling user behavior, namely likes, shares and comments, the effectiveness of the selected methods with such a data set with many parameters is studied.

Based on the data obtained, a basic predictive analytics system was developed, which can be further developed through integration into recommendation systems to increase content personalization and improve the analysis of user activities in social networks. In addition, the VADER lexical analyzer was used to assess the emotional tone of users and their virality on social networks was investigated. The results obtained allowed us to display the distribution of likes, comments and expansions in social networks and to carry out predictive analytics, which can be integrated into recommender systems based on the results of modeling. The process of modeling virality and expansion of content is also shown and the principle of operation of the program is schematically presented. The results of the performance evaluation confirm the feasibility of using a combination of quasi-Monte Carlo methods and show their effectiveness and accuracy of prediction in the analysis of user activity for a comprehensive study of behavior in social networks. Virality prediction reflects the spread of viral content.

Key words: *Monte Carlo simulation, social media analytics, user behavior modeling, machine learning, predictive analytics.*

ЗАСТОСУВАННЯ МЕТОДІВ МОНТЕ-КАРЛО ДЛЯ МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ ПОВЕДІНКИ КОРИСТУВАЧІВ У СОЦІАЛЬНИХ МЕРЕЖАХ

Ірина Мисюк, *аспірантка, Львівський національний університет імені Івана Франка, Львів, Україна. https://orcid.org/0000-0002-3641-4518, iruna.musyk8a@gmail.com*

Роман Мисюк, *д.ф., Львівський національний університет імені Івана Франка, Львів, Україна. https://orcid.org/0000-0002-7843-7646, mysyukr@ukr.net*

Анотація. *Розуміння та прогнозування поведінки користувачів у соціальних мережах є складним завданням через динамічний характер онлайн-взаємодій користувачів. Методи Монте-Карло є одним із методів моделювання залученості користувачів, вірусності контенту, еволюції настроїв і персоналізованих рекомендацій у динамічних середовищах.*

У статті досліджено застосування методів Монте-Карло для моделювання та прогнозування поведінки користувачів у соціальних мережах. Розглянуто декілька варіантів цього підходу, серед яких – квазі-Монте-Карло, класичний метод Монте-Карло й Метрополіс-Гастингс методи продемонстрували кращу ефективність для обраного набору даних завдяки рівномірнішому охопленню сценаріїв поведінки користувачів, ураховуючи обрані параметри їхніх активностей. Також відображено збіжності методами Монте-Карло з вибіркою у 10000 значень і на основі основних обраних параметрів для моделювання поведінки користувачів, а саме лайків, поширень і коментарів, досліджено ефективність обраних методів при такому наборі даних з багатьма параметрами.

Базуючись на отриманих даних, розробили базову систему прогнозної аналітики, яка може отримати подальший розвиток завдяки інтеграції в рекомендаційні системи для підвищення персо-

налізації контенту й покращення аналізу активностей користувачів у соціальних мережах. Крім того, застосовано лексичний аналізатор VADER для оцінювання емоційного тону користувачів і досліджено їх вірусність у соціальних мережах. Отримані результати дали змогу відобразити розподіл лайків, коментарів і поширень у соціальних мережах. Також відображено процес моделювання вірусності й розповсюдження контенту, схематично подано принцип роботи програми. Результати оцінок ефективності підтверджують доцільність використання комбінації квазі-Монте-Карло методів і показують їх ефективність, точність прогнозування при аналізі активностей користувачів для комплексного дослідження поведінки в соціальних мережах. Прогнозування вірусності відображає поширення вірусного контенту.

Ключові слова: *моделювання методом Монте-Карло, аналітика соціальних мереж, моделювання поведінки користувачів, машинне навчання, прогнозна аналітика.*

Introduction

Social media platforms have become almost the main method of communication in the development of digital communication, influencing public opinion, marketing strategies and online interactions, etc. Modeling user behavior using digital platforms is crucial for businesses, researchers and policymakers looking to predict engagement trends, optimize content delivery and identify anomalies such as misinformation or bot activity and match users with better content.

However, user interactions, such as likes, shares, comments, and their changing moods, are inherently uncertain and depend on various external factors, making it difficult to make more accurate predictions when analyzing social media or displaying incompatible content on digital platforms. The Monte Carlo method is widely used in probabilistic modeling and analysis of results and offers a powerful approach to modeling behavior of social media users. By generating multiple results based on probabilistic distributions, Monte Carlo modeling allows you to estimate, for example, the probability of different patterns of interaction, the spread of viral content, and changes in user sentiment over time. This method will help eliminate uncertainty in the dynamics of social media, making it significant in forecasting and analyzing data, recommendation systems, and fraud detection. This article explores the application of the method Monte Carlo for modeling and predicting user behavior in social networks, as well as highlighting the advantages and problems of using probabilistic modeling in the dynamic environment of social networks. By using the Monte Carlo method, it is possible to improve decision-making in the platform's content strategy, advertising, and moderation, ultimately improving user experience and online security. Predictive analytics and modeling allows you to optimize your posting time for maximum user engagement for social media marketers and analysts.

Social media user behavior modeling research mostly focuses on machine learning, statistical modeling, and relationship analysis in graph structures. However, a small number of studies have used probabilistic methods, including Monte Carlo methods, for some cases of predicting patterns of interaction, dissemination of information, and mood dynamics in social networks. In general, the areas studied in this area can be divided into the use of probabilistic models for the analysis of social networks, Monte Carlo methods in predicting social networks, the study of user behavior and network dynamics and application in recommendation systems and personalization.

The first group of studies from the above used probabilistic approaches to analyze social media data. For example, Bayesian networks and hidden Markov models (HMM) were used to simulate the evolution of user sentiments and online discussions, in particular in the papers [1; 2]. Similarly, Monte Carlo Markov chain techniques (MCMCs) have been used to estimate the probability of specific user behaviors, such as content distribution and the likelihood of interaction at work [3]. As a result, the methods described allow us to investigate the stochastic nature of online interactions and improve behavioral predictions, which is important for general understanding.

The second group is the Monte Carlo methods in social network forecasting, used to estimate uncertainty in social media analytics. Studies [4] have demonstrated their effectiveness in predicting content virality by modeling different distribution scenarios based on historical interaction data. The papers [5; 6] integrated Monte Carlo methods with deep learning to refine trend predictions in social networks, particularly in dynamic environments where deterministic models face difficulties due to rapid changes in content.

The next group of related works can be attributed to studies of user behavior and network dynamics, namely how users interact with content on social networks using stochastic and networked approaches. Agent-based modeling and Monte Carlo simulations were combined to study cascades of information and the role of influencers in content distribution as implemented in the work [7].

Regarding the use in recommendation systems and personalization, the paper [8; 9] shows that Monte Carlo-based reinforcement methods can improve adaptive recommendation systems by predicting how users will react to different types of content.

Presentation of the main material and discussion of the results

To collect potential data for the study, publicly available social media datasets were selected, including user interaction (likes, shares, comments) and content functions (text, images, and metadata), and the time of publication was performed based on [10]. The main purpose of the data collected is to display

user engagement metrics across different types of publications, including news articles, user-generated entertainment content.

The collected data has been pre-processed to remove irrelevant information, such as spam or duplicate posts. Also, emotions or empty comments appear on social networks most often. Data preprocessing provides a clean and representative set of real-world data that is well standardized for display and modeling. For example, the collected dataset on the social network Instagram mostly contains collections of photos and comments, likes and shares, that is, engagement indicators that are associated with popular hashtags. The data obtained from the social network Facebook contains publications from public pages and groups that are not combined with each other by type of content, but are active in user interaction with publications.

The Monte Carlo method was used to model user behavior under different conditions and aspects of user behavior. The Monte Carlo method involves the use of not only based simulations of user activities in social networks, but also improved approaches of this method using Markov chains. Basic Monte Carlo simulations estimate engagement levels by generating random samples using probability distributions. When simulating more advanced methods. For example, using the Monte Carlo method with Markov chains (MCMC). This method is used to simulate complex user behavior and check the virality of content.

In Fig. 1 it is shown that the data collected by likes, shares on comments from the social network Instagram and Facebook, as the main parameters of engagement when modeling user behavior. This parameter allows you to simulate user interactions for posts with different content characteristics and time of publication dependence, which shows the amount of potentially engaged for this type of content based on a random sample of engagement distributions of previous historical data.

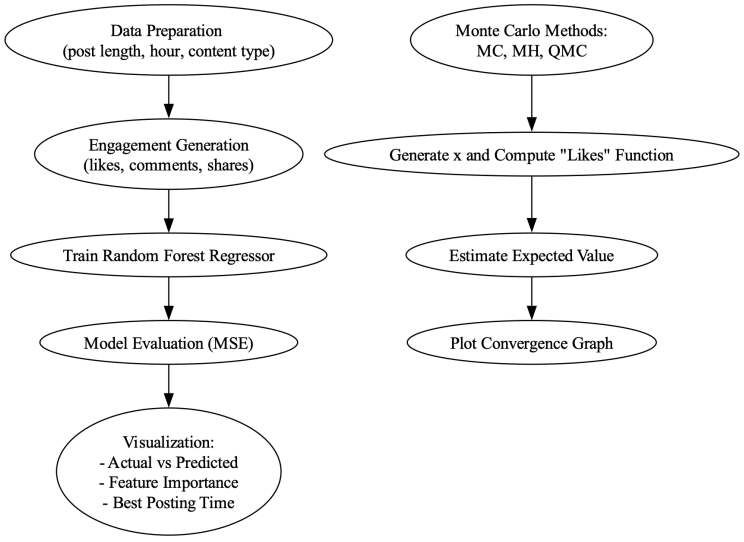

Fig. 1. Scheme of the program

The paper uses popular Python libraries, including NumPy and Pandas for generating and processing synthetic data, as well as Matplotlib for visualizing the results. Scikit-learn was used to build and evaluate the machine learning model, including the Random Forest Regressor algorithm and the standard error (MSE) metric.

In the first step, a dataset [10] is created that simulates social posts with three characteristics: length, time of publication, and type of content. Based on these parameters, interaction values are formed – likes, comments, and shares – with the addition of random noise. The built Random Forest Regressor model learns to predict the number of likes using these parameters. The accuracy of the prediction is estimated using the standard error. Additionally, graphs are created that visualize the correspondence between actual and predicted likes, the importance of signs and the optimal time to publish a post. The second stage analyzes three methods for estimating the expected number of likes: classic Monte Carlo, Metropolis-Hastings and Quasi-Monte Carlo. The classic Monte Carlo is simple and fast, but less accurate. Metropolis-Hastings is more accurate, but slower due to the use of Markov chains. Quasi-Monte Carlo provides the best balance between accuracy and efficiency through the use of Sobol sequences.

The number of values for research was chosen to be 10000. Organic user engagement and focused importance sampling, the basic Monte Carlo method, the improved Monte Carlo method using Markov chains, the Metropolis-Hastings and Quasi-Monte Carlo method are also displayed.

At the same time, the value of the error in each of the methods was significantly correlated. The estimation results show the values obtained by three different methods: the conventional Monte Carlo method (MC), the

Metropolis-Hastings method (MH), and the quasi-Monte Carlo (QMC) method. In this case, the convergence of methods is shown in Fig. 2. The MC method gave a score of 0.4064 and worked very quickly (0.0017 seconds). This shows that it is effective in simple tasks, although its accuracy depends on the number of simulations. The Metropolis-Hastings method, although it belongs to the improved stochastic methods, showed a significantly higher score (0.7124) and worked the longest (0.0888 seconds). This may indicate an insufficient number of iterations or an imperfect setting of the probabilistic transition, resulting in a less accurate result. This method requires careful adjustment to ensure stability and convergence to the correct solution. The quasi-Monte Carlo (QMC) method showed a score of 0.3929, which is close to the result of a conventional MC, but at the same time shows slightly higher stability. Although it took longer to complete (0.017 seconds), it often provides a lower variance in estimates, especially with fewer simulations.

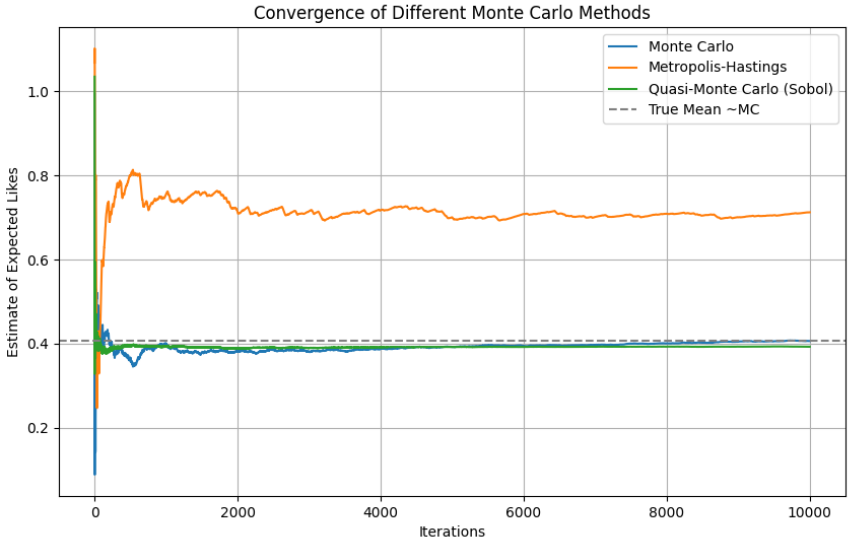

Fig. 2. Display of convergence by Monte Carlo methods (10000 values)

The next parameter that helps assess the likelihood of viral content due to active distribution and user engagement at the initial levels is the virality of the content.

Another parameter that allows you to check the dependence of sentiments on a certain topic or post are popularized over time, taking into account the activities of users, is the evolution of sentiments.

The simulation was carried out using the Monte Carlo method in the Python programming language. In addition, several basic NumPys were used to randomly sample the data and SciPy to statistically analyze the data obtained. And directly for data visualization, the most popular library matplotlib was used, which provides and provides an integrated approach to modeling uncertainty and predicting social media trends.

Data for modeling user behavior was obtained over 10,000 iterations to account for a wide range of possible outcomes and provide reliable statistical reliability. The user behavior model was built on the basis of a probabilistic structure that took into account user interaction and the engagement model from the distribution with the parameters evaluated from the dataset. The process of modeling content diffusion is used to investigate branching in the distribution of a publication and is determined by the level of engagement and its impact on the social network. Sentiment tracking simulation is also carried out. When tracking changes in user sentiment over time, sentiment analysis is applied on pre-trained models such as VADER (Valence Aware Dictionary and sEntiment Reasoner) to analyze the textual content of posts and comments. At the same time, the sentiment score is updated at each step of the simulation based on user activities (likes and comments, shares) to reflect the emotional tone of user interactions in social networks.

Modeling virality and content distribution (Fig. 3) is carried out by changing the initial number of interactions and the number of users and determining the factors that are responsible for the spread of viral content on social networks. Also, from the collected data, you can determine the dynamics of user sentiment among the provided post texts. Then the resulting parameter shows the evolution of sentiment regarding a certain trend (negative area of the Sentiment Score is negative, positive area is positive) or topics in social networks, taking into account external influences. For example, how holidays in Ukraine or marketing campaigns or news affect sentiment and how they fluctuate depending on user behavior and engagement.

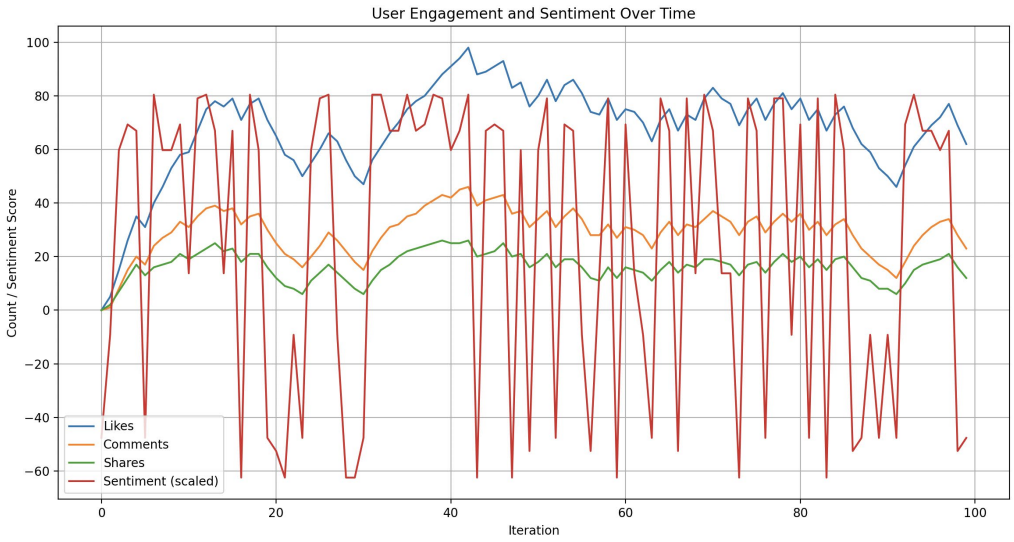

Fig. 3. Displaying the distribution of content: likes, comments and shares

Modeling virality and content distribution is done by changing the initial number of interactions and the number of users and determining the factors that are responsible for the spread of viral content on social networks. Also, from the collected data, you can determine the dynamics of user sentiment. Then the resulting parameter will show the evolution of sentiment regarding a particular trend or topic in social networks, taking into account external influences. For example, how holidays in Ukraine or marketing campaigns or news affect sentiment and how they fluctuate depending on user behavior and engagement.

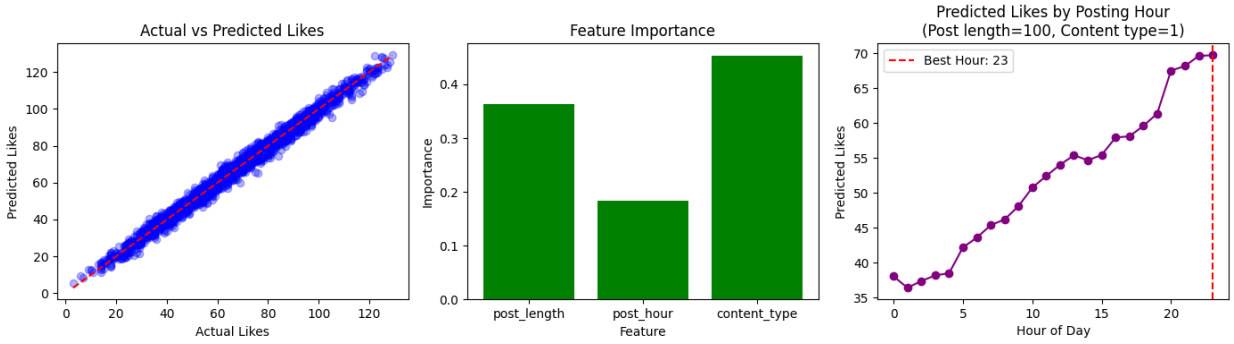

Fig. 4. Display of the Monte Carlo Efficiency Assessment

To evaluate the effectiveness of Monte Carlo modeling in predicting user behavior, as can be seen in Fig. 4, several performance indicators were used, such as: prediction accuracy, virality prediction, and exposure threshold. Prediction accuracy reflects the model's ability to predict actual engagement levels (e.g., the number of likes, shares, or comments) compared to real data, and virality prediction shows the accuracy of propagation modeling viral content, measured by the percentage of content that exceeded a predetermined engagement level. First, a dataset of 10.000 entries from prepared publications [10] and their interaction metrics such as likes, comments, and shares are presented. Next, this data is broken down into training and test samples to train the model. The Random Forest Regressor algorithm is used as a model, which learns to predict the number of likes. After training, the quality of the model is evaluated using the standard error (MSE), which shows the accuracy of the predictions. To better understand the results, three graphs are constructed: a comparison of actual and predicted likes, the importance of traits in the model, and recommendations for the best time to publish a post. The recommendation of the time of publication is based on the model's forecasts for different hours of the day, taking into account the fixed length of the post.

Conclusion

As part of the study, several variants of the Monte Carlo method were applied to model and predict user behavior in social networks. In particular, quasi-Monte Carlo methods turned out to be quasi-optimal for the selected dataset – they provided higher accuracy and stability of the results due to the reduction of sample

variance and more uniform coverage of the space of possible scenarios. Basic predictive analytics was developed, which, based on the results of simulations, can be integrated into recommendation systems. Such a system is potentially able to adapt the content delivery according to the intended level of user engagement, which opens up opportunities for personalizing marketing, optimizing news delivery and improving interaction with the audience. To analyze the emotional tone and identify the potential virality of the content, the lexical tool VADER (Valence Aware Dictionary and sEntiment Reasoner) was used, which made it possible to effectively classify the sentiments of messages in a large sample. This made it possible to better model the evolution of community sentiment and its impact on the dissemination of information. In general, the results of the study confirm the feasibility of using Monte Carlo methods, especially their quasi-deterministic variants, for a comprehensive analysis of user behavior in social networks.

References

1. Zheng, C. (2025). A comprehensive review of probabilistic and statistical methods in social network sentiment analysis. Advances in Engineering Innovation, 16(3), P. 38–43. DOI: https://doi.org/10.54254/2977-3903/2025.21918.

2. Li, H., Wang, Y., Zhang, J., & Liu, F. (2022). A Modified Attention Mechanism Powered by Bayesian Network for User Activity Analysis and Prediction. *Information Sciences, 603*, P. 115–130. DOI: https://doi.org/10.1016/j.ins.2022.02.003.

3. Wang, X., & Li, Q. (2022). Personality Trait Identification Based on Hidden Semi-Markov Model in Online Social Networks. *Proceedings of the 2022 7th International Conference on Intelligent Information Technology*, P. 30–35. DOI: https://doi.org/10.1145/3524889.3524898.

4. Fong, S. J., Li, G., Dey, N., Crespo, R. G., & Herrera-Viedma, E. (2020). Composite Monte Carlo Decision Making under High Uncertainty of Novel Coronavirus Epidemic Using Hybridized Deep Learning and Fuzzy Rule Induction. arXiv preprint arXiv:2003.09868. DOI: https://doi.org/10.48550/arXiv.2003.09868.

5. Zhang, J. (2020). Modern Monte Carlo Methods for Efficient Uncertainty Quantification and Propagation: A Survey. arXiv preprint arXiv:2011.00680. DOI: https://doi.org/10.48550/arXiv.2011.00680.

6. Sosa, J., & Martínez, C. (2025). Bayesian Sociality Models: A Scalable and Flexible Alternative for Network Analysis. arXiv preprint arXiv:2503.14697. DOI: https://doi.org/10.48550/arXiv.2503.14697.

7. Garibay, I. et al. (2021). Deep Agent: Studying the Dynamics of Information Spread and Evolution in Social Networks. In: Yang, Z., von Briesen, E. (eds) *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*. CSSSA 2020. Springer Proceedings in Complexity. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-77517-9_11.

8. Meshram, R., & Kaza, K. (2021). Monte Carlo Rollout Policy for Recommendation Systems with Dynamic User Behavior. *2021 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, P. 86–89. DOI: https://doi.org/10.1109/COMSNETS51098.2021.9352741.

9. Tommasel A., Diaz-Pace J. A. (2024). Leveraging Monte Carlo Tree Search for Group Recommendation. In Proceedings of the 18th ACM Conference on Recommender Systems (RecSys '24). *Association for Computing Machinery*, New York, NY, USA, P. 1136–1141. DOI: https://doi.org/10.1145/3640457.3691713.

10. Mysiuk I., Mysiuk R., Shuvar R. (2023). Collecting and analyzing news from newspaper posts in Facebook using machine learning. *Artificial Intelligence*. 28(1), P. 147–154 DOI: https://doi.org/10.15407/jai2023.01.147.

Дата першого надходження рукопису до видання: 26.05.2025
Дата прийнятого до друку рукопису після рецензування: 24.06.2025
Дата публікації: 25.07.2025