

УДК 004.8

DOI <https://doi.org/10.32782/IT/2024-1-5>

Артем КОНОТОПЧИК

студент, Луцький національний технічний університет, вул. Львівська, 75, м. Луцьк, Волинська обл., Україна, 43018

ORCID: 0009-0002-4770-8164

Катерина МЕЛЬНИК

кандидат технічних наук, доцент, доцент кафедри комп'ютерної інженерії та кібербезпеки, Луцький національний технічний університет, вул. Львівська, 75, м. Луцьк, Волинська обл., Україна, 43018

ORCID: 0000-0002-9991-582X

Scopus Author ID: 56366911300

Світлана ЛАВРЕНЧУК

кандидат технічних наук, доцент, доцент кафедри комп'ютерної інженерії та кібербезпеки, Луцький національний технічний університет, вул. Львівська, 75, м. Луцьк, Волинська обл., Україна, 43018

ORCID: 0000-0002-5453-3924

Scopus Author ID: 57209243001

Бібліографічний опис статті: Конотопчик, А., Мельник, К., Лавренчук, С. (2024). Методики CRO на основі машинного навчання. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 1, 35–40, doi: <https://doi.org/10.32782/IT/2024-1-5>

МЕТОДИКИ CRO НА ОСНОВІ МАШИННОГО НАВЧАННЯ

У даній статті проводиться аналіз методик оптимізації конверсій (CRO) на основі машинного навчання. Детально розглядаються існуючі системи CRO, зокрема їх функціонал та ефективність. В статті надається огляд різних методів, які вже використовуються в CRO системах, таких як A/B-тестування, веб-аналітика та персоналізація контенту.

Окрема увага приділяється впровадженню методів машинного навчання в сферу CRO. Зазначаються переваги використання ML, такі як здатність аналізувати великі обсяги даних, швидкість прийняття рішень та автоматизація процесів. Подано конкретний план впровадження ML в систему CRO, включаючи використання алгоритмів для прогнозування змін у конверсіях, ідентифікації патернів користувачів та оптимізації контенту.

Стаття висвітлює важливість розвитку та сучасності в галузі CRO, а також вказує на потенційні переваги використання машинного навчання для підвищення ефективності оптимізації конверсій.

Мета роботи полягає у вдосконаленні існуючих методів CRO шляхом використання методів машинного навчання для точнішого прогнозування поведінки користувачів та ефективнішої персоналізації контенту.

Методологія дослідження базується на аналізі великих даних, використанні нейронних мереж та кластерного аналізу. Зокрема, застосовано метод кластерного аналізу для групування текстових описів товарів та рекомендаційна система на основі алгоритму колаборативної фільтрації реалізованому глибокою нейронною мережею.

Наукова новизна полягає у використанні машинного навчання для адаптивної рекомендаційної та пошукової системи на основі нейронних мереж та кластерного аналізу.

Висновком є значний потенціал використання машинного навчання у сфері CRO, демонструючи, що впровадження глибоких нейронних мереж та алгоритмів кластерного аналізу можуть бути ефективними у прогнозуванні конверсій та персоналізації контенту.

Ключові слова: оптимізація коефіцієнту конверсій, машинне навчання, оптимізація контенту.

Artem KONOTOPCHUK

student, Lutsk National Technical University, 75, Lvivska Str., Lutsk, Ukraine, 43018, artemkonotopchik2003@gmail.com

ORCID: 0009-0002-4770-8164

Kateryna MELNYK

PhD, Associate Professor, Associate Professor at Department of Computer Engineering and Cybersecurity, Lutsk National Technical University, 75, Lvivska Str., Lutsk, Ukraine, 43018, ekaterinamelnik@gmail.com

ORCID: 0000-0002-9991-582X

Scopus Author ID: 56366911300

Svitlana LAVRECHUK

PhD, Associate Professor, Associate Professor at Department of Computer Engineering and Cybersecurity, Lutsk National Technical University, 75, Lvivska Str., Lutsk, Ukraine, 43018, LavrSveet@gmail.com

ORCID: 0000-0002-5453-3924

Scopus Author ID: 57209243001

To cite this article: Konotopchyk, A., Melnyk, K., Lavrenchuk, S. (2024). Metodyky CRO na osnovi mashynnoho navchannia [CRO methods based on machine learning]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 1, 35–40, doi: <https://doi.org/10.32782/IT/2024-1-5>

CRO METHODS BASED ON MACHINE LEARNING

This article analyzes conversion rate optimization (CRO) methodologies based on machine learning. Existing CRO systems, including their functionality and efficiency, are examined in detail. The article provides an overview of various methods already used in CRO systems, such as A/B testing, web analytics, and content personalization.

Special attention is paid to the implementation of machine learning methods in the CRO field. The advantages of using ML are highlighted, such as the ability to analyze large volumes of data, speed of decision-making, and process automation. A specific plan for implementing ML in the CRO system is presented, including the use of algorithms for predicting changes in conversions, identifying user patterns, and optimizing content.

The article highlights the importance of development and modernity in the CRO field, as well as points out the potential advantages of using machine learning to enhance the efficiency of conversion optimization.

The purpose of the work is to improve existing CRO methods by using machine learning methods for more accurate prediction of user behavior and more effective content personalization.

The research methodology is based on the analysis of big data, the use of neural networks, and cluster analysis. In particular, a cluster analysis method is applied for grouping text descriptions of goods and a recommendation system based on a collaborative filtering algorithm implemented by a deep neural network.

The scientific novelty lies in the use of machine learning for an adaptive recommendation and search system based on neural networks and cluster analysis.

The conclusion is the significant potential of using machine learning in the field of CRO, demonstrating that the implementation of deep neural networks and cluster analysis algorithms can be effective in predicting conversions and personalizing content.

Key words: conversion rate optimization, machine learning, content optimization.

Постановка наукової проблеми. В епоху цифрового маркетингу оптимізація коефіцієнта конверсії (CRO) стала вирішальним аспектом для бізнесу, щоб покращити свої сервіси в Інтернеті та збільшити дохід. Зростання конкуренції та постійні зміни відгуків користувачів вимагають більш ефективних підходів до CRO, щоб забезпечити стабільні та значущі результати. Машинне навчання надає системам CRO можливість аналізувати величезні набори даних і отримувати бажаний результат. Використовуючи алгоритми, які навчаються та адаптуються з часом, машинне навчання дозволяє системам CRO визначати шаблони, розуміти вподобання користувачів і прогнозувати потенційні бажання користувача. Цей підхід закладає основу для більш персоналізованої та ефективною взаємодії з користувачем.

Аналіз досліджень. Вибір методики проведення дослідження конверсії на веб-сайтах

залежить від багатьох факторів. Існуючі методики, такі як: A/B-тестування (Quin, 2023, с. 1-3), веб-аналітика (Purnomo, 2023, с. 59) та персоналізація контенту, є дієвими та мають свої переваги та обмеження. Однак, у зв'язку з стрімким розвитком інтернет маркетингу кількість даних, що необхідно брати до уваги для проведення аналітики, зростає швидше, ніж можливо обробити стандартними методиками. У зв'язку з цим, пропонується поєднання методів аналізу великих даних, машинного навчання та CRO методів. Одним з способів використання технологій машинного навчання задля оптимізації коефіцієнту конверсії з передбаченням рекомендацій користувачу описані в статті (Zaki, 2024), де автори розглядають такі методики машинного навчання, як: SGD класифікатор, класифікатор k-найближчих сусідів та класифікатор випадкового лісу. Також, для

рішення завдання оптимізації рейтингу конверсії застосовуються гібридні методи машинного навчання, що описані в статті (Kagurusamy, 2020). Використання нейронних мереж в економічній сфері було розглянуто авторами Y. Rajihu, K. Nermend та A. Alsakaa у статті (Rajihu, 2017), а саме, нейронні мережі зворотного поширення, показали їх ефективність у сфері фінансового прогнозування, провівши дослідження трьох моделей з різними гіперпараметрами.

В даній статті розглядаються глибокі нейронні мережі та алгоритм кластеризації.

Метою роботи є вдосконалення систем CRO шляхом впровадження методів машинного навчання для прискорення аналізу даних, забезпечення швидкого прийняття рішень та автоматизації ключових етапів оптимізації конверсії.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. В сучасних CRO системах використовуються різні методики для підвищення рівня конверсії, такими методиками є: A/B тестування, веб-аналітика та персоналізація контенту.

A/B тестування – це методика, при якій веб-сайт, або його частина, відображається по-різному для декількох окремих груп користувачів (Яковленко, 2024). Особливістю даної методики є те, що результатом її проведення буде саме маркетинговий ефект від реальних користувачів. Такими методами можна визначати коректність появи оголошення в конкретних подіях. Недоліком даної системи є відсутність можливості отримати результат швидко.

Веб-аналітика – це використання інструментарію аналітичних інструментів (Preston, 2024). Прикладом такого інструменту може бути сервіс Google Analytics, що збирає інформацію про користувача та його дії. Дана модель оптимізації конверсії дозволяє відслідковувати поведінку користувача в часі, проведеному на ресурсі, кількості приділеної уваги до елементів веб-сайту, шлях, що проходить користувач на ресурсі.

Персоналізація контенту – це методика візуалізації контенту та його змісту на основі попереднього досвіду користувача на ресурсі (Носенко, 2020). Це дозволяє чітко розуміти фокус уваги користувача за рахунок його попередніх самовільних дій, однак має проблему холодного старту та залежність від так званого «шуму».

Кожна з перелічених методик є невід'ємною частиною побудови CRO систем, однак характеризуються взаємодією з великим об'ємом даних. У зв'язку з цим, пропонується використання

машинного навчання задля вдосконалення ефективності використання класичних методик.

В даній статті розглядаються такі методики, як:

- Колоборативна фільтрація по рейтингу оцінки продуктів користувачами засобами глибокої нейронної мережі.

- Рекомендація товарів на основі кластерного текстового аналізу опису товару без історії покупок користувачів.

Глибока нейронна мережа. Дана технологія машинного навчання використовується для досягнення встановлення взаємозв'язків між великими наборами даних тоді, коли існує багато вхідних даних. Одним з таких випадків є рейтингова система, коли користувачі ставлять оцінки товарам. В такому випадку, якщо кількість даних є великою, використання глибокої нейронної мережі дозволяє реалізувати колоборативну фільтрацію. Колоборативна фільтрація – це алгоритм, що дозволяє знайти вподобання користувача в залежності від його оцінок, та оцінок інших користувачів на конкретні товари (Srifi, 2020, с. 4-5). Дана методика CRO системи є реалізацією рекомендаційної системи на основі оцінок інших користувачів.

Для реалізації колоборативної фільтрації на основі глибокої нейронної мережі був використаний датасет «Amazon – Ratings (Beauty Products)» (Kaggle: сайт, 2018). Передбачається, що кожен користувач виставив оцінку чи декілька оцінок різним товарам, які вже були оцінені іншими користувачами. В такому випадку, навчальною вибіркою буде набір користувачів та товарів, та оцінок, виставлених ними. Задля покращення ефективності навчання нейронної мережі була проведена нормалізація даних.

Архітектура нейронної мережі схематично відображена на рисунку 1. Вхідними параметрами є користувач та товар, що він оцінює, маркерованими даними є виставлена оцінка. Вихідним результатом є прогноз оцінки товару.

Результат навчання моделі зображений на рисунку 2. У зв'язку з тим, що в даному випадку розглядається задача регресії, були обрані такі метрики оцінки втрат, як: середньоквадратична похибка та середня абсолютна похибка.

На основі побудованої рекомендаційної функції отримуємо наступний рейтинг товарів, що зображений на рисунку 3.

Результат роботи мережі (рисунок 3) представляє собою список товарів, відсортованих таблично по прогнозованому рейтингу товару, що визначає можливу зацікавленість користувача в кожному з товарів.

Реалізація нейронної мережі потребує маркерованих даних для проведення навчання та

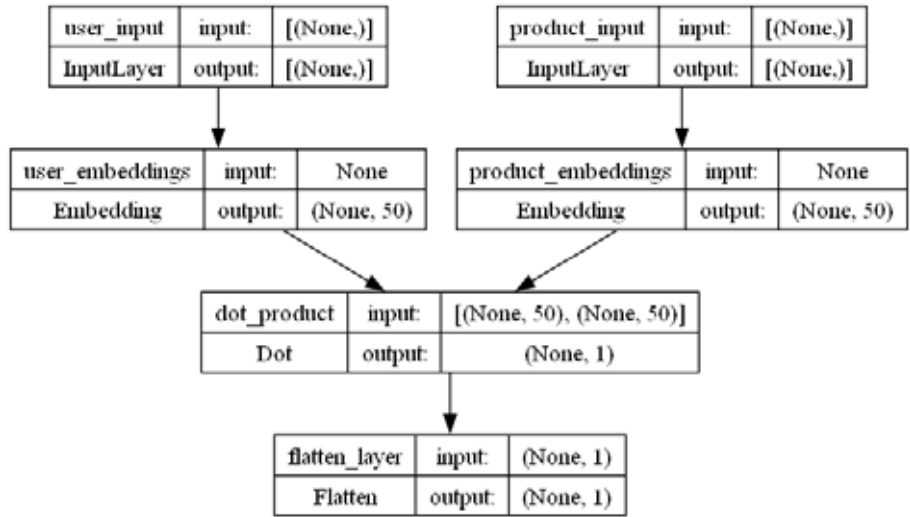


Рис. 1. Архітектура глибокої нейронної мережі

```

35 Epoch 4/10
36 6329/6329 [=====] - 1286s 203ms/step - loss: 0.0572 - mean_squared_error: 0.0572
37 - mean_absolute_error: 0.1891 - val_loss: 0.7364 - val_mean_squared_error: 0.7364 - val_mean_absolute_error: 0.7882
38 1563/1563 [=====] - 1s 628us/step - loss: 0.7366 - mean_squared_error: 0.7366
39 - mean_absolute_error: 0.7946
40 Test Metrics: {'loss': 0.7366313934326172, 'mean_squared_error': 0.7366313934326172,
41               'mean_absolute_error': 0.7946288715438843}
42 Кількість користувачів: 708895
43 Кількість продуктів: 92740
44
  
```

Рис. 2. Результат навчання моделі та вивід тестових метрик

```

71      UserId  ProductId  Rating
72      3641     45         3641  1.283130
73      13647    45        13647  1.214035
74      53843    45        53843  1.212302
75      84376    45        84376  1.212063
76      69831    45        69831  1.209296
77      ...     ...         ...     ...
78      42821    45        42821  0.797776
79      64205    45        64205  0.795509
80      22621    45        22621  0.791986
81      30729    45        30729  0.780751
82      1230     45         1230  0.772023
83
84      [92740 rows x 3 columns]
85
86      Process finished with exit code 0
  
```

Рис. 3. Рейтинг товарів для конкретного користувача на основі навченої нейронної мережі

отримання ефективного результату. Крім того, дані повинні бути підготовлені, маркеровані та збалансовані. У випадку невиконання умов підготовки даних, а саме їх оптимального об'єму, з'являється проблема, яка називається «холодний старт». Проблема «холодного старту» – це ситуація, коли наявних даних недостатньо для проведення аналітичних висновків з неї

(Орехова, 2019). Дана проблема може бути вирішена природнім шляхом, за рахунок тимчасового очікування набору достатньої кількості даних. Однак, є методи для запобігання очіванню природнього вирішення даної проблеми. Одним з таких методів є задача кластеризації, що не потребують маркерованих даних. Опис датасета «Home Depot product» наведено на (Data World: сайт, 2021). Користувач вважається новим та немає взаємозв'язків з іншими користувачами та товарами. Пропонується використати кластерний аналіз, де алгоритм TF-IDF визначає важливість кожного слова в кожному документі, а алгоритм K-Means визначає кластери (групи) документів на основі їхніх TF-IDF представлень (рисунок 4).

Реалізація кластеризації та формування кластерів з виводом популярних термів продемонстровано на рисунку 5.

Результатом роботи (рисунок 6) розробленої програми є можливість передбачити кластер, до якого відноситься пошуковий запит користувача, що дозволяє більш точно визначати можливий товар чи список товарів, що варто запропонувати користувачу на основі описів товарів, вже доданих на онлайн-ресурс (веб-сайт).

```
# Використання TfIdfVectorizer для перетворення текстових даних у матрицю Term
Frequency-Inverse Document Frequency (TF-IDF)
vectorizer = TfIdfVectorizer(stop_words='english')
X = vectorizer.fit_transform(product_descriptions["description"])

# Виконання кластеризації алгоритмом KMeans з 15-ма кластерами
kmeans = KMeans(n_clusters=15, init='k-means++')
y_kmeans = kmeans.fit_predict(X)
```

Рис. 4. Програмний код стадії препроцесінгу кластеризації текстових даних

```
url ... specifications
72 https://www.homedepot.com/3Green-Its-Start-Kit-... [{"shape": "Round"}, {"package type": "Set"}, ...
74 https://www.homedepot.com/211mm-Primer-1-... [{"Minimum Temperature For Use (F)": "32"}, {"...
76 https://www.homedepot.com/2KPA-PAINT-1-2-... [{"Minimum Temperature For Use (F)": "32"}, {"...
78 https://www.homedepot.com/2KMB-104-3-1-1-... [{"Minimum Temperature For Use (F)": "32"}, {"...
80 https://www.homedepot.com/2KMB-111-3-1-1-... [{"shape": "Rectangular"}, {"Number of Seps"...

15 rows x 17 columns
Кластер 0: plow, precision, 3rd, mowed, lawn, vests, independent, requirements, party, benefits
Кластер 1: scuff, defense, ultra, paint, durability, primer, bear, settle, breakthrough, oxide
Кластер 2: pop, diamond, outstanding, resistance, provides, beautifies, protects, paint, ultralast, exceptional
Кластер 3: exterior, start, base, mesh, trembly, great, weeps, restoring, doors, ease
Кластер 4: gear, machine, interior, delivering, fresh, performance, stay, solar, coverage, guarantee
Кластер 5: exterior, excellent, ultra, fences, roof, nichies,kehr, paint, siding, look
Кластер 6: latex, exterior, severe, rays, conditions, acrylic, uv, prefer, application, weather
Кластер 7: outdoor, withstand, wear, sleek, radiant, appearance, exterior, shutters, repulsed, doors
Кластер 8: area, paint, hide, interior, certified, ultra, latex, application, painted, woc
Кластер 9: sample, covers, sq, sq, latex, idea, is, true, buy, tested
Кластер 10: rugs, rug, area, collection, nuLoom, living, home, decor, pile, woven
Кластер 11: timeless, pop, coat, coverage, complete, colors, paint, specified, guarantee
Кластер 12: 1, wipes, moisture, glass, like, spore, resists, enamel, plus, creates
Кластер 13: diamond, results, essentials, precision, value, woc, paint, coating, years, exceptional
Кластер 14: design, use, light, easy, style, wall, room, home, perfect, collection
```

Рис. 5. Результат виконання кластеризації текстових описів товарів

```
Передбачений кластер W по запиту rugs: 10
Кластер 10: rugs, rug, area, collection, nuLoom, living, home, decor, pile, woven
Передбачений кластер W по запиту latex: 6
Кластер 6: latex, exterior, severe, rays, conditions, acrylic, uv, prefer, application, weather
Передбачений кластер W по запиту color: 14
Кластер 14: design, use, light, easy, style, wall, room, home, perfect, collection
Передбачений кластер W по запиту spray paint: 14
Кластер 14: design, use, light, easy, style, wall, room, home, perfect, collection
Передбачений кластер W по запиту steel drill: 14
Кластер 14: design, use, light, easy, style, wall, room, home, perfect, collection

Process finished with exit code 0
```

Рис. 6. Передбачення кластеру по пошуковому запису користувача

Висновки. У даній статті було розглянуто два методи оптимізації рейтингу конверсії на основі машинного навчання та реалізована можливість їх використання. Були розглянуті методи глибоких нейронних мереж на основі алгоритму колаборативної фільтрації, що враховує спільні оцінки користувачів для формування

потенційних рекомендацій, та рекомендація товарів на основі пошукового запиту користувача по кластеризованим даним по алгоритму k-найближчих сусідів. Зазначено важливість розвитку CRO та показані потенційні переваги використання машинного навчання для підвищення ефективності оптимізації конверсії.

ЛІТЕРАТУРА:

1. A/B Testing: A Systematic Literature Review / Federico Quin та ін. 2023. URL: <https://arxiv.org/pdf/2308.04929.pdf> (дата звернення 19.02.2024).
2. Digital Marketing Strategy to Increase Sales Conversion on E-commerce Platforms. Yudiyanto Joko Purnomo. 2023. URL: <https://doi.org/10.61100/adman.v1i2.23> (дата звернення 20.02.2024).
3. Predictive Analytics and Machine Learning in Direct Marketing for Anticipating Bank Term Deposit Subscriptions / A. M. Zak та ін. 2024. URL: <http://surl.li/rxrue> (дата звернення: 26.02.2024).
4. P. D. K. Analysis of Neural Network Based Language Modeling. *March 2020*. 2020. Т. 2, № 1. С. 53–63. URL: <http://surl.li/ryndd> (дата звернення: 25.03.2024).
5. Rajihy, Y., Nermend, K. and Alsakaa, A. (2017). Back-propagation artificial neural networks in stock market forecasting. An application to the Warsaw Stock Exchange WIG20, AESTIMATIO, The IEB International Journal of Finance, 15, pp. 88-99. doi: 10.5605/IEB.15.5. URL: <http://surl.li/rxrue> (дата звернення: 26.02.2024).

6. A/B-тестування: що це таке та як проводити спліт-тест. *Блог хостера HOSTiQ.ua*. URL: <http://surl.li/rxrwh> (дата звернення: 04.03.2024).
7. Preston J. digital marketing institute. URL: <http://surl.li/rxrvb> (дата звернення: 20.03.2024).
8. Носенко Ю. Knewton analytical review as a platform for personalization of learning content. *Information Technologies in Education*. 2020. № 44. С. 65–76. URL: <https://doi.org/10.14308/ite000727> (дата звернення: 25.03.2024).
9. Recommender Systems Based on Collaborative Filtering Using Review Texts—A Survey / M. Srifi та ін. 2020. URL: <https://doi.org/10.3390/info11060317> (дата звернення: 20.02.2024).
10. Amazon – Ratings (Beauty Products). Kaggle: Your Machine Learning and Data Science Community. URL: <http://surl.li/rxrvk> (дата звернення: 19.02.2024).
11. Орехова І. В. _Дослідження методів вирішення проблеми холодного старту в проектах побудови рекомендаційних систем. URL: <http://surl.li/rxrwp> (дата звернення: 13.03.2024).
12. Home Depot Product Datasets. URL: <http://surl.li/rxrwy> (дата звернення: 21.02.2024).

REFERENCES:

1. A/B Testing: A Systematic Literature Review / Federico Quin та ін. 2023. Retrieved from: <https://arxiv.org/pdf/2308.04929.pdf>.
2. Digital Marketing Strategy to Increase Sales Conversion on E-commerce Platforms / Yudiyanto Joko Purnomo. 2023. Retrieved from: <https://doi.org/10.61100/adman.v1i2.23>.
3. Predictive Analytics and Machine Learning in Direct Marketing for Anticipating Bank Term Deposit Subscriptions / A. M. Zak та ін. 2024. Retrieved from: <http://surl.li/ryndd>.
4. P. D. K. Analysis of Neural Network Based Language Modeling. March 2020. 2020. Т. 2, № 1. С. 53–63. Retrieved from: https://web.archive.org/web/20210813230021id_/https://www.irojournals.com/aicn/V2/I1/06.pdf.
5. Rajihy, Y., Nermend, K. and Alsakaa, A. (2017). Back-propagation artificial neural networks in stock market forecasting. An application to the Warsaw Stock Exchange WIG20, AESTIMATIO, The IEB International Journal of Finance, 15, pp. 88-99. doi: 10.5605/IEB.15.5. Retrieved from: <http://surl.li/rxrwo>.
6. A/B-testuvannya: shcho tse take ta yak provodyty split-test. Bloh khostera HOSTiQ.ua. Retrieved from <http://surl.li/rxrwh>.
7. Preston J. digital marketing institute. Retrieved from: <http://surl.li/rxrvb>.
8. Nosenko Yu. (2020). Knewton analytical review as a platform for personalization of learning content. *Information Technologies in Education*. № 44. С. 65–76. Retrieved from: <https://doi.org/10.14308/ite000727>.
9. Srifi, M. та ін. (2020). Recommender Systems Based on Collaborative Filtering Using Review Texts—A Survey / Retrieved from: <https://doi.org/10.3390/info11060317>.
10. Amazon – Ratings (Beauty Products). Kaggle: Your Machine Learning and Data Science Community. Retrieved from: <http://surl.li/rxrvk>.
11. Oriekhova I. V. _Doslidzhennia metodiv vyrishennia problemy kholodnoho startu v proektakh pobudovy rekomendatsiinykh system [Research of methods of solving the problem of cold start in projects of building recommender systems]. Retrieved from: <http://surl.li/rxrwp> (data zvernennia: 13.03.2024).
12. Home Depot Product Datasets. Retrieved from: <http://surl.li/rxrwy>.