

І. В. Піх¹
Ю. Ю. Меренич^{1,2}

ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ КЛАСТЕРИЗАЦІЇ ЦІЛЬОВОЇ АУДИТОРІЇ ВЕБ-ДОДАТКІВ

¹Національний університет «Львівська політехніка»;

²Державний вищий навчальний заклад «Ужгородський національний університет»

У дослідженні використано метод машинного навчання для кластеризації даних щодо цільової аудиторії веб-додатків електронної комерції. Машинне навчання є потужним інструментом аналізу, що дає змогу автоматично виокремлювати закономірності у великих вибірках, підвищуючи точність прогнозування поведінки користувачів. Вибрано основні показники взаємодії з веб-додатками: коефіцієнт відмов, тривалість сесії та коефіцієнт конверсії. Виконано нормалізацію вхідних даних. Для нормалізації та коректної роботи алгоритмів машинного навчання використано метод, що забезпечує масштабування значень у діапазоні від нуля до одиниці. Оптимальна чисельність кластерів отримана за методом «ліктя», який аналізує залежність між кількістю кластерів та внутрішньокластерною сумою квадратів відстаней. Застосовано метод k -середніх для аналізу поведінкових параметрів, що мінімізує суму квадратів відстаней між точками та центроїдами кластерів, використовуючи евклідову метрику. Візуалізацію результатів здійснено за допомогою тривимірного графіка. Представлено розподіл кластерів за аналізованими параметрами.

Результати кластеризації дозволили виділити чотири групи користувачів з різними характеристиками взаємодії з веб-ресурсами. Користувачі першого кластера мають низьку залученість, коротку тривалість сесії та високий коефіцієнт відмов, що вказує на недостатню релевантність контенту. Другий кластер демонструє тривалу взаємодію. Однак високий рівень відмов може свідчити про складність навігації. Третій характеризується високою конверсією за помірної тривалості сесії, що вказує на ефективний користувацький досвід. Останній має найнижчий коефіцієнт відмов і максимальну конверсію, що відображає високу відповідність контенту потребам користувачів. Практичне значення отриманих результатів полягає у можливості застосування методів кластерного аналізу для адаптації UX/UI-рішень, оптимізації контенту та підвищення рівня конверсії. Запропонований підхід може бути використаний у сфері електронної комерції, цифрового маркетингу та вебаналітики для вдосконалення стратегії взаємодії з користувачами.

Ключові слова: машинне навчання, кластеризація, вебдодаток, цільова аудиторія, електронна комерція.

Вступ

Машинне навчання є революційною технологією у сфері штучного інтелекту. З кожним роком воно стає все популярнішим завдяки здатності до прогнозування, аналізу великих масивів даних і надання інформації, яку раніше було складно чи неможливо отримати. Машинне навчання змінює спосіб взаємодії з технологіями та відкриває нові можливості для інновацій [1]. Для класифікації об'єктів на основі набору характеристик успішно використовується кластеризація, зокрема метод k -середніх. Такі дослідження активно проводяться у багатьох сферах людської діяльності: у сільському господарстві [2], медицині [3], освіті [4], сфері інформаційних технологій [5] та ін.

У сучасному цифровому середовищі, де веб-ресурси є основними інструментами взаємодії між бізнесом та споживачами, розуміння поведінки користувачів набуває критичного значення для оптимізації користувацького досвіду та підвищення ефективності бізнес-процесів. Аналіз поведінкових характеристик відвідувачів дозволяє виявити приховані закономірності та сегментувати користувачів для розроблення цільових стратегій взаємодії.

Попередні дослідження показали, що кластеризація цільової аудиторії на основі поведінкових метрик може суттєво покращити розуміння потреб та підвищити ефективність маркетингових рішень. Наприклад, у [6] розроблено модель рекомендаційної системи для комерційних вебсайтів, основою якої є аналіз поведінки користувачів. Проте відсутнє порівняння запропонованої моделі з іншими сучасними рекомендаційними системами, що ускладнює оцінювання відносної ефективності. У [7] представлено фреймворк рекомендацій веб-сторінок для електронних комерційних сайтів. Використовуються алгоритми кластеризації для групування схожих веб-сторінок або користувачів згідно з поведінковими характеристиками. Цікавим є також дослідження [8], яке зосереджене на моделюванні вподобань відвідувачів інтернет-магазинів на основі аналізу діяльності пристрою введення користувацьких команд. Виявлені патерни уможливили прогнозування ймовірності купівлі. Проте в аналізованих літературних джерелах недостатньо уваги приділено таким ключовими метрикам, як коефіцієнт відмов, тривалість сесії та коефіцієнт конверсії, що здійснюють неопосередкований вплив на ефективність електронної комерції.

Мета дослідження полягає саме у кластеризації цільової аудиторії, що може бути застосовано як один з етапів майбутнього аналізу впливу різних факторів, зокрема й веб-дизайну на ефективність електронної комерції.

Результати досліджень

Для аналізу використано набір даних про поведінку користувачів, зібраних з платформ електронної комерції [9]. Виокремлено три ключові параметри: тривалість сеансу, коефіцієнт відмов і коефіцієнт конверсії. Тривалість сеансу відображає глибину взаємодії користувача з контентом, коефіцієнт відмов є індикатором негайного залишення сторінки, а коефіцієнт конверсії є прямим показником ефективності взаємодії.

Вибір обумовлений важливістю даних факторів у визначенні рівня залученості користувачів, ефективності взаємодії з веб-сайтом і ймовірності досягнення комерційних цілей. Тривалість сесії відображає загальний час, витрачений на взаємодію з контентом. Висока тривалість сесії часто корелюється із зацікавленістю користувачів. Проте її надмірне збільшення за відсутності конверсій може свідчити про складність навігації або нерелевантність інформації. Коефіцієнт відмов вказує частку користувачів, які покинули веб-ресурс після перегляду лише однієї сторінки. Високий коефіцієнт відмов може сигналізувати про проблеми з релевантністю контенту, зручністю користування або швидкістю завантаження сторінок. Коефіцієнт конверсії відображає частку користувачів, які виконали цільову дію, таку як здійснення покупки або заповнення форми. Низький коефіцієнт конверсії за високої тривалості сесії може свідчити про перешкоди у процесі ухвалення рішень або неефективність закликів до дії.

Первинним етапом опрацювання даних є нормалізація, яка виконується для приведення значень змінних до єдиного масштабу. Це запобігає домінуванню одних параметрів над іншими під час виконання алгоритмів машинного навчання. Трансформацію значень змінних у діапазон $[0, 1]$ здійснено за формулою

$$D_{norm} = \frac{d - d_{min}}{d_{max} - d_{min}}, \quad (1)$$

де d — значення змінної, d_{min} та d_{max} — мінімальне та максимальне значення відповідної змінної.

Фрагмент даних, отриманих внаслідок нормалізації, подано у табл. 1.

Таблиця 1

Фрагмент нормалізованих даних

№	Коефіцієнт відмов	Тривалість сеансу	Коефіцієнт конверсії
1	0,649253	0,360857	0,197616
2	0,738218	0,157792	0,714384
3	0,558675	0,646467	0,315775
4	0,665473	0,058005	0,014434
5	0,982772	0,857408	0,566322

Кількість кластерів визначено за методом «ліктя» (рис. 1) [2]. Він базується на аналізі залежності між кількістю кластерів та сумою квадратів відстаней між точками та їхніми відповідними центрами:

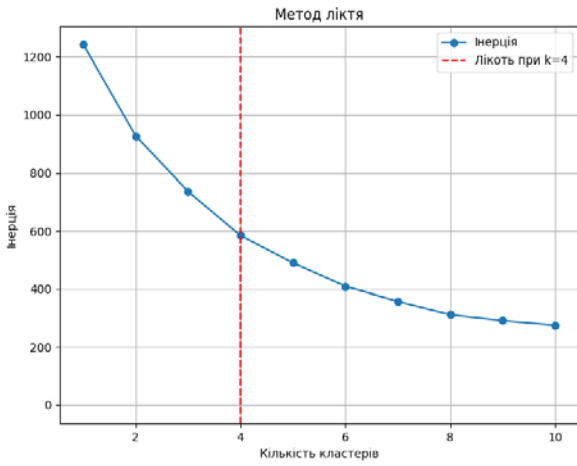


Рис. 1. Визначення оптимальної кількості кластерів

чати природні групи серед даних. При цьому цільова функція має такий вигляд:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^j - \mu_j\|^2, \tag{3}$$

де величина $\|x_i^j - \mu_j\|^2$ — міра визначення відстані між конкретною точкою даних x_i^j та центром кластера μ_j .

Результати кластеризації показані на рис. 2.

$$WCSS = \sum_{i=1}^n (x_i^j - \mu_j)^2, \tag{2}$$

де x_i^j — точка даних; μ_j — центр кластера; j — номер кластера.

Отже, оптимальною є така кількість кластерів: $k = 4$.

Для кластеризації використано метод k -середніх [2], [10], що полягає у розподілі об'єктів у k групи на основі їхньої схожості. Алгоритм мінімізує внутрішньокластерну дисперсію шляхом ітеративного оновлення центрів кластерів. Метод широко застосовується в аналізі поведінки користувачів, сегментації ринку та інших сферах, де необхідно визна-

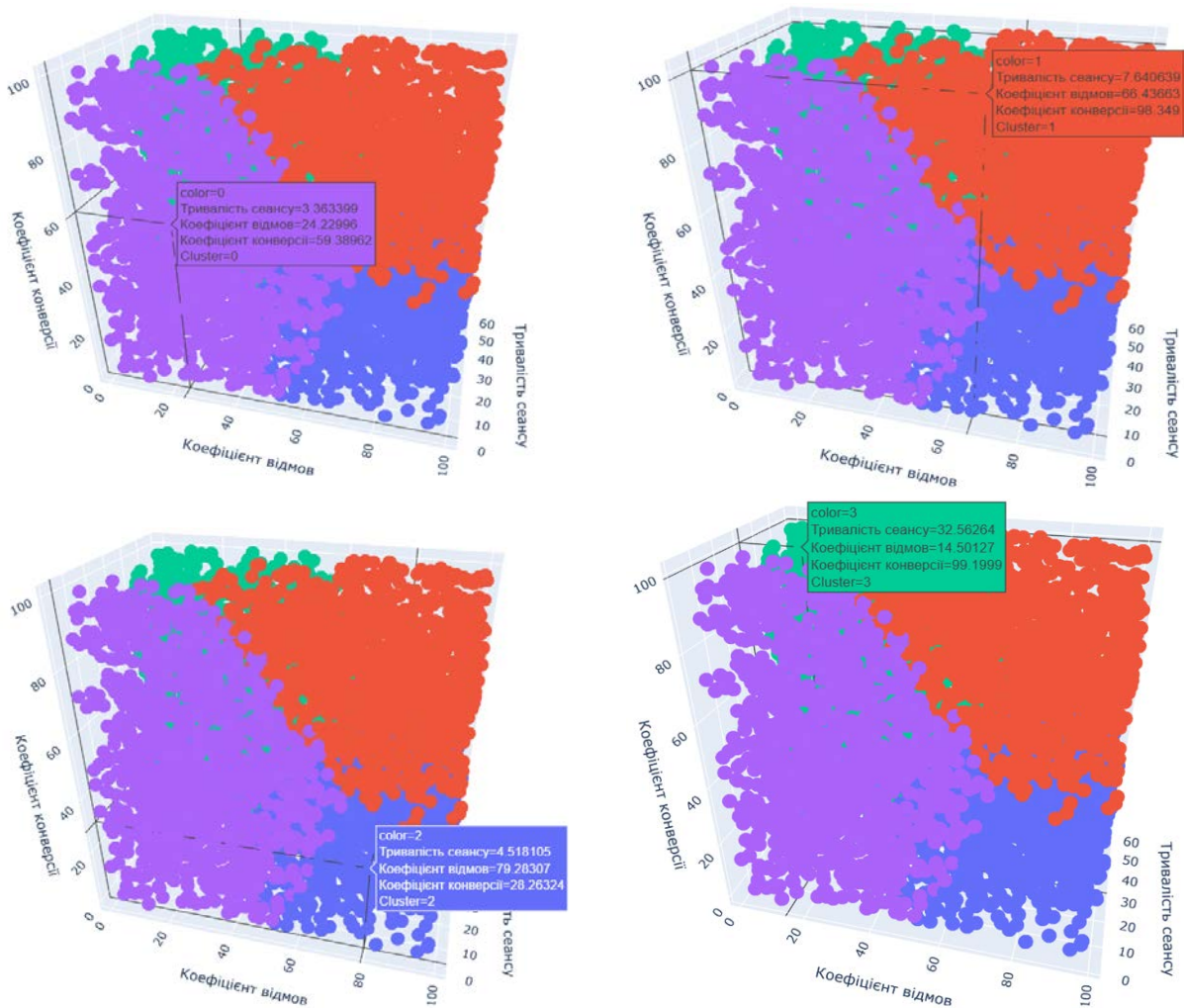


Рис. 2. Візуалізація результатів кластеризації

Цільова аудиторія, представлена у кластері 0 демонструє середню тривалість сесії (3,36), середній коефіцієнт відмов (24,22) та високий коефіцієнт конверсії (59,38). Це вказує на ефективний процес взаємодії з веб-додатком, що сприяє досягненню цілей без зайвих перешкод. Користувачі кластера 1 демонструють високу тривалість сесії (7,64), високий коефіцієнт відмов (66,43) та високий коефіцієнт конверсії (98,34). Така поведінка свідчить про неоднорідність даного сегмента: частина користувачів стикається з труднощами у навігації або прийнятті рішень, тоді як інші здійснюють покупки з високою ймовірністю. Кластер 2 характеризується такими параметрами: коротка тривалість сесії (4,51), високий коефіцієнт відмов (79,28) та низький коефіцієнт конверсії (28,26). Незначний рівень залучення користувачів обумовлює короткочасність взаємодії з веб-ресурсом без здійснення цільових дій. Кластер 3: довга тривалість сесії (32,56), низький коефіцієнт відмов (14,15) та найвищий коефіцієнт конверсії (99,19). Такі характеристики вказують на високу зацікавленість та активну взаємодію із контентом, що призводить до максимальної ефективності конверсійних процесів. Таким чином, можна спостерігати чіткий вплив дизайну веб-додатків на залученість цільової аудиторії. Використання кластеризації сприяє обґрунтованому прийняттю рішень щодо вдосконалення вебдодатків та підвищення їх ефективності.

Висновки

Виокремлено три основні фактори, що характеризують якість проєктування та дизайну вебдодатків електронної комерції: тривалість сеансу, коефіцієнт відмов і коефіцієнт конверсії. Для зручності подальшого опрацювання проведено нормалізацію даних у діапазон $[0, 1]$. Використано методи машинного навчання для кластеризації цільової аудиторії вебдодатків. За методом «ліктя» визначено оптимальну кількість класів. Для кластеризації використано метод k -середніх. Проаналізувавши результати кластеризації, можна зробити висновок, що:

- для кластера 0 важливо підтримувати ефективність процесу залучення та аналізувати впливи на швидкість конверсії;
- для кластера 1 необхідно покращити навігацію та надати користувачам додаткові стимули;
- для кластера 2 варто аналізувати причини швидких відмов і вдосконалювати користувацький досвід;
- для кластера 3 важливо підтримувати наявний рівень взаємодії та залучати користувачів до повторних візитів.

Результати дослідження мають важливе значення для оптимізації роботи вебдодатків у сфері електронної комерції та цифрового маркетингу. Вони дозволяють виявити закономірності в поведінці користувачів, що сприяє персоналізації контенту, покращенню користувацького досвіду та підвищенню ефективності комерційних процесів. Запропонована кластеризація дає змогу розробляти таргетовані маркетингові стратегії для кожної групи відвідувачів. Зокрема покращувати взаємодію з користувачами, які мають високий коефіцієнт відмов, а також стимулювати конверсійні дії для сегментів із середнім рівнем залученості. Отримані результати можуть бути використані для вдосконалення інтерфейсу веб-додатків, підвищення релевантності рекламних кампаній та впровадження персоналізованих пропозицій.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] K. Sharifani, and M. Amini, "Machine learning and deep learning: A review of methods and applications," *World Information Technology and Engineering Journal*, vol. 10, no. 07, pp. 3897-3904, 2023.
- [2] S. M. Javidan, A. Banakar, K. A. Vakilian, and Y. Ampatzidis, "Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning," *Smart agricultural technology*, vol. 3, pp. 100081, 2023. <https://doi.org/10.1016/j.atech.2022.100081>.
- [3] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data," *Quality & quantity*, vol. 56, no. 3, pp. 1283-1291. 2022. <https://doi.org/10.1007/s11135-021-01176-w>.
- [4] R. Vankayalapati, K. B. Ghutugade, R. Vannapuram, and B. P. S. Prasanna, "K-Means algorithm for clustering of learners performance levels using machine learning techniques," *Rev. d'Intelligence Artif.*, vol. 35, no. 1, pp. 99-104, 2021. <https://doi.org/10.18280/ria.350112>.
- [5] B. D. Puspasari, L. L. Damayanti, A. Pramono, and A. K. Darmawan, "Implementation K-means clustering method in job recommendation system," in *2021 IEEE 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, Malang, Indonesia, 2021, pp. 1-6. <https://doi.org/10.1109/ICEEIE52663.2021.9616654>.
- [6] P. Chaipornkaew, and T. Banditwattanawong, "A recommendation model based on user behaviors on commercial websites using TF-IDF, KMeans, and Apriori algorithms," in *International Conference on Computing and Information Technology*. Cham: Springer International Publishing, 2021, pp. 55-65. https://doi.org/10.1007/978-3-030-79757-7_6.

- [7] H. Singh, and Kaur, P, “An effective clustering-based web page recommendation framework for e-commerce websites,” *SN Computer Science*, vol. 2, no. 4, pp. 339, 2021. <https://doi.org/10.1007/s42979-021-00736-z> .
- [8] S. SadighZadeh, and M. Kaedi, “Modeling user preferences in online stores based on user mouse behavior on page elements,” *Journal of Systems and Information Technology*, vol. 24, no. 2, pp. 112-130. 2022. <https://doi.org/10.1108/JSIT-12-2019-0264> .
- [9] Behavior data for engagement & conversion metrics. [Electronic resource]. Available: <https://www.kaggle.com/datasets/ashishsharma150102/behavior-data-for-engagement-and-conversion-metrics> .
- [10] Т. Желдак, і О. Владико, «Використання машинного навчання для кластеризації з використанням індексу небезпеки хвостосховищ та їх впливу на навколишнє середовище», *Information Technology: Computer Science, Software Engineering and Cyber Security*, № 4, с. 81-91. 2024. <https://doi.org/10.32782/IT/2024-4-11> .

Рекомендована кафедрою комп'ютерних наук ВНТУ

Стаття надійшла до редакції 25.03.2025

Піх Ірина Всеволодівна — д-р. техн. наук, професор, професор кафедри систем віртуальної реальності, e-mail: iryua.v.pikh@lpnu.ua .

Національний університет «Львівська політехніка», Львів;

Меренич Юліан Юліанович — аспірант кафедри систем віртуальної реальності Національного університету «Львівська політехніка»; асистент кафедри інформатики та фізико-математичних дисциплін Державного вищого навчального закладу «Ужгородський національний університет», e-mail: merenich.julian@uzhnu.edu.ua

I. V. Pikh¹
Yu. Yu. Merenych^{1,2}

Application of Machine Learning for Target Audience Clustering in Web Applications

¹Lviv Polytechnic National University;

²Uzhhorod National University

In this study, a machine learning method was applied for clustering data on the target audience of e-commerce web applications. Machine learning is a powerful analytical tool that enables the automatic identification of patterns in large datasets, improving the accuracy of user behavior prediction. Key interaction metrics with web applications were selected, including bounce rate, session duration, and conversion rate. The input data were normalized. To ensure proper normalization and the correct operation of machine learning algorithms, a method was used to scale values within the range from zero to one. The optimal number of clusters was determined using the "elbow" method, which analyzes the relationship between the number of clusters and the within-cluster sum of squared distances. The k-means method was applied to analyze behavioral parameters, minimizing the sum of squared distances between data points and cluster centroids using the Euclidean metric. The results were visualized using a three-dimensional plot, representing the distribution of clusters based on the analyzed parameters.

The clustering results identified four groups of users with different interaction characteristics with the web resource. Users in the first cluster exhibited low engagement, short session durations, and high bounce rates, indicating insufficient content relevance. The second cluster demonstrated prolonged interaction with the web resource, but the high bounce rate may suggest navigation difficulties. The third cluster was characterized by a high conversion rate with moderate session duration, indicating an efficient user experience. The last cluster had the lowest bounce rate and the highest conversion rate, reflecting a strong alignment between content and user needs.

The practical significance of the obtained results lies in the possibility of applying clustering methods to adapt UX/UI solutions, optimize content, and enhance conversion rates. The proposed approach can be utilized in e-commerce, digital marketing, and web analytics to improve user interaction strategies.

Keywords: machine learning, clustering, web application, target audience, e-commerce.

Pikh Iryna V. — Dr. Sc. (Eng.), Professor, Professor of the Chair of Virtual Reality Systems, e-mail: iryua.v.pikh@lpnu.ua ;

Merenych Yulian Yu. — Post-Graduate Student of the Chair of Virtual Reality Systems, Lviv Polytechnic National University; Assistant of the Chair of Informatics and Physical and Mathematical Disciplines, Uzhhorod National University, e-mail: merenich.julian@uzhnu.edu.ua