

DOI: <https://doi.org/10.32782/2308-1988/2023-47-9>

УДК 338.242

Жуковський Дмитро Миколайович

аспірант кафедри економічної інформатики,
Український державний університет науки і технологій
ORCID: <https://orcid.org/0009-0006-1105-7434>

Лозовська Людмила Іванівна

кандидат фізико-математичних наук,
доцент кафедри економічної інформатики,
Український державний університет науки і технологій
ORCID: <https://orcid.org/0000-0003-2119-6703>

Dmytro Zhukovskiy, Ludmila Lozovska

Ukrainian State University of Science and Technologies

**АНАЛІЗ ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ
В АНАЛІТИЦІ ПОКАЗНИКІВ ІНТЕРНЕТ РЕСУРСІВ****ANALYSIS OF THE USE OF MACHINE LEARNING METHODS
IN THE ANALYSIS OF INDICATORS OF INTERNET RESOURCES**

Анотація. У сучасних умовах в результаті цифровізації соціально-економічних явищ все більше підприємств переводять свій бізнес до мережі Інтернет. Сучасні веб-технології дозволяють збирати великі масиви статистичних даних для аналізу ефективності економічної діяльності інтернет-ресурсів. Для прийняття більш правильних управлінських рішень поряд з класичними статистичними методами доцільно використовувати методи машинного навчання. У статті описані базові методи машинного навчання та різні приклади застосування їх для розв'язання задач у сфері аналітики веб-ресурсів. Проаналізовано проблеми, пов'язані з недостатньою ефективністю класичних статистичних методів для прийняття оптимальних управлінських рішень. Досліджено різні напрями цифрової економіки, де можуть застосовуватись методи машинного навчання як альтернатива до класичних статистичних методів. Наведено приклади впровадження методів машинного навчання для підвищення ефективності реалізації різного роду задач у цифровому бізнес середовищі діяльності підприємств. У науковій статті описані такі види задач наукового навчання як: 1) застосування методів навчання з вчителем для прогнозування доходів у проєктах електронної комерції; 2) використання методів навчання без вчителя для сегментації користувачів; 3) впровадження методів машинного навчання для розробки систем рекомендацій; 4) використання алгоритмів штучного інтелекту для прогнозування та виявлення аномалій; 5) інтеграція генетичних алгоритмів для оптимізації онлайн-рекламних кампаній; 6) застосування методу моделювання підвищення для оптимізації витрат на маркетингову комунікацію; 7) впровадження алгоритму багаторукого бандита для оптимізації A/B-тестування; 8) проєктування чат-ботів за допомогою різних типів нейронних мереж для обробки природної мови, таких як багатопаровий перцептрон, згортоква нейронна мережа, рекурсивна нейронна мережа, рекурентна нейронна мережа та короткострокова пам'ять. Доведено доцільність застосування методів штучного інтелекту для розв'язання широкого спектру задач з аналізу показників інтернет ресурсів.

Ключові слова: веб аналітика, методи машинного навчання, цифрова економіка, математична статистика, штучний інтелект, діджиталізація, інтелектуальні системи прийняття рішень.

Summary. In modern conditions, due to the digitization of socio-economic phenomena, more and more businesses are moving their activity to the Internet. Modern web technologies allow collecting large amounts of statistical data for analyzing the effectiveness of economic activities of internet resources. To make more effective management decisions, it is appropriate to use machine learning methods alongside classical statistical methods. This article describes basic machine learning methods and various examples of their application to solve problems in web resource analytics. The issues related to the insufficient effectiveness of classical statistical methods for making optimal management decisions are analyzed. Different directions of the digital economy where machine learning methods can be applied as an alternative to classical statistical methods are investigated. Examples of implementing machine learning methods to enhance the efficiency of various tasks in the digital business environment of enterprises are provided. In the scientific article described cases of applying such kind of task as: 1) application of supervised machine learning methods for revenue forecasting in e-commerce projects; 2) utilization of unsupervised machine learning methods for user segmentation; 3) implementation of machine learning methods for developing

recommendation systems; 4) deployment of artificial intelligence algorithms for prediction and anomaly detection tasks; 5) integration of genetic algorithms for optimizing online advertising campaigns; 6) application of Uplift modeling method to optimize marketing communication expenses; 7) implementation of the multi-armed bandit algorithm for optimizing A/B testing; 8) designing chatbots using various types of neural networks for natural language processing, such as multi-layer perceptron, convolutional neural network, recursive neural network, recurrent neural network, and Long Short-Term Memory. The expediency of applying artificial intelligence methods to solve a wide range of internet resource analysis tasks is demonstrated.

Keywords: web analytics, machine learning methods, digital economy, mathematical statistics, artificial intelligence, digitization, intelligent decision-making systems.

Постановка проблеми. У сучасних умовах в результаті цифровізації соціально-економічних явищ все більше підприємств переводять свій бізнес до мережі Інтернет. Сучасні веб-технології дозволяють збирати великі масиви статистичних даних для аналізу ефективності економічної діяльності інтернет-ресурсів. Для прийняття більш правильних управлінських рішень поряд з класичними статистичними методами доцільно використовувати методи машинного навчання.

Однією з найбільших проблем при роботі з великими обсягами даних у цифровій економіці є недостатня ефективність класичних статистичних методів для пошуку найбільш раціональних управлінських рішень. Тому в статті буде проведено огляд перспектив впровадження методів машинного навчання в проектах цифрової економіки та електронної комерції.

Аналіз останніх досліджень і публікацій. Для дослідження питань застосування методів машинного навчання у сфері електронної комерції було розглянуто та проаналізовано роботи таких учених як Andrew Ng, LeCun Y., Hinton G., Poutanen R., Roy S., Sharma M. та інші. Також було опрацьовано технічну документацію провідних компаній у сфері застосування штучного інтелекту у digital економіці, наприклад Google, Facebook.

Мета статті. Основною метою наукової роботи є пошук можливостей застосування математичних методів машинного навчання та їх можливих сфер застосування у цифровій економіці та проектах big data.

Виклад основного матеріалу дослідження. Наведемо найбільш популярні методи машинного навчання, які використовуються на практиці:

1) Навчання з вчителем (Supervised learning) – цей метод машинного навчання полягає в тому, що алгоритм навчається на підставі комплексу прикладів і як результат побудована модель може класифікувати невідомі їй до того приклади.

Метод навчання з вчителем може використовуватись, наприклад, для задач прогнозування доходів підприємства на проектах з електронної комерції [1].

Також застосування методів машинного навчання з вчителем можливе для задач виявлення неправдивих відгуків на веб ресурсах [2, с. 317].

2) Навчання без вчителя (Unsupervised learning) – на відміну від попередньо описаного

методу, навчання в моделі проходить без набору прикладів, тобто без вчителя. В цьому випадку метод використовується для пошуку закономірностей та кореляцій в досліджуваній вибірці даних.

Як правило у e-commerce проектах метод навчання без вчителя застосовується для сегментації або кластеризації користувачів, товарів або відгуків користувачів.

Так, наприклад, unsupervised learning методи використовуються при RFM-аналізі для розбиття на кластери користувачів. При реалізації алгоритму RFM-аналізу найбільш часто використовується метод машинного навчання k-середніх [3].

До того математичні моделі для навчання без вчителя є ефективними для розв'язання задач при аналізі відгуків користувачів в інтернеті [4].

3) Навчання з частковим залученням вчителя (Semi supervised learning). Використання даної математичної моделі передбачає розмітку невеликої кількості даних у нерозміченому масиві з великою кількістю спостережень. Одним із напрямів використання є автоматична розмітка графічного контенту з присвоюванням алгоритмом зображенням певних міток [5].

На практиці також відомо багато прикладів вдалого застосування semi supervised learning методу при розробці рекомендаційних систем на інтернет ресурсах, наприклад, на веб сайті з продажу відео контенту [6].

4) Навчання з підкріпленням (Reinforcement learning) є однією з трьох парадигм машинного навчання поряд з навчанням з вчителем та навчанням без вчителя. Метод Reinforcement learning є окремим випадком навчання з учителем, але для навчання використовується віртуальне або реальне середовище. Машина або агент не має попередньої інформації про середовище, але вона може здійснювати в середовищі певні дії. При цих діях в середовищі, агент, на підставі аналізу отриманих даних з середовища, вчиться приймати правильні рішення.

Серед методів, що відносяться до парадигми навчання з підкріпленням, більше за всіх на практиці застосовується метод ланцюгів Маркова, який може вдало застосовуватись для оптимізації онлайн реклами [7].

5) Алгоритми машинного навчання для виявлення аномалій відхилень статистичних показників. При дослідженнях перед експертами з

веб-аналітики та науковцями дуже часто постає задача щодо пошуку аномалій даних. Виявлення аномалій – це вивчення конкретного статистичного показника та виявлення рідких випадків, які можуть вважатися підозрілими, оскільки сильно відрізняються від встановленої математичної моделі поведінки статистичного показника. Серед найбільш відомих методів, які в тому числі використовуються на практиці в продуктах веб-аналітики компанії Google, є байєсовська модель просторово-часового ряду (Bayesian state-space time series model) та метод головних компонент (Principal component analysis). Байєсовська модель використовується при пошуку аномалій часових рядів, а метод головних компонент при пошуку аномалій в сегментах даних [8].

6) Генетичний алгоритм – це клас еволюційних алгоритмів пошуку. Ідея генетичних алгоритмів ґрунтується на еволюційній теорії Чарльза Дарвіна. Цей алгоритм симулює процес природного відбору, коли сильніші особини з популяції переживають слабших і виробляють наступне покоління особин.

В проектах з електронної комерції генетичні алгоритми більш за всього застосовуються при оптимізації онлайн реклами [9, с. 39].

7) Метод Uplift-моделювання (Uplift modelling). Uplift-моделювання містить набір математичних методів, які дозволяють передбачити зміни в поведінці певного сегменту цільової аудиторії в результаті якого-небудь цільового впливу, наприклад, у відео маркетингової комунікації. В індустрії цифрового маркетингу є безліч способів маркетингового впливу на споживачів, і Uplift-моделювання, що використовується для відбору клієнтів, через комунікацію з якими цільовий маркетинг виглядає найбільш ефективним. Uplift-моделювання допомагає компаніям оптимізувати маркетинговий бюджет [10, с. 645].

8) Алгоритм багаторукого бандита (multi-armed bandit algorithms) відноситься до одного із видів методів машинного навчання. Алгоритм отримав назву через аналогію з азартними іграми, де гравець стоїть перед рядом слот-машин (одноруких бандитів) і вибирає на якого з них зробити ставку, щоб максимізувати свій виграш.

Найбільш за все Multi-Armed Bandit Algorithms застосовується для оптимізації А/В тестів. А/В тестування (split testing) – це метод маркетингових досліджень, в якому контрольна група елементів порівнюється з набором тестових груп з тим, щоб з'ясувати які зі змін покращують показники ефективності. В digital маркетингу це може бути, наприклад, порівняння різних рекламних оголошень або веб сторінок з певним товаром.

Суть задачі оптимізації А/В тестів полягає в тому, щоб визначити оптимальний варіант тесту мінімізуючи при цьому кількість експериментів та грошові витрати на ці експерименти.

На відміну від традиційних А/В тестів, де для визначення результатів тестування необхідно зібрати статистично значимий набір даних, в методі «багаторукого бандита» трафік розподіляється таким чином, щоб той варіант, який має більше шансів бути кращим, вже під час тестувань отримав більше трафіку. Таким чином, при використанні алгоритму multi-armed bandit можна визначити найкращий варіант тесту з проведенням на порядок меншої кількості експериментів ніж при застосуванні класичних методів математичної статистики [11].

Існує кілька видів алгоритмів багаторукого бандита, таких як ϵ -Greedy, UCB (Upper Confidence Bound), Thompson Sampling та інші, які використовують різні стратегії для вибору варіанту, який слід тестувати.

Загальний механізм дії алгоритмів багаторукого бандита допомагає ефективно використовувати трафік та мінімізувати витрати часу, прискорюючи процес виявлення оптимального варіанту та збільшуючи ефективність А/В-тестування.

Щодо застосування на практиці, то алгоритм багаторукого бандита є революційним рішенням для оптимізації онлайн веб ресурсів [12, с. 37].

При розв'язанні задач інтернет маркетингу метод multi-armed bandit algorithms застосовується для проведення А/В тестування рекламних оголошень [14, с. 500], оптимізації онлайн рекламних компаній [13].

Широке визнання алгоритмів багаторукого бандиту серед експертів-практиків доводить факт того, що цей метод дуже часто застосовується в онлайн-платформах для проведення А/В тестів [15].

9) Алгоритми нейронних мереж для чат ботів. Одним з основних трендів застосування штучного інтелекту у сфері інтернет маркетингу є розробка такого класу програмного забезпечення як чат-боти, які використовуються для ведення діалогів з користувачами веб-сайтів.

Для розробки архітектури діалогових систем використовуються такі методи машинного навчання, як нейронні мережі для обробки природної мови – NLP. Тоді далі розглянемо найбільш відомі види нейронних мереж, які застосовуються як математичний апарат для створення чат-ботів.

1) Багатошаровий перцептрон складається з трьох або більше шарів. Він використовує нелінійну функцію активації, часто тангенціальну або логістичну, яка дозволяє класифікувати лінійно нероздільні дані. Кожен вузол у шарі з'єднаний з кожним вузлом у наступному шарі, що робить мережу повністю пов'язаною. Така архітектура знаходить застосування у завданнях розпізнавання мови та машинному перекладі.

2) Згортоква нейронна мережа (Convolutional neural network, CNN) містить один або більше об'єднаних або з'єднаних згорткових шарів.

Convolutional neural network використовує варіацію багатшарового перцептронну, розглянутого вище. Згорткові шари використовують операцію згортки для вхідних даних і передають результат наступного шару. Ця операція дозволяє мережі бути глибше з меншою кількістю параметрів. Згорткові мережі показують гарні результати у додатках для класифікації картинок та мови [16, с. 99].

3) Рекурсивна нейронна мережа – тип глибокої нейронної мережі, сформований при застосуванні тих самих наборів терезів рекурсивно над структурою, щоб зробити скалярне чи структуроване передбачення над вхідний структурою змінного розміру через активацію структури в топологічному порядку. У найпростішій архітектурі нелінійність, така як тангенціальна функція активації, і матриця терезів, що розділяється всією мережею, використовуються для об'єднання вузлів батьківські об'єкти.

4) Рекурентна нейронна мережа, на відміну від прямої нейронної мережі, є варіантом рекурсивної нейронної мережі, де зв'язки між нейронами – спрямовані цикли. Останнє означає, що вихідна інформація залежить не тільки від поточного входу, але також станів нейрона на попередньому кроці. Така пам'ять дозволяє користувачам вирішувати завдання Natural Language Processing: розпізнавання рукописного тексту чи мови.

5) Мережа довгої короткострокової пам'яті (Long Short-Term Memory, LSTM) – різновид архітектури рекурентної нейромережі, створена

для більш точного моделювання часових послідовностей та їх довгострокових залежностей, ніж традиційна рекурентна мережа. LSTM-мережа не використовує функцію активації в рекурентних компонентах, збережені значення не модифікуються, а градієнт не прагне зникнути під час тренування. Часто LSTM застосовується у блоках по кілька елементів. Ці блоки складаються з 3 або 4 затворів (наприклад, вхідного, вихідного та гейту забування), які контролюють побудову інформаційного потоку за логістичною функцією.

Висновки. Отже, в статті проведено аналіз найбільш популярних методів машинного навчання, які можуть застосовуватись у веб-аналітиці для прийняття більш якісних управлінських рішень. Тому для вдосконалення якості бізнес-процесів на проєктах з електронної комерції метою подальших досліджень буде апробація алгоритмів машинного навчання для збільшення показника KPI інтернет ресурсів різного типу.

В результаті огляду та аналізу перспектив застосування алгоритмів машинного навчання можна дійти висновків, що алгоритми machine learning є одним з найбільш перспективних напрямів застосування штучного інтелекту при обробці великих масивів даних на проєктах цифрової економіки. В результаті огляду та проведення досліджень був виявлений позитивний тренд значного зростання кількості наукових праць щодо вдалого застосування математичних методів машинного навчання у різних сферах digital-економіки.

Список використаних джерел:

1. Gomez-Cravioto, D.A., Diaz-Ramos, R.E., Hernandez-Gress, N. et al. (2022) Supervised machine learning predictive analytics for alumni income. *J Big Data*, 9, 11. DOI: <https://doi.org/10.1186/s40537-022-00559-6>
2. Y. Li, X. Feng and S. Zhang (2016) Detecting fake reviews utilizing semantic and emotion model, in 2016 3rd Int. Conf. on Information Science and Control Engineering, Beijing, China, pp. 317–320.
3. R. Ait Daoud, A. Amine, B. Bouikhalene and R. Lbibb (2015) Combining RFM model and clustering techniques for customer value analysis of a company selling online, 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA), Marrakech, Morocco, pp. 1-6. DOI: <https://doi.org/10.1109/AICCSA.2015.7507238>
4. Jai Prakash Verma, Atul Patel (2017) Evaluation of Unsupervised Learning based Extractive Text Summarization Technique for Large Scale Review and Feedback Data. *Indian Journal of Science and Technology*, vol. 10(17). DOI: <https://doi.org/10.17485/ijst/2017/v10i17/106493>
5. Semi-Supervised Machine Learning Algorithms [Semi-Supervised Machine Learning Algorithms]. Available at: <https://hackernoon.com/semi-supervised-machine#learning-algorithms-fnm32cw> (accessed 04.12.2022).
6. S. Roy, M. Sharma and S. K. Singh (2019) Movie Recommendation System Using Semi-Supervised Learning *Global Conference for Advancement in Technology (GCAT)*. Bangalore, India, pp. 1–5. DOI: <https://doi.org/10.1109/GCAT47503.2019.8978353>
7. Riku Poutanen, (2020) Analysis of online advertisement performance using Markov chains Master's thesis Tampere University Industrial Engineering and Management.
8. GA4 Anomaly detection [GA4 Anomaly detection]. Available at: <https://support.google.com/analytics/answer/9517187?hl=en> (accessed 04 July 2023).
9. Luis Miralles-Pechuán, Hiram Ponce, Lourdes Martínez-Villaseñor (2018) A novel methodology for optimizing display advertising campaigns using genetic algorithms, *Electronic Commerce Research and Applications*, vol. 27, pp. 39–51. ISSN 1567-4223. DOI: <https://doi.org/10.1016/j.elerap.2017.11.004>
10. Baier, D., Stöcker, B. (2022) Profit uplift modeling for direct marketing campaigns: approaches and applications for online shops. *J Bus Econ.*, 92, 645–673. DOI: <https://doi.org/10.1007/s11573-021-01068-3>
11. Xiang, Ding, et al. (2022) Multi Armed Bandit vs. A/B Tests in E-commerce—Confidence Interval and Hypothesis Test Power Perspectives. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
12. Scott, Steven L. (2015) Multi-armed bandit experiments in the online service economy. *Applied Stochastic Models in Business and Industry*, 31.1: 37–45.

13. Lu, Tyler, Dávid Pál, and Martin Pál. (2009) Showing relevant ads via context multi-armed bandits. *Proceedings of AISTATS*.

14. Schwartz, Eric M., Eric T. Bradlow, and Peter S. Fader. (2017) Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, 36.4: 500–522.

15. Multi-Armed Bandit (MAB) – A/B Testing Sans Regret [Multi-Armed Bandit (MAB) – A/B Testing Sans Regret]. Available at: <https://vwo.com/blog/multi-armed-bandit-algorithm/> (accessed 10 October 2023).

16. Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou (2022) A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects, in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 99–119. DOI: <https://doi.org/10.1109/TNNLS.2021.3084827>

Стаття надійшла до редакції 21.11.2023