

МОРОЗ ДМИТРО

Національний університет "Львівська політехніка"

<https://orcid.org/0009-0001-2971-1493>e-mail: dmytro.a.moroz@lpnu.ua**ЮЗЕФОВИЧ РОМАН**

Національний університет "Львівська політехніка",

Фізико-механічний інститут ім. Г.В. Карпенка НАН України

<https://orcid.org/0000-0001-5546-453X>e-mail: roman.yuzefovych@gmail.com**ОДРЕХІВСЬКИЙ МИКОЛА**

Національний університет "Львівська політехніка"

<https://orcid.org/0000-0003-3165-4384>e-mail: mykola.v.odrekhivskiy@lpnu.ua

ЗАСТОСУВАННЯ МАТЕМАТИЧНИХ МЕТОДІВ У ВЕБАНАЛІЗУВАННІ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ

Сучасна вебаналітика є важливим інструментом для дослідження поведінки користувачів, оцінювання ефективності цифрових платформ і підтримки прийняття управлінських рішень у соціально-економічних системах. Зі зростанням обсягів даних і ускладненням цифрової взаємодії традиційні підходи аналізу виявляються недостатніми, що зумовлює необхідність використання математичних методів для моделювання, прогнозування та оптимізації вебаналітики.

У статті розглянуто широкий спектр математичних методів, що застосовуються у вебаналізуванні, зокрема кореляційний і регресійний аналіз, методи часових рядів, байєсові мережі, марковські ланцюги, кластеризацію та алгоритми машинного навчання. Здійснено порівняльний аналіз ефективності різних методів, визначено їхні переваги та обмеження. Окрему увагу приділено використанню графових моделей для аналізування соціальних зв'язків, прогнозування поведінки користувачів та персоналізації контенту.

Також у статті висвітлено основні виклики, пов'язані з вебаналізуванням, зокрема проблему обчислювальної складності, необхідність ефективної обробки великих обсягів даних, а також питання конфіденційності та етичних аспектів збору й аналізу інформації. Окремо розглянуто вплив міжнародних нормативних актів, таких як GDPR і CCPA, на використання вебаналітики, а також обмеження, з якими стикаються компанії під час збирання та оброблення персональних даних. Отримані результати демонструють значний потенціал інтеграції математичних методів для підвищення ефективності вебаналітики та точності прогнозування соціально-економічних процесів. Висновки підкреслюють необхідність подальших досліджень щодо поєднання класичних статистичних підходів із сучасними методами машинного навчання та штучного інтелекту для розвитку інтелектуальних вебаналітичних систем. Окреслено перспективи вдосконалення методів вебаналізування з урахуванням технологічних змін, що відбуваються в галузі.

Ключові слова – вебаналізування, математичні методи, соціально-економічні системи, марковські ланцюги, байєсові методи, графові моделі.

MOROZ DMYTRO

Lviv Polytechnic National University

YUZEFOVYCH ROMAN

Lviv Polytechnic National University,

Karpenko Physico-mechanical Institute of National Academy of Sciences of Ukraine

ODREKHIVSKYI MYKOLA

Lviv Polytechnic National University

APPLICATION OF MATHEMATICAL METHODS IN WEBANALYSIS OF SOCIO-ECONOMIC SYSTEMS

Modern webanalytics is an important tool for studying user behavior, evaluating the effectiveness of digital platforms, and supporting management decision-making in socioeconomic systems. With the growth of data volumes and the complexity of digital interaction, traditional analysis approaches are proving insufficient, which necessitates the use of mathematical methods for modeling, forecasting, and optimizing webanalytics.

The article examines various mathematical methods in webanalytics, including correlation and regression analysis, time series methods, Bayesian networks, Markov chains, clustering, and machine learning algorithms. A comparative analysis of the effectiveness of various methods is carried out, and their advantages and limitations are identified. Particular attention is paid to the use of graph models for analyzing social connections, predicting user behavior, and personalizing content.

The article also highlights the main challenges associated with webanalytics, including the problem of computational complexity, the need for efficient processing of large amounts of data, as well as privacy and ethical aspects of information collection and analysis. The impact of international regulations, such as GDPR and CCPA, on the use of webanalytics, as well as the limitations faced by companies in collecting and processing personal data, are also considered. The findings demonstrate the significant potential of integrating mathematical methods to improve the efficiency of webanalytics and the accuracy of forecasting socio-economic processes. The conclusions emphasize the need for further research on combining classical statistical approaches with modern methods of machine learning and artificial intelligence for the development of intelligent webanalytics systems. The article outlines the prospects for improving webanalysis methods about technological changes taking place in the industry.

Keywords - webanalytics, mathematical methods, socio-economic systems, Markov chains, Bayesian methods, graph models.

Стаття надійшла до редакції / Received 11.04.2025

Прийнята до друку / Accepted 26.04.2025

Постановка проблеми

Зростання обсягу даних та ускладнення цифрової поведінки користувачів зумовлює необхідність застосування комплексних аналітичних методів для точного прогнозування дій, визначення основних факторів, що впливають на ефективність вебресурсів, та оптимізації взаємодії з користувачами. Одним із ключових завдань є визначення ефективних математичних моделей, що дозволяють отримати достовірні результати та забезпечити релевантні висновки. Однак більшість підходів мають обмеження, пов'язані з припущеннями щодо розподілу даних, стійкості до шуму та адаптивності до змінних умов. Особливо актуальним є питання інтегрування традиційних методів (кореляційний [1] та регресійний [2] аналіз, ARIMA [3], A/B тестування [4]) із сучасними підходами, такими як графові моделі, байєсові методи та алгоритми глибокого навчання, для підвищення ефективності вебаналітики.

Аналіз останніх досліджень та публікацій

Питанням вебаналізування присвячена велика кількість робіт. У роботах зарубіжних та українських вчених значна увага приділяється аналізуванню інструментів вебаналітики, таким, наприклад, як Google Analytics, їхнім порівняльним можливостям [5–8].

У роботі [9] розглядається застосування вебаналітики для підвищення ефективності вебсайтів і цифрових платформ. Запропоновано використання методів машинного навчання, марковських процесів, байєсових підходів та сучасних агентів штучного інтелекту.

У дослідженні [10] пропонується використання вебаналітики для оцінювання ефективності передплатної діяльності наукової бібліотеки. Отримані результати підтверджують доцільність такого підходу та демонструють його переваги для університетських бібліотек.

У праці [11] було продемонстровано використання вебаналізування для оцінювання ефективності онлайн-карт. Для розв'язання дослідницьких завдань використовуються методи регресії, а для моделювання взаємозв'язків даних – методи просторового аналізу.

Розвиток нейронних мереж призвів до появи досліджень щодо їхнього використання у системах вебаналітики. У роботі [12] досліджуються методи машинного навчання, які можна застосовувати для вебаналізування, наводяться архітектури нейронних мереж, які можна використовувати для рішення різних задач вебаналізування.

З появою великих мовних моделей (LLM) розпочалось їх використання для розв'язання задач вебаналізування. У [13] запропоновано систему, що використовує моделі GPT (Generative Pre-trained Transformer) для вилучення структурованих ознак із неструктурованих заголовків і описів новин. Запропонований метод використовує здатність GPT виявляти семантичні нюанси в тексті, забезпечуючи точну класифікацію новин на основі контексту вмісту.

У роботах [14, 15] розглядається методика використання Google Analytics для вебаналізування наукових установ. У роботі [16] проводиться дослідження вдосконалення сайту туристичної фірми засобами вебаналізування.

У [17] аналізуються математичні методи, що використовуються у вебаналітиці для оцінювання вартості цифрових медіа-ресурсів. Пропонується система економічних показників та алгоритми для визначення рентабельності медіа-ресурсів. Авторами також пропонується алгоритм “однорукого бандита”, який забезпечує вищу точність та ефективність методу A/B тестування. Автори також наводять архітектури різних нейронних мереж організації чат-ботів для комунікації та збору даних.

У роботі [18] пропонується метод аналізу вебграфа на основі алгоритмів кластеризації. Такий підхід дозволяє підвищити якість дослідження вебсторінок та вибудувати їхню ієрархію.

Загалом, проблема вдосконалення та вибору релевантного методу вебаналізування для конкретної прикладної задачі залишається актуальною.

Цілі статті

Метою статті є систематизація та порівняння сучасних методів аналізу у вебсередовищі, визначення їхніх переваг та обмежень, а також обґрунтування можливостей інтеграції різних математичних підходів для покращення точності прогнозування та оптимізації вебаналітики.

Виклад основного матеріалу

Сучасні соціально-економічні системи активно використовують мережу Інтернет для збору, оброблення та аналізування даних, що дозволяє приймати обґрунтовані управлінські рішення. Соціальні мережі та онлайн-платформи впливають на суспільні процеси, формуючи громадську думку, підтримуючи комунікацію та сприяючи розвитку цифрових спільнот. Вони стали одним з основних засобів комунікації компаній зі своїми клієнтами, формування аудиторії прихильників тощо. Також використання технологій великих даних та алгоритмів персоналізації в онлайн-сервісах допомагає бізнесу краще розуміти споживачів і пропонувати їм релевантні продукти та послуги.

Все це формує сучасні інструменти управління компаніями, орієнтовані на аналіз даних, які можна отримати за допомогою різних технічних засобів інформаційно-телекомунікаційних технологій.

Одним з таких інструментів є методи та засоби вебаналітики, які здійснюють відбір, збір, аналіз і звітування Інтернет-даних з метою розуміння та оптимізування використання Інтернету [19]. Вебаналізування відноситься до оцінювання поведінки користувача на вебсайті. За допомогою вебаналізування можна визначити джерело відвідувачів, які частини вебсайту переглядають відвідувачі та як довго відвідувачі залишаються на окремих розділах вебсайту.

Вебаналізування допомагає отримати детальну інформацію, яка дозволяє:

- визначити джерела трафіку, які приводять відвідувачів;
- аналізувати поведінку відвідувачів та їхню реакцію на певні елементи сайту;
- визначити найпопулярніші сторінки та ті, які взагалі не відвідуються;
- зрозуміти, як користувачі переміщуються між сторінками вебсайту, та що спонукає їх виконувати цілеспрямовану дію.

Вебаналізування – це процес, який пропонує ефективний цикл для оптимізації вебсайту. Він включає кроки, що забезпечують покращення користувацького досвіду, підвищення конверсії та ефективності вебресурсу (рис. 1) [15]:

- постановка бізнес-цілей, тобто визначення ключових показників, які задаватимуть успіх організації та вебсайту;
- збирання даних – збір інформації, статистики та даних про відвідувачів вебсайту за допомогою інструментів аналітики;
- оброблення даних, тобто перетворення необроблених даних, які були зібрані, у значущі коефіцієнти, KPI та іншу інформацію, яка визначає стан вебсайту;
- аналізування даних на основі звітів за обробленими даними;
- дослідження станів вебсайту;
- розроблення онлайн-стратегії;
- впровадження оновлень вебсайту на основі отриманих висновків та оцінювання їхнього впливу на ключові показники ефективності.

Отримані результати можна використовувати для оптимізування неефективних сторінок та подальшого просування більш ефективних сторінок на вебсайті.

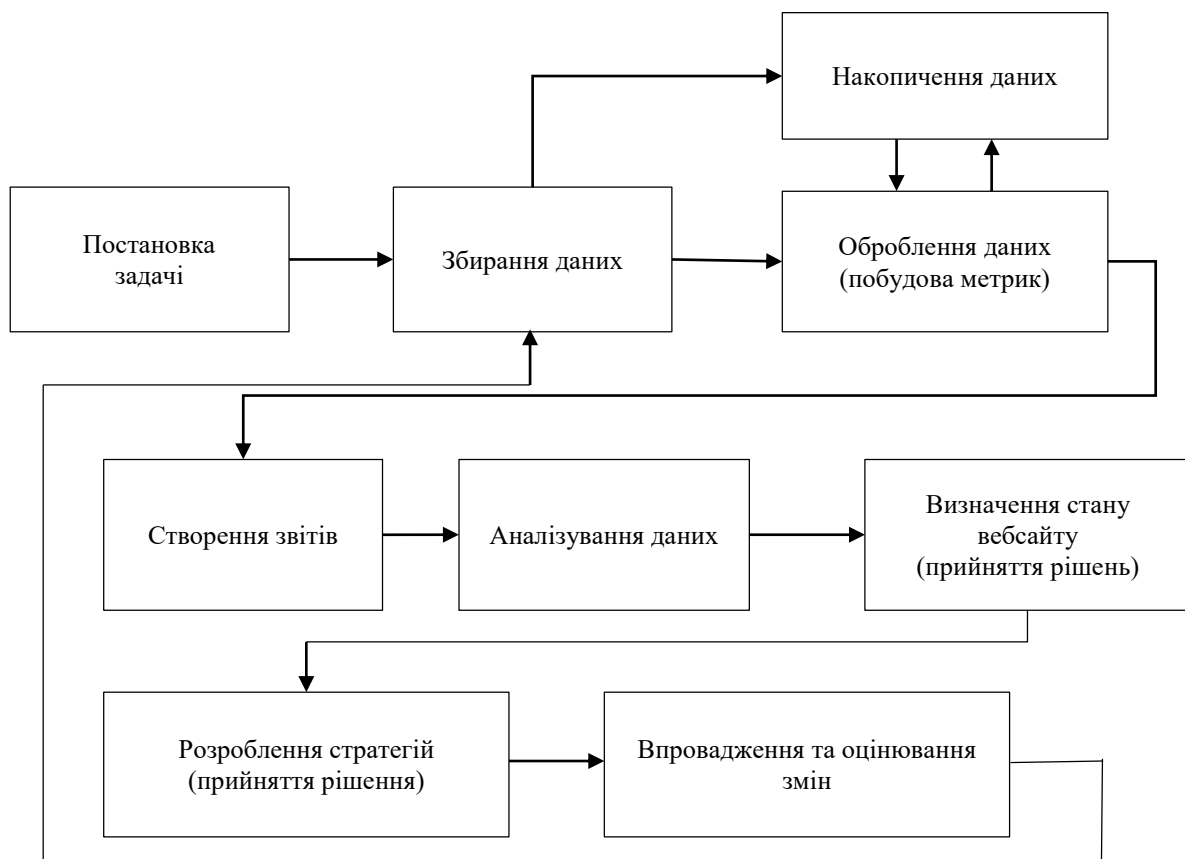


Рис.1. Алгоритм вебаналізування

В основі більшості інструментів вебаналізування лежать математичні методи, які дозволяють отримати кількісні характеристики показників ефективності функціонування вебсайту або виявити певні закономірності щодо поведінки користувачів та їхніх дій [12]. Фундаментальною основою вебаналізування є статистичні методи оброблення даних.

Загалом математичні методи у вебаналізуванні можна класифікувати таким чином:

1. Формальні методи оброблення вебданих:

- Статистичні методи: описова статистика, кореляційний аналіз, регресійний аналіз.

- Методи часових рядів для аналізування економічних тенденцій.
 - Байєсові мережі для ймовірнісного аналізу поведінки користувачів.
2. Машинне навчання:
- Класифікація (логістична регресія; Random Forest – ансамблевий метод, що використовує набір дерев рішень для покращення точності класифікації; SVM – метод, що знаходить гіперплощину, яка найкраще розділяє дані на класи).
 - Підходи, засновані на нейронних мережах (глибоке навчання; LSTM – тип рекурентної нейромережі, що ефективно працює з послідовними даними, такими як часові ряди; трансформери – архітектура, що використовується для роботи з текстами та послідовними даними).
3. Математичні моделі прогнозування:
- Марковські ланцюги для оцінювання та прогнозування станів поведінки користувачів.
 - Графові моделі для аналізування соціальних зв'язків.

На основі описової статистики здійснюється оброблення так званих “сирих” даних та формуються їхні показники та метрики. Їхні усереднені величини використовуються для визначення середнього часу перебування користувачів на сайті. На основі коефіцієнта асиметрії визначаються показники відмов від подальшої взаємодії з вебресурсом (наприклад, частка користувачів, які залишили сайт після першого перегляду сторінки), а на основі коефіцієнта ексцесу – популярність певних сторінок.

Для визначення наявності залежності між кількісними параметрами функціонування вебсайту використовується кореляційний аналіз [1]. Відтак дослідження кореляції між середнім часом перебування користувачів та коефіцієнтом конверсії дозволяє оцінити, чи сприяє довше перебування на сайті зростанню частки відвідувачів, які виконують взаємодію з вебресурсом. Коефіцієнт конверсії (Conversion Rate) – це відсоток користувачів, які здійснили певну цільову дію (наприклад, покупку, реєстрацію чи підписку), від загальної кількості відвідувачів сайту. Висока позитивна кореляція між рівнем відмов та швидкістю завантаження сторінок може свідчити про те, що повільне завантаження контенту сприяє зростанню показника відмов.

Регресійний аналіз дозволяє моделювати залежність між однією залежною змінною Y та однією або кількома незалежними змінними X_i . Такий аналіз дозволяє прогнозувати коефіцієнт конверсії на основі таких факторів, як тривалість сесії, джерело трафіку, кількість переглянутих сторінок та частота взаємодії з контентом. Регресійний аналіз є одним з найбільш поширених інструментів в цифровому маркетингу, який дозволяє оцінити майбутні результати рекламних кампаній.

А/В тестування є потужним методом оцінювання змін у вебаналітиці, що дозволяє приймати обґрунтовані рішення на основі даних. Використання коректних статистичних методів та уникнення типових помилок дозволяє забезпечити високу достовірність результатів. Успішна інтеграція А/В тестування в процес прийняття рішень дозволяє підвищити ефективність маркетингових кампаній, покращити користувацький досвід та збільшити ключові показники ефективності вебресурсів [21].

Основний принцип А/В тестування полягає у випадковому розподілі користувачів на дві (або більше) групи, де одна група взаємодіє з контрольним варіантом (А), а інша – з тестовим (В). Далі статистичні методи дозволяють визначити, чи є різниця між показниками ефективності (конверсії, часу перебування тощо) статистично значущою [5].

Основні етапи А/В тестування:

1. Визначення гіпотези. Передбачає формулювання чіткої гіпотези щодо можливого покращення метрики. Наприклад, "Зміна кольору кнопки заклик до дії з синього на червоний збільшить кількість натискань на 5%". "Оптимізація тексту заголовка на головній сторінці збільшить середній час перебування користувачів".

2. Вибір метрики. Для оцінювання результатів тестування необхідно вибрати відповідну метрику, яка буде показником успішності змін.

3. Розподіл трафіку та експериментальний дизайн. Користувачі випадковим чином розподіляються між варіантами А (контрольний) і В (тестовий), що забезпечує об'єктивність експерименту. Варіанти можуть відрізнятися кольоровим оформленням інтерфейсу, розташуванням елементів, текстовим контентом, швидкістю завантаження тощо.

4. Збирання даних та моніторинг. Залежно від обсягу трафіку та необхідного рівня статистичної значущості тест триває певний період (зазвичай від кількох днів до тижнів). Впродовж цього періоду аналітики контролюють стабільність трафіку та фіксують значення метрик для кожного варіанту.

5. Статистичний аналіз. Після збирання достатньої кількості даних проводиться статистичне аналізування результатів тесту. Використовуються наступні методи:

- Z-тест або t-тест для порівняння середніх значень метрик.
- Критерій Пірсона для аналізу пропорцій (наприклад, частки конверсій).

Для перевірки статистичної значущості використовується p -значення, яке показує ймовірність отримання такої ж або більш екстремальної різниці між групами за умови, що нульова гіпотеза правдива. Якщо $p < 0.05$, то зміни вважаються статистично значущими.

Обмеженнями методу є необхідність забезпечення мінімального обсягу вибірки та мінімальної тривалості тесту. Проведення великої кількості А/В тестів збільшує ймовірність отримання хибних позитивних результатів.

Одним з ефективних математичних апаратів для аналізування послідовності дій користувачів є марковські ланцюги, які дозволяють описувати процеси, де наступний стан системи залежить лише від поточного стану, незалежно від попередніх подій.

Марковські ланцюги застосовуються для розв'язання широкого спектру завдань у вебаналітиці, включаючи аналізування послідовності натискань (clickstream analysis), прогнозування ймовірності відтоку користувачів (churn prediction), побудову персоналізованих рекомендацій та оптимізування навігації вебресурсів.

Марковський ланцюг – це стохастичний процес $\{X_t\}$, що набуває значень з дискретної множини станів $S = \{s_1, s_2, \dots, s_n\}$, де ймовірність переходу між станами визначається лише поточним станом:

$$P(X_{t+1} = s_j | X_t = s_i, X_{t-1}, \dots, X_0) = P(X_{t+1} = s_j | X_t = s_i) \quad (1)$$

Основними елементами марковського ланцюга є простір станів S – набір можливих подій або дій користувачів (наприклад, перегляд сторінки, додавання товару в кошик, оформлення замовлення).

Матриця ймовірностей переходу P – квадратна матриця $n \times n$, де кожен елемент p_{ij} позначає ймовірність переходу зі стану S_i у стан S_j :

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{bmatrix} \quad (2)$$

Ймовірності станів – ймовірність того, що користувач перебуває у певному стані в момент часу t , що визначається як вектор стану $\pi(t)$:

$$\pi(t) = \pi(0)P^t \quad (3)$$

Якщо існує стаціонарний розподіл, то вектор станів збігається до граничного розподілу π^* , який є розв'язком рівняння:

$$\pi^*P = \pi^* \quad (4)$$

Марковські ланцюги використовуються для моделювання послідовностей дій користувачів на вебсайті. При цьому простір станів складається з різних сторінок сайту, а ймовірності переходів визначаються інтенсивністю переходів між ними [25].

Використовуючи марковські ланцюги, можна оцінити ймовірність того, що користувач на певному етапі покине сайт. Граничний розподіл станів дозволяє оцінити ймовірність знаходження користувача у кожному з станів у довгостроковій перспективі. Якщо стан "відтоку" (наприклад, вихід із сайту) має високу ймовірність у граничному розподілі, то необхідно вжити заходів щодо його зменшення (наприклад, покращити зручність інтерфейсу або персоналізувати контент).

Таким чином, методи вебаналізування, з використанням марковських ланцюгів, дозволяють врахувати невизначеність в поведінці користувачів, що дає змогу побудувати більш точні системи вебаналітики.

Сучасні вебаналітичні дослідження потребують ефективних статистичних методів для оцінювання динаміки користувацької поведінки, аналізування ефективності маркетингових кампаній та оптимізування вебінтерфейсів. Одним із перспективних підходів є застосування байєсових методів, які забезпечують адаптивний підхід до оновлення знань про параметри системи на основі нових даних [20].

Байєсова статистика базується на використанні апостеріорного висновку, що оновлюється в міру отримання нової інформації. Це дозволяє отримувати більш точні оцінки параметрів навіть за обмеженого обсягу даних, що є особливо важливим у контексті вебаналітики, де часто необхідно приймати рішення в режимі реального часу.

Основне рівняння байєсової статистики визначається наступним чином:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}, \quad (5)$$

де $P(\theta|D)$ – апостеріорний розподіл параметра θ після отримання даних D ; $P(D|\theta)$ – функція правдоподібності, яка відображає ймовірність спостережуваних даних D за умови параметра θ ; $P(\theta)$ – апріорний розподіл, що містить початкову інформацію про параметр θ ; $P(D)$ – нормувальна константа, яка гарантує, що апостеріорний розподіл є коректним розподілом ймовірностей.

Байєсовий підхід дозволяє гнучко комбінувати попередні знання (апріорний розподіл) з новими даними, що є особливо цінним у динамічних умовах вебаналітики. Таким прикладом може виступити байєсове А/В тестування, яке, на відміну від частотного підходу, дозволяє оновлювати оцінки ефективності змін у реальному часі.

Для байєсового А/В тесту використовується наступна модель:

Конверсії у варіантах А та В розглядаються як біноміальні розподіли:

$$X_A \sim \text{Bin}(n_A, p_A), \quad X_B \sim \text{Bin}(n_B, p_B), \quad (6)$$

де p_A і p_B – ймовірності конверсії для контрольної та тестової групи.

Як апіорний розподіл для p_A і p_B зазвичай використовується бета-розподіл:

$$p_A, p_B \sim \text{Beta}(\alpha, \beta). \quad (7)$$

Це дозволяє легко оновлювати апостеріорні ймовірності за допомогою отриманих даних.

Після збору достатньої кількості даних визначається ймовірність того, що тестовий варіант В є кращим за контрольний А:

$$P(p_A > p_B | D). \quad (8)$$

Якщо ця ймовірність перевищує наперед заданий поріг (наприклад, 0,95), приймається рішення про впровадження змін.

Графові моделі є потужним інструментом для аналізу зв'язків між об'єктами у вебсередовищі. Вебаналітика широко використовує такі моделі для розуміння поведінки користувачів, структурних властивостей вебсайтів, соціальних мереж та аналізу взаємодій. В основі цих моделей лежить концепція графа, де вершини представляють об'єкти (наприклад, користувачів, сесії, вебсторінки), а дуги чи ребра – зв'язки між ними, що можуть відображати переходи, взаємодії або відносини. Графові підходи дозволяють досліджувати структуру вебресурсів, визначати впливові вузли, прогнозувати поведінку користувачів та оптимізувати навігацію.

Одним з найбільш відомих методів є алгоритм PageRank [24] для визначення найбільш впливових вузлів, який дозволяє виявити лідерів думок, оптимізувати маркетингові стратегії та аналізувати динаміку поширення інформації. У випадку онлайн-спільнот та платформ електронної комерції такі графові підходи допомагають краще розуміти структуру зв'язків між користувачами та рекомендувати товари на основі схожих покупок або переглядів. Метод ґрунтується на ідеї, що важливість вебсторінки визначається не лише кількістю посилань на неї, а й якістю цих посилань. Формально PageRank визначається як розв'язок наступного рівняння:

$$PR(A) = (1 - d) + d \sum_{i=1}^n \frac{PR(L_i)}{C(L_i)}, \quad (9)$$

де $PR(A)$ – значення PageRank сторінки А;

d – коефіцієнт демпфування (зазвичай 0,85), що враховує ймовірність того, що користувач перейде на випадкову сторінку замість переходу за посиланням;

L_i – сторінки, які посилаються на А;

$PR(L_i)$ – значення PageRank сторінки L_i ;

$C(L_i)$ – кількість вихідних посилань зі сторінки L_i .

Алгоритм PageRank ітеративно обчислює ваги всіх вузлів графа, поки вони не досягнуть стабільних значень.

Ще одним важливим напрямком застосування графових моделей у вебаналітиці є побудова графів рекомендацій. У таких моделях вершини можуть відповідати користувачам та продуктам, а дуги – взаємодіям між ними, наприклад, переглядам або покупкам. Використання графових алгоритмів, зокрема випадкових процесів з перезапуском (Random Walk with Restart), дозволяє генерувати персоналізовані рекомендації, підвищуючи релевантність контенту чи товарів.

Незважаючи на всі переваги, графові моделі мають певні обмеження, серед яких висока обчислювальна складність при аналізованні великих графів, необхідність очищення даних від шуму та складність у роботі з динамічними змінами в структурі графа. Проте їхня здатність відображати складні взаємозв'язки та адаптуватися до специфіки вебаналітики робить їх незамінним інструментом у дослідженні поведінки користувачів та оптимізованні цифрових платформ.

Розвиток машинного навчання в останні роки призвів до того, що ці методи почали широко впроваджуватися в системи вебаналізування, дозволяючи автоматично виявляти закономірності в поведінці користувачів, прогнозувати тенденції та оптимізувати процеси ухвалення рішень [12].

Класифікація є одним із фундаментальних методів контрольованого навчання, що використовується у вебаналізі для автоматичної категоризації користувачів або контенту. Цей метод базується на побудові моделей, які навчаються на попередньо підготовлених даних, після чого можуть здійснювати прогнозування класів нових об'єктів.

Популярними алгоритмами класифікації є логістична регресія, дерева рішень, градієнтний бустинг (XGBoost, LightGBM) та нейронні мережі. У вебаналітиці класифікаційні моделі застосовуються для виявлення шахрайської активності, прогнозування відтоку користувачів та аналізу рівня задоволення клієнтів на основі текстових відгуків. Наприклад, використання Support Vector Machines (SVM) дозволяє ефективно розділяти групи користувачів залежно від їх поведінки.

Класифікація також відіграє ключову роль у фільтрації спаму, модерації контенту та визначенні тональності текстових матеріалів за допомогою методів Natural Language Processing (NLP).

Глибоке навчання (Deep Learning) є частиною машинного навчання, що використовує багаторівневі штучні нейронні мережі для автоматичного виявлення складних закономірностей у великих обсягах даних. Вебаналізування значно виграє від застосування таких методів. Воно дозволяє вдосконалювати персоналізацію контенту, прогнозувати поведінкові тренди, оптимізувати маркетингові кампанії та виявляти аномалії в користувацьких діях.

Глибокі нейронні мережі (DNN) використовуються для моделювання складних взаємозв'язків між різними характеристиками вебтрафіку. Вони можуть прогнозувати коефіцієнти конверсії, ідентифікувати найбільш релевантні сегменти аудиторії та аналізувати ефективність рекламних кампаній.

Рекурентні нейронні мережі (RNN), зокрема архітектура LSTM (Long Short-Term Memory), застосовуються для аналізу послідовностей дій користувачів на вебсайтах. Це допомагає визначати найбільш типові сценарії взаємодії, прогнозувати відтік користувачів та будувати персоналізовані рекомендаційні системи.

Таким чином, для вебаналізування зараз використовується цілий набір методів від досить простих, які дають можливість описати загальну структуру отриманих показників та їх агреговані значення, до складних методів прогнозування, виявлення прихованих закономірностей в поведінці користувачів, функціонуванні вебсайтів. Порівняння основних математичних методів наведено в таблиці 1.

Таблиця 1

Порівняльний аналіз математичних методів вебаналізування

Назва моделі	Для чого застосовується	Переваги	Обмеження (недоліки)
Марковські моделі	Аналіз послідовності переходів користувачів між сторінками	Легкість інтерпретації, можливість прогнозування поведінки	Ігнорує далекі попередні стани, може вимагати великих даних
Графові моделі (PageRank, Random Walks)	Аналіз структури переходів між сторінками	Добре працює для великих вебсайтів, ефективний пошук важливих вузлів	Не враховує часову динаміку та порядок дій користувача
Кластеризація (k-means, DBSCAN, Hierarchical Clustering)	Групування користувачів за поведінковими патернами	Виявляє схожі групи користувачів, корисний для персоналізації	Вимагає правильного вибору кількості кластерів, може бути чутливий до шуму
LSTM (Long Short-Term Memory)	Прогнозування майбутніх дій користувача на основі історії	Враховує довгі послідовності, добре підходить для часових рядів	Вимагає багато обчислювальних ресурсів і великих обсягів даних
Байєсові методи (Bayesian Inference)	Аналіз ймовірностей різних сценаріїв поведінки	Гнучкість, можливість оновлення оцінок у реальному часі	Може бути складним для реалізації при великій кількості параметрів

Підсумовуючи, можна зауважити, що розглянуті методи дозволяють вирішувати різні задачі вебаналізування. Їх використання залежить від конкретної прикладної задачі. Зазвичай, найкращим варіантом є використання комбінації методів, таким чином підсилюючи позитивні характеристики сукупності методів.

Незважаючи на поширеність різних інструментів вебаналізування, необхідно відмітити, що навіть у країнах із високим рівнем цифрової трансформації їх використання стикається з низкою проблем, що зумовлені як технічними, так і регуляторними факторами.

Однією з ключових проблем є конфіденційність та захист персональних даних. У зв'язку з посиленням міжнародних норм, зокрема Загального регламенту про захист даних (GDPR) у Європейському Союзі [22] та Закону про конфіденційність споживачів Каліфорнії (CCPA) у США [23], компанії змушені переглядати свої підходи до збирання, зберігання та оброблення інформації про користувачів. Це призводить до ускладнення процесів збирання даних та обмеження доступу до важливих аналітичних метрик, що раніше широко використовувалися для бізнес-аналітики.

Наступною значною проблемою є залежність від глобальних аналітичних платформ. Більшість компаній використовує рішення таких технологічних гігантів, як Google (Google Analytics), Adobe (Adobe Analytics) та Microsoft, що створює ризики монополізації ринку. Внаслідок цього малий і середній бізнес змушений адаптуватися до змін у політиці цих компаній, зокрема обмежень у відстеженні поведінки користувачів через блокування сторонніх файлів cookie.

Також важливим викликом є оброблення та аналізування великих обсягів даних (Big Data). Сучасні вебаналітичні платформи працюють з величезними потоками інформації, що вимагає високої продуктивності обчислювальних ресурсів та ефективних алгоритмів для інтерпретації результатів. В умовах швидкої цифровізації традиційні методи аналізу виявляються недостатньо ефективними, що зумовлює необхідність впровадження штучного інтелекту та машинного навчання для автоматизації процесів.

Ще однією проблемою є точність та достовірність зібраних даних. У багатьох країнах користувачі активно використовують засоби захисту своєї цифрової ідентичності, зокрема VPN,

блокувальники реклами та функції браузерів для обмеження збору даних. Це призводить до спотворення аналітичної інформації та ускладнює побудову ефективних стратегій на основі отриманих результатів.

Останнім, але не менш важливим викликом є етичні аспекти використання вебаналітики. В умовах зростаючого суспільного інтересу до цифрової етики компанії змушені шукати баланс між ефективністю вебаналітики та дотриманням прав користувачів. Використання алгоритмів персоналізації, профілювання та поведінкового таргетингу викликає дискусії щодо меж допустимості аналізування даних, що регулюється національними і міжнародними правовими актами.

Таким чином, вирішення проблем вебаналітики на міжнародному рівні потребує комплексного підходу, що включає вдосконалення технологічних методів оброблення даних, розроблення нових регуляторних механізмів та впровадження етичних стандартів у сфері цифрової аналітики.

Для України зазначені проблеми мають свої особливості, що обмежують ефективне застосування вебаналізування.

Однією з основних проблем є низький рівень цифрової грамотності на підприємствах та у державних установах. Багато організацій недостатньо використовують сучасні аналітичні інструменти або обмежуються базовими метриками, що ускладнює глибокий аналіз даних та прийняття обґрунтованих рішень. Це частково пов'язано з відсутністю кваліфікованих фахівців у сфері вебаналітики та недостатнім рівнем інтеграції вебаналітичних платформ у бізнес-процеси.

Другою суттєвою проблемою є захист персональних даних та відповідність процесу їх оброблення міжнародним стандартам. Україна знаходиться в процесі узгодження законодавства з вимогами ЄС, зокрема із Загальним регламентом про захист даних (GDPR). Проте, на практиці, існують труднощі у впровадженні необхідних заходів для дотримання стандартів конфіденційності та прозорості збору інформації. Це створює правові ризики для компаній, що працюють з європейськими клієнтами та партнерами.

Ще однією проблемою є відсутність централізованих державних ініціатив щодо розвитку вебаналітики у державному секторі. Хоча деякі органи влади починають використовувати аналітичні інструменти для управління електронними послугами, загалом рівень використання вебаналітики у сфері державного управління залишається низьким. Це обмежує можливості моніторингу ефективності державних сервісів, оптимізації адміністративних процесів та покращення комунікації з громадянами.

Останньою, але не менш важливою проблемою є обмеженість доступу до великих обсягів даних для аналізу. Хоча в Україні активно розвиваються ініціативи з відкритих даних, багато критично важливої інформації залишається недоступною для аналітичних досліджень. Це ускладнює прогнозування економічних тенденцій, оптимізацію маркетингових стратегій та формування персоналізованих рекомендацій для користувачів.

Таким чином, для ефективного розвитку вебаналітики в Україні необхідно впроваджувати освітні програми для фахівців, адаптувати законодавство до міжнародних стандартів, розвивати державні ініціативи у сфері цифрової аналітики та стимулювати створення національних вебаналітичних рішень.

Висновки

Вебаналізування є потужним інструментом, який забезпечує ефективність прийняття рішень щодо управління різними соціально-економічними системами. На даний час існує досить велика кількість різноманітних інструментів, які можуть забезпечити проведення вебаналізування. Вони використовують в своїй роботі різні методи від простих описових статистик до методів глибокого навчання. Однак, інформація про конкретну реалізацію часто прихована від користувачів таких систем. В той же час, розробка власних систем вебаналізування на основі сучасних методів, може стикатися з різними проблемами: нестача даних, перенавчання, обчислювальна складність, нерелевантність. Оптимальним підходом є використання комбінацій різних методів, що забезпечить синергетичний ефект. Тому розробка таких комбінаційних методів є задачею майбутніх досліджень.

Література

1. Яворський, І. М., Кравець, І. Б., Мацько, І. Й., & Юзефович, Р. М. (2011). Інваріантний кореляційний аналіз векторних періодично корельованих випадкових процесів. Відбір і обробка інформації, 35, 22–31. <https://doi.org/10.20535/S0021347014090039>
2. Юзефович, Р. М. (2017). Моделювання та статистичний аналіз взаємопов'язаних періодично нестационарних вібраційних сигналів для виявлення дефектів механізмів. *Дисертація доктора технічних наук*
<https://lpnu.ua/sites/default/files/2020/dissertation/1433/disertationyuzefovychrm.pdf>
3. Кишакевич, Б., & Нахаева, М. (2018). Прогнозування прямих іноземних інвестицій в Україну за допомогою ARIMA-моделей. *Вісник соціально-економічних досліджень*. https://journals.uran.ua/vsed_oneu/article/view/168781

4. Quin, F., Weyns, D., Galster, M., & Costa Silva, C. (2023). A/B Testing: A Systematic Literature Review. <https://arxiv.org/abs/2308.04929>
5. Jyothi, P., Bonthu, S., & Prasanthi, B. V. (2017). A study on raise of web analytics and its benefits. *International Journal of Computer Sciences and Engineering*, 5(10), 59-64.
6. Пахольчук, Я. (2018). Інструменти веб-аналітики для аналізу відвідувачів сайтів. *Суспільні комунікації*, 1(3), 71-77.
7. Camacho, D., Panizo-Lledot A., Bello-Orgaz G., Gonzalez-Pardo A., & Cambria E. (2020). The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion*, 63, 88-120.
8. Пономаренко, І. В. Веб-аналітика як важливий інструмент цифрового маркетингу. (2018). *Східна Європа: економіка, бізнес та управління*, 4 (15), 231-235.
9. Жуковський, Д., & Лозовська, Л. (2023). Аналіз використання методів машинного навчання в аналітиці показників інтернет-ресурсів. *Сталий розвиток економіки*, 2(47), 65-69. <https://doi.org/10.32782/2308-1988/2023-47-9>
10. Coughlin, D.M., Campbell, M.C., & Jansen, B.J. (3 March 2016). A web analytics approach for appraising electronic resources in academic libraries. *Journal of the Association for Information Science and Technology*, 67, 518-534.
11. Shaheen, A. (2023). Maximizing Website Performance with Google Analytics. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 14(03), 1273–1277. <https://doi.org/10.61841/turcomat.v14i03.14229>
12. Лозовська, Л. (2023). Аналіз використання методів машинного навчання в аналітиці показників інтернет ресурсів. *Сталий розвиток економіки*, 2 (47), 65-69.
13. Sufi, F. (2014). Advances in Mathematical Models for AI-Based News Analytics. *Mathematics*, 12(23), 3736. <https://doi.org/10.3390/math12233736>
14. Кільченко, А. В., Шиненко, М. А., & Яськова, Н. В. (2019). Методика використання системи Google Analytics для розвитку інформаційно-дослідницької компетентності наукових і науково-педагогічних працівників. *Наукова молодь-2019 : матеріали VII Всеукр. наук.-практ. конф., (м. Київ, 04 жовт. 2019 р.), 110–116.*
15. Кільченко, А. В. (2019). Google Analytics як засіб для здійснення аналітики веб-ресурсів наукової установи. *Звітна наук. конф. ІТЗН НАПН України: матеріали наук.-практ. конф., присвяч. 20-річчю ІТЗН НАПН України*, 109-117.
16. Beri, B., & Singh, P. (2013). Web analytics: Increasing website's usability and conversion rate. *International Journal of Computer Applications*, 72(6), 35-38.
17. Шмиголь, Н., Бирський, В., & Антонюк А. (2024). Математичний інструментарій веб-аналітики для оцінки вартості медіа-ресурсів у цифровій економіці. *Управління змінами та інновації*, (11), 5-9. <https://doi.org/10.32782/CMI/2024-11-1>
18. Гук, Н. А., Диханов, С. В., & Сірик С. Ф. (2019). Аналіз структури сайту за допомогою web-графа. *Питання прикладної математики і математичного моделювання*, (19), 88-95
19. Burby, J., Brown A. & WAA Standards Committee (2007) Web Analytics Definitions – Version 4.0. *Web Analytics Association*. <https://www.webanalytiker.dk/wp-content/logo/blog/WAA-Standards-Analytics-Definitions.pdf>
20. Малієнко, С.С., & Селівьорстова, Т.В. (2023). Огляд математичних моделей та інформаційних технологій бізнес аналізу великих web даних. *Системні технології. Регіональний міжвузівський збірник наукових праць*, 4(147), 21–28
21. Вальковський, Б. (n.d.) Від даних до стратегії: як веб-аналітика може трансформувати ваш бізнес. <https://inproject.org/yak-veb-analytika-mozhe-transformuvaty-vash-biznes/>
22. *European Parliament & Council of the European Union*. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union*, L119, 1–88. <https://gdpr-info.eu>
23. *California State Legislature*. (2018). California Consumer Privacy Act of 2018 (CCPA). <https://oag.ca.gov/privacy/ccpa>
24. Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
25. Одрехівський, М., Пшик-Ковальська, О., Жежуха, В., Іваночко, І. (2023). Інтелектуальне управління бізнес-процесами підприємства. *Математика*, 11(1), 78. <https://doi.org/10.3390/math11010078>

References

- 1.
2. Yavorskyi, I. M., Kravets, I. B., Matsko, I. Y., & Yuzevovich, R. M. (2011). Invariant correlation analysis of vector periodically correlated random processes. *Information Selection and Processing*, 35, 22–31. <https://doi.org/10.20535/S0021347014090039>
3. Yuzevovich, R. M. (2017). Modeling and statistical analysis of interrelated periodically non-stationary vibration signals for detecting mechanism defects. Doctoral dissertation. <https://lpnu.ua/sites/default/files/2020/dissertation/1433/dissertationyuzevovichrm.pdf>
4. Kyshakevych, B., & Nakhaeva, M. (2018). Forecasting foreign direct investments in Ukraine using ARIMA models. *Bulletin of Socio-Economic Research*. https://journals.urau.ua/vsed_oneu/article/view/168781
5. Quin, F., Weyns, D., Galster, M., & Costa Silva, C. (2023). A/B Testing: A Systematic Literature Review. <https://arxiv.org/abs/2308.04929>
6. Jyothi, P., Bonthu, S., & Prasanthi, B. V. (2017). A study on raise of web analytics and its benefits. *International Journal of Computer Sciences and Engineering*, 5(10), 59-64.
7. Pakholchuk, Ya. (2018). Web analytics tools for website visitor analysis. *Public Communications*, 1(3), 71-77.
8. Camacho, D., Panizo-Lledot A., Bello-Orgaz G., Gonzalez-Pardo A., & Cambria E. (2020). The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion*, 63, 88-120.
9. Ponomarenko, I. V. Web analytics as an important tool of digital marketing. (2018). *Eastern Europe: Economics, Business and Management*, 4 (15), 231-235.
10. Zhukovsky, D., & Lozovska, L. (2023). Analysis of the use of machine learning methods in the analytics of internet resource indicators. *Sustainable Economic Development*, 2(47), 65-69. <https://doi.org/10.32782/2308-1988/2023-47-9>
11. Coughlin, D.M., Campbell, M.C., & Jansen, B.J. (3 March 2016). A web analytics approach for appraising electronic resources in academic libraries. *Journal of the Association for Information Science and Technology*, 67, 518-534.
12. Shaheen, A. (2023). Maximizing Website Performance with Google Analytics. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 14(03), 1273–1277. <https://doi.org/10.61841/turcomat.v14i03.14229>
13. Lozovska, L. (2023). Analysis of the use of machine learning methods in the analysis of Internet resource indicators. *Sustainable Development of Economy*, 2 (47), 65-69.
14. Sufi, F. (2014). Advances in Mathematical Models for AI-Based News Analytics. *Mathematics*, 12(23), 3736. <https://doi.org/10.3390/math12233736>
15. Kilchenko, A. V., Shynenko, M. A., & Yaskova, N. V. (2019). Methodology for using the Google Analytics system to develop information and research competence of scientific and scientific-pedagogical workers. *Scientific Youth-2019: Proceedings of the VII All-Ukrainian Scientific-Practical Conference*, (Kyiv, 04 Oct. 2019), 110–116.
16. Kilchenko, A. V. (2019). Google Analytics as a means of conducting web resource analytics of a scientific institution. Report of the Scientific Conference of the Institute of Information Technology of the National Academy of Sciences of Ukraine: Proceedings of the Scientific-Practical Conference, dedicated to the 20th anniversary of the Institute of Information Technology of the National Academy of Sciences of Ukraine, 109-117.
17. Beri, B., & Singh, P. (2013). Web analytics: Increasing website's usability and conversion rate. *International Journal of Computer Applications*, 72(6), 35-38.
18. Shmygol, N., Byrsky, V., & Antonyuk A. (2024). Mathematical tools of web analytics for assessing the value of media resources in the digital economy. *Change and Innovation Management*, (11), 5-9. <https://doi.org/10.32782/CMI/2024-11-1>
19. Guk, N. A., Dykhanov, S. V., & Siryk S. F. (2019). Analysis of the structure of the site using a web graph. *Issues of applied mathematics and mathematical modeling*, (19), 88-95
20. Burby, J., Brown A. & WAA Standards Committee (2007) Web Analytics Definitions – Version 4.0. Web Analytics Association. <https://www.webanalytiker.dk/wp-content/logo/blog/WAA-Standards-Analytics-Definitions.pdf>
21. Malienko, S. E., & Seliverstova, T. V. (2023). Review of mathematical models and information technologies for business analysis of big web data. *System technologies. Regional Interuniversity Collection of Scientific Papers*, 4(147), 21–28
22. Valkovsky, B. (n.d.) From Data to Strategy: How Web Analytics Can Transform Your Business. <https://inproject.org/yak-veb-analytika-mozhe-transformuvaty-vash-biznes/>
23. European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union*, L119, 1–88. <https://gdpr-info.eu>
24. California State Legislature. (2018). California Consumer Privacy Act of 2018 (CCPA). <https://oag.ca.gov/privacy/ccpa>
25. Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
26. Odrekhivskiy, M., Pshyk-Kovalska, O., Zhezukha, V., Ivanochko, I. (2023). Intelligent Management of Enterprise Business Processes. *Mathematics*, 11(1), 78. <https://doi.org/10.3390/math11010078>