МІНІСТЕРСТВО ФІНАНСІВ УКРАЇНИ
ДЕРЖАВНИЙ ПОДАТКОВИЙ УНІВЕРСИТЕТ

ФАКУЛЬТЕТ ФІНАНСІВ ТА ЦИФРОВИХ ТЕХНОЛОГІЙ
КАФЕДРА КОМП'ЮТЕРНИХ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ І СИСТЕМ

КВАЛІФІКАЦІЙНА РОБОТА

на тему:
РОЗРОБКА СИСТЕМИ АНАЛІЗУ ПОВЕДІНКИ КОРИСТУВАЧІВ ІНТЕРНЕТ-МАГАЗИНУ НА БАЗІ ORACLE APEX ТА OML

Здобувача вищої освіти другого (магістерського) рівня,
ІІ курсу, групи КМІ-24-1, спеціальності 122 «Комп'ютерні науки»,
освітня програма «Інформаційні управляючі системи і технології в економіці»



Небоги Сергія Вікторовича
(підпис) (прізвище, ім'я, по батькові)

Науковий керівник Грищенко Світлана Миколаївна, канд. пед. наук, ст. досл.
(підпис) (прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Керівник освітньо-професійної програми Омельчук Антон Анатолійович, канд. техн. наук
(підпис) (прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Допущено до захисту за рішенням кафедри протокол від «02» січня 2026 № 10

Завідувач кафедри Філоненко Михайло Миколайович, канд. фіз.-мат. наук, доцент
(підпис) (прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Ірпінь – 2026

Реєстрація № 13 20.01.2026 _____
(дата) (підпис відповідальної особи)

Результати перевірки до захисту _____ 02.01.2026 _____
(до захисту, на доопрацювання) (дата) (підпис)

Результат захисту 5 (відмінно) 09.02.2026 _____
(оцінка) (дата) (підпис)

Голова комісії: _____ О.С. Дюбанов
(підпис)

Члени комісії:

_____ А.Ю. Горбовий
(підпис)

_____ М.М. Філоненко
(підпис)

_____ І.Д. Погорєловська
(підпис)

БЛАНК ОЦІНЮВАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ

Об'єкти контролю	Діапазон оцінки в балах	Фактична кількість балів	Примітки	Дата, П.І.Б.: керівника, голови комісії, членів комісії, секретаря, (підписи)
ОЦІНЮВАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ КОМІСІЄЮ ПО ЗАХИСТУ				
1. Зміст кваліфікаційної роботи (якість теоретичного матеріалу, його зв'язок з практичним, вміння представляти матеріал в логічній послідовності, наявність висновків та пропозицій)	від 0 до 30	29		09.02.2026 Дюбанов О.С. Горбовий А.Ю. Філоненко М.М. Погорєловська І.Д. Суховецька Л.П.
2. Дотримання вимог щодо оформлення	від 0 до 10	9		
3. Вчасне виконання роботи на всіх етапах у встановлені кафедрою терміни	від 0 до 5	5		
4. Рецензент	від 0 до 5	5		
РАЗОМ БАЛІВ	від 0 до 50	48		

Оцінювання кваліфікаційної роботи науковим керівником - 48 балів.

ПЕРЕВЕДЕННЯ ОЦІНКИ ЗА 100-БАЛЬНОЮ ШКАЛОЮ ДО НАЦІОНАЛЬНОЇ ШКАЛИ ТА ШКАЛИ ЄКТС

Оцінка за 100-бальною шкалою	Оцінка за шкалою ЄКТС	Оцінка за національною шкалою	Дата	П.І.Б.: голови, членів комісії, секретаря, (підпис)
96	A	5	09.02.2026	Дюбанов О.С. Горбовий А.Ю. Філоненко М.М. Погорєловська І.Д. Суховецька Л.П.

ПРАВИЛА ПЕРЕВЕДЕННЯ ОЦІНОК ДО НАЦІОНАЛЬНОЇ ШКАЛИ ТА ШКАЛИ ЄКТС

Оцінка за 100-бальною шкалою	Оцінка за шкалою ЄКТС	Оцінка за національною шкалою
90-100	A	5 (відмінно)
80-89	B	4 (добре)
70-79	C	
60-69	D	3 (задовільно)
50-59	E	
35-49	FX	2 (незадовільно)

Факультет фінансів та цифрових технологій
Кафедра комп'ютерних та інформаційних технологій і систем
Галузь знань 12 «Інформаційні технології»
Спеціальність 122 «Комп'ютерні науки»
Освітня програма «Інформаційні управляючі системи і технології в економіці»

Затверджую
Завідувач кафедри _____ Філоненко М.М.
«11» грудня 2024

Завдання на кваліфікаційну роботу
здобувачу вищої освіти
Небозі Сергію Вікторовичу
(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи
Розробка системи аналізу поведінки користувачів Інтернет-магазину на базі Oracle APEX та OML
Затверджена наказом ректора від «10» грудня 2024 № 1852
2. Строк здачі закінченої роботи «02» січня 2026
3. Мета роботи Розробка системи збору та аналізу поведінкових даних користувачів Інтернет-магазину на основі Oracle APEX та Oracle Machine Learning.
Об'єкт дослідження Поведінка користувачів Інтернет-магазину
Предмет дослідження Методи та засоби збору, інтеграції та кластеризації поведінкових даних
Орієнтовна база дослідження Власний вебсайт Інтернет-магазину, Oracle APEX та Oracle Machine Learning
4. Вихідні дані до роботи Дані про поведінку користувачів Інтернет-магазину
5. Перелік графічного матеріалу Скріншоти коду та інтерфейсу системи
6. Зміст кваліфікаційної роботи (перелік питань за кожним розділом)
ВСТУП
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ПОВЕДІНКИ КОРИСТУВАЧІВ ІНТЕРНЕТ-МАГАЗИНУ
1.1. Методи збору та структуризації даних.
1.2. Огляд сучасних підходів та інструментів аналізу.
РОЗДІЛ 2. РОЗРОБКА СИСТЕМИ ЗБОРУ ТА АНАЛІЗУ ПОВЕДІНКОВИХ ДАНИХ
2.1. Архітектура системи та реалізація механізмів збору даних.
2.2. Використання Oracle Machine Learning для аналізу поведінкових даних.
РОЗДІЛ 3. ВІЗУАЛІЗАЦІЯ ТА ІНТЕРПРЕТАЦІЯ РЕЗУЛЬТАТІВ В ORACLE APEX
ВИСНОВКИ
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ
ДОДАТКИ

7. Календарний план виконання роботи

№ п/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	
		За планом	Фактично
1	2	3	4
1	Вибір теми, підбір літературних джерел, формування структури кваліфікаційної роботи	До 15.12.2024	До 15.12.2024
2	**І розділ** (розкрити сутність і стан дослідження наукової проблематики роботи, після чого навести її загальну характеристику; огляд літератури за темою і вибір конкретних напрямів досліджень)	28.05.2025	28.05.2025
3	**ІІ розділ** (аналіз експериментальної частини про методики досліджень)	01.09.2025	20.08.2025
4	**ІІІ розділ** (проведені теоретичні і (або) експериментальні дослідження та проектні рішення)	10.11.2025	10.11.2025
5	**Висновки**	01.12.2025	01.12.2025
6	Доопрацювання роботи з урахуванням зауважень наукового керівника	30.12.2025	25.12.2025
7	Отримання відгуку керівника	20.01.2026	05.01.2026
8	Отримання зовнішньої рецензії	23.01.2026	23.01.2026

8. Дата видачі завдання «11» грудня 2024

9. Керівник кваліфікаційної роботи _____ (підпис) Грищенко С.М. (прізвище, ініціали)

10. Керівник освітньо-професійної програми _____ (підпис) Омельчук А.А. (прізвище, ініціали)

11. Завдання до виконання отримав здобувач вищої освіти _____ (підпис) Небога С.В. (прізвище, ініціали)

АНОТАЦІЯ

РОЗРОБКА СИСТЕМИ АНАЛІЗУ ПОВЕДІНКИ КОРИСТУВАЧІВ ІНТЕРНЕТ-МАГАЗИНУ НА БАЗІ ORACLE APEX ТА OML

Мета роботи полягає в розробці системи збору та аналізу поведінкових даних користувачів Інтернет-магазину на основі Oracle APEX та Oracle Machine Learning.

Завданнями роботи є аналіз предметної області, розробка механізму збору clickstream-даних, формування поведінкових ознак, побудова моделей кластеризації K-Means та візуалізація результатів в APEX.

Предметом дослідження є методи та засоби збору, інтеграції та кластеризації поведінкових даних.

Об'єктом дослідження виступає поведінка користувачів Інтернет-магазину.

Базою дослідження виступають власний вебсайт Інтернет-магазину, Oracle APEX та Oracle Machine Learning.

За результатами дослідження сформульовані рекомендації щодо побудови аналітичних панелей, сегментації користувачів та використання поведінкових ознак для оптимізації маркетингових стратегій.

Одержані результати можуть бути використані для аналізу поведінки користувачів та підтримки прийняття рішень у сфері електронної комерції.

Кваліфікаційна робота містить 56 сторінок, 14 рисунків, список використаних джерел з 40 найменувань, 4 додатки.

Ключові слова: поведінкові дані, clickstream, кластеризація, Oracle Machine Learning, Oracle APEX, електронна комерція.

Рік виконання кваліфікаційної роботи 2025

Рік захисту роботи 2026

ABSTRACT

DEVELOPMENT OF A SYSTEM FOR ANALYSING THE BEHAVIOR OF ONLINE STORE USERS BASED ON ORACLE APEX AND OML

The purpose of this work is to develop a system for collecting and analyzing the behavioral data of online store users based on Oracle APEX and Oracle Machine Learning.

The tasks of the work are to analyze the subject area, develop a mechanism for collecting clickstream data, form behavioral characteristics, build K-Means clustering models, and visualize the results in APEX.

The subject of the research is methods and means of collecting, integrating, and clustering behavioral data.

The object of the research is the behavior of online store users.

The research is based on the online store's own website, Oracle APEX, and Oracle Machine Learning.

The results of the research include the recommendations for building analytical panels, segmenting users, and using behavioral features to optimize marketing strategies.

The obtained results can be used to analyze user behavior and support decision-making in the field of e-commerce.

The qualification work consists of 56 pages, 14 images, the list of 40 references, and 4 appendices.

Keywords: behavioral data, clickstream, clustering, Oracle Machine Learning, Oracle APEX, e-commerce.

Year of qualification work: 2025

Year of work defense: 2026

ЗМІСТ

ВСТУП 3
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ПОВЕДІНКИ КОРИСТУВАЧІВ ІНТЕРНЕТ-МАГАЗИНУ 6
1.1. Методи збору та структуризації даних. 6
1.2. Огляд сучасних підходів та інструментів аналізу. 11
РОЗДІЛ 2. РОЗРОБКА СИСТЕМИ ЗБОРУ ТА АНАЛІЗУ ПОВЕДІНКОВИХ ДАНИХ 21
2.1. Архітектура системи та реалізація механізмів збору даних. 21
2.2. Використання Oracle Machine Learning для аналізу поведінкових даних. .. 30
РОЗДІЛ 3. ВІЗУАЛІЗАЦІЯ ТА ІНТЕРПРЕТАЦІЯ РЕЗУЛЬТАТІВ В ORACLE APEX 36
ВИСНОВКИ 48
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 51
ДОДАТКИ 57

ВСТУП

У сучасному цифровому середовищі, де Інтернет став повноцінною платформою для взаємодії між бізнесом і користувачами, поведінка відвідувачів вебсайтів набула вирішального значення для успішного розвитку електронної комерції. Онлайн-магазини вже давно перестали бути лише інструментом для продажу товарів – вони перетворилися на складні інтерактивні системи, де кожен клік, кожен перехід між сторінками, тривалість перебування на сайті та послідовність дій користувача формують цінні дані, які можуть бути використані для оптимізації маркетингових стратегій, покращення інтерфейсу, персоналізації контенту та підвищення відсотку конверсії[39, 11].

У міру зростання конкуренції на ринку електронної комерції бізнеси дедалі частіше звертаються до аналітики поведінкових даних, що дозволяє глибше розуміти наміри клієнтів, прогнозувати їхні дії та вчасно реагувати на зміни у споживчих патернах. Поширення мобільних пристроїв, збільшення частки онлайн-покупок та розвиток інструментів штучного інтелекту лише посилюють потребу у точному зборі та якісному аналізі такої інформації[4, 13, 38, 25].

Одним із сучасних підходів до вирішення цієї задачі є використання вбудованих аналітичних можливостей баз даних, зокрема Oracle Machine Learning, яке дозволяє проводити обчислення поблизу даних, забезпечуючи високу продуктивність і зручність інтеграції. У поєднанні з Oracle APEX, що виступає як low-code платформа для створення вебінтерфейсів та аналітичних панелей, це відкриває можливість побудови повноцінної системи збору, обробки та аналізу поведінки користувачів без необхідності використання зовнішніх сервісів або складних інфраструктурних рішень.

Актуальність даної теми зумовлена глобальною тенденцією переходу бізнесу до орієнтованих на дані моделей управління, де інформація про поведінку клієнтів стає основою для прийняття рішень. Ефективний аналіз clickstream-даних дозволяє виявляти проблемні місця у структурі сайту, визначати фактори відтоку користувачів, сегментувати аудиторію за схожими

патернами поведінки та підвищувати загальну ефективність електронної комерції. Дослідження в цій сфері ґрунтуються на роботах науковців, які розкривають методи обробки поведінкових даних, технології машинного навчання та принципи побудови аналітичних систем[3, 5, 14, 17, 20, 34].

Метою кваліфікаційної роботи є розробка системи збору та аналізу поведінкових даних користувачів Інтернет-магазину на основі Oracle APEX та Oracle Machine Learning з використанням методів кластеризації.

Для досягнення поставленої мети необхідно виконати такі завдання:

- провести аналіз предметної області та вивчити сучасні підходи до збору й аналізу поведінкових даних;
- розробити архітектуру системи збору clickstream-даних та реалізувати механізм їх передачі й збереження за допомогою JavaScript-трекера;
- створити представлення даних, що агрегують сесії та формують поведінкові ознаки користувачів;
- побудувати та протестувати моделі кластеризації за допомогою інструментів Oracle Machine Learning, реалізувати візуалізацію результатів аналізу в середовищі Oracle APEX;
- оцінити отримані результати та визначити можливі напрями подальшого вдосконалення системи.

Об’єктом дослідження є поведінка користувачів Інтернет-магазину.

Предмет дослідження – методи та засоби збору, інтеграції та кластеризації поведінкових даних у середовищі Oracle.

Методи дослідження – теоретичні – аналіз наукових джерел, систематизація підходів до збору та обробки clickstream-даних, порівняння методів кластеризації; *емпіричні* – розробка механізму збору даних, формування ознак в Oracle ADW, побудова та навчання моделей OML, статистична обробка та візуалізація результатів.

База дослідження – власний вебсайт Інтернет-магазину, аналітична платформа Oracle APEX та обчислювальне середовище Oracle Machine Learning.

Науково-практична значимість одержаних результатів. Розроблена система реалізує повноцінний цикл аналізу поведінкових даних без необхідності використання зовнішніх інструментів. Вона може бути використана для підтримки прийняття рішень у сфері електронної комерції, оптимізації інтерфейсу вебсайту та підвищення ефективності маркетингових кампаній. Крім того, реалізований підхід до формування поведінкових ознак та побудови моделей кластеризації може бути масштабований та адаптований до інших типів вебресурсів.

Результати роботи можуть бути використані як основа для подальших досліджень у галузі поведінкової аналітики та для впровадження практичних рішень у бізнес-процеси електронної комерції.

Апробація результатів роботи (додаток Г). Основні результати дослідження висвітлено в І Міжнародній науково-практичній конференції здобувачів вищої освіти та молодих вчених «ЕКОНОМІКА І БІЗНЕС: НОВІ РЕАЛІЇ, ТЕХНОЛОГІЧНІ ТРЕНДИ, ЦИФРОВІ ІННОВАЦІЇ» (15 травня 2025 р.) та в XVIII міжнародній науково-практичній конференції «ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ І АВТОМАТИЗАЦІЯ – 2025» (30 - 31 жовтня 2025 р.).

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ПОВЕДІНКИ КОРИСТУВАЧІВ ІНТЕРНЕТ-МАГАЗИНУ

1.1. Методи збору та структуризації даних.

В епоху цифровізації підприємства змушені швидко адаптуватися, щоб залишатися актуальними та конкурентоспроможними. Центральним елементом цієї адаптації є розуміння поведінки споживачів та стратегічне впровадження передових технологій[25]. Цифрова трансформація змінила підхід до взаємодії між бізнесом і клієнтами, створивши умови, у яких саме дані про поведінку користувачів стають ключовим ресурсом для прийняття ефективних управлінських рішень[29].

Оскільки платформи електронної комерції стають дедалі більш невід'ємною частиною повсякденного життя, споживачі шукають не лише зручних та ефективних транзакцій, але й персоналізованого та збагачуючого досвіду[25]. Це свідчить про те, що у цифрову добу успіх онлайн-бізнесу залежить не лише від пропозиції товарів чи послуг, а й від якості досвіду користувача (UX), який формується під впливом численних взаємодій між клієнтом і платформою[28].

Зміна очікувань споживачів зумовлена розширенням можливостей платформ електронної комерції. На ранніх етапах онлайн-шопінгу споживачів приваблювала, перш за все, зручність здійснення покупок, не виходячи з дому. Однак, з розвитком електронної комерції, очікування вийшли за рамки простої зручності та охопили персоналізований досвід[26].

Споживачі зараз вимагають більше, ніж звичайної операції купівлі-продажу; вони шукають захопливої та персоналізованої взаємодії з онлайн-платформами. Ця еволюція була підживлена такими факторами, як швидші варіанти доставки, зручні інтерфейси та доступність широкого спектру продуктів і послуг. Зростання популярності мобільних пристроїв ще більше прискорило цю еволюцію, зробивши електронну комерцію доступною в дорозі та посиливши потребу в адаптивному та інтуїтивно зрозумілому інтерфейсі[25].

Водночас із підвищенням очікувань користувачів відбувається й зростання рівня конкуренції серед різних брендів. Оскільки електронна комерція стала практикою, яку застосовує більшість роздрібних торговців та продавців, масштаби конкуренції значно зросли, оскільки електронна комерція пом'якшила бар'єри географічного розташування[36].

Однак, складність поведінки споживачів – це проблема, з якою стикаються всі роздрібні торговці та продавці у світі електронної комерції. Щодня звичайні споживачі зараз схильні приймати надзвичайно складні рішення і це призвело до того, що платформи електронної комерції, які не здатні надати споживачеві вагомих причин використовувати їх замість конкурентів, поступово втрачатимуть актуальність і будуть витіснені конкурентами. Тому для бізнесу життєво важливо стимулювати поведінку споживачів, яка відповідає їхнім найкращим інтересам, тобто поведінка, яку демонструють ці споживачі, повинна мати можливість конвертувати в прибутковість або дохід для бізнесу[35].

Технології на базі штучного інтелекту, включаючи машинне навчання та аналітику даних, здійснили революцію в розумінні компаніями своїх клієнтів, взаємодії з ними та обслуговуванні. Одним із найважливіших внесків цих технологій в електронну комерцію є персоналізовані системи рекомендацій. Передові алгоритми аналізують величезні набори даних, включаючи історію переглядів та моделі покупок, щоб надавати індивідуальні рекомендації щодо продуктів. Такий рівень персоналізації не лише збільшує ймовірність покупки товару, а й підвищує задоволеність клієнтів[25]. Саме завдяки цим підходам аналіз поведінки користувачів переходить від пасивного спостереження до активного формування пропозицій, що визначає конкурентоспроможність бізнесу в умовах цифрової економіки.

Одним із ключових інструментів для такого аналізу є clickstream («потік кліків») – послідовність дій, здійснених користувачами за допомогою кліків протягом усіх сеансів перегляду вебсторінок[15]. Дані кліків надають безліч поведінкових даних, які фіксують дії користувачів Інтернету, зокрема потік кліків на продукт, кількість переглядів цього продукту, покупки на вебсайті,

тривалість кожного відвідування, кількість відвіданих сторінок або кліків на рекламу тощо[14]. Ці дані відображають не тільки фінальні результати взаємодії користувача, наприклад, здійснену покупку, але й проміжні дії, які дозволяють побачити весь шлях користувача на вебсайті. Саме ця детальна послідовність дій створює можливість для аналітичного дослідження поведінкових патернів, виявлення точок входу і виходу, а також оцінки ефективності елементів інтерфейсу та маркетингових кампаній.

Завдяки даним clickstream аналітики можуть отримати комплексну картину взаємодії користувачів із сайтом, що відкриває широкі можливості для оптимізації UX та підвищення ефективності бізнес-процесів[22, 24, 23]. Зокрема, аналіз послідовностей кліків дозволяє зрозуміти, як користувачі рухаються по вебсайту, у якому порядку переглядають сторінки, які елементи інтерфейсу викликають найбільшу або найменшу активність, а також які переходи сприяють завершенню покупки. Ця інформація надає можливість оцінити швидкість реагування сайту, виявити точки, де користувачі натискають кнопки «Назад» або залишають сторінку, та визначити обсяг даних, який обробляється під час переходу між етапами перегляду. На основі цього аналізу можна оптимізувати навігацію сайту, покращити розміщення елементів і кнопок, скоротити кількість випадків, коли товари додаються до кошика, але не оформлюються, і підвищити відсоток покупок.

Створення персоналізованих механізмів взаємодії, таких як відображення нещодавно переглянутих товарів, дозволяє підтримувати безперервність користування вебсайтом, нагадуючи користувачам про раніше переглянуті продукти і стимулюючи повторні відвідування або покупки. Завдяки швидкому збору подій і формуванню потоку даних у практично реальному часі стає можливим оперативно оновлювати рекомендаційні блоки або адаптувати інтерфейс під поточну активність аудиторії, навіть у випадку великого обсягу користувачів та високого навантаження на систему[7].

Крім того, clickstream-дані дозволяють здійснювати аналіз, орієнтований на вивчення поведінки користувачів, зокрема, виявляти можливі ризику відтоку

клієнтів, передбачати майбутній вибір покупок та адаптувати пропозиції на основі індивідуальних вподобань. Це дозволяє компаніям оптимізувати процес ціноутворення, контроль запасів та маркетингові стратегії, що призводить до збільшення доходів та покращення досвіду користувача[16].

Ще однією важливою можливістю, яку відкриває clickstream, є аналіз наступного найкращого продукту (Next Best Product, NBP). Вивчення історичних моделей поведінки користувачів дозволяє прогнозувати наступну покупку, визначати товари, які з високою ймовірністю будуть придбані разом, і надавати персоналізовані рекомендації.

Також clickstream дає змогу здійснювати глибоку мікрорівневу сегментацію користувачів. На основі моделей переглядів і покупок можна формувати кластери користувачів із подібними патернами поведінки, враховуючи середню вартість кошика, частоту відвідувань або порядок перегляду продуктів. Це дозволяє пропонувати персоналізовані рекомендації, які мають вищу ймовірність прийняття, підвищуючи ефективність маркетингових стратегій та покращуючи загальний досвід користувача на вебсайті[27].

Таким чином, clickstream виступає не просто джерелом подій, а потужним аналітичним інструментом, який надає широкі можливості для вивчення поведінкових закономірностей, оптимізації інтерфейсу, прогнозування попиту та побудови персоналізованих стратегій взаємодії з користувачами[21].

Одним із поширених підходів є використання JavaScript-трекера як інструмента початкового збору подій, які далі передаються в обробний конвеєр[21, 12]. Такий трекер вставляється безпосередньо у код веб-сторінки і здатний відстежувати різноманітні події: натискання на кнопки, переходи за посиланнями, зміну URL, прокручування сторінки, взаємодію з формами та інші дії користувача. JavaScript-трекер виконується на стороні клієнта, тобто у браузері користувача, що дозволяє фіксувати кожну дію в момент її здійснення, зберігаючи контекст сеансу і точний час події. Це забезпечує високу точність та деталізацію даних у порівнянні з методами серверного логування, де часто втрачаються проміжні дії користувача або контекст їхнього виконання.

Процес збору clickstream-даних через JavaScript-трекер базується на принципі формування структурованих подій, де кожна взаємодія користувача перетворюється на окремий об’єкт події з чітко визначеними атрибутами. До таких атрибутів належать часові мітки, ідентифікатори сесії та користувача, URL сторінки джерела та сторінки призначення, а також характеристики взаємодії з конкретними елементами сайту. Кожен об’єкт події включає контекст дії, що дозволяє аналізувати не лише сам факт кліку, але і умови, за яких він був здійснений, та його взаємозв’язок із попередніми або наступними діями користувача. Таким чином, збір даних перетворюється на створення потоку подій, який відображає послідовність дій користувачів у часі та дозволяє побудувати комплексну модель поведінки.

Завдяки такому підходу можна не тільки виявляти типові шляхи користувачів на сайті, але й аналізувати проміжні точки взаємодії, що раніше залишалися поза увагою. Наприклад, можливо визначити моменти, коли користувачі затримуються на сторінках, повертаються на попередні етапи або переривають сеанс. Це відкриває перспективи для аналітичних досліджень, де дані використовуються для побудови моделей навігації, прогнозування поведінки, виявлення аномалій або оптимізації шляху користувача. Крім того, збір подій у структурованому вигляді дозволяє виконувати агрегацію, групування та фільтрацію даних у будь-якому обраному часовому масштабі, що створює основу для глибшого аналізу – від короткострокових патернів переглядів до довгострокових тенденцій поведінки аудиторії[33].

Таким чином, використання JavaScript-трекера дозволяє організувати неперервний, деталізований та структурований процес збору даних, що створює фундамент для подальшої агрегації, трансформації та аналітичної обробки в середовищі бази даних. Цей підхід дозволяє отримати повну картину взаємодії користувачів із веб-сайтом, оцінити ефективність інтерфейсу, виявити типові патерни поведінки та закономірності, а також сформувати дані для прогнозних і сегментаційних моделей, які зможуть підвищити якість користувацького досвіду та ефективність бізнес-процесів.

1.2. Огляд сучасних підходів та інструментів аналізу.

Розуміння поведінки користувачів в онлайн-шопінгу має вирішальне значення для покращення досвіду споживачів та збільшення продажів, особливо у стрімко мінливому секторі електронної комерції. З розвитком технологій з'являється можливість дедалі точніше фіксувати та аналізувати журнали сеансів і поведінкові сліди користувачів на вебсайтах. Використовуючи дані кліків, можна ефективніше аналізувати купівельну поведінку споживачів, пропонуючи нову перспективу та збагачуючи розуміння їхніх процесів прийняття рішень[3].

Сучасні системи переходять від простого фіксування подій до глибшого розуміння динаміки взаємодії, що дозволяє будувати персоналізовані рекомендації та знаходити приховані закономірності у поведінці. Одним із показових прикладів є дослідження A. Jain та A. Kumawat "Analysis of Clickstream Data"[14], яке зосереджене на структуризації, очищенні та інтерпретації поведінкових подій.

У роботі автори підкреслюють, що сучасні вебплатформи генерують великі обсяги даних про переходи між сторінками, час перегляду та взаємодію з інтерфейсом, тому важливою є якісна попередня обробка даних: очищення від шумів, нормалізація сесій та формування узгоджених логів. Без цього неможливо створити надійні вибірки для алгоритмів машинного навчання, оскільки clickstream-дані часто є фрагментарними та неоднорідними.

Ключовим аспектом роботи є застосування методів інтелектуального аналізу для виявлення закономірностей у послідовностях подій. Кластеризація, послідовні моделі та моделі прогнозування дозволяють ідентифікувати типові сценарії поведінки, аналізувати шляхи переходів і визначати потенційні проблемні точки у навігації. Алгоритми групування допомагають виділяти схожі патерни взаємодії та сегментувати користувачів за інтересами.

Автори також акцентують увагу на практичній користі clickstream-аналітики для електронної комерції. Навіть базові показники, такі як частота переходів чи тривалість сесій, можуть лягти в основу моделей, що прогнозують повернення користувачів, ймовірність покупки або ефективність рекламних

кампаній. Таким чином, аналітика має не лише описовий, а й прогнозний характер.

Дослідження показує, що точність аналізу залежить від якісного збору та структурування clickstream-даних і застосування алгоритмів, здатних виявляти приховані поведінкові патерни. Це дозволяє інтегрувати результати у процес прийняття рішень, оптимізувати інтерфейс і підвищити ефективність взаємодії користувачів із вебсайтом[14].

Продовжуючи огляд сучасних наукових підходів до аналізу поведінки користувачів, варто звернутися до робіт, що демонструють застосування машинного навчання для прогнозування дій на основі clickstream-даних. Такі дослідження формують фундамент для систем, здатних не лише ідентифікувати існуючі патерни, а й передбачати майбутні кроки користувачів, що є особливо важливим для електронної комерції, про що, зокрема, свідчить робота Y. F. Altayeb “Predicting Consumer Behavior in Online Shopping Using Clickstream Data and Machine Learning Algorithms”[3], у якій продемонстровано застосування сучасних моделей для прогнозування ймовірності покупки, повернення на сайт чи виконання ключових дій у межах сесії.

У дослідженні clickstream-дані розглядаються як точне відображення інтересів та намірів користувача. Автор наголошує, що послідовності подій формують значно ширшу картину, ніж традиційні метрики: вони допомагають оцінити проходження воронки продажів, динаміку навігації та зміни інтересу між категоріями товарів. Також підкреслюється важливість якісного структурування поведінкових даних – формування сесій, визначення ключових подій і стандартизації часових інтервалів, оскільки ці характеристики надалі стають основою для алгоритмів машинного навчання.

Частина роботи присвячена порівнянню різних моделей для прогнозування поведінки. Автор аналізує логістичну регресію, дерева рішень, випадкові ліси й градієнтний бустинг, визначаючи їхню здатність враховувати нелінійні взаємозв’язки між ознаками. Найкращі результати продемонстрували ансамблеві

моделі, зокрема градієнтний бустинг, що забезпечив найвищу точність завдяки здатності адаптуватися до складної та мінливої поведінки користувачів.

У роботі також розглянуто підходи до оцінювання якості моделей, а також акцентується важливість метрик точності, F1-міри та ROC-AUC, які відображають збалансованість класифікації, особливо для рідкісних подій на кшталт здійснення покупки після тривалої сесії. Найвищі показники точності досягнуто саме моделями градієнтного бустингу[3].

Інформація про поведінку користувачів не завжди є повною або однорідною, що створює складнощі для її аналізу та прогнозування намірів. Це особливо актуально для Інтернет-магазинів, де значна частина відвідувачів не авторизується, що ускладнює роботу систем персоналізації. Дослідження “Machine-Learning-Based Approach for Anonymous Online Customer Purchase Intentions Using Clickstream Data” авторів Z. Wen, W. Lin та H. Liu[34] є прикладом того, як методи машинного навчання можуть компенсувати відсутність профільних даних.

Автори підкреслюють, що анонімні відвідувачі становлять значну частку трафіку, але їхню поведінку складніше класифікувати через нестабільність патернів. Clickstream-дані стають основним джерелом інформації, адже саме вони фіксують кожен крок користувача і можуть замінити демографічні чи історичні дані. У роботі наголошується, що структурування послідовностей подій – аналіз переходів між категоріями товарів, тривалості сесій, повторних дій – є ключовою умовою для побудови моделей, здатних прогнозувати ймовірність покупки на рівні окремої сесії.

Особливу увагу приділено використанню моделей машинного навчання для класифікації намірів користувачів. Дослідження охоплює нейронні мережі, моделі для часових послідовностей та ансамблеві алгоритми, демонструючи їхні результати на «коротких» сесіях, коли кількість подій мінімальна. Автори показують, що навіть за умов обмежених даних моделі можуть досягати високої точності, якщо використовуються поведінкові ознаки, пов’язані з темпом навігації, швидкістю переходів і ритмом взаємодії.

Традиційні підходи, орієнтовані на профілі користувачів, виявляються неефективними, тому автори пропонують метрики, що описують інтенсивність поведінки, швидкість прийняття рішень та повторюваність дій. Саме такі ознаки підвищують точність моделей і мінімізують залежність від історичних даних.

Таким чином, дослідження доводить, що навіть анонімні поведінкові дані можуть бути ефективним ресурсом для прогнозування намірів користувачів і прийняття рішень у сфері електронної комерції.[34].

У дослідженнях сегментації користувачів за clickstream-даними особливе значення має підхід, який поєднує методи кластеризації з маркетинговою персоналізацією, дозволяючи виділяти сегменти користувачів із різним рівнем купівельного наміру, навіть у сесіях без здійснених покупок. Важливий внесок у цю сферу зробила робота Jennifer-Zhu Lezu “Analyzing and Clustering Clickstream Data for Marketing Personalization”[17], в якій показано, як кластеризація поведінкових сесій дає змогу виявити сегменти з різним рівнем купівельного наміру.

На початку формується задача класифікації not-buy-сесій за потенційним наміром купити, спираючись на анонімний набір даних clickstream. Вилідяються ключові ознаки – кількість подій, тривалість сесії, частоту переглядів компонентів і деталей товарів – обґрунтовуючи їх у літературі як релевантні індикатори наміру. Надалі ці ознаки використовуються як основа для сегментації.

У роботі застосовано алгоритм HDBSCAN, який дає змогу виявляти кластери різної щільності та виокремлювати аномальні сесії. Оцінювання кластерів проводиться на основі обмежень щодо кількості кластерів і перевірки відмінностей між ними за рівнем купівельного наміру. Додатково висувається гіпотеза про подібність клікової поведінки у buy- та not-buy-сесіях: якщо поведінка подібна, то й рівень наміру має бути близьким. Для перевірки цієї гіпотези застосовується як лінійна регресія (OLS), так і Random Forest Regression, моделюючи зв’язок між подібністю поведінки та наміром.

Результати показують наявність двох чітких кластерів not-buy-сесій, які суттєво відрізняються за купівельним наміром: один демонструє високий рівень наміру, інший – низький. Найінформативнішими ознаками виявилися кількість подій, тривалість сесії та частота переглядів. Аналіз buy- та not-buy-сесій підтвердив наявність нелінійної залежності: зі зростанням подібності поведінки наміри зближуються, але не лінійно. В дослідженні зазначається, що включення додаткових ознак, таких як навігаційні графи чи мережеві метрики, могло б підвищити цінність аналізу, проте це не було можливим через малий обсяг вибірки[17].

Системний аналіз clickstream-даних у сфері електронної комерції активно доповнюється методами глибокого навчання для відстеження та передбачення часових патернів поведінки. Саме до таких належить дослідження Iqra Batool “A Deep Learning Approach for the Online User Behaviour Prediction Using Clickstream Data”[5], яке демонструє приклад системного застосування рекурентних нейронних мереж для прогнозування подій у послідовності користувацьких дій.

Дослідження орієнтується на використанні LSTM-мереж для ідентифікації часових шаблонів у сесіях користувачів. На етапі попередньої обробки даних формуються структуровані послідовності подій: взаємодії на кшталт перегляду товару, пошуку чи додавання до кошика доповнюються часовими інтервалами між кліками, що дозволяє моделі враховувати як зміст взаємодій, так і їхню періодичність. Після цього конструюються кілька варіантів LSTM-архітектур, які навчаються прогнозувати як ключові цільові дії (наприклад, купівлю), так і проміжні події, що формують «траєкторію наміру» користувача.

У процесі навчання застосовується байєсова оптимізація гіперпараметрів, що дозволяє адаптувати модель до специфіки послідовностей clickstream-даних та дозволяє досягти високої точності прогнозування. За результатами експериментальної частини оптимізована модель демонструє понад 85% точності при прогнозуванні наступних подій у послідовності, що є переконливим показником ефективності використання LSTM у сфері поведінкової аналітики. Водночас зазначається, що традиційні моделі, такі як логістична регресія чи

дерева рішень, не враховують часовий порядок подій, а тому втрачають значну частину послідовних характеристик взаємодії користувачів.

LSTM моделює не лише статичні показники активності, а й характерні переходи між подіями, що дозволяє виявляти неочевидні закономірності – наприклад, повторювані послідовності переходів між переглядом товару, пошуком та взаємодією з кошиком. Отримана інформація може бути цінною для побудови персоналізованих рекомендацій, раннього визначення купівельного наміру або оптимізації інтерфейсу, але важливо правильно поділити дані на сесії та усунути шуми, оскільки якість сегментації часових залежностей істотно впливає на результат.

Попри високу точність моделі, у такого підходу наявні обмеження, пов’язані з обчислювальною складністю та потребою у великій кількості якісних даних. Для застосування LSTM необхідне потужне обчислювальне та програмне середовище, здатне обробляти значні обсяги потокових подій[5].

Серед класичних підходів до аналізу поведінкових даних особливу роль відіграє дослідження Noviantoro та Huang “Applying data mining techniques to investigate online shopper purchase intention based on clickstream data”[20], яке показує потенціал традиційних алгоритмів data mining у задачах прогнозування без опори на глибинні нейронні моделі. Такий підхід демонструє, що грамотно підібрані ознаки та продумана методологія обробки даних здатні забезпечити високий рівень точності прогнозів навіть без використання глибинних нейронних мереж.

Автори працюють з однорічними clickstream-логами інтернет-магазину, зосереджуючись на виявленні поведінкових факторів, які найбільше впливають на ймовірність покупки. На початковому етапі вони застосовують комплексну процедуру відбору ознак, що дозволяє усунути малозначущі параметри та виділити ключові метрики, серед яких важливу роль відіграли показники Google Analytics. Особливе місце займає метрика “page value”, яка виявилася одним із найсильніших предикторів купівельного наміру, оскільки вона відображає

внесок кожної сторінки в загальну конверсію та пов’язує поведінку відвідувача з цінністю окремих етапів його навігації.

Для побудови моделей класифікації автори застосовують широкий спектр стандартних алгоритмів машинного навчання, включаючи дерева рішень, Random Forest, нейронні мережі, метод k-найближчих сусідів, наївний Байєс, логістичну регресію та моделі, засновані на індукції правил. Порівнюючи їх за метриками точності, F-міри та ROC-AUC, дослідники встановлюють, що нейронні мережі забезпечують найвищу точність та F-міру, у той час як Random Forest демонструє найкращу стабільність в умовах неоднорідного набору даних і найвищу здатність до розрізнення класів навіть за наявності дисбалансу класів.

З огляду на це, дослідження показує, що традиційні методи data mining досі відіграють важливу роль у поведінковій аналітиці. Їхні результати можуть бути застосовані в інтернет-магазинах для прогнозування ймовірності покупки в реальному часі, раннього визначення цінного трафіку чи формування сегментів користувачів на основі їхніх дій, що робить це дослідження актуальним у контексті розвитку систем поведінкового аналізу[20].

Під час визначення вимог до технологічної інфраструктури підхід, представлений у дослідженні Y. A. Singgalen “Customer Experience Analysis for Marketing Strategy Optimization Using CRAF Framework”[30], демонструє, що клієнтський досвід слід розглядати як багатоступеневий процес, у якому цілісність аналітики має ключове значення для прийняття управлінських рішень. Це дослідження робить акцент на тому, що ефективна аналітика неможлива без наскрізної технологічної підтримки, у якій різні етапи – від фіксації подій до моделювання – реалізуються в межах єдиної екосистеми.

У рамках запропонованого автором фреймворку CRAF (Collect, Reflect, Adapt, Forward) продемонстровано, що повноцінна система роботи з клієнтським досвідом має включати механізми для збору поведінкових сигналів, їхнього аналізу, адаптації бізнес-стратегій та формування рекомендацій. Підкреслюється, що інструменти, здатні забезпечити взаємопов’язаність цих етапів, мають очевидні переваги над розподіленими рішеннями, які вимагають

складних інтеграцій, додаткових витрат на підтримку та створюють ризики неузгодженості даних.

Саме в цьому контексті Oracle APEX та Oracle Machine Learning відповідають концепції CRAF Framework: Oracle APEX забезпечує швидке створення інтерфейсів, збір і візуалізацію даних у вигляді інтерактивних сторінок, тоді як OML реалізує функції Reflect і Adapt, надаючи можливість виконувати моделі кластеризації, прогнозування та оцінки поведінкових намірів безпосередньо в базі даних. Завдяки цьому аналітичний процес стає не просто автоматизованим, а цілісним: дані не покидають сервер, моделі виконуються максимально близько до джерела інформації, а результати одразу доступні.

Як висновок, дослідження демонструє, що інтегровані рішення не лише спрощують архітектуру, а й забезпечують універсальність, надійність та швидкість аналітичного процесу – ключові характеристики, що визначають успішність аналітики у сфері електронної комерції[30].

З позиції вибору відповідної технологічної платформи особливо корисними є дослідження, у яких аналізуються можливості low-code рішень – платформ, що дозволяють створювати застосунки з мінімальним обсягом ручного кодування завдяки використанню візуальних компонентів та автоматизованих механізмів – у корпоративному середовищі, зокрема щодо роботи з поведінковими даними. Таке порівняння не лише дозволяє оцінити функціональні характеристики окремих платформ і визначити їхню відповідність вимогам проєктів із збору, обробки та аналізу поведінкових даних, але й є основою для системного аналізу ефективності Oracle APEX порівняно з іншими популярними low-code середовищами, що застосовуються у корпоративних IT-проєктах, як показано у дослідженні A. Syed “Evaluating Oracle APEX with Competing Low-Code Platforms for Enterprise Development”[31].

У роботі підкреслюється, що ключовими критеріями порівняння low-code платформ є швидкість розробки, можливість реалізації складних бізнес-процесів, інструменти для гнучкої побудови інтерфейсів, продуктивність системи при роботі з великими обсягами даних, безпека та масштабованість. Автор детально

аналізує, як Oracle APEX відповідає цим критеріям, звертаючи увагу на основну перевагу – тривимірну інтегрованість з екосистемою Oracle, де робоче середовище, база даних та машинне навчання функціонують як єдина інфраструктура без зовнішніх залежностей. Це дозволяє розробникам працювати безпосередньо з великими масивами даних, оптимізувати SQL-запити, використовувати PL/SQL для бізнес-логіки та застосовувати моделі Oracle Machine Learning без додаткових модулів чи API.

Особливу увагу в дослідженні приділено можливостям APEX у створенні аналітичних додатків. Зокрема, зазначається, що платформа добре пристосована до побудови інтерактивних панелей моніторингу, звітів і компонентів візуалізації, необхідних для аналізу поведінкових даних, включаючи clickstream-події. Особливої актуальності цей аспект набуває в системах електронної комерції, де аналітика має бути доступною в режимі реального часу, а зміни у поведінці користувачів потребують оперативного оновлення логіки обробки й моделювання. APEX показує стабільно високу продуктивність при обробці великих обсягів транзакційних даних завдяки використанню потужностей Oracle Database, що виділяє її серед low-code рішень, які покладаються на сторонні або слабше інтегровані сховища даних.

Важливою є також оцінка платформи з погляду безпеки. Так, APEX забезпечує комплексні механізми автентифікації, авторизації, сегментації доступу, моніторингу активності та шифрування, що робить її придатною для проєктів з обробки чутливої інформації. Це важливо для корпоративних систем, особливо в галузях електронної комерції, фінансів і телекомунікацій, де вимоги щодо конфіденційності та цілісності даних є високими.

Узагальнюючи результати порівняння, автор дослідження доходить висновку, що Oracle APEX поєднує властивості швидкої low-code розробки з широкими аналітичними та інтеграційними можливостями, характерними для повноцінних корпоративних платформ. Завдяки оптимальній взаємодії з Oracle Database та вбудованій підтримці Oracle Machine Learning, APEX дозволяє реалізувати повний цикл аналізу поведінки користувачів – від збору даних за

допомогою трекерів, їх попередньої обробки та візуалізації до побудови моделей прогнозування, сегментації та виявлення патернів поведінки[31].

Дослідження різних авторів показали, що аналіз послідовностей дій користувачів дозволяє не лише класифікувати сесії за типом поведінки, але й прогнозувати намір покупки, оцінювати потенційну цінність користувачів та визначати поведінкові патерни, що мають значення для персоналізації маркетингових стратегій.

Узагальнюючи проаналізовані вище дослідження, можна виділити кілька ключових принципів, які вплинули на вибір технологій для побудови системи аналізу поведінки користувачів. По-перше, необхідність роботи з потоковими даними clickstream визначила вибір платформи Oracle APEX як засобу інтеграції та зберігання структурованих подій, що дозволяє ефективно управляти сесіями користувачів та підготовкою даних для аналітики. По-друге, застосування алгоритмів машинного навчання та глибокого навчання обґрунтувало використання Oracle Machine Learning, що забезпечує роботу класифікаційних моделей, ансамблів, нейронних мереж та функцій трансформації ознак, необхідних для прогнозування поведінкових сценаріїв. По-третє, наголос на сегментації та виділенні цінних користувачів обумовив використання методів кластеризації, які можуть бути реалізовані безпосередньо у середовищі OML або шляхом попередньої обробки даних у APEX.

Таким чином, огляд сучасних досліджень підтвердив ефективність поєднання clickstream-аналітики та машинного навчання для прогнозування поведінки користувачів, виділення сегментів та персоналізації маркетингових стратегій. Він визначив технологічну основу системи аналізу поведінки користувачів, де збір та зберігання даних виконується в Oracle APEX, а побудова прогнозних моделей відбувається в Oracle Machine Learning, що дозволяє реалізувати інтегровану систему аналізу поведінки користувачів Інтернет-магазину.

.

РОЗДІЛ 2. РОЗРОБКА СИСТЕМИ ЗБОРУ ТА АНАЛІЗУ ПОВЕДІНКОВИХ ДАНИХ

2.1. Архітектура системи та реалізація механізмів збору даних.

Сучасні вебплатформи та онлайн-сервіси генерують великі обсяги поведінкових даних, що охоплюють широкий спектр дій користувачів – від перегляду сторінок до здійснення покупок. Ефективність бізнес-процесів дедалі більше залежить від здатності організацій застосовувати інтелектуальні засоби аналізу даних для виявлення закономірностей та прогнозування потреб користувачів[18]. Тому особливо важливо забезпечити узгоджену інтеграцію між модулями збору поведінкових даних на клієнтській стороні та аналітичними інструментами серверної частини, що дозволяє отримувати достовірну інформацію у режимі реального часу[39].

Під час виконання кваліфікаційної роботи було розроблено архітектуру інтегрованої системи збору, передачі та аналітичної обробки даних у хмарному середовищі Oracle Autonomous Data Warehouse (рис. 2.1). В основі лежить створення стабільного та захищеного каналу комунікації між JavaScript-трекером, який функціонує у браузері користувача, та серверною платформою Oracle ADW. Для забезпечення безпечної передачі даних реалізовано механізм автентифікації через Oracle Wallet, що гарантує шифрування та захист від несанкціонованого доступу. Кожна подія користувача – перегляд сторінок, взаємодія з елементами інтерфейсу тощо – фіксується JS-трекером, перетворюється у формат JSON та безпосередньо надсилається до ADW.

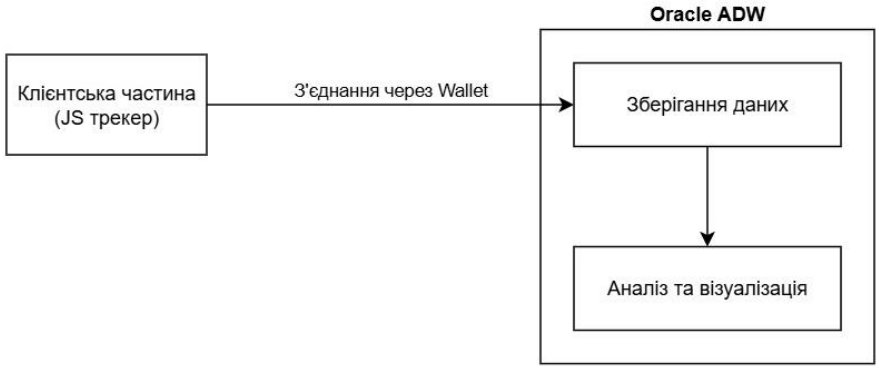


Рис. 2.1. Архітектура передачі та обробки даних в Oracle ADW. Джерело:[39]

В Oracle ADW відбувається прийом, перевірка та зберігання даних. Механізми бази забезпечують коректність структури записів, унікальність ідентифікаторів сесій та відсутність дублікатів. Далі дані доступні для аналітичної обробки у рамках Oracle Machine Learning (OML), де здійснюється агрегація, побудова ознак для моделей, кластеризація користувачів, прогнозування ймовірності покупок та аналіз шляхів навігації.

Платформа Oracle APEX інтегрується з ADW та OML, дозволяючи створювати вебдодатки для інтерактивної візуалізації результатів аналізу. Це забезпечує відображення ключових показників активності користувачів у реальному часі, таких як кількість сесій, середня тривалість відвідувань, взаємодія з контентом та інші метрики. Завдяки поєднанню збору, зберігання, обробки та візуалізації даних, забезпечується оперативна підтримка прийняття управлінських рішень та створюється основа для масштабованих рішень у веборієнтованих системах[39].

В сучасних умовах електронної комерції практично кожна взаємодія користувача з вебсайтом залишає цифровий слід у вигляді кліків, переглядів сторінок, переходів між категоріями товарів, додавання продуктів у кошик та інших дій, що можуть бути в подальшому проаналізовані для покращення досвіду користувача та підвищення ефективності бізнес-процесів. Кожна така подія містить важливу інформацію про поведінку користувача: коли і на якій сторінці відбулася взаємодія, який шлях проходив користувач між сторінками, з яких джерел трафіку він потрапив на сайт, і куди він переходив далі[37].

Для системного збору та зберігання даних про кліки користувачів було створено таблицю CLICKS, яка виступає основним джерелом даних для подальшого аналізу у середовищі Oracle APEX та Oracle Machine Learning. Таблиця дозволяє накопичувати великий обсяг даних, зберігати історію кожного кліку та організовувати інформацію так, щоб вона могла бути безпосередньо використана для побудови аналітичних моделей, машинного навчання та прогнозування поведінки користувачів.

Окрім цього, в ній реалізовано концепцію журналу подій (event log), що дозволяє фіксувати дії користувача на сайті. Кожен запис журналу відповідає одиничній дії – кліку по посиланню, який призводить до переходу між сторінками сайту[32]. У поєднанні з часовою міткою та унікальним ідентифікатором сесії такі дані дозволяють відтворити логіку взаємодії користувача з Інтернет-магазином, включно з послідовністю дій, глибиною навігації, тривалістю сесії, а також з'ясовувати, які саме переходи призводять до цільових дій (наприклад, додавання товару до кошика чи оформлення замовлення).

В структурі таблиці CLICKS, наведеній на рисунку 2.2, кожен стовпець має чітко визначене призначення, тип даних і набір обмежень. Основним ключем є CLICK_ID – числове поле типу NUMBER(22), яке слугує унікальним ідентифікатором кожного запису. Значення цього поля генерується автоматично за допомогою послідовності, що гарантує відсутність дублювання. Такий підхід особливо важливий у системах, де дані можуть надходити паралельно з великої кількості клієнтських пристроїв, що створює ризик одночасного запису. Автоматична генерація ідентифікаторів у межах самої бази усуває цей ризик і забезпечує цілісність даних.

Table	Column	Data Type	Length	Precision	Scale	Primary Key	Nullable	Default	Comment
CLICKS	CLICK_ID	NUMBER	22	-	-	1	-	"MY_SCHEMA"."ISEQ$$_105940".nextval	-
	SESSION_ID	VARCHAR2	32	-	-	-	-	-	-
	SOURCE_URL	VARCHAR2	150	-	-	-	-	-	-
	TARGET_URL	VARCHAR2	150	-	-	-	-	-	-
	CLICK_TIMESTAMP	TIMESTAMP(6) WITH TIME ZONE	13	-	6	-	-	-	-

Рис. 2.2. Метадані таблиці CLICKS

Поле SESSION_ID, визначене як VARCHAR2(32), зберігає унікальний ключ сесії користувача, який створюється сервером Django під час першого звернення користувача до сайту. Саме цей атрибут дозволяє групувати кілька кліків в одну логічну сесію, аналізуючи поведінку конкретного користувача. Цей ідентифікатор є неперсоніфікованим, тобто не містить жодних особистих даних, що відповідає принципам анонімності та вимогам щодо зберігання інформації про користувачів.

Поля SOURCE_URL і TARGET_URL, які призначені для фіксації маршруту користувача, мають тип VARCHAR2(150). Перше поле зберігає поточну адресу сторінки, з якої користувач здійснив клік, а друге – сторінку, на яку він перейшов. Такий підхід дозволяє будувати граф переходів у межах сайту, виявляючи найчастіші напрями руху між сторінками, «вузькі місця» у навігації, а також ключові сторінки, що спонукають до покупки. Крім того, фіксація цих даних у явному вигляді спрощує інтеграцію з інструментами машинного навчання, де така інформація може бути перетворена у вигляді пар переходів тощо.

Поле CLICK_TIMESTAMP, визначене як TIMESTAMP(6) WITH TIME ZONE, зберігає дату та час здійснення кліку з точністю до шести знаків після секунди, а також включає часову зону, що дозволяє враховувати географічні відмінності у поведінці користувачів. Це надзвичайно важливо для аналітики активності – зокрема, для виявлення пікових годин відвідувань, розподілу трафіку за часовими поясами або відстеження сезонних тенденцій. Збереження часу з часовою зоною забезпечує єдину хронологічну послідовність незалежно від того, звідки здійснювався запит.

Щоб забезпечити цілісність даних і уникнути ситуацій, коли час кліку передається в некоректному форматі або взагалі не передається, в базі даних було створено тригер clicks_timestamp_trg, код якого наведено на рисунку 2.3. Цей тригер автоматично встановлює значення поля CLICK_TIMESTAMP, якщо воно не було передано під час вставки запису.

```
CREATE OR REPLACE TRIGGER clicks_timestamp_trg
BEFORE INSERT ON CLICKS
FOR EACH ROW
WHEN (NEW.click_timestamp IS NULL)
BEGIN
  :NEW.click_timestamp := SYSTIMESTAMP AT TIME ZONE 'Europe/Kyiv';
END;
/
```

Рис. 2.3. Код тригеру click_timestamp_trg.

Тригер спрацьовує перед кожною операцією вставки (BEFORE INSERT) для кожного нового рядка (FOR EACH ROW). Умова WHEN (NEW.click_timestamp IS NULL) гарантує, що значення часу буде згенеровано лише тоді, часову мітку не було передано. Це особливо корисно у випадках, коли кліки імітуються сторонніми скриптами, або коли користувач взаємодіє із системою в офлайн-режимі, а дані надсилаються пізніше. Завдяки цьому тригеру таблиця завжди містить повну часову інформацію, що робить можливим відтворення подій у хронологічному порядку без втрат.

Використання SYSTIMESTAMP AT TIME ZONE 'Europe/Kyiv' забезпечує фіксацію часу відповідно до київського поясу, що важливо для узгодженості аналітики у регіональному контексті. Таким чином, навіть якщо сервер чи клієнт мають різні часові зони, усі події записуються у спільному часовому стандарті.

Така структура дає змогу не лише аналізувати популярні джерела переходів і найбільш відвідувані цільові сторінки, а й виявляти вузькі місця у структурі сайту, де користувачі можуть залишати сторінку або зупиняти навігацію. Дослідження таких переходів є основою для оптимізації структури сайту, включаючи UX/UI, оскільки дозволяє зрозуміти, які елементи сайту працюють ефективно, а які потребують доопрацювання.

Завдяки такій організації таблиці CLICKS забезпечується не лише надійне накопичення даних, а й ефективна основа для подальшої аналітики за допомогою Oracle Machine Learning. На основі комбінацій SESSION_ID, SOURCE_URL, TARGET_URL та CLICK_TIMESTAMP можна будувати моделі кластеризації поведінки, визначати закономірності переходів, а також прогнозувати наступні дії користувача.

Після того, як структура таблиці CLICKS створена і визначені всі необхідні поля для збереження інформації про кліки користувачів, наступним ключовим етапом є організація механізму передачі даних з вебсайту до бази даних Oracle APEX. Даний механізм був реалізований шляхом інтеграції JavaScript-трекера на стороні клієнта, серверної логіки на Django та таблиці CLICKS. Така архітектура дозволяє збирати всі події користувача у реальному часі, зберігати їх із

максимальною точністю та робити дані доступними для подальшого аналітичного оброблення і побудови моделей машинного навчання.

На стороні клієнта реалізований JavaScript-трекер (рис. 2.4), який відстежує події кліку на елементах <a> (гіперпосиланнях) вебсторінок. Коли користувач здійснює клік, скрипт формує об’єкт data, що містить такі поля: адресу поточної сторінки (source_url), адресу цільової сторінки (target_url), ідентифікатор продукту (product_id), якщо відбувся перехід на сторінку продукту, а також CSRF-токен для безпечного відправлення даних. Важливо, що source_url завжди фіксує поточну сторінку, на якій користувач перебуває, а target_url автоматично отримується з атрибуту href елемента, на який було здійснено клік. Це дозволяє точно відтворювати навігаційні шляхи користувача та формувати послідовність кліків у межах однієї сесії.

```
$(document).ready(function () {
  $(document).on('click', 'a', function () {
    data = {
      "source_url": document.location.pathname,
      "target_url": new URL(this.href).pathname,
      "csrfmiddlewaretoken": $("[name=csrfmiddlewaretoken]").val(),
      "product_id": $(this).data("product-id"),
    };
    console.log(data);
    $.ajax({
      type: "POST",
      url: '/track/',
      data: data,
      success: function (data) {
        console.log("Success")
      },
      error: function (data) {
        console.log("Error");
      },
    });
  });
});
```

Рис. 2.4. Код JS-трекера.

Для передачі даних використовується асинхронний AJAX-запит методом POST на серверний ендпоінт «/track/». Асинхронність забезпечує те, що

користувач не відчуває затримок під час взаємодії з сайтом, і процес збору даних відбувається у фоновому режимі. Завдяки цьому навіть у разі високого навантаження або великої кількості одночасних користувачів не порушується робота сайту, а всі кліки фіксуються без втрат.

На боці сервера Django створено представлення track_click (рис. 2.5), яке приймає ці дані, обробляє їх і записує у таблицю CLICKS. Важливою особливістю є використання декоратора @csrf_exempt. Його наявність не випадкова – частина подій може надходити не лише від браузера користувача, а й від сторонніх скриптів, що імітують дії. Звичайна перевірка CSRF-токенів у таких випадках призвела б до блокування запитів, тому вимкнення перевірки на цьому етапі дозволяє приймати всі дані для повного аналізу.

```
@csrf_exempt
def track_click(request):
    if request.method == 'POST':
        source_url = request.POST.get('source_url')
        target_url = request.POST.get('target_url')

        if '/cart/cart_' in target_url:
            target_url += '?product_id=' + request.POST.get('product_id')

        session_id = request.session.session_key

        Clicks.objects.using('oracle').create(
            source_url=source_url,
            target_url=target_url,
            session_id=session_id
        )

        return JsonResponse({'status': 'ok'})
    return JsonResponse({'error': 'invalid request'}, status=400)
```

Рис. 2.5. Код функції track_click.

Спершу отримуються значення полів source_url і target_url. Якщо користувач додає, видаляє або редагує кількість товарів в кошику (/cart/cart_add/, /cart/cart_remove/, /cart/cart_change/), до target_url додається параметр product_id, що дозволяє точно визначати, стосовно якого продукту виконані вищеперелічені

дії. Це важливо для аналізу поведінки щодо конкретних товарів та побудови моделей аналізу кошиків.

Фіксація сесії здійснюється за допомогою session_id, який отримується з ключа сесії Django. Це поле дозволяє групувати всі кліки одного користувача в межах однієї сесії, що необхідно для аналізу повних шляхів користувачів, виявлення повторюваних сценаріїв поведінки, визначення типових переходів між сторінками та прогнозування наступних дій користувача. Всі дані передаються до таблиці CLICKS у базі Oracle APEX через ORM Django з використанням підключення using('oracle'), що забезпечує правильну роботу із зовнішньою базою та автоматичне формування нового запису з усіма необхідними полями: source_url, target_url, session_id. Для цього у конфігураційному файлі settings.py (рис. 2.6) використовується окреме підключення ‘oracle’, де підключення виконується з використанням Oracle Wallet – програмного контейнера, який зберігає облікові дані для автентифікації та підписання[10].

```
DATABASES = {
    'default': {
        'ENGINE': 'django.db.backends.sqlite3',
        'NAME': BASE_DIR / 'db.sqlite3',
    },
    'oracle': {
        'ENGINE': 'django.db.backends.oracle',
        'NAME': 'connection_name',
        'USER': 'MY_SCHEMA',
        'PASSWORD': 'V5mu33HEt4dm',
        'OPTIONS': {
            'config_dir': BASE_DIR / 'wallet_folder',
            'wallet_location': BASE_DIR / 'wallet_folder',
            'wallet_password': 'm1~o/ak1Y{9O'
        }
    }
}
```

Рис. 2.6. Конфігурація підключення до Oracle ADW

Цей механізм зберігає зашифровані облікові дані та сертифікати, необхідні для з’єднання з хмарною базою Oracle Autonomous Database. Використання параметрів config_dir, wallet_location та wallet_password гарантує безпечну передачу даних і конфіденційність автентифікації. Інше підключення – ‘default’, призначене для локального зберігання даних вебсайту. Для деяких значень конфігурації використано умовні приклади задля захисту конфіденційної інформації.

Після налаштування безпечного каналу з’єднання між додатком Django і базою даних Oracle наступним кроком стало впровадження механізмів, що забезпечують інтеграцію системи збору кліків із аналітичним середовищем Oracle Application Express (APEX). Основна ідея впровадження механізму передачі даних полягає в тому, що всі кліки, зафіксовані на рівні веб-інтерфейсу, потрапляють у таблицю CLICKS, розміщену в схемі MY_SCHEMA у базі Oracle. Система APEX підключається до тієї ж самої схеми, тому одразу отримує доступ до даних у режимі реального часу без необхідності дублювання чи проміжного копіювання. Така архітектура значно спрощує обробку та мінімізує затримку між моментом збору інформації та її візуалізацією.

Окрім цього, особливу увагу було приділено забезпеченню узгодженості часових зон. Оскільки тригер clicks_timestamp_trg зберігає часові мітки у часовій зоні Europe/Kyiv, усі SQL-запити в APEX налаштовано з використанням функції AT TIME ZONE 'Europe/Kyiv'. Це гарантує, що відображення даних у графіках та звітах відповідає локальному часу аналітичного середовища.

Для відображення результатів аналізу поведінки користувачів у розробленій системі використовується інтеграція Oracle Machine Learning з платформою Oracle APEX, що дозволяє поєднати аналітичну обробку даних і їх візуальне представлення в єдиному середовищі. Після виконання попередньої агрегації clickstream-даних та формування ознак на рівні підсесій, моделі кластеризації в OML здійснюють аналіз поведінкових патернів і зберігають результати у вигляді службових представлень бази даних, які стають джерелом даних для аналітичної панелі в APEX.

2.2. Використання Oracle Machine Learning для аналізу поведінкових даних.

Для підготовки даних до аналізу в середовищі Oracle Machine Learning було використано таблицю CLICKS, що містить записи кліків користувачів на сторінках Інтернет-магазину. Оскільки вебсайт не має реальних відвідувачів, для наповнення бази даних використовувався спеціальний Python-скрипт, код якого наведений в додатку А.

Скрипт імітує поведінку користувачів на сайті, відстежує перехід між сторінками та формує часові мітки кліків із урахуванням перерв і нічного часу. Він здійснює завантаження сторінок сайту, парсинг HTML для пошуку активних посилань, випадковий вибір наступного URL для переходу, а також обробку спеціальних сторінок, таких як каталоги товарів і кошик. Кожна сесія отримує унікальний ідентифікатор (SESSION_ID), що забезпечує коректну агрегацію даних, а також формується інформація про джерело кліка (SOURCE_URL) та цільову сторінку (TARGET_URL). Для забезпечення безпечної передачі даних скрипт підключається до бази Oracle через wallet, вставляючи записи безпосередньо у таблицю CLICKS.

В результаті формується набір реалістичних clickstream-даних, який включає в себе як короткі, так і тривалі сесії, а також переходи на сторінки товару, додавання товарів до кошику, повернення на головну сторінку тощо.

Оскільки довгі сесії часто містять перерви, не пов’язані зі справжньою взаємодією з сайтом, наприклад, відволікання користувача, перехід до інших вкладок або повернення до пристрою через тривалий час, для більш точного аналізу поведінки користувачів було введено поняття підсесій. Підсесія формується, коли інтервал між двома послідовними кліками в межах однієї сесії перевищує 30 хвилин, що дозволяє враховувати паузи у взаємодії користувача та розділяти довгі сесії на менші сегменти активності. Окрім цього, розбиття сесії на підсесії дає змогу уникнути спотворення статистичних показників, таких як середній час між кліками чи тривалість сесії, і забезпечує більш коректне формування поведінкових ознак.

Скриптом було згенеровано 5007 кліків, на основі яких сформовано 692 унікальні сесії, що у свою чергу були розділені на 990 підсесій. Такий обсяг даних є достатнім для побудови моделей поведінки користувачів, оскільки включає різні сценарії взаємодії: короткі та довгі сесії, переходи на головну сторінку, відвідування каталогу та роботи з кошиком, а також випадки відскоків, коли користувач не здійснює більше одного кліку на сайті.

Для підготовки даних до навчання створено представлення USER_BEHAVIOR_VIEW (додаток В), в якому для кожної підсесії обчислюються базові метрики, що слугують первинними ознаками для аналізу: загальна кількість кліків (total_clicks), кількість унікальних цільових сторінок (unique_targets) та джерел трафіку (unique_sources), тривалість підсесії у хвилинах (session_duration_minutes), а також ознака відскоку (is_bounce), яка приймає значення 1 для підсесій із одним кліком. Ці базові метрики дозволяють оцінити активність користувача та визначити інтенсивність його взаємодії з сайтом.

На основі зібраних даних формується розширений набір ознак для подальшої кластеризації:

- clicks_per_minute – середня кількість кліків на хвилину, що відображає активність підсесії. Цей показник відображає інтенсивність взаємодії користувача з вебсайтом, тобто наскільки активно він переміщується між сторінками. Високе значення може свідчити про швидке перегортання сторінок або поверхневий огляд, тоді як низьке – про більш спокійний, уважний перегляд або тривалі паузи між діями;
- avg_time_between_clicks – середній час між кліками, який дозволяє оцінити ритмічність поведінки: короткі проміжки часу означають динамічну взаємодію, а довші можуть вказувати на читання, аналіз товарів або очікування завантаження сторінки. Ознака також допомагає виявляти підсесії з потенційно аномальною поведінкою;
- product_view_ratio – характеризує частку переходів на сторінки товарів відносно загальної кількості кліків у межах підсесії. Цей відношення дає

змогу визначити ступінь потенційної зацікавленості користувача покупкою: високе значення зазвичай сигналізує про орієнтовану на покупку поведінку, тоді як низьке – про навігацію по сайту, не звертаючи увагу на конкретні товари;

- cart_visits –кількість кліків, пов’язаних із додаванням товарів у кошик. Це одна з найбільш значущих ознак, оскільки вона прямо свідчить про формування наміру здійснити покупку. Навіть поодинокі додавання можуть бути індикатором високої зацікавленості, тому цей показник є важливим для сегментації користувачів за рівнем залученості;
- session_start_hour – година доби, коли почалась підсесія. Включення цієї ознаки дозволяє враховувати добові патерни поведінки: наприклад, вечірні сесії можуть бути більш активними в контексті купівельних намірів, тоді як ранкові спрямовані виключно на перегляд товарів;
- subsession_id – відображає порядковий номер підсесії в межах основної сесії користувача. Вона дозволяє простежити структуру взаємодій: перші підсесії часто характеризуються первинним знайомством із сайтом, тоді як наступні можуть відображати повернення до товарів або здійснення цільових дій.

Ці ознаки формують комплексний профіль підсесії та слугують вхідними даними для побудови моделей кластеризації в Oracle Machine Learning. Завдяки такому підходу можна оцінювати поведінкові патерни користувачів на рівні підсесій, виділяти типові сценарії взаємодії та готувати основу для подальшого прогнозування і сегментації аудиторії.

Для подальшого аналізу поведінки користувачів в Oracle Machine Learning було реалізовано процес кластеризації сесій за допомогою алгоритму K-Means. Основною метою цього підходу є виділення груп користувачів із подібними поведінковими патернами, що дозволяє виявляти типові сценарії взаємодій, оцінювати рівень залученості та інтенсивність активності.

На основі представлення USER_BEHAVIOR_VIEW, що містить сформовані ознаки підсесій, було створено серію моделей кластеризації K-Means

із різною кількістю кластерів (від 2 до 10) за допомогою скрипта, код якого наведений в додатку В. Для кожної моделі встановлюються однакові параметри алгоритму:

- ALGO_NAME = 'ALGO_KMEANS' – вибір алгоритму K-Means, який розбиває дані на попередньо визначену кількість кластерів за схожістю ознак;
- PREP_AUTO = 'ON' – автоматична попередня обробка даних перед навчанням моделі, включаючи нормалізацію та обробку пропущених значень;
- KMNS_DISTANCE = 'KMNS_EUCLIDEAN' – використання Евклідової відстані для вимірювання подібності між об’єктами;
- KMNS_DETAILS = 'KMNS_DETAILS_ALL' – збереження детальної інформації про кластери та ознаки;
- KMNS_ITERATIONS = 50 – максимальна кількість ітерацій алгоритму для досягнення збіжності кластерів;
- KMNS_NUM_BINS = 10 – кількість інтервалів для розподілу ознак всередині кластерів;
- CLUS_NUM_CLUSTERS = TO_CHAR(k) – кількість кластерів, яка визначається змінною k, що змінюється від 2 до 10[11].

Параметри ALGO_NAME і PREP_AUTO застосовуються для всіх моделей, тоді як інші – для алгоритму кластеризації K-Means. Після виконання кластеризації результати моделей зберігаються у службових представленнях Oracle Machine Learning:

- DM$VD<назва_моделі> – містить інформацію про дисперсію об’єктів у кожному кластері. Для кожного кластера зберігаються значення LEFT_CHILD_ID та RIGHT_CHILD_ID, які дозволяють визначити вузли ієрархії кластерів, а також поле DISPERSION, що показує середню відстань об’єктів від центру кластера. Ці дані використовуються для оцінки компактності кластера: менша дисперсія свідчить про більш однорідну групу підсесій;

- DM$VA<назва_моделі> – містить статистичні характеристики кожного кластеру для всіх ознак. Для кожного атрибуту підсесії зберігаються середнє значення (MEAN), дисперсія (VARIANCE) та мода (MODE_VALUE). Це дозволяє порівнювати кластери між собою за рівнем активності користувачів, інтенсивністю кліків, часткою переходів на товари або кошик та іншими поведінковими ознаками. За допомогою цих показників можна визначити типовий профіль підсесії для кожного кластера.;
- DM$VH<назва_моделі> – містить інформацію про розподіл значень ознак у межах кластерів. Кожен інтервал (BIN) визначається за нижньою (LOWER_BIN_BOUNDARY) та верхньою (UPPER_BIN_BOUNDARY) межами, і для кожного інтервалу вказується кількість підсесій, які потрапили в нього (COUNT). Таким чином, можна побачити розподіл кількості кліків на підсесію чи додавань у кошик за інтервалами[19];
- CLUSTER_ID() та CLUSTER_PROBABILITY() – функції, що дозволяють для кожної підсесії отримати ідентифікатор кластера та ймовірність належності до нього. У представленнях ці значення зберігаються у вигляді числових стовпців: CLUSTER_ID показує, до якого кластера віднесена підсесія, а PROBABILITY відображає ступінь впевненості алгоритму у приналежності підсесії до цього кластера. Завдяки цим даним можна отримати повну таблицю підсесій із їх кластерними характеристиками і деталізованими показниками ймовірності[8].

Оскільки службові представлення створюються і підтримуються на рівні бази даних, управління результатами моделей здійснюється централізовано: адміністратор або аналітик може оновлювати моделі, змінювати їх параметри, переглядати якість кластеризації та контролювати актуальність вихідних ознак без змін у клієнтських застосунках.

Використання службових представлень дозволяє аналізувати поведінкові патерни безпосередньо у базі даних без додаткової обробки за межами Oracle ADW, забезпечуючи оперативний доступ до результатів кластеризації для

інтеграції у вебдодатки та дозволяючи централізовано управляти моделями і результатами, що підвищує безпеку та точність аналітики.

Завдяки інтеграції службових представлень із аналітичними інструментами Oracle APEX результати кластеризації можуть бути використані для побудови інтерактивних панелей, графіків, таблиць із фільтрами та персоналізованих рекомендацій без потреби у написанні великого обсягу коду. Це дозволяє створювати аналітичні інтерфейси, у яких зміни в моделях та даних автоматично відображаються на рівні візуалізації, забезпечуючи актуальність даних у режимі реального часу.

РОЗДІЛ 3. ВІЗУАЛІЗАЦІЯ ТА ІНТЕРПРЕТАЦІЯ РЕЗУЛЬТАТІВ В ORACLE APEX

Ефективне використання поведінкових даних неможливе без їх інтерпретації у формі зрозумілих та наочних аналітичних представлень. Навіть за наявності якісних моделей машинного навчання, самі по собі числові результати кластеризації, статистичні характеристики ознак або матриці ймовірностей не дають цінності доти, доки не будуть інтегровані у процес прийняття рішень.

У сучасних інформаційних системах візуалізація відіграє роль інструмента когнітивного посередництва: вона перетворює багатовимірні дані, приховані закономірності та статистичні патерни на структури, які легко оцінюються з позицій бізнес-цілей[6]. Саме тому важливо забезпечити перехід від результатів обчислень, отриманих у середовищі Oracle Machine Learning, до візуалізованих елементів інформаційної панелі, доступної для кінцевих користувачів і аналітиків.

Oracle APEX забезпечує пряму інтеграцію зі сховищем Oracle Autonomous Database та службовими уявленнями моделей OML, що створює безперервний аналітичний цикл: збір даних – виділення ознак – моделювання в OML – візуалізація та інтерпретація в APEX. Завдяки цьому аналітичні панелі стають частиною єдиної платформи, а процес дослідження поведінкових патернів – інтерактивним і гнучким.

Аналітична панель була обрана як основний інструмент для відображення та аналізу поведінкових даних користувачів Інтернет-магазину, оскільки вона дозволяє поєднати збір, зберігання, обробку та візуалізацію даних у єдиному середовищі. Використання аналітичної панелі виправдане тим, що вона забезпечує швидкий доступ до результатів моделей, інтерактивність у роботі з даними та можливість динамічного аналізу без потреби у додаткових зовнішніх інструментах. Такий підхід особливо ефективний для веборієнтованих систем,

де поведінка користувачів змінюється динамічно і важливо оперативно відслідковувати тенденції та аномалії[2].

Після визначення аналітичної панелі як основного інструменту для відображення та аналізу поведінкових даних, наступним кроком стало проєктування структури додатку в середовищі Oracle APEX. Головна сторінка аналітичної панелі, зображена на рисунку 3.1, виконує роль центрального навігаційного вузла, що дозволяє отримати доступ до двох основних функціональних блоків: сторінки аналізу кліків та сторінки аналізу кластерів.

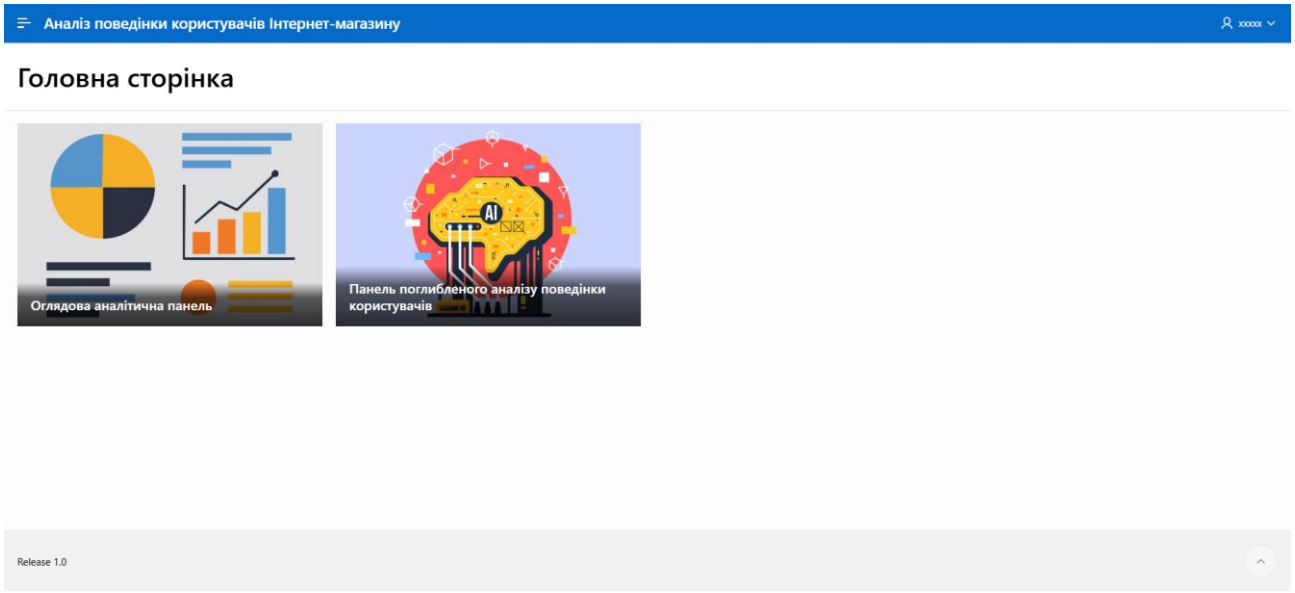


Рис. 3.1. Головна сторінка аналітичної панелі.

Головна сторінка містить інтерактивні картки для переходу, кожна з яких відповідає за доступ до конкретного функціоналу: перша картка веде до сторінки аналізу кліків, де відображаються дані про їх кількість та характер взаємодій користувачів із сайтом, а друга – до сторінки аналізу кластерів, де результати сегментації користувачів на основі моделей K-Means відображаються у вигляді графіків та таблиць. Картки реалізовані через стандартні APEX Regions із застосуванням динамічних дій, які реагують на клік і виконують переадресацію на потрібну сторінку.

Під час розробки аналітичної панелі було виявлено, що часті запити до представлень з великим обсягом даних, зокрема до USER_BEHAVIOR_VIEW призводять до значного навантаження на пул підключень і можуть спричиняти

помилки підключення до бази даних (HTTP 571, HTTP 572). Це пов’язано з тим, що при кожному виклику представлення його SQL-код виконується заново, повторно обчислюючи всі агреговані показники та ознаки підсесій. Так, під час розробки додатку число викликів (Execute Count) в пікові моменти перевищувало 50 тисяч, а за добу могло досягати понад 4 мільйонів.

Для зменшення навантаження та підвищення стабільності роботи було прийнято рішення замінити динамічне представлення USER_BEHAVIOR_VIEW на фізичну таблицю USER_BEHAVIOR_TABLE, наповнену агрегованими даними та ознаками сесій, що дозволить зменшити час відповіді запитів, оптимізувати використання пулу та забезпечити стабільне відображення аналітики в Oracle APEX.

Окрім цього, для оптимізації запитів до таблиці було створено індекси на стовпцях SESSION_ID та SUBSESSION_ID. Використання індексів дозволить значно прискорити виконання операцій вибірки та мінімізувати кількість повних сканувань таблиці.

Перша функціональна сторінка – сторінка аналізу переходів та активності користувачів (рис. 3.2), яка призначена для швидкої оцінки активності відвідувачів, виявлення найпопулярніших сторінок, а також розуміння шляхів навігації користувачів.

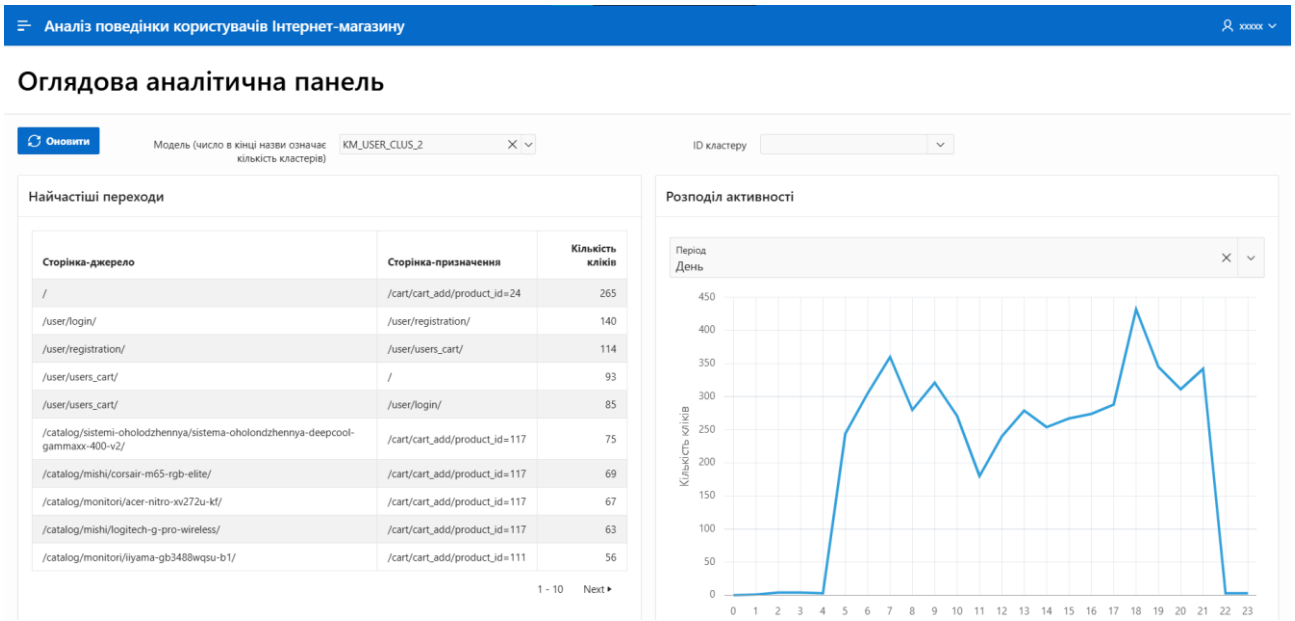


Рис. 3.2. Сторінка «Аналіз переходів та активності користувачів».

Першим компонентом сторінки є таблиця найчастіших переходів, який відображає найчастіші пари «джерело – ціль». Дані для цього блоку надходять із таблиці USER_BEHAVIOR_TABLE, де групуються переходи за парою «сторінка-джерело – сторінка-призначення». Цей компонент дозволяє виявляти найбільш навантажені маршрути навігації, визначати «гарячі точки» інтерфейсу та оцінювати, які сторінки відіграють роль основних точок входу або переходу. Отримані дані можуть сигналізувати як про вдало сконструйовані UX-шляхи, так і про потенційні проблеми, наприклад, циклічні переходи або слабко структуровану архітектуру вебсайту.

Результати топ-10 переходів показують найбільш часті маршрути серед згенерованих сесій: найвища активність припадає на головну сторінку сайту (/) із переходами до сторінок додавання товарів у кошик, а також до сторінок реєстрації та входу. Серед інших поширених маршрутів виділяються переходи між різними товарами та їх додавання до кошика. Це свідчить про те, що скрипт відтворює реалістичні сценарії взаємодії з вебсайтом і реагує на запропоновані елементи навігації та рекомендації товарів.

Другим центральним компонентом є інтерактивний графік розподілу кліків у часових інтервалах. Він реалізований у вигляді лінійного графіка з можливістю вибору інтервалу (день, тиждень, місяць, рік). Цей механізм базується на використанні APEX-параметра P1_PERIOD, який передає вибраний інтервал у SQL-запит, що будує актуальний розподіл. Тут використовується концепція вибору параметрів, що визначають, які саме дані буде завантажено й оброблено.

Вибір періоду реалізовано у вигляді стандартного Select List, який взаємодіє з відповідними SQL-запитами через механізм підстановки значень. Для кожного варіанту періоду – «День», «Тиждень», «Місяць» або «Рік» – налаштовано відповідне перетворення у часовий інтервал у WHERE-умові основних запитів. Завдяки цьому панель може миттєво перебудовувати агреговані таблиці та графічні відображення даних, без повторного завантаження сторінки.

Оновлення компонентів сторінки реалізовано через Dynamic Action, прив’язану до зміни обраного періоду, після чого запускається послідовність: застосування вибраного значення P1_PERIOD – оновлення регіонів даних – візуалізація оновленої інформації. Таким чином, процес оновлення відбувається одразу після визначення параметру.

Дія Refresh використовує стандартний механізм Partial Page Refresh, що забезпечує оновлення лише тих компонентів, які потребують цього, роблячи процес швидким і не перевантажуючи сервер непотрібними запитами. На рисунку 3.3 зображено технічну організацію механізму оновлення, що включає елемент P1_PERIOD у структурі параметрів сторінки і конфігурацію дії Refresh, прив’язану до події Change всередині Dynamic Action.

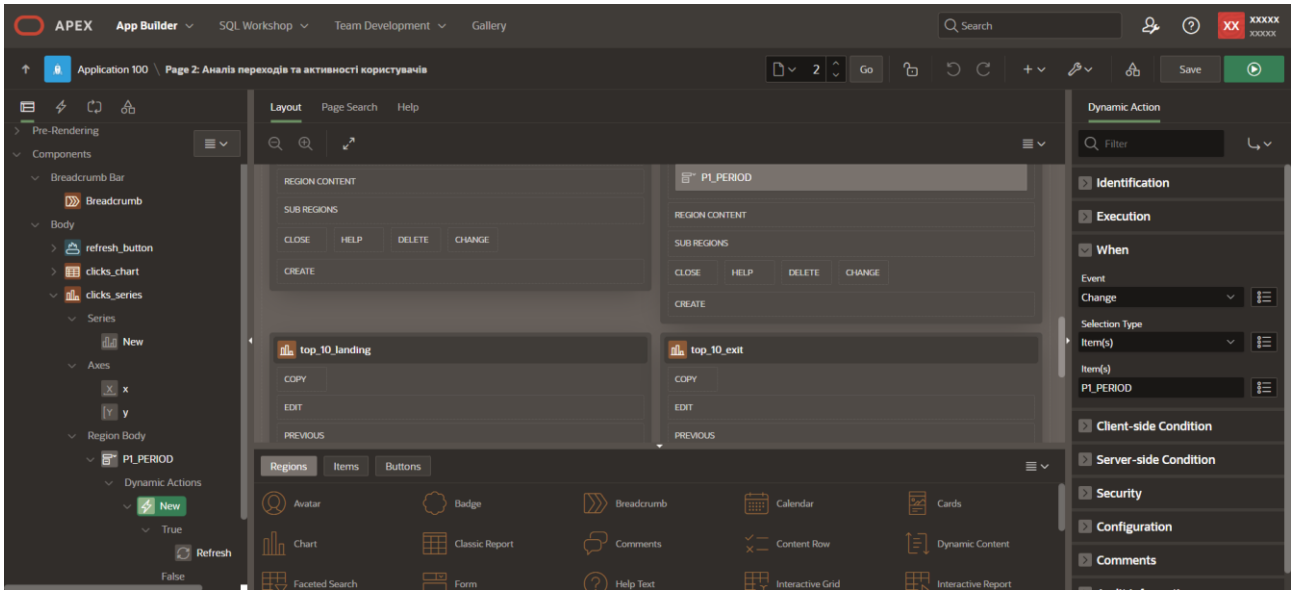


Рис. 3.3. Інтерфейс розробки додатку Oracle APEX

Графік розподілу активності користувачів демонструє зміну кількості кліків у різні часові інтервали. Оскільки дані згенеровані скриптом, ці зміни дозволяють лише умовно оцінити тренди поведінки користувачів і не відображають фактичну взаємодію. Водночас, використаний механізм візуалізації повністю придатний для подальшого аналізу реальних даних.

Наступний аналітичний блок сторінки складається з двох панелей: «топ-10 сторінок входу» та «топ-10 сторінок виходу», які зображені на рисунку 3.4. На відміну від попереднього рейтингу переходів, ці панелі зосереджуються на

межах сесій і дозволяють виявляти закономірності початку та завершення взаємодії. Дані для цього блоку формуються через аналіз мінімальних і максимальних часових міток у межах кожної підсесії, після чого вони агрегуються. Ці рейтинги є важливими для оцінки ефективності контенту, який впливає на перші враження користувача, а також для розуміння того, на яких сторінках найчастіше припиняється взаємодія.

Дані для блоків формуються на основі таблиці USER_BEHAVIOR_TABLE, яка містить агреговані дані про сесії та підсесії користувачів, включаючи поля SESSION_ID, SOURCE_URL та TARGET_URL. SQL-запити, що лежать в основі побудови графіків, використовують функції COUNT і GROUP BY для визначення кількості переходів на кожну сторінку та кількості виходів з неї, після чого відбираються лише перші 10 позицій за рейтингом популярності

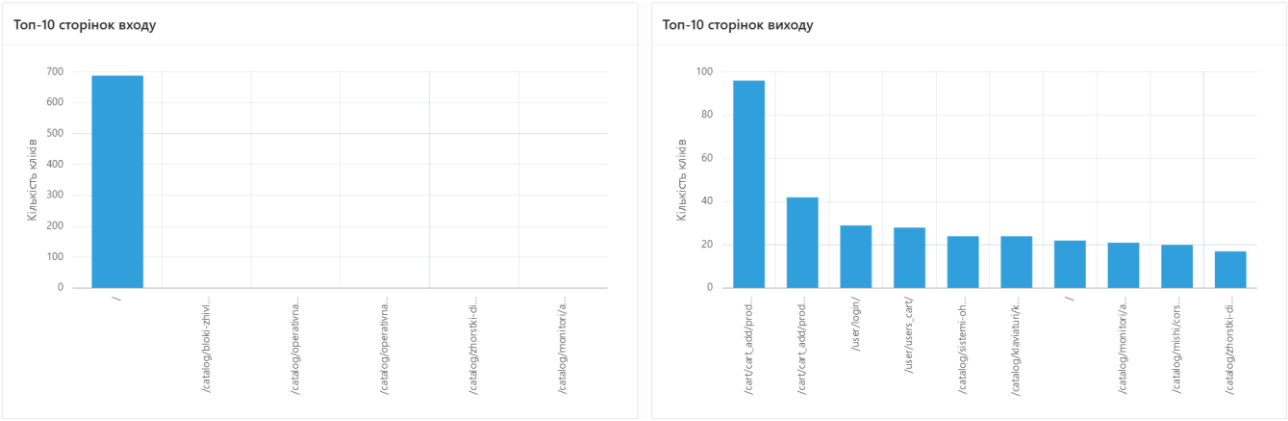


Рис. 3.4. Таблиці з переліком сторінок входу і виходу

Результатом реалізації цього елемента є наочне відображення найбільш відвідуваних сторінок і точок виходу користувачів із сайту. Діаграми та таблиці дозволяють миттєво оцінити, які сторінки залучають відвідувачів, а які сприяють відтоку користувачів.

Слід зазначити, що у рейтингу «топ-10 сторінок входу» відображається лише головна сторінка (/). Це зумовлено особливістю скрипта генерації даних, який ініціює кожну сесію з головної сторінки вебсайту. Відповідно, усі сесії починаються з одного й того ж URL, і в цьому блоці аналітичної панелі не спостерігається різноманіття початкових точок входу користувачів. Натомість

рейтинг «топ-10 сторінок виходу» демонструє більш рівномірний розподіл, причому більшість сесій завершуються додаванням товару до кошику.

Таким чином, сторінка аналізу клікової активності виконує роль аналітичного ядра першого рівня, візуалізуючи ключові патерни взаємодії. Вона поєднує статистичні, графічні та рейтингові індикатори, створюючи повноцінне середовище для вивчення активності користувачів.

Друга функціональна сторінка аналітичної панелі призначена для детального дослідження результатів кластеризації та надання аналітику інструментів для інтерпретації отриманих груп сесій. Її головна мета – показати, як змінюється якість кластеризації при різній кількості кластерів, дати візуальне уявлення про розподіл сесій у просторі ознак, надати можливість швидко порівняти характеристики кожного кластера і перейти від загальної картини до конкретних сесій з високою або низькою ймовірністю належності.

На сторінці відображаються п’ять взаємопов’язаних блоків: графік дисперсії моделей (для вибору оптимальної кількості кластерів), стовпчаста діаграма розподілу кластерів з вибором моделі, таблиця характеристик кластерів з можливістю вибору кластеру, гістограма розподілу обраної ознаки всередині кластеру та таблиця приналежності сесій до кластеру зі ймовірностями. Зовнішній вигляд сторінки та двох перших блоків представлено на рисунку 3.5.

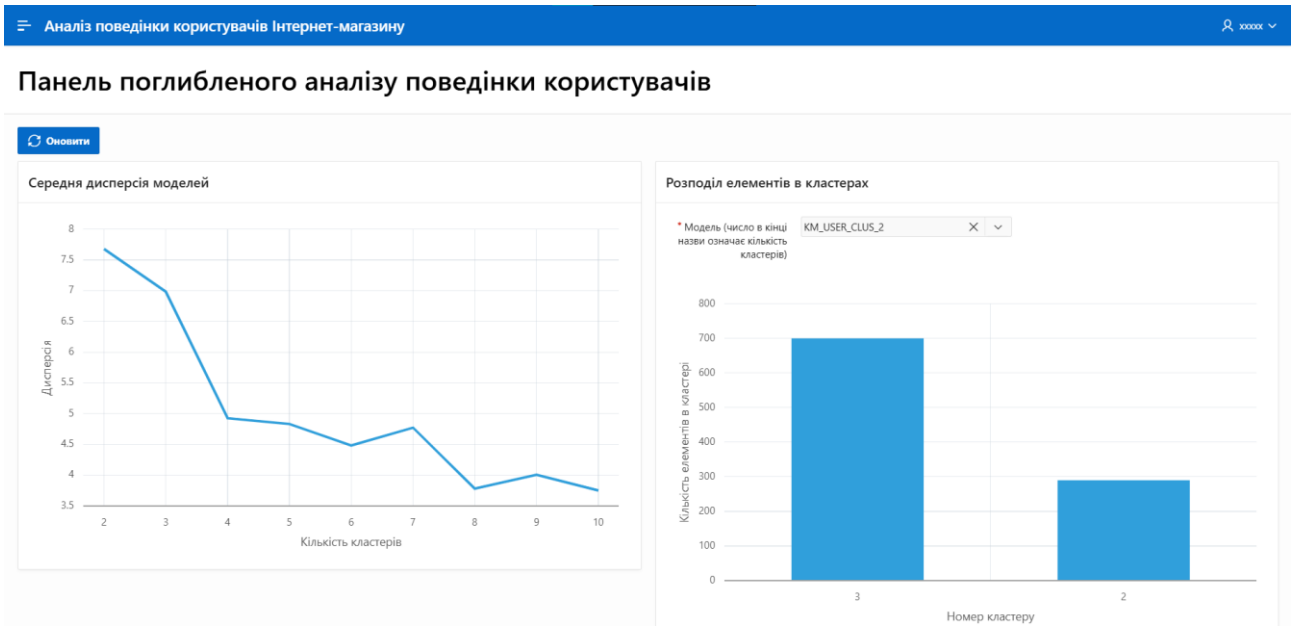


Рис. 3.5. Сторінка результатів кластеризації

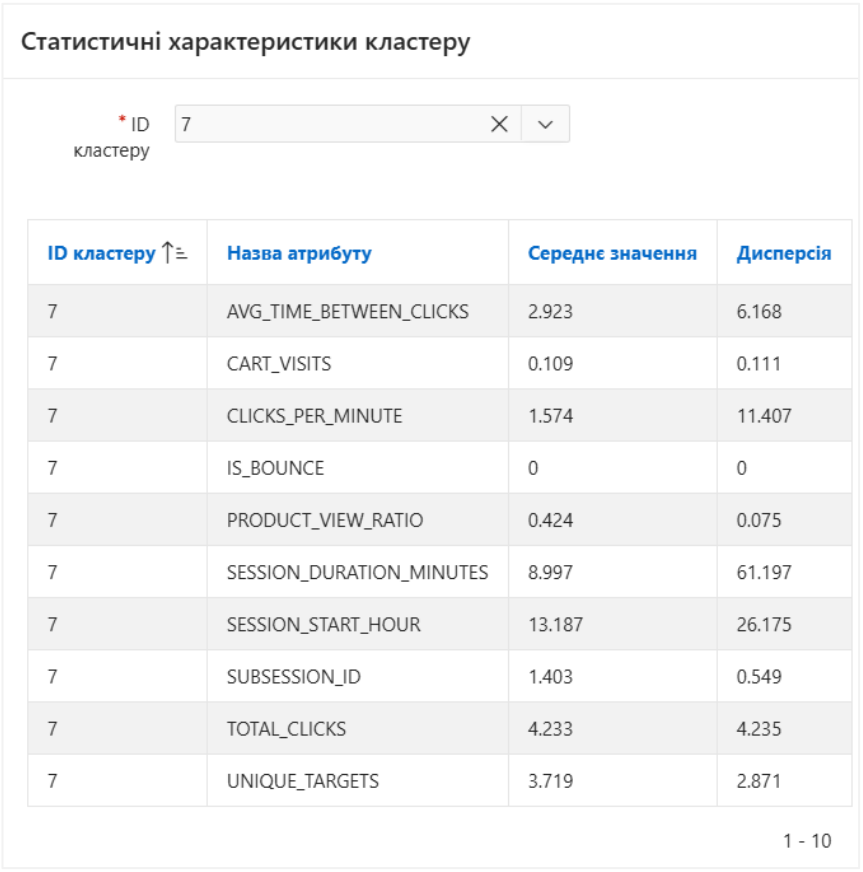
Для побудови графіка дисперсії використовується службове представлення DM$VA<назва_моделі>. На цьому графіку відображено середню дисперсію об’єктів у кожному кластері для моделей із 2-10 кластерами, що дозволяє візуально оцінити оптимальну кількість кластерів. Одним із поширених підходів є метод «ліктя»: на графіку обирається точка, де середня дисперсія зменшується найшвидше, після чого додавання нових кластерів не дає суттєвого покращення[9]. В даному випадку оптимальною кількістю кластерів визначено 4, оскільки зменшення дисперсії при переході від 3 до 4 кластерів є помітним, тоді як подальше збільшення числа кластерів не призводить до значного покращення.

Для відображення даних у всіх інших віджетах сторінки – діаграмі розподілу кластерів, таблиці характеристик та таблиці приналежності сесій до кластерів – необхідно обрати конкретну модель кластеризації. Для демонстрації результатів розглянуто модель KM_USER_CLUS_4, яка містить 4 кластери.

Стовпчаста діаграма розподілу кластерів слугує візуальним інструментом для оцінки того, як саме підсесії розподіляються між кластерами. Кожен стовпець на діаграмі відповідає окремому кластеру, а його висота відображає загальну кількість підсесій, що до нього належать. Додатково у діаграму інтегровано інтерактивні підказки: при наведенні курсору на певний стовпець з’являється точне числове значення кількості підсесій у відповідному кластері.

Для обраної моделі спостерігаються чотири кластери, один із яких (кластер 3) містить лише дві підсесії. Така мала кількість може вказувати на аномальні або нетипові поведінкові патерни, що потребує додаткового аналізу.

Таблиця характеристик кластерів, зображена на рисунку 3.6, формується на основі службового представлення моделі DM$VA<назва_моделі>, що містить статистику по кожному атрибуту: середнє, дисперсію, та модальне значення. Оскільки модальне значення застосовується лише для категоріальних ознак, яких у представленні USER_BEHAVIOR_VIEW немає, то даний атрибут розглядатись не буде.



Статистичні характеристики кластеру

* ID кластеру: 7

ID кластеру	Назва атрибуту	Середнє значення	Дисперсія
7	AVG_TIME_BETWEEN_CLICKS	2.923	6.168
7	CART_VISITS	0.109	0.111
7	CLICKS_PER_MINUTE	1.574	11.407
7	IS_BOUNCE	0	0
7	PRODUCT_VIEW_RATIO	0.424	0.075
7	SESSION_DURATION_MINUTES	8.997	61.197
7	SESSION_START_HOUR	13.187	26.175
7	SUBSESSION_ID	1.403	0.549
7	TOTAL_CLICKS	4.233	4.235
7	UNIQUE_TARGETS	3.719	2.871

1 - 10

Рис. 3.6. Таблиця характеристик кластеру 7 в моделі з 4 кластерами

В інтерфейсі APEX ця таблиця реалізована як інтерактивна сітка з можливістю вибору кластеру. Інші елементи сторінки, зокрема діаграма розподілу ознак і таблиця підсесій, отримують дані саме для обраного кластеру.

Розглядаючи кластер 7, який містить найбільше підсесій, можна відзначити такі особливості: середня тривалість підсесії складає близько 9 хвилин, користувачі виконують у середньому 4 кліки з частотою 1,57 кліків за хвилину, при цьому 42 % кліків припадають на перегляди товарів. Кількість відскоків дорівнює нулю, що свідчить про активну взаємодію. Середній час між кліками близько 2,9 хвилин, середня кількість унікальних цільових сторінок – 3,7. Показники вказують на типовий поведінковий патерн користувачів з помірною активністю та регулярними переглядами товарів.

Гістограма розподілу ознак (рис. 3.7) дає змогу оцінити внутрішню структуру вибраного кластеру по будь-якій з десяти ознак, обраній з випадаючого списку. Для побудови цієї діаграми APEX виконує SQL-запит до представлення DM$VH<назва_моделі>, яке містить інформацію про розподіл

ознак в інтервалах для кожного кластеру моделі. Після вибору кластеру та ознаки зі списку система динамічно формує діаграму, де по осі X відкладено значення інтервалів ознаки, а по осі Y – кількість підсесій, що потрапляють у відповідний інтервал. Це дозволяє оцінити, наскільки однорідний кластер за конкретною ознакою, а також виявити домінуючі значення та потенційні аномалії.

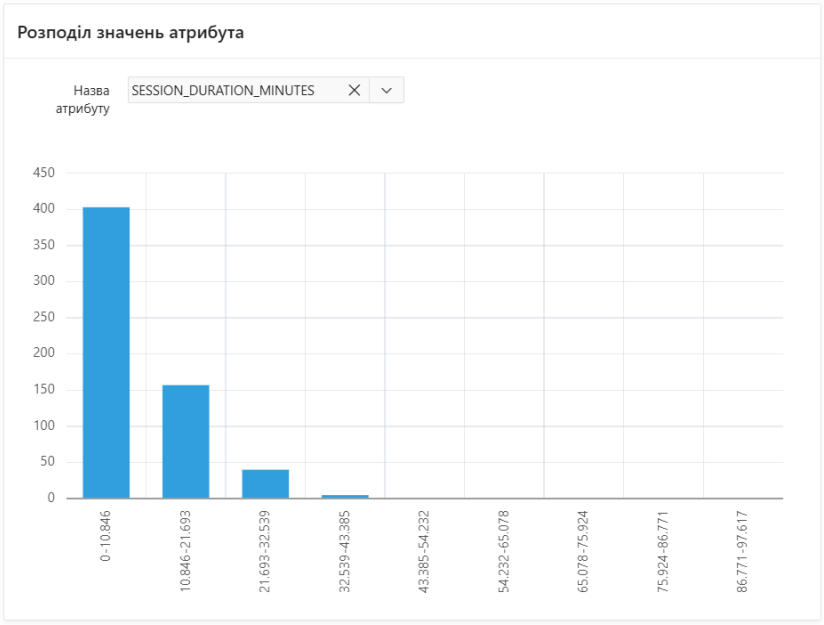


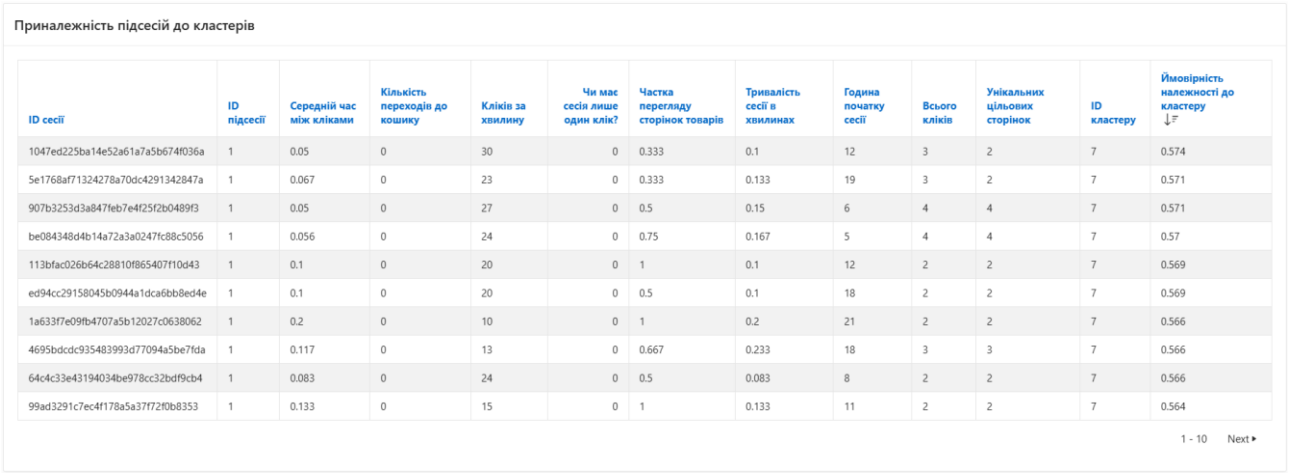
Рис. 3.7. Гістограма розподілу ознак.

Гістограма демонструє, що атрибут SESSION_DURATION_MINUTES має виражену позитивну асиметрію, характерну для тривалості сесій: більшість підсесій зосереджена в діапазоні коротких значень. Основна частка припадає на перший інтервал (0–10,846 хв), де зосереджено близько 400 спостережень, другий інтервал (10,846–21,693 хв) містить приблизно 150–160 випадків. Частота спостережень різко зменшується у наступних інтервалах, і сесії довше 43 хв практично відсутні.

Такий розподіл свідчить про те, що для згенерованих даних характерні короткі сесії, і більшість взаємодій завершується швидко, що відповідає типовим сценаріям активності користувачів.

Таблиця приналежності підсесій до кластеру (рис. 3.8) містить перелік SESSION_ID, значення ключових ознак, CLUSTER_ID і CLUSTER_PROBABILITY. Ця таблиця отримує дані з динамічного запиту до моделі з використанням функцій CLUSTER_ID(<назва_моделі> USING *) та

CLUSTER_PROBABILITY(<назва_моделі> USING *), що відображує належність для нових сесій.



Приналежність підсесій до кластерів

ID сесії	ID підсесії	Середній час між кліками	Кількість переходів до кошику	Кліків за хвилину	Чи має сесія лише один клік?	Частка перегляду сторінок товарів	Тривалість сесії в хвилинах	Година початку сесії	Всього кліків	Унікальних цільових сторінок	ID кластеру	Ймовірність належності до кластеру
1047ed225ba14e52a61a7a5b674f036a	1	0.05	0	30	0	0.333	0.1	12	3	2	7	0.574
5e1768af71324278a70dc4291342847a	1	0.067	0	23	0	0.333	0.133	19	3	2	7	0.571
907b3253d3a847feb7e4f25f2b0489f3	1	0.05	0	27	0	0.5	0.15	6	4	4	7	0.571
be084348d4b14a72a3a0247fc88c5056	1	0.056	0	24	0	0.75	0.167	5	4	4	7	0.57
113bfac026b64c28810f865407f10d43	1	0.1	0	20	0	1	0.1	12	2	2	7	0.569
ed94cc29158045b0944a1dca6bb8ed4e	1	0.1	0	20	0	0.5	0.1	18	2	2	7	0.569
1a633f7e09fb4707a5b12027c0638062	1	0.2	0	10	0	1	0.2	21	2	2	7	0.566
4695bdcdc935483993d77094a5be7fda	1	0.117	0	13	0	0.667	0.233	18	3	3	7	0.566
64c4c33e43194034be978cc32bdf9cb4	1	0.083	0	24	0	0.5	0.083	8	2	2	7	0.566
99ad3291c7ec4f178a5a37f72f0b8353	1	0.133	0	15	0	1	0.133	11	2	2	7	0.564

1 - 10 Next ▸

Рис. 3.8. Таблиця приналежності підсесій до кластеру 7.

На функціональних сторінках реалізовано механізм оновлення: кнопка «Оновити» дає змогу перезапустити генерацію службових представлень та автоматично оновити в APEX усі віджети. Окрім цього, забезпечено узгодження часових зон та форматів (всі дати відображаються з використанням AT TIME ZONE 'Europe/Kyiv'), а також передбачено механізм безпеки: доступ до аналітичної панелі мають лише користувачі з відповідними правами доступу.

Таким чином, друга функціональна сторінка аналітичної панелі надає комплексний інструментарій для інтерпретації кластеризації: від вибору оптимальної кількості кластерів до переходу до конкретних сесій, а технічні рішення щодо оновлення та безпеки гарантують практичну придатність та продуктивність інтерфейсу.

В результаті реалізації аналітичної панелі в Oracle APEX було створено інтерфейс для інтерактивного відображення поведінкових даних користувачів Інтернет-магазину, що включає сторінку з оглядом активності, аналізом топ-переходів та детальним вивченням кластерів. Використання службових представлень, згенерованих моделями кластеризації K-Means, дозволило оцінити внутрішню структуру підсесій, характер розподілу ознак і виявити потенційні аномалії.

Хоча дані були згенеровані скриптом, отримані результати демонструють можливості системи щодо відстеження та аналізу поведінкових патернів: від домінування коротких сесій і основних маршрутів переходів до розподілу підсесій по кластерам. Сформовані графіки та таблиці надають можливості аналізу як тестових, так і реальних даних, що демонструє ефективність інтеграції Oracle Machine Learning та APEX для візуалізації та оцінки поведінкових моделей користувачів.

Отримати доступ до аналітичної панелі можна за посиланням: https://g0679acf523b3ac-txmvsjuktn3shcil.adb.eu-frankfurt-1.oraclecloudapps.com/ords/r/my_wksp/userbehavioranalytics/main-page

Дані для входу:

- ім’я користувача – app_user
- пароль – 5B3hZ2v6GdFK

ВИСНОВКИ

Під час виконання кваліфікаційної роботи було розроблено систему збору та аналізу поведінкових даних користувачів Інтернет-магазину із застосуванням Oracle APEX та Oracle Machine Learning. Створення такої системи є актуальним завданням у зв’язку зі зростанням онлайн-шопінгу та необхідністю бізнесу в отриманні детальної аналітики для підвищення ефективності маркетингових кампаній, оптимізації навігації та персоналізації пропозицій для користувачів. Аналіз поведінкових даних дозволяє своєчасно виявляти аномальні дії, оцінювати активність клієнтів та планувати заходи для покращення взаємодії з сайтом.

В ході роботи було реалізовано механізм збору даних про кліки користувачів за допомогою JavaScript-трекера, який фіксує всі події DOM та передає їх у базу даних через асинхронні запити. Для тестування системи та заповнення бази даних було створено спеціальний скрипт на Python, який імітує поведінку користувачі, генеруючи сесії.

Для подальшого аналізу ці сесії розбиваються на підсесії за допомогою представлення USER_BEHAVIOR_VIEW, яке агрегує підсесії та формує десять ознак для кожної підсесії, таких як кількість кліків за хвилину, середній час між кліками, відсоток переглядів каталогу, година початку сесії, ознака bounce, загальна тривалість сесії, порядковий номер підсесії, загальна кількість кліків, унікальні цільові сторінки та кількість відвідувань кошика. Сукупність цих ознак дозволяє детально охарактеризувати поведінковий профіль кожної підсесії та формує основу для подальшого аналізу в середовищі Oracle Machine Learning.

Використання скрипта дозволило отримати 5007 кліків, що сформували 692 сесії та 990 підсесій, забезпечуючи достатній обсяг даних для подальшого аналітичного опрацювання.

Для аналізу поведінки користувачів в OML було реалізовано процес кластеризації сесій за допомогою алгоритму K-Means. Було створено серію моделей кластеризації з кількістю кластерів від 2 до 10, де кожна модель

формувалась автоматично за допомогою PL/SQL-блоку. Цей блок задає параметри алгоритму, включаючи метрику Евкліда, кількість ітерацій, налаштування попередньої обробки даних, а також автоматично видаляє попередні моделі з тим самим ім’ям. Використання серії моделей дозволяє порівняти результати кластеризації для різної кількості кластерів та обрати оптимальну кількість кластерів на основі метрик якості, таких як середня дисперсія об’єктів у кластерах.

Результати кластеризації інтегровані у середовище Oracle APEX через службові представлення, що формуються моделями. Дані цих представлень відображаються у вигляді інтерактивних таблиць та аналітичних діаграм, що дозволяє оцінювати активність користувачів, розподіл сесій за кластерами, а також поведінкові характеристики різних груп користувачів.

Водночас, відсутність реальної клієнтської бази наклала обмеження на проведення повноцінного аналізу поведінки користувачів у реальних умовах. Це зумовило необхідність генерації тестових даних за допомогою скрипта, що дозволяло моделювати різні сценарії користувацької активності. Хоча такий підхід забезпечує достатню кількість даних для перевірки механізмів збору та аналізу, він не відображає повної варіативності поведінки реальних користувачів, що слід враховувати при інтерпретації результатів.

Також під час розробки було виявлено, що без оптимізації SQL-запитів та представлень значна кількість складних обчислень, виконуваних при кожному зверненні до аналітичних діаграм, призводить до підвищеного навантаження на базу даних і збільшення часу відгуку інтерфейсу APEX. Для покращення продуктивності було використано агреговані таблиці та проведено індексацію стопвців, що дозволяє зменшити час виконання запитів та стабілізувати роботу системи при великій кількості одночасних звернень.

Економічна ефективність розробленої системи полягає в можливості прийняття обґрунтованих управлінських рішень на основі аналізу поведінкових даних без залучення високовартісних сторонніх аналітичних платформ. Використання Oracle APEX та Oracle Machine Learning у межах єдиної

екосистеми дозволяє зменшити витрати на інтеграцію та супровід, оскільки всі етапи – від зберігання даних до їх аналізу та візуалізації – реалізовані в одному середовищі.

Даний підхід дає змогу підвищити ефективність роботи Інтернет-магазину за рахунок виявлення типових поведінкових патернів, оптимізації UX-шляхів та зменшення кількості неефективних або «зайвих» переходів користувачів. Це, в свою чергу, потенційно сприяє зростанню конверсії та скороченню часу прийняття рішення про покупку.

Таким чином, у ході виконання кваліфікаційної роботи було отримано практичні навички збору, обробки та аналізу поведінкових даних користувачів Інтернет-магазину. Розроблена система дозволяє досліджувати поведінкові патерни, виявляти аномалії, оцінювати ефективність функціональних елементів вебсайту та інтегрувати результати аналізу у зручні інтерфейси для аналітиків і менеджменту. Отримані знання та навички можуть бути використані для подальшої професійної діяльності у сфері веброзробки, електронної комерції та аналітики даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Adeniran I. A., Efunniyi C. P., Osundare O. S., Abhulimen A. O. Transforming marketing strategies with data analytics: A study on customer behavior and personalization. *International Journal of Management & Entrepreneurship Research*. 2024. Vol. 6, No 8. P. 41–51. URL: https://doi.org/10.56781/ijsret.2024.4.1.0022 (дата звернення: 16.02.2025).
2. Alhamadi M., Alsayahani H., Clinch S., Vigo M. Behavioural Indicators of Usability in Visual Analytics Dashboards. *ACM Transactions on Interactive Intelligent Systems*. 2025. Vol. 15, No 2. P. 1–35. URL. https://dl.acm.org/doi/abs/10.1145/3715710 (дата звернення: 02.10.2025).
3. Altayeb Y. F. Predicting consumer behavior in online shopping using clickstream data and machine learning algorithms. *Tilburg University*. 2024. URL: http://arno.uvt.nl/show.cgi?fid=170699 (дата звернення: 26.05.2025).
4. Anand B., Chakravarty H., Athalye M. S. G., Varalaxmi P., Mishra A. K. Understanding consumer behaviour in the digital age: a study of online shopping habits. *Shodha Prabha (UGC Care J.)*. 2023. Vol. 48, No 3. P. 84–93. URL: https://www.academia.edu/download/108187784/UNDERSTANDING_CONSUMER_BEHAVIOUR_IN_THE_DIGITAL_AGE.pdf (дата звернення: 16.03.2025).
5. Batool I. A. A deep learning approaches for online shopping behavior prediction using clickstream data. *University Institute of Information Technology, PMAS Arid Agriculture University*. 2021. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4662627 (дата звернення: 23.06.2025).
6. Bharadiya J. P. The role of machine learning in transforming business intelligence. *International Journal of Computing and Artificial Intelligence*. 2023. Vol. 4, No 1. P. 16–24. URL: https://doi.org/10.33545/27076571.2023.V4.I1A.60 (дата звернення: 09.04.2025).

7. Bouali S., Khelia I., Douha S. Optimizing ad targeting and content personalization in e-stores through machine learning algorithms. *International Journal of Economic Perspectives*. 2024. Vol. 18, No 3. P. 66–77. URL: https://ijeponline.org/index.php/journal/article/view/927 (дата звернення: 18.10.2025).
8. CLUSTER_PROBABILITY. URL: https://docs.oracle.com/en/database/oracle/oracle-database/26/sqlrf/CLUSTER_PROBABILITY.html (дата звернення: 10.11.2025).
9. Clustering Use Case Scenario. URL: https://docs.oracle.com/en/database/oracle/machine-learning/oml4sql/23/mlsql/clustering1.html#GUID-CBF62963-4AE6-4CB0-901A-8ACEE9849EBE (дата звернення: 09.11.2025).
10. Configuring Authentication. URL: https://docs.oracle.com/en/database/oracle/oracle-database/21/dbseg/configuring-authentication.html#GUID-2239A690-3F63-4B1D-A09B-94B8411700D5 (дата звернення: 18.03.2025).
11. DBMS_DATA_MINING. URL: https://docs.oracle.com/en/database/oracle/oracle-database/21/arpls/DBMS_DATA_MINING.html#GUID-8987EE6F-41A9-4DF9-997C-129B41FDC59A (дата звернення: 20.09.2025).
12. de Paula A. F. Behavioral analysis and bot detection in e-commerce navigation data. 2025. URL: https://hdl.handle.net/1843/83638 (дата звернення: 11.05.2025).
13. Guo Ge. A Study on User Behavior of E-commerce Platforms based on Data Analysis. *Journal of Global Humanities and Social Sciences*. 2023. Vol. 4, No 4. P. 181–185. URL: https://doi.org/10.61360/BoniGHSS232014180806 (дата звернення: 23.04.2025).
14. Jain A., Kumawat A. ANALYSIS OF CLICKSTREAM DATA. 2022. URL: https://www.academia.edu/download/90685274/IRJET_V9I4566.pdf (дата

звернення: 25.03.2025).

15. Khan T. Conversion Rate Optimization in E-commerce: Implementing ML algorithms to identify clickstream patterns. Doctoral dissertation / Dublin, National College of Ireland, 2025. URL: https://norma.ncirl.ie/id/eprint/7991 (дата звернення: 22.07.2025).
16. Kolluru V., Chintakunta A. N., Nuthakki Y., Koganti S., Mungara S. Exploring Consumer Behaviors In E-Commerce Using Machine Learning. 2023. URL: https://doi.org/10.2139/ssrn.4988848 (дата звернення: 15.03.2025).
17. Lezu J. Z. Analyzing and Clustering Clickstream Data for Marketing Personalization. Doctoral dissertation / Tilburg University, 2021. URL: http://arno.uvt.nl/show.cgi?fid=156283 (дата звернення: 20.05.2025).
18. Miah S. J. Impact of business analytics and decision support systems on e-commerce in SMEs. *arXiv preprint arXiv:2212.00016*. 2022. URL: https://doi.org/10.48550/arXiv.2212.00016 (дата звернення: 24.07.2025).
19. Model Detail Views. URL: https://docs.oracle.com/en/database/oracle/machine-learning/oml4sql/23/dmprg/model-detail-views.html#GUID-98B328E3-3FEB-4E84-B8A9-D4E578C2EFAB (дата звернення: 25.03.2025).
20. Noviantoro T., Huang J. P. Applying data mining techniques to investigate online shopper purchase intention based on clickstream data. *Rev Bus Account Financ*. 2021. Vol. 1, No 2. P. 130–159. URL: https://www.researchgate.net/profile/Tri-Noviantoro/publication/352794957_APPLYING_DATA_MINING_TECHNIQUES_TO_INVESTIGATE_ONLINE_SHOPPER_PURCHASE_INTENTION_BASED_ON_CLICKSTREAM_DATA/links/60d9edb6a6fdccb745ed7fe0/APPLYING-DATA-MINING-TECHNIQUES-TO-INVESTIGATE-ONLINE-SHOPPER-PURCHASE-INTENTION-BASED-ON-CLICKSTREAM-DATA.pdf (дата звернення: 22.01.2025).
21. Pal G., Atkinson K., Li G. Real-time user clickstream behavior analysis based on apache storm streaming. *Electronic Commerce Research*. 2023. Vol. 23, No 3. P. 1829–1859. URL: https://doi.org/10.1007/s10660-021-09518-4 (дата

звернення: 09.09.2025).

22. Patel, S., Mendoza, R., Cruz, I., Kim, N., Singh, P., & Martinez, A. Machine Learning Driven Adaptive Accessibility Interfaces for Enhanced User Experience. 2025. URL: https://doi.org/10.13140/RG.2.2.12545.36965 (дата звернення: 21.04.2025).
23. Pimenov D., Solovyov A., Askarbekuly N., Mazzara M. Data-driven approaches to user interface design: A case study. *Journal of Physics: Conference Series*. 2021. Vol. 2134, No 1. P. 012020. URL: https://doi.org/10.1088/1742-6596/2134/1/012020 (дата звернення: 25.10.2025).
24. Prasad H. K., Kalhapure B., Mohanapriya T., Rella M. K., Devendran A., Thavara S. S. Beyond the Click: AI-Powered UX Strategies Driving Consumer Confidence and Conversion. *Advances in Consumer Research*. 2025. Vol. 2. P. 3787–3793. URL: https://acr-journal.com/article/beyond-the-click-ai-powered-ux-strategies-driving-consumer-confidence-and-conversion-1474/ (дата звернення: 15.10.2025).
25. Raji M. A., Olodo H. B., Oke T. T., Addy W. A., Ofodile O. C., Oyewole A. T. E-commerce and consumer behavior: A review of AI-powered personalization and market trends. *GSC Advanced Research and Reviews*. 2024. Vol. 18, No 3. P. 066–077. URL: https://doi.org/10.30574/gscarr.2024.18.3.0090 (дата звернення: 09.08.2025).
26. Sagar S. The impact of digital transformation on retail management and consumer behavior. *Journal of Business and Management*. 2024. Vol. 26, No 1. P. 06–14. URL: https://www.academia.edu/download/110983275/B2601010614.pdf (дата звернення: 06.08.2025).
27. Sakalauskas V., Kriksciuniene D. Personalized advertising in e-commerce: Using clickstream data to target high-value customers. *Algorithms*. 2024. Vol. 17, No 1. P. 27. URL: https://doi.org/10.3390/a17010027 (дата звернення: 06.01.2025).
28. Sharma R. Data-driven personalization: revolutionizing user experience.

International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2024. Vol. 10, No 5. P. 868–877. URL: https://doi.org/10.32628/CSEIT241051075 (дата звернення: 13.03.2025).

29. Sharma R., Srivastva S., Fatima S. E-commerce and digital transformation: Trends, challenges, and implications. *International Journal of Multidisciplinary Research (IJFMR)*. 2023. Vol. 5. P. 1–9. URL: https://doi.org/10.36948/ijfmr.2023.v05i05.7128 (дата звернення: 01.08.2025).
30. Singgalen Y. A. Customer Experience Analysis for Marketing Strategy Optimization Using CRAF Framework. *Journal of Business and Economics Research (JBE)*. 2024. Vol. 5, No 2. URL: https://doi.org/10.47065/jbe.v5i2.5314 (дата звернення: 25.03.2025).
31. Syed A. Evaluating Oracle APEX with Competing Low-Code Platforms for Enterprise Development. *IJLRP – International Journal of Leading Research Publication*. 2025. Vol. 6, No 8. URL: https://doi.org/10.70528/IJLRP.v6.i8.1689 (дата звернення: 26.09.2025).
32. Varshitha R., Aparna T., Shravani P., Reddy A. V., Reddy V. S. N. A Multi-Perspective Fraud Detection Method for Multi-Participant E-Commerce Transactions. *Fringe Multi-Engineering Proceedings (FMEP)*. 2025. Vol. 1, No 3. P. 39–52. URL: https://doi.org/10.69996/fmep.2025017 (дата звернення: 28.10.2025).
33. Waldmann F. Prediction of Cart Abandonment Using Clickstream Data in Online Shopping. Doctoral dissertation / Tilburg University, 2024. URL: https://doi.org/10.1007/s10660-025-10021-3 (дата звернення: 18.10.2025).
34. Wen Z., Lin W., Liu H. Machine-learning-based approach for anonymous online customer purchase intentions using clickstream data. *Systems*. 2023. Vol. 11, No 5. P. 255. URL: https://doi.org/10.3390/systems11050255 (дата звернення: 20.04.2025).
35. Yong S. C. S. C., Huan R. T., Poh W. S., Osman M., Ng D. C. W. Assessing the Factors Influencing Consumer Behaviour in E-Commerce Platforms.

International Journal of Multidisciplinary: Applied Business and Education Research. 2023. Vol. 4, No 10. P. 3725–3735. URL: https://doi.org/10.11594/ijmaber.04.10.25 (дата звернення: 28.07.2025).

36. Yu J., Zhao J., Zhou C., Ren Y. Strategic business mode choices for e-commerce platforms under brand competition. *Journal of Theoretical and Applied Electronic Commerce Research*. 2022. Vol. 17, No 4. P. 1769–1790. URL: https://doi.org/10.3390/jtaer17040089 (дата звернення: 06.10.2025).
37. Yue N. Identify potential loyalists in shopping festival: repeat buyer prediction for e-commerce based on feature engineering and ensemble learning. Doctoral dissertation / MS thesis, Erasmus School of Economics, Erasmus University Rotterdam, 2024. URL: http://hdl.handle.net/2105/72505 (дата звернення: 13.06.2025).
38. Zhai J., Le F. Research on Cross-National E-commerce User Behavior Analysis and Conversion Rate Improvement Based on the Improved XLSTM Algorithm. *Applied Mathematics and Nonlinear Sciences*. 2025. Vol. 10, No 1. URL: https://doi.org/10.2478/amns-2025-0828 (дата звернення: 28.05.2025).
39. Небога С.В. Інтеграція модулів збору та аналізу даних у веборієнтованих інформаційних системах з використанням Oracle APEX. *Інформаційні технології і автоматизація – 2025* : матеріали XVIII міжнародної науково-практичної конференції. Одеса, 30-31 жовтня 2025 р. ОНТУ, 2025. С. 778–780. URL: https://www.ontu.edu.ua/download/konfi/2025/Collection-of-abstracts-of-the-conference-ITIA-2025.pdf (дата звернення: 21.11.2025).
40. Небога С.В. Інтелектуальний аналіз поведінки користувачів як чинник підвищення ефективності електронної комерції. *Економіка і бізнес: нові реалії, технологічні тренди, цифрові інновації* : збірник матеріалів Міжнародної науково-практичної конференції здобувачів вищої освіти та молодих вчених, Львів-Ірпінь, 15 травня 2025 р. СПДФО Омельченко В. Г., 2025. С. 207–209. URL: https://financial.lnu.edu.ua/wp-content/uploads/2025/06/Zbirnyk-konferentsii-15-travnia-2025.pdf (дата звернення: 10.07.2025).

ДОДАТКИ

Додаток А

```
import requests
from bs4 import BeautifulSoup
import random
import oracledb
import time
from datetime import datetime, timedelta
import uuid
from urllib.parse import urljoin
from tqdm import tqdm


def get_random_link(url, click_probability=1.0, visited=[], session_data=[], session_id=None):
    try:
        response = requests.get(url, timeout=5)
        response.raise_for_status()
    except Exception as e:
        print(f"Помилка: {e}")
        return

    soup = BeautifulSoup(response.text, "html.parser")

    links = []
    # збір активних посилань, пропускаючи "#" та футер
    if soup.select("a[href]"):
        for a in soup.select("a[href]"):
            if a.get("href") != "#":
                if a.find_parent("footer"): continue # пропускати посилання у футері
                link = urljoin(url, str(a.get("href")))
                # для сторінок кошика додається product_id
                if "/cart/" in str(a.get("href")):
                    link += "product_id=" + str(a.get("data-product-id"))
            links.append(link)
        # виключення поточної сторінки зі списку
        links = [link for link in links if link != url]
    else:
        return

    # додавання поточного URL до списку відвіданих
    visited.append(url)

    # випадковий вибір наступного URL
    next_url = random.choice(links)

    time.sleep(0.01)
```

Продовження додатка А

```
    # імовірність здійснення кліку
    if random.random() < click_probability:
        # генерація часу кліку
        if session_data:
            timestamp = datetime.strptime(session_data[-1]["timestamp"], "%Y-%m-%d %H:%M:%S")
        else:
            # початковий час
            timestamp = datetime.strptime("2025-01-01 12:00:00", "%Y-%m-%d %H:%M:%S") \
                    + timedelta(days=random.randint(0, 200)) \
                    + timedelta(hours=random.randint(0, 23)) \
                    + timedelta(minutes=random.randint(0, 59)) \
                    + timedelta(seconds=random.randint(0, 59))
        # додаткові випадкові затримки
        if random.random() < 0.1:
            timestamp += timedelta(hours=random.randint(0, 23))
        if random.random() < 0.5:
            timestamp += timedelta(minutes=random.randint(0, 10))
        if random.random() < 0.7:
            timestamp += timedelta(seconds=random.randint(0, 59))

        # перевірка на нічний час (00:00–06:00)
        if 0 <= timestamp.hour <= 6:
            if random.random() < 0.95:
                return

        # форматування URL для зручності (короткий шлях)
        url_to_print = url.replace("http://127.0.0.1:8000/", "/") if
url.startswith("http://127.0.0.1:8000/") else url
        next_url_to_print = next_url.replace("http://127.0.0.1:8000/", "/") if
next_url.startswith("http://127.0.0.1:8000/") else next_url

        # додавання запису про клік у session_data
        session_data.append({
            "timestamp": datetime.strftime(timestamp, '%Y-%m-%d %H:%M:%S'),
            "source_url": url_to_print,
            "target_url": next_url_to_print,
            "session_id": session_id
        })

        # рекурсивний перехід до наступного URL
        get_random_link(next_url if '/cart/' not in next_url else url, random.uniform(0.85, 0.95),
visited, session_data, session_id)

    return session_data


def send_click(settings, data, max_retries=5, delay=10):
```

Продовження додатка А

```
    attempt = 0
    while attempt < max_retries:
        try:
            with oracledb.connect(**settings) as connection:
                with connection.cursor() as cursor:
                    cursor.execute(
                        """
                        INSERT INTO CLICKS (CLICK_TIMESTAMP, SOURCE_URL, TARGET_URL,
SESSION_ID)
                        VALUES (
                            TO_TIMESTAMP(:timestamp, 'YYYY-MM-DD HH24:MI:SS'),
                            :source_url,
                            :target_url,
                            :session_id
                        )
                        """,
                        data
                    )
                    connection.commit()
            return
        except oracledb.Error as error:
            attempt += 1
            print(f"Error connecting to Oracle (спроба {attempt}/{max_retries}):", error)
            if attempt < max_retries:
                time.sleep(delay)
            else:
                print("Максимальна кількість спроб вичерпана. Дані не надіслано.")


if __name__ == "__main__":
    settings = {
        "user": "MY_SCHEMA",
        "password": "xxxxx",
        "dsn": "xxxxx_high",
        "config_dir": "xxxxx\\wallet_folder",
        "wallet_location": "xxxxx\\wallet_folder",
        "wallet_password": "xxxxx",
    }

    SITE_URL = "http://127.0.0.1:8000/"
    try:
        response = requests.get(SITE_URL)
        if not response.ok:
            raise requests.exceptions.RequestException
        else:
            for i in tqdm(range(200)):
                click_probability = random.uniform(0.85, 0.95)
                session_id = str(uuid.uuid4().hex)
```

Продовження додатка А

```
            session_data = get_random_link(SITE_URL, click_probability=click_probability,
visited=[], session_data=[], session_id=session_id)
            if session_data:
                for click_data in session_data:
                    send_click(settings, click_data)
    except requests.exceptions.RequestException:
        print("Site is not available")
    input('Press Enter...')
```

Додаток Б

```
-- Створення представлення USER_BEHAVIOR_VIEW із поділом сесій на підсесії (subsessions),
-- де нова підсесія починається після паузи >30 хвилин між кліками.
-- Для кожної підсесії обчислюються: загальна кількість кліків, кількість унікальних джерел та цільових сторінок,
-- тривалість сесії у хвилинах і ознака відскоку (bounce), якщо сесія містить лише один клік.
CREATE OR REPLACE VIEW USER_BEHAVIOR_VIEW AS
WITH subsessions AS (
   SELECT
      SESSION_ID,
      CLICK_TIMESTAMP,
      SOURCE_URL,
      TARGET_URL,
      SUM(new_session_flag)
         OVER (PARTITION BY SESSION_ID ORDER BY CLICK_TIMESTAMP) AS subsession_id
   FROM (
      SELECT
         SESSION_ID,
         CLICK_TIMESTAMP,
         SOURCE_URL,
         TARGET_URL,
         CASE
            WHEN LAG(CLICK_TIMESTAMP) OVER (PARTITION BY SESSION_ID ORDER BY CLICK_TIMESTAMP) IS NULL THEN 1
            WHEN (CAST(CLICK_TIMESTAMP AS DATE) - CAST(LAG(CLICK_TIMESTAMP) OVER (PARTITION BY SESSION_ID ORDER BY CLICK_TIMESTAMP) AS DATE)) * 24*60 > 30 THEN 1
            ELSE 0
         END AS new_session_flag
      FROM MY_SCHEMA.CLICKS
   )
)
SELECT
   SESSION_ID,
   subsession_id,
   COUNT(*) AS total_clicks,
   COUNT(DISTINCT SOURCE_URL) AS unique_sources,
   COUNT(DISTINCT TARGET_URL) AS unique_targets,
   ROUND(((CAST(MAX(CLICK_TIMESTAMP) AS DATE) - CAST(MIN(CLICK_TIMESTAMP) AS DATE)) * 24 * 60), 2) AS session_duration_minutes,
   CASE WHEN COUNT(*) = 1 THEN 1 ELSE 0 END AS bounce
FROM subsessions
GROUP BY SESSION_ID, subsession_id;
```

Додаток В

```
-- Блок PL/SQL виконує автоматизоване створення серії моделей кластеризації K-Means
-- у діапазоні від 2 до 10 кластерів. Для кожної ітерації задаються параметри алгоритму,
-- попередня модель з таким ім'ям видаляється (якщо існувала), після чого створюється нова
-- модель на основі агрегованих сесій у поданні USER_BEHAVIOR_VIEW.
%script
DECLARE
   v_setlist       DBMS_DATA_MINING.SETTING_LIST;
   v_model_name    VARCHAR2(50);
BEGIN
   FOR k IN 2..10 LOOP  -- від 2 до 10 кластерів
      v_model_name := 'KM_USER_CLUS_' || TO_CHAR(k);

      BEGIN
         DBMS_DATA_MINING.DROP_MODEL(v_model_name);
      EXCEPTION WHEN OTHERS THEN NULL;
      END;

      v_setlist('ALGO_NAME')        := 'ALGO_KMEANS';
      v_setlist('PREP_AUTO')        := 'ON';
      v_setlist('KMNS_DISTANCE')    := 'KMNS_EUCLIDEAN';
      v_setlist('KMNS_DETAILS')     := 'KMNS_DETAILS_ALL';
      v_setlist('KMNS_ITERATIONS')  := '50';
      v_setlist('KMNS_NUM_BINS')    := '10';
      v_setlist('CLUS_NUM_CLUSTERS'):= TO_CHAR(k);

      DBMS_DATA_MINING.CREATE_MODEL2(
         MODEL_NAME          => v_model_name,
         MINING_FUNCTION     => 'CLUSTERING',
         DATA_QUERY          => 'select * from USER_BEHAVIOR_VIEW',
         SET_LIST            => v_setlist,
         CASE_ID_COLUMN_NAME => 'SESSION_ID');
   END LOOP;
END;
/
```

Додаток Г



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
МІНІСТЕРСТВО ФІНАНСІВ УКРАЇНИ

Сертифікат

Сергій НЕБОГА

взяв(ла) участь у роботі

І Міжнародної науково-практичної конференції здобувачів вищої освіти та молодих вчених
«ЕКОНОМІКА І БІЗНЕС: НОВІ РЕАЛІЇ, ТЕХНОЛОГІЧНІ ТРЕНДИ, ЦИФРОВІ ІННОВАЦІЇ»
ECONOMY AND BUSINESS: NEW REALITIES, TECHNOLOGICAL TRENDS, DIGITAL INNOVATIONS

Андрій ГУКАЛЮК
Перший проректор

15 травня 2025 р.

Олександра СМІРНОВА
Проректор з наукової роботи