НАЦІОНАЛЬНИЙ УНІВЕРСТИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Факультет інформаційних технологій

ПОГОДЖЕНО
Декан факультету інформаційних технологій

_____ Ігор Болбот
(підпис) (ім'я та прізвище)

"___"_____2025 р.

ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ
Завідувач кафедри комп'ютерних наук

_____ Белла Голуб
(підпис) (ім'я та прізвище)

"__"_____2025 р.

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему Інтелектуальна система аналізу поведінки користувачів на комерційних платформах.

Спеціальність 122 «Комп'ютерні науки»
(Код і найменування)

Освітня програма Інформаційні управляючі системи та технології
(Назва)

Орієнтація освітньої програми Освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Гарант освітньої програми

к.т.н., доцент _____ Белла Голуб
(науковий ступінь та вчене звання) (підпис) (ім'я та прізвище)

Керівник магістерської кваліфікаційної роботи

к.т.н., доцент _____ Белла Голуб
(науковий ступінь та вчене звання) (підпис) (ім'я та прізвище)

Виконала _____ Марія Саяпіна
(підпис) (ім'я та прізвище здобувача)

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Факультет інформаційних технологій

ЗАТВЕРДЖУЮ
Завідувач кафедри
комп'ютерних наук
(назва кафедри)
доцент, к.т.н. _____ Белла Голуб
(науковий ступінь, вчене звання) (підпис) (ім'я та прізвище)

«01» Листопада 2024р.

З А В Д А Н Н Я

ДО ВИКОНАННЯ МАГІСТЕРСЬКОЇ КВАЛІФІКАЦІЙНОЇ РОБОТИ СТУДЕНТУ

Саяпіній Марії Сергіївні
(прізвище, ім'я, по батькові)

Спеціальність 122 «Комп'ютерні науки»
(код і назва)
Освітня програма Інформаційні управляючі системи та технології
(назва)
Орієнтація освітньої програми освітньо-професійна

Тема магістерської кваліфікаційної роботи Інтелектуальна система аналізу поведінки користувачів на комерційних платформах.
затверджена наказом ректора НУБіП України від "01" листопада 2024р. №1964 «С»
Термін подання завершеної роботи на кафедру 01 грудня 2025 р.
(число, місяць, рік)
Вихідні дані до магістерської кваліфікаційної роботи: набори даних відкритого доступу з комерційних онлайн-платформ, а також інформація про відгуки та покупки з платформи Amazon.
Перелік питань, що підлягають дослідженню:

1. Аналіз ключових показників поведінки користувачів на комерційних онлайн-платформах.
2. Дослідження можливостей застосування OLAP та Data Mining методів для підвищення ефективності електронної комерції.
3. Визначення факторів, що впливають на активність користувачів.
4. Виявлення закономірностей та сегментація користувачів за поведінковими характеристиками.

Перелік графічного матеріалу (за потреби)
Дата видачі завдання "01" листопада 2024 р.

Керівник магістерської кваліфікаційної роботи _____ Белла Голуб
(підпис) (ім'я та прізвище)

Завдання прийняла до виконання _____ Марія Саяпіна
(підпис) (ім'я та прізвище студента)

Календарний план

№ з/п	Назва етапів виконання магістерської кваліфікаційної роботи	Строк виконання етапів магістерської кваліфікаційної роботи	Примітка
1	Видача завдання	01.11.2024	
2	Аналіз предметної області	02.11-24.11.2024	
3	Проєктування системи	25.11-31.12.2024	
4	Розробка системи	01.01-30.04.2025	
5	Аналіз результатів	01.05-31.07.2025	
6	Оформлення записки	01.08-10.11.2025	
7	Оформлення постеру	05.10-18.10.2025	
8	Написання тез до постеру	19.10-27.10.2025	
9	Постерна сесія	28.10-29.10.2025	
10	Перевірка на плагіат	13.11.2025	
11	Попередній захист	01.12.2025	
12	Захист	16.12.2025	

Студент _____ Марія Саяпіна
(підпис) (ім'я та прізвище)

Керівник магістерської кваліфікаційної роботи _____ Белла Голуб
(підпис) (ім'я та прізвище)

РЕФЕРАТ

Робота присвячена створенню інтелектуальної системи для виявлення поведінкових закономірностей користувачів на комерційних онлайн-платформах з метою підвищення якості аналітики та підтримки управлінських рішень.

Об'єкт дослідження: поведінка користувачів на комерційних онлайн-платформах.

Предмет дослідження: інтелектуальна система аналізу поведінки користувачів на комерційних онлайн-платформах для виявлення закономірностей у взаємодії з товарами й сервісами.

Використані методи: сховище даних у зірковій схемі з ETL-процесами (SQL/SSIS), OLAP-модель та підходи Data Mining: 1-Rule, Наївний Баєс, Apriori (асоціативні правила), K-Means (кластеризація з PCA).

Мета роботи – розробити рішення, яке дозволяє проводити глибокий аналіз взаємодії користувачів із товарами за допомогою OLAP-структур і алгоритмів Data Mining.

Наукова складова полягає у тому, що у ході дослідження було запропоновано інтеграцію OLAP-аналізу та Data Mining, що забезпечує отримання чітких правил і сегментів користувачів.

Рекомендації щодо впровадження результатів: результати можуть бути застосовані для персоналізації, виявлення товарів із низькою конверсією, підвищення продажів і вдосконалення клієнтського досвіду.

Прикладна значимість роботи: запропонована система забезпечує гнучкі інструменти для сегментації аудиторії, моніторингу KPI та ухвалення обґрунтованих маркетингових рішень у сфері електронної комерції.

Кількість сторінок – 63.

Кількість ілюстрацій – 41.

Кількість таблиць – 1.

Кількість додатків – 3.

Кількість джерел – 26.

ABSTRACT

The thesis is dedicated to the development of an intelligent system for identifying behavioral patterns of users on commercial online platforms, aiming to improve the quality of analytics and support managerial decision-making.

Object of research: user behavior on commercial online platforms.

Subject of research: an intelligent system for analyzing user behavior on e-commerce platforms to detect patterns in interactions with products and services.

Methods used: data warehouse in a star schema with ETL processes (SQL/SSIS), OLAP model, and data mining approaches: 1-Rule, Naive Bayes, Apriori (associative rules), K-Means (clustering with PCA).

Purpose of the work is to develop a solution that allows for in-depth analysis of user interaction with products using OLAP structures and Data Mining algorithms.

The scientific component consists in the fact that during the research, the integration of OLAP analysis and Data Mining was proposed, which provides clear rules and user segments.

Recommendations for implementing the results: results can be used for personalization, identifying low-conversion products, increasing sales and improving customer experience.

Applied significance of the work: the proposed system provides flexible tools for audience segmentation, KPI monitoring, and informed marketing decisions in the field of e-commerce.

Number of pages – 63.

Number of illustrations – 41.

Number of tables – 1.

Number of appendices – 3.

Number of sources – 26.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

СД – сховище даних.

БД – база даних.

ШІ – штучний інтелект.

1R – One Rule алгоритм.

СУБД – система управління базами даних.

OLAP – On-Line Analytical Processing.

SQL – Structured Query Language.

SSAS – SQL Server Analysis Services.

BI - Business intelligence.

SSIS – SQL Server Integration Services.

SSRS – SQL Server Reporting Services.

DWH – Data Warehouse, сховище даних.

PCA – Principal Component Analysis (зниження розмірності).

ETL/ELT – Extract-Transform-Load.

KPI – Key Performance Indicators (ключові показники ефективності).

Apriori – алгоритм пошуку асоціативних правил.

K-Means – кластеризація k-середніх.

MLXTend – бібліотека Python для Data Mining та машинного навчання.

Scikit-learn – бібліотека Python для реалізації алгоритмів машинного навчання.

LMS (Learning Management System) – система керування навчанням.

WinForms – технологія створення GUI-додатків у C#.

UserBehaviorDW – аналітичне сховище поведінки користувачів.

MetricValue – числовий показник активності користувача.

ActivityLevel – рівень активності користувача (Low, Medium, High).

ВСТУП

Актуальність. Онлайн-комерція стрімко зростає, а конкуренція за увагу користувача при цьому лише посилюється. Успіх платформ дедалі більше залежить від здатності розуміти реальну поведінку їх аудиторії: що саме переглядають, у який момент приймають рішення, які фактори впливають на повернення покупок та лояльність. Зі збільшенням обсягів даних критично важливо мати рішення, що перетворюють розрізнену інформацію на зрозумілі висновки для щоденного управління.

Практична потреба полягає у своєчасній оцінці динаміки попиту, виявленні активних і ризикових сегментів, визначенні популярних товарів і сезонних коливань, а також у відстеженні показників, пов'язаних із конверсією, відмовами/поверненнями та утриманням клієнтів. Такий підхід дає змогу персоналізувати взаємодію з користувачем, оптимізувати товарну базу й підвищувати результативність маркетингових рішень.

Об'єктом дослідження виступає поведінка користувачів на комерційних платформах.

Предметом дослідження виступає інтелектуальна система аналізу, яка дозволяє ідентифікувати ключові показники ефективності товарів та їх категорій, а також аналізувати дії користувачів.

Мета роботи – створити прикладне рішення для аналізу поведінки користувачів онлайн-комерції з виділенням інтерпретованих правил і сегментів, корисних для управління показниками ефективності.

Методи та засоби, котрі були використані: OLAP-аналіз, методи інтелектуального аналізу даних – 1-Rule, Наївний Баєс, асоціативні правила, кластеризація, SQL і SSIS для ETL та підготовки даних, SSAS для побудови OLAP-моделі, засоби звітності (Reporting Services та Power BI) для візуалізації. Джерелом для даного дослідженням став синтетично згенерований набір даних, сформований на основі попереднього планування, аналізу, та частково

реальних датасетів (зокрема платформи «Amazon»), організований у зірковій схемі «факт/виміри».

Наукова складова полягає у формуванні цілісної моделі аналізу поведінкових даних користувачів, яка узгоджує багатовимірні показники з результатами Data Mining, забезпечуючи однозначне трактування виявлених зразків поведінки і сегментів користувачів для подальшого можливого впровадження для реальних комерційних платформ.

Апробація:

1. М.С. Саяпіна, Б.Л. Голуб: Інтелектуальна система аналізу поведінки користувачів на комерційних платформах. Програма конференції VII Всеукраїнської науково-практичної інтернет конференції студентів і аспірантів “Теоретичні та прикладні аспекти розробки комп'ютерних систем 2025”. 24 квітня 2025 року, НУБіП України, Київ – С. 62.

2. М.С. Саяпіна, Б.Л. Голуб: Застосування OLAP-технологій у поведінковому аналізі користувачів на комерційних платформах. Збірник матеріалів II Міжнародно науково-практичної конференції «АКТУАЛЬНІ ПИТАННЯ РОЗВИТКУ НАУКИ ТА ТЕХНІКИ В УМОВАХ ГЛОБАЛІЗАЦІЇ», 14.05.2025. НУБіП України, ВСП «Боярський фаховий коледж НУБіП України», Боярка.С. 140-142.

3. XVI Міжнародна науково-практична конференція молодих вчених «Інформаційні технології: економіка, техніка, освіта» (м. Київ, 2025 р.).

Структура роботи складається з чотирьох розділів. Перший розділ подає постановку задачі, огляд сучасних рішень і контекст аналізу поведінки користувачів; другий – містить моделювання та проєктування системи (зіркова схема даних, сутності, діаграми процесів і сценарії); третій – описує основні методи й засоби, а також надає опис надходження джерела даних; четвертий розділ присвячено узагальненню результатів, інтерпретацію отриманих правил і сегментів користувачів та демонструє висновки.

Робота містить 63 сторінки, 41 малюнок та 26 джерел.

1 АНАЛІЗ ПОВЕДІНКИ КОРИСТУВАЧІВ НА КОМЕРЦІЙНИХ ПЛАТФОРМАХ

1.1 Ключові процеси та явища

У сучасному цифровому світі комерційні онлайн-платформи відіграють дедалі важливішу роль у повсякденному житті людей. Від вибору товарів до формування звичок споживання – користувачі взаємодіють із цифровими сервісами щодня, залишаючи за собою численні цифрові сліди. Дані сліди є цінним джерелом інформації, що дозволяє не лише зрозуміти, як саме користувачі поводяться на платформі, а й передбачати їхні потреби, виявляти проблемні зони, оптимізувати процеси взаємодії та приймати стратегічно важливі рішення на основі даних.

Поведінка користувача в онлайн-середовищі – це складне явище, що формується під впливом дуже багатьох різноманітних факторів: емоційного стану, попереднього досвіду, довіри до платформи, дизайну інтерфейсу, рекомендацій та соціального контексту, тощо. Поведінковий слід користувача включає такі дії, як пошук товарів, перегляд категорій, додавання в кошик, покидання кошика, написання/читання відгуків, порівняння характеристик, здійснення або скасування покупки, оцінювання продукту після отримання і т.д.

Дослідження ж цих дій допомагає компаніям краще розуміти свого клієнта, адаптувати свої сервіси до очікувань цільової аудиторії та відповідати сучасним вимогам персоналізації. Саме поведінкові дані виступають основою для аналізу конверсій, виявлення вузьких місць у клієнтському шляху, а також для розробки моделей передбачення попиту.

Типові сценарії поведінки на комерційних онлайн-платформах включають кілька ключових етапів:

- пошук і ознайомлення з товарами: на цьому етапі користувачі найчастіше користуються внутрішнім пошуком, переходять у категорії, відкривають фільтри та читають заголовки;
- ознайомлення з товарами: сюди входить вивчення опису, технічних характеристик, перегляд фотографій чи відеооглядів;
- читання відгуків та рейтингових оцінок: відгуки попередніх покупців є критично важливими для ухвалення рішення;
- додавання до кошика/обране: це може свідчити як про зацікавленість, так і про спробу «відкласти» покупку на потім;
- придбання або покидання кошика: це критичний момент, від якого залежить кінцева конверсія;
- поведінка після покупки: включає додавання відгуку, здійснення повторних покупок, звернення до служби підтримки або відмова від товару.

Кожна з цих дій залишає цифровий слід, який за умови правильного збору та обробки може бути використаний для подальшого вдосконалення сервісу.

Значення аналізу поведінки для бізнесу. Інтернет-компанії, що займаються торгівлею або послугами, стикаються з високою конкуренцією, тож розуміння користувача є не просто перевагою, а умова виживання на ринку.

Аналітика споживчої поведінки дозволяє:

- покращити користувацький досвід (UX) за рахунок виявлення точок напруги або нерозуміння;
- персоналізувати інтерфейс, рекламні кампанії та рекомендаційні системи;
- оптимізувати структуру веб-сайту або застосунку;
- виявляти «вузькі місця» у процесі замовлення, які заважають завершенню покупки;

- оцінювати ефективність маркетингових кампаній на основі реакцій користувачів.

Наприклад, якщо дані показують, що значна частина користувачів додає товар у кошик, але не завершує покупку, це сигнал для перевірки цінової політики, доступності варіантів оплати або простоти оформлення замовлення.

Сучасні виклики та тренди. З кожним роком змінюються не лише патерни взаємодії з продуктами, а й самі умови функціонування цифрових платформ. Користувачі стають більш вимогливими, чекаючи блискавичного завантаження сторінок, релевантних рекомендацій, безперебійної роботи мобільних додатків.

У відповідь на це компанії впроваджують нові механізми відстеження поведінки: теплові карти, аналітику кліків, інтеграцію із CRM-системами, всеканальні стратегії тощо.

Крім того, все частіше піднімається питання етики – наскільки допустимо збирати ті чи інші типи моделей поведінкових даних, як забезпечити прозорість і захист персональної інформації, як уникати дискримінаційних рішень на основі автоматизованого аналізу.

Поведінкові шаблони та сегментація аудиторії. Одним з важливих напрямів дослідження є ідентифікація типових моделей поведінки – тенденцій, які повторюються серед певних груп користувачів. Наприклад:

- користувачі, які часто читають відгуки, але рідко купують;
- клієнти, які реагують тільки на знижки;
- постійні покупці, які діють швидко та цілеспрямовано.

Сегментація аудиторії за такими ознаками дає змогу створювати таргетовані кампанії, точніше передбачати реакції на зміни у сервісі, а також покращувати утримання клієнтів.

Загалом, дослідження поведінки на комерційних онлайн-платформах – це не лише про аналітику. Це про вміння слухати користувача, адаптуватися до його реальності, вдосконалювати цифрове середовище так, щоб воно

відповідало сучасним очікуванням і сприяло ефективній взаємодії обох сторін: продавця й покупця.

1.2 Аналіз існуючих програм-аналогів

Комерційні програмні рішення. На сьогодні на ринку представлено значну кількість програмних продуктів і сервісів, які дозволяють проводити аналіз поведінки користувачів у середовищі електронної комерції. Проте ці інструменти, попри свою популярність, мають низку обмежень – зокрема, фокусуються на поверхневих метриках або не забезпечують комплексної інтеграції з аналітичними моделями.

Одним із найбільш розповсюджених інструментів є Google Analytics (рис. 1.1), який пропонує зручну візуалізацію трафіку, конверсій, демографії, але його функціональність у частині виявлення закономірностей або сегментації користувачів обмежена[1].

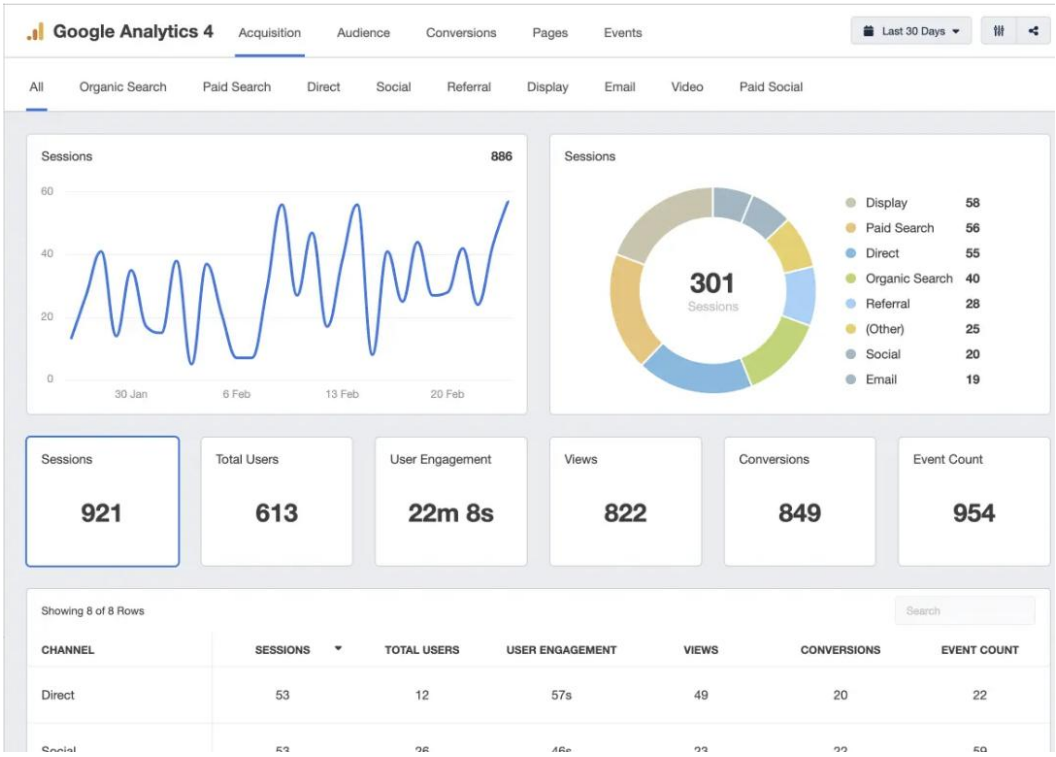

Рис 1.1. Демонстраційна аналітика з сервісу Google Analytics

Аналітичні платформи Mixpanel (рис. 1.2) [2] і Amplitude (рис. 1.3) [3] вже орієнтовані на подієвий аналіз, тобто дозволяють відстежувати послідовності дій, будувати «воронки» (funnels), сегментувати аудиторію за атрибутами поведінки.

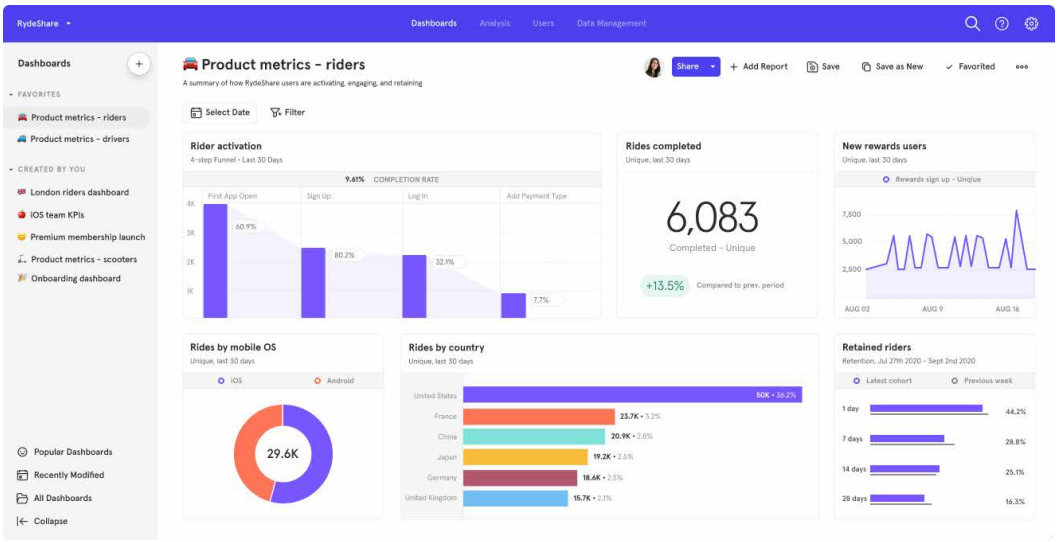

Рис. 1.2. Приклад аналітики з платформа Mixpanel

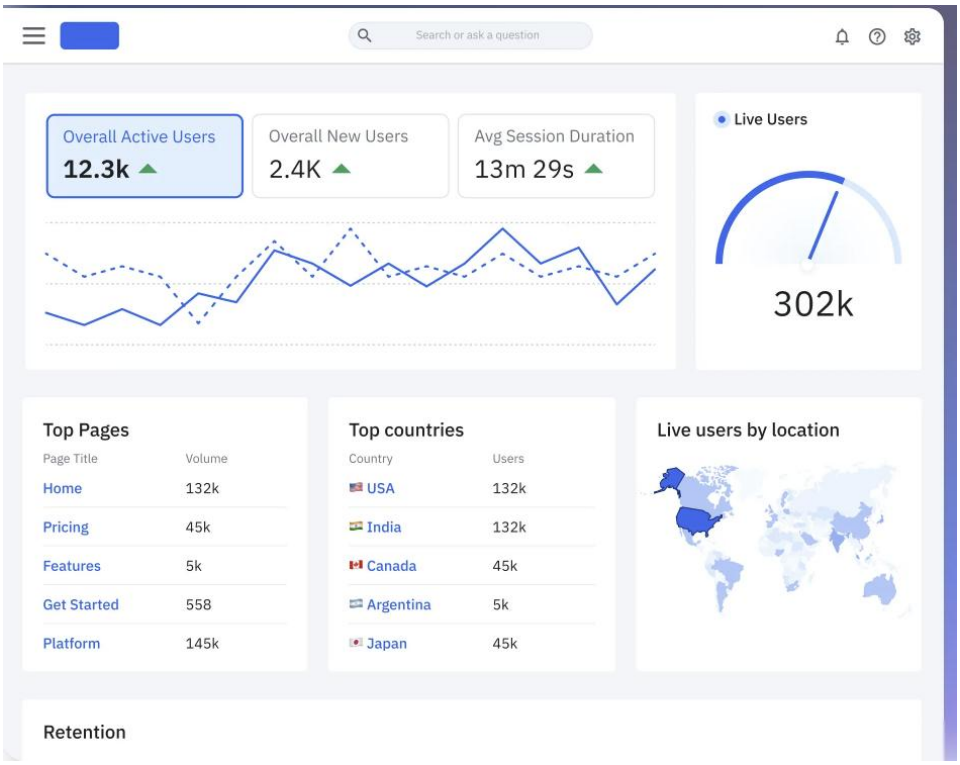

Рис. 1.3. Приклад аналітичного аналізу Amplitude

Проте їх застосування потребує певного рівня технічної інтеграції та базових знань у сфері аналітики, а можливості реалізації кастомних моделей

або роботи зі складними вимірами є обмеженими або недоступними без дорогих підписок.

У сфері машинного навчання також розвиваються окремі рішення – наприклад, рекомендаційні системи на основі Apache Mahout, кластеризація у RapidMiner, або використання бібліотек на кшталт scikit-learn чи ML.NET. Однак ці інструменти вимагають значних технічних знань та не завжди легко інтегруються з платформами.

На противагу вищезгаданих рішень, розроблена система вирізняється своєю цілісністю та гнучкістю. Вона поєднує потужності сховища даних для зберігання і обробки великого обсягу інформації з алгоритмами Data Mining, які дозволяють проводити повний цикл поведінкового аналізу. Всі етапи дослідження реалізовано в єдиному інтегрованому середовищі, без потреби у зовнішніх сервісах, що забезпечує прозорість, адаптивність до конкретного завдання і повний контроль над аналітикою. Такий підхід робить систему не лише технічно ефективною, а й зручною для подальшого впровадження у реальні бізнес-процеси.

Аналіз наукових підходів. У сучасних ж наукових дослідженнях поведінка користувачів у цифровому середовищі розглядається як багатовимірний феномен, що поєднує психологічні, соціальні та технологічні фактори. Одним із ключових напрямів є вивчення закономірностей покупки через великі обсяги даних з платформ електронної комерції.

Так, у роботі «Research on Purchasing Behavior Pattern of E-commerce Platform Consumers Based on Big Data Analysis» («Дослідження моделей купівельної поведінки споживачів електронної комерції на основі аналізу великих даних») Yanyi Qu 2025 р. було застосовано методи аналізу великих даних для виявлення поведінкових шаблонів споживачів на основі реальних транзакцій із маркетплейсів. Авторка підкреслює важливість інтеграції історичних дій користувача для побудови індивідуалізованих рекомендацій та покращення конверсій на платформі [4].

Інше дослідження «Exploring the Role of Personalization in E-commerce: Impacts on Consumer Trust and Purchase Intentions» («Дослідження ролі персоналізації в електронній комерції: вплив на довіру споживачів та наміри щодо покупки»), проведене авторами – Ananthaneni Madhuri, s. Melchior Reddy, Shireesha Manchem та B.R. Kumar 2024 р., зосереджено на впливі персоналізації на довіру до бренду та наміри покупки [5]. У дослідженні виявлено, що глибше розуміння індивідуальних уподобань і адаптація контенту значно підвищують зацікавленість користувачів, особливо на етапі порівняння товарів.

Також важливим напрямом вважається сегментація користувачів. Зокрема, у праці «Personalization in personalized marketing: Trends and ways forward» («Персоналізація в персоналізованому маркетингу: тенденції та шляхи розвитку») авторів: Shobhana Chandra, Sanjeev Verma, Weng Marc Lim, Satish Kumar, Naveen Donthu 2022 р. розглядається персоналізований маркетинг як системна практика, що базується на точковому аналізі поведінкових та контекстуальних змін [6]. Дослідження демонструє ефективність класифікаційних моделей при адаптації пропозицій до окремих сегментів клієнтів.

Цікаво, що в аналітичній роботі «Analysis of E-Commerce Purchase Patterns Using Big Data: An Integrative Approach to Understanding Consumer Behavior» («Аналіз моделей покупок в електронній комерції з використанням Big Data: інтегрований підхід до розуміння поведінки споживачів»), опублікованій 2023 р. [7], дослідники запропонували інтегрований підхід до класифікації та прогнозування покупок із використанням як історичних, так і контекстуальних змінних, включаючи частоту відвідувань, інтерес до категорій і взаємодію з карткою товару, що дозволяє розпізнавати не лише сегменти, а й наміри споживачів.

Додатково, у праці «Categorizing Online Shopping Behavior from Cosmetics to Electronics: An Analytical Framework» («Категоризація поведінки покупців в інтернет-магазинах від косметики до електроніки: аналітична

модель») 2020 р. [8] наведено аналітичну модель для класифікації користувачів за типами товарів, які їх цікавлять. Автори обґрунтовують ефективність моделювання «шаблонів зацікавлення» як основи для таргетованої реклами.

Загалом, попри значний обсяг академічних досліджень у сфері вивчення цифрової поведінки, більшість з них або зосереджуються на теоретичних аспектах, або не охоплюють повного циклу трансформації сирих даних у практичні рекомендації, що створює розрив між науковими напрацюваннями та реальними потребами бізнесу в гнучких, адаптивних аналітичних рішеннях.

Дане дослідження і було спрямоване на подолання цього розриву шляхом створення цілісної інтелектуальної системи, яка базується на кращих наукових підходах, але реалізована в зручному та ефективному інструменті для реального ефективного аналізу користувацької поведінки.

1.3 Постановка завдання

У межах цього магістерського дослідження було поставлено завдання створити комплексну інтелектуальну систему для всебічного аналізу користувацької поведінки на комерційних онлайн-платформах. Така система має забезпечувати глибоке розуміння взаємодії користувачів із товарами та послугами, виявляти ключові закономірності у їхній поведінці, а також сприяти прийняттю обґрунтованих рішень у сфері електронної комерції.

Основним завданням дослідження є розробка рішення, яке б дозволяло здійснювати багаторівневий аналіз поведінки користувачів на основі єдиного джерела правдивих даних – сховища даних у зірковій структурі, що підтримує інтеграцію з OLAP-кубом та методами Data Mining.

Серед запланованих алгоритмічних підходів передбачено використання:

- методів класифікації для визначення рівня активності користувачів та поведінкових груп (One-Rule, Наївний Байєс);
- асоціативного аналізу для виявлення зв’язків між переглядами, вподобаннями та подальшими діями;
- кластеризації для формування сегментів на основі схожості дій або характеристик (K-Means у поєднанні з PCA для зменшення розмірності).

При цьому важливо не лише технічно реалізувати перелічені підходи, а й забезпечити логічну узгодженість процесу – від моменту підготовки даних до формування висновків, які можна використовувати у реальному бізнес-середовищі. Передбачено також реалізацію інструментів для підрахунку основних KPI.

Окремим завданням стало забезпечення візуалізації результатів у зручному середовищі: частина реалізації виконується в Microsoft SQL Server з використанням SSAS для побудови куба та подальшої обробки, інша частина – у Python і C# WinForms для реалізації моделей та відображення висновків.

Таким чином, розроблена система повинна підтримувати повний аналітичний цикл:

- обробку сирих даних у структурованому вигляді (ETL-процеси, перетворення, очищення);
- побудову багатовимірного сховища та OLAP-куба;
- застосування алгоритмів класифікації, кластеризації та асоціативного аналізу;
- формування звітів, графіків, інтерпретацій та рекомендацій у зручному форматі.

У підсумку, дослідження мало на меті створення комплексної системи, що допомагатиме на основі єдиного набору даних здійснювати поглиблений аналіз поведінки клієнтів, враховуючи як бізнес-потреби, так і можливості сучасних засобів інтелектуальної аналітики.

2 МОДЕЛЮВАННЯ ПРЕДМЕТНОЇ ОБЛАСТІ

2.1 Загальна характеристика положень моделювання

Моделювання є фундаментальним етапом у процесі створення інформаційних систем. Воно забезпечує формалізоване уявлення про структуру, функціональність і взаємодію елементів системи ще до її фактичної реалізації. Завдяки йому розробники отримують змогу абстрагуватися від технічних деталей і зосередитися на логіці системи, її архітектурі, бізнес-процесах і поведінкових сценаріях.

Моделі слугують інструментом комунікації між усіма учасниками розробки – замовниками, аналітиками, розробниками, тестувальниками. Саме через графічне і логічне представлення компонентів системи можна ефективно узгодити вимоги, визначити критичні точки, спростити подальшу реалізацію і супровід проєкту. На практиці моделювання дозволяє суттєво зменшити ризики проєктних помилок і знижує витрати на внесення змін на пізніших етапах.

Використання уніфікованої мови моделювання UML (Unified Modeling Language) [9] стало стандартом у галузі, оскільки вона дозволяє створювати діаграми, які чітко описують як функціональні аспекти (через прецеденти, активності, послідовності), так і статичну структуру системи (класи, компоненти, об’єкти). Дані діаграми допомагають виявити основні сутності, їхні атрибути, зв’язки між ними, а також сценарії взаємодії користувачів із системою.

У межах даного дослідження моделювання застосовується як засіб для формування чіткої логічної основи розроблюваної системи інтелектуального аналізу поведінки користувачів. Створені моделі дозволяють описати

функціональну структуру, визначити взаємозв’язки між елементами системи, а також закласти основу для реалізації ключових модулів аналітики.

Підсумовуючи, моделювання виконує не лише описову, але й прогностичну функцію, допомагаючи побачити потенційну поведінку системи ще до її створення. Воно виступає своєрідною «віртуальною репетицією» реального функціонування, що підвищує якість розробки, сприяє узгодженості дій між учасниками проєкту та забезпечує системний підхід до реалізації складних інформаційних рішень.

2.2 Діаграма прецедентів

Діаграма прецедентів (Use-Case diagram) [10] є важливим інструментом у процесі моделювання функціональних вимог. Вона дозволяє візуалізувати основні сценарії використання системи, визначити межі її функціоналу, а також встановити взаємозв’язки між користувачами (акторами) та системою. Такий тип діаграми особливо корисний на етапі системного аналізу, оскільки допомагає узгодити бачення функціональності між замовником, аналітиком та командою розробки.

Діаграма фокусується на зовнішніх взаємодіях, що дозволяє зосередитись не на реалізації, а саме на тому, що повинна робити система з точки зору її користувачів, що дає змогу уникнути двозначностей у вимогах, забезпечити повноту охоплення очікуваних функцій та покращити якість подальшого моделювання.

У межах цього дослідження було створено діаграму прецедентів для інформаційної системи аналізу поведінки користувачів на комерційних платформах (рис. 2.1). Вона охоплює основні сценарії роботи системи та взаємодії з нею різних груп користувачів.

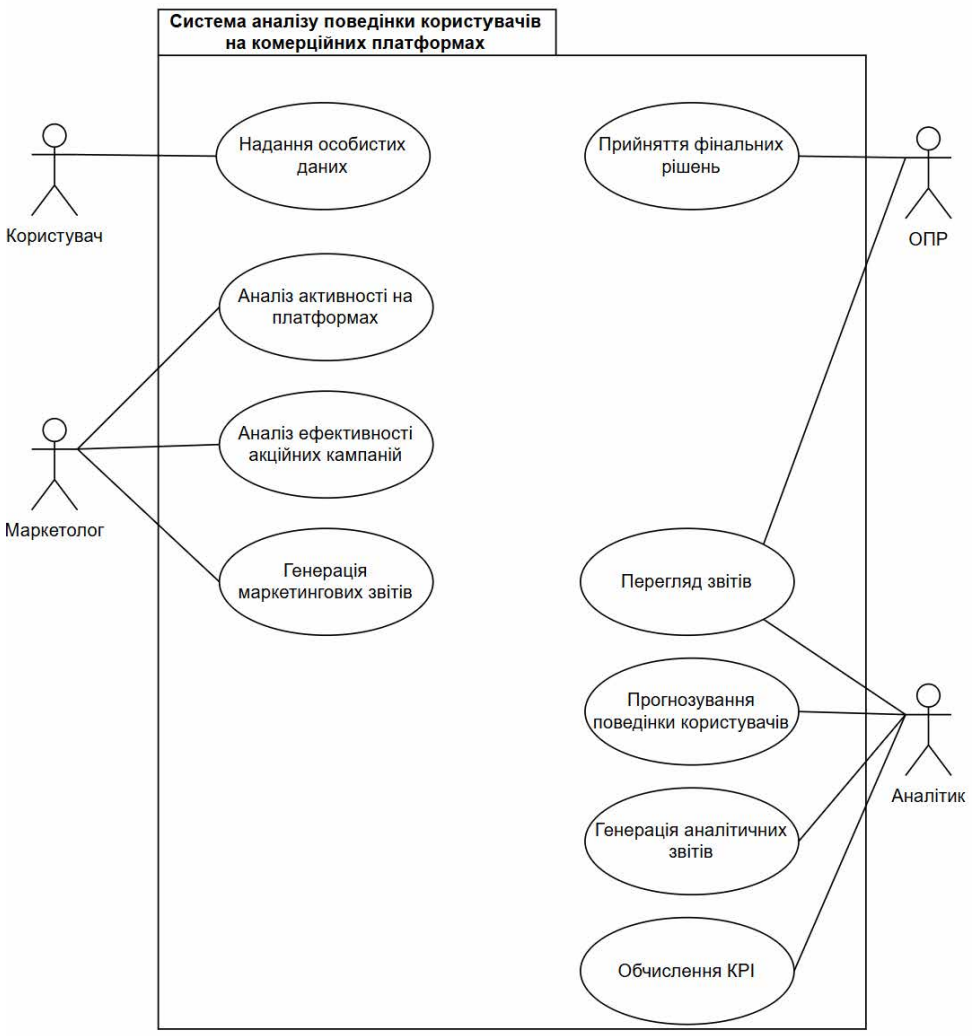

Рис. 2.1. Діаграма прецедентів системи

Таблиця 2.1

Актори, які діють у системі

Актор	Опис
Користувач	Взаємодіє із платформою, надає особисті дані, залишає сліди у вигляді дій (перегляди, покупки).
Маркетолог	Аналізує поведінку користувачів, ефективність кампаній та створює маркетингові звіти.
Аналітик	Прогнозує поведінку користувачів, генерує аналітичні звіти та обчислює KPI.
ОПР (Особа, що приймає рішення)	Переглядає аналітичні звіти та ухвалює стратегічні рішення для розвитку платформи.

Побудована діаграма дозволяє чітко визначити, які саме функції реалізуються для кожного типу користувача. Наприклад, аналітик не має доступу до редагування персональних даних користувача, а користувач, зі свого боку, не бере участі у генерації звітів. Такий поділ ролей забезпечує розмежування прав доступу та підвищує безпеку системи.

Крім того, використання діаграми прецедентів дозволило сформувати базову структуру майбутньої реалізації системи та закласти основу для подальшого проектування діаграм розгортання, послідовностей та класів.

Завдяки чіткому візуальному уявленню про функціональні можливості кожного з акторів, діаграма стала ефективним засобом комунікації між технічними та нетехнічними сторонами проєкту, дозволяючи узгодити вимоги ще до початку етапу реалізації.

2.3 Діаграма послідовності

Діаграма послідовності (Sequence diagram) є одним з ключових інструментів візуалізації сценаріїв взаємодії між об’єктами в рамках певного процесу [11]. Вона дозволяє простежити хронологію обміну повідомленнями, акцентуючи увагу на послідовності викликів методів, операцій або запитів у рамках визначеного сценарію. Завдяки цьому інструменту можливо чітко зрозуміти, як саме відбувається комунікація між актором, системними компонентами, базами даних, сервісами обробки даних та зовнішніми модулями.

Використання таких діаграм є особливо важливим при розробці систем, які мають справу зі складною логікою обробки даних. У таких випадках чітке розуміння порядку дій і умов гілкування дозволяє уникнути логічних помилок, забезпечити коректну взаємодію між компонентами процесів та пришвидшити етап фактичної програмної реалізації.

У межах даного дослідження було побудовано діаграми послідовності, які відображають ключові процеси функціонування розробленої інтелектуальної системи аналізу поведінки користувачів на комерційних платформах. Особливу увагу приділено двом сценаріям: аналіз ефективності акційних кампаній та прогнозування поведінки користувачів.

На першій діаграмі (рис. 2.2) змодельовано процес запуску аналізу маркетингової кампанії, що починається з ініціативи маркетолога:

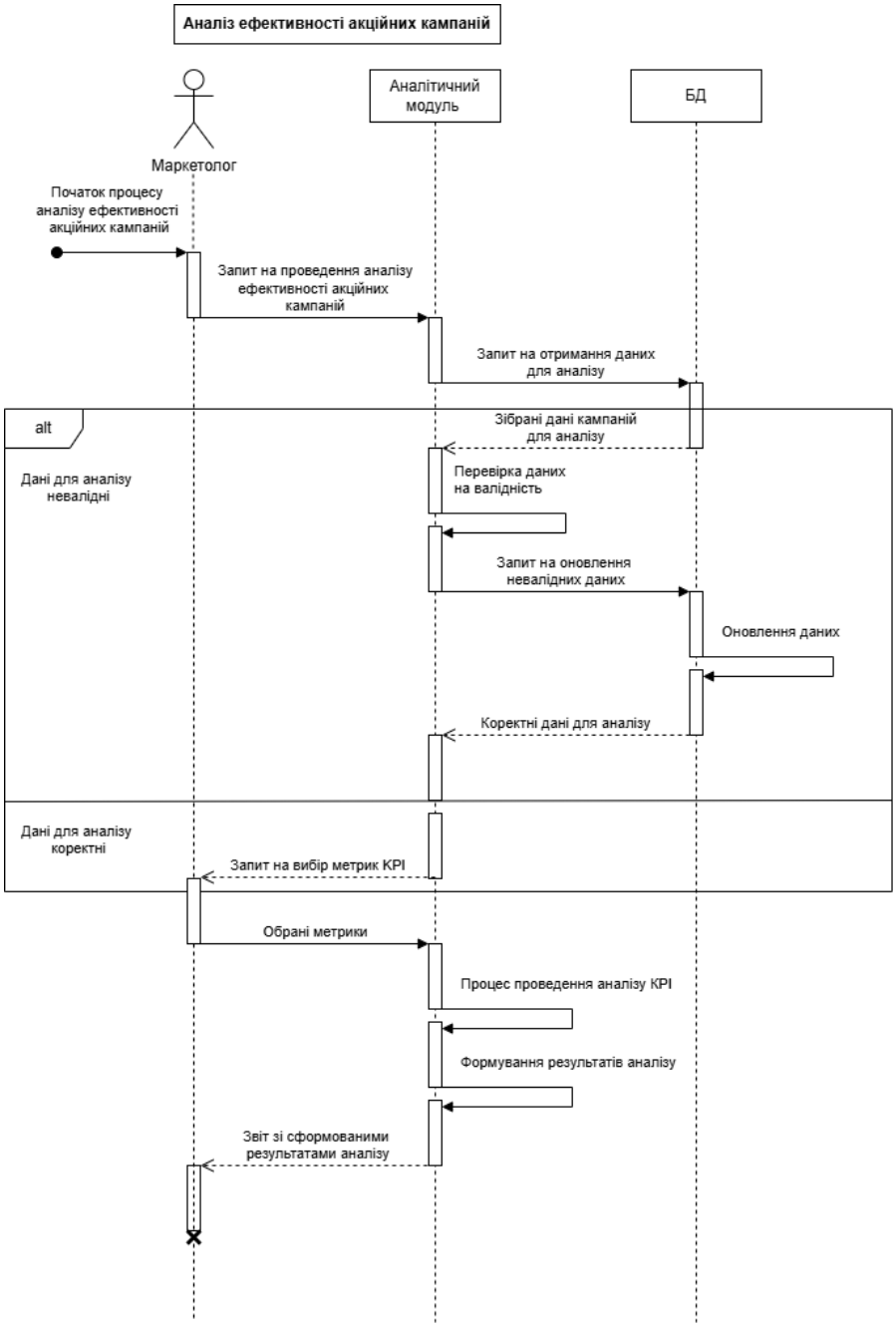

Рис. 2.2. Діаграма послідовності. Аналіз ефективності акційних кампаній

Аналітичний модуль платформи надсилає запит на отримання релевантних даних до бази даних. У випадку виявлення невалідних даних реалізовано сценарій їхнього оновлення. Далі відбувається вибір KPI-метрик, обчислення показників ефективності та формування звіту.

У другому випадку (рис. 2.3) представлено сценарій взаємодії аналітика із аналітичним модулем при запуску процесу прогнозування поведінки користувачів:

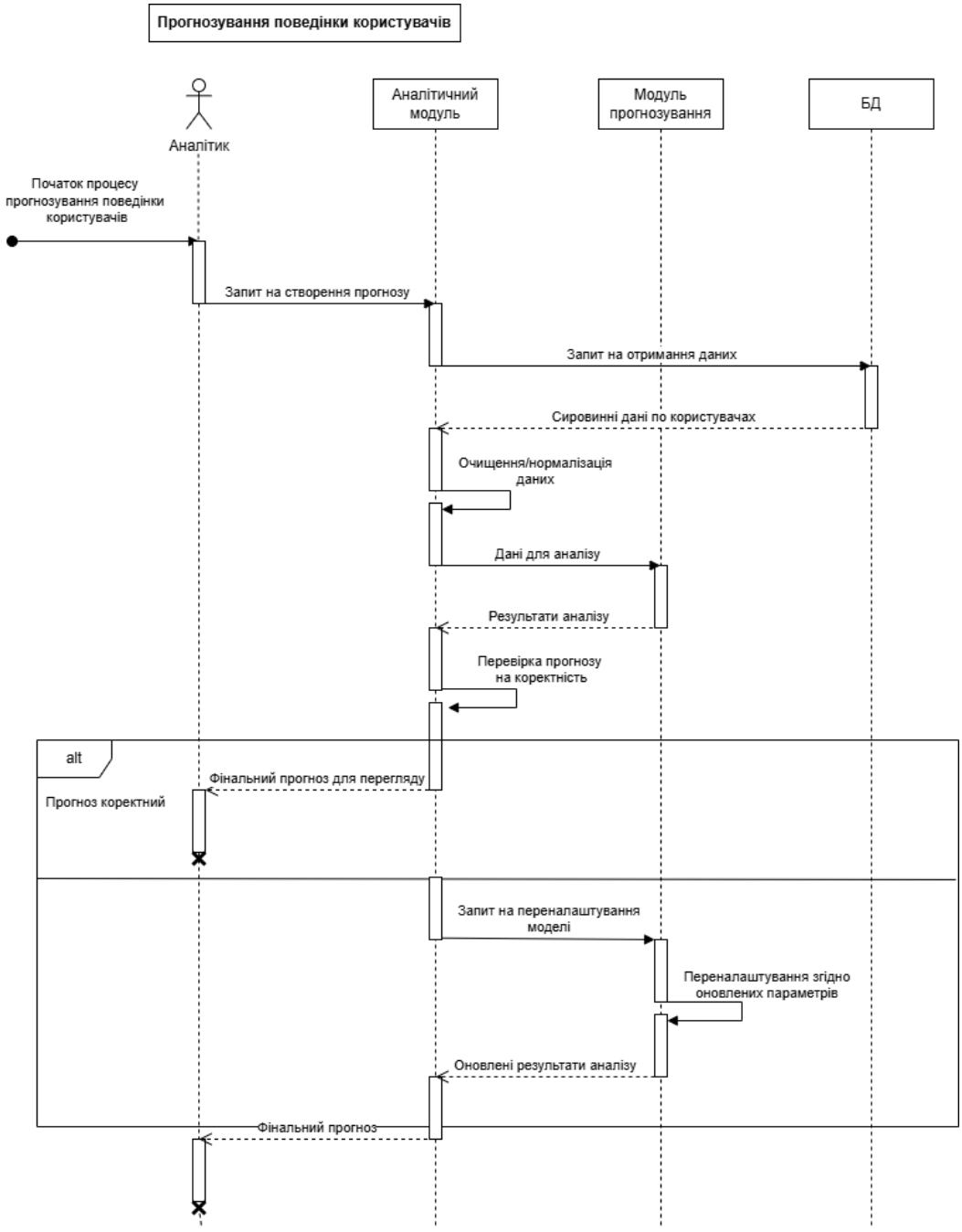

Рис. 2.3. Діаграма послідовності. Прогнозування поведінки користувачів

Після формування запиту на створення прогнозу, модуль аналізу отримує сирі дані з бази, проводить їх очистку, нормалізацію та передає до модуля прогнозування. Отриманий прогноз перевіряється, і в разі потреби, здійснюється переналаштування з подальшим оновленням результатів.

Візуалізація таких процесів є важливим етапом проектування, що забезпечує цілісне розуміння поведінки впроваджуваної системи і спрощує подальшу реалізацію її функціоналу.

2.4 Діаграма класів

Діаграма класів належить до фундаментальних інструментів об’єктно-орієнтованого аналізу та проєктування. Вона слугує графічним поданням структури предметної області, де кожен клас відображає тип об’єкта з відповідним набором характеристик (атрибутів – attributes) та операцій (методів – methods) [12].

Такий підхід дозволяє формалізовано представити зв’язки між класами (classes), встановити їхню ієрархію, асоціації, агрегації чи композиції, що в сукупності забезпечує цілісність архітектури програмної системи. Застосування діаграм класів є важливим етапом у побудові надійного програмного забезпечення, оскільки вони створюють чітку основу для реалізації бізнес-логіки та полегшують подальшу розробку, тестування і підтримку системи.

У межах дослідження було побудовано діаграму класів для інтелектуальної системи аналізу поведінки користувачів на комерційних платформах. Структура діаграми охоплює функціональні компоненти, що реалізують основні процеси збору, обробки, аналізу та інтерпретації поведінкових даних користувачів, з подальшою генерацією звітності (reporting) та підтримкою прийняття управлінських рішень (рис. 2.4):

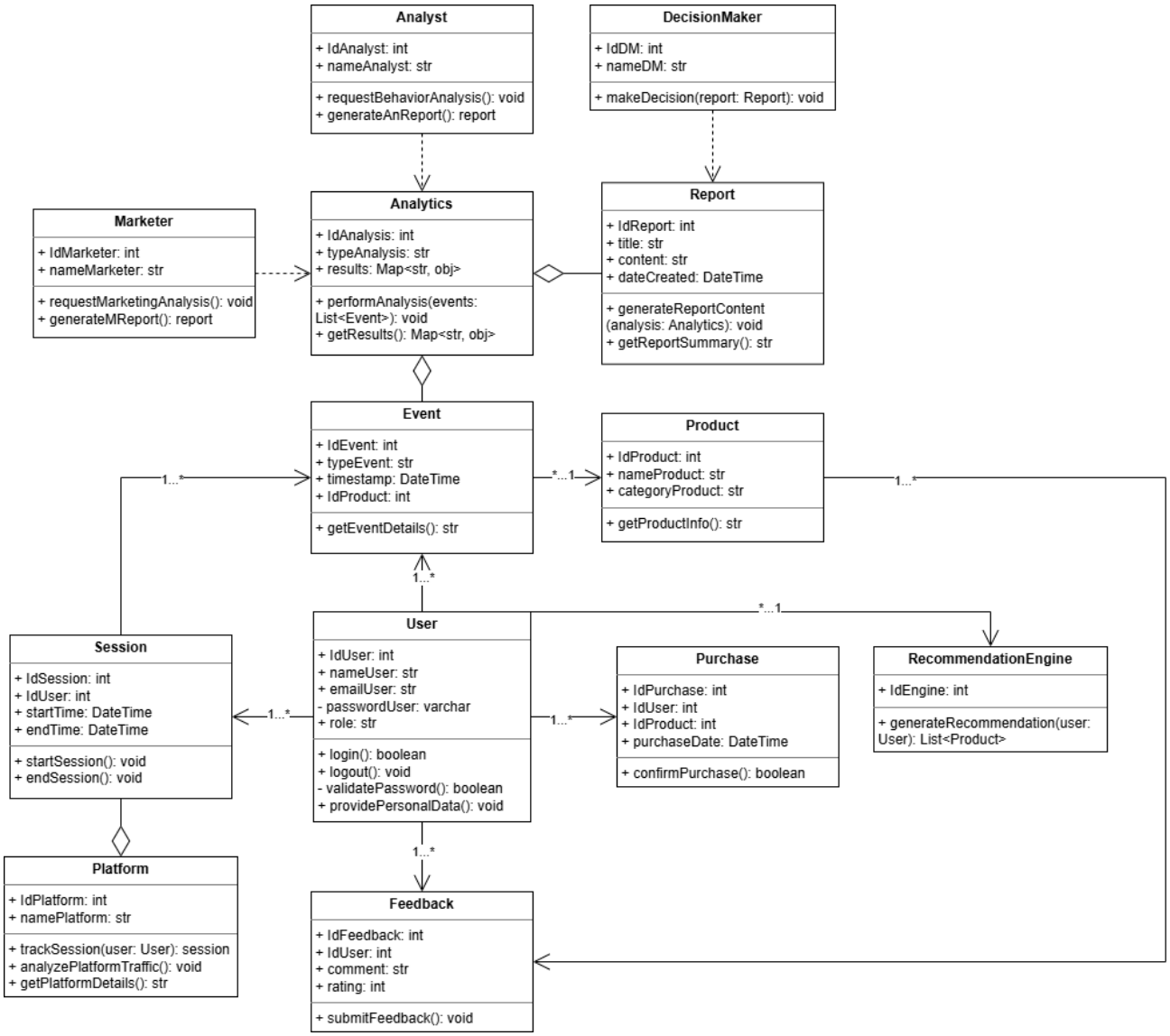

Рис. 2.4. Діаграма класів

Центральним елементом діаграми є клас User (Користувач), який представляє користувача платформи з відповідними атрибутами (ідентифікатор – IdUser, електронна пошта – emailUser, роль – role тощо) та методами авторизації (login()), валідації пароля (validatePassword()) та надання персональних даних (providePersonalData()).

З боку користувача ініціюється взаємодія з платформою (Platform – Платформа), що фіксується у класі Session (Сеанс), а сама активність відображається через множину подій (Event – Подія), пов’язаних із товарами (Product – Товар) та часовими мітками (timestamp – час події).

Клас Analytics (Аналітика) реалізує механізми аналітичної обробки подій та інших даних. Він агрегує інформацію, виконує обчислення та надає результати (getResults()) в структурованому вигляді. Аналітик (Analyst) ініціює аналітичний запит (requestBehaviorAnalysis()), а також генерує звіти (generateAnReport()). Маркетолог (Marketer) викликає функціонал оцінки ефективності маркетингових кампаній (requestMarketingAnalysis(), generateMReport()).

Результати обробки формуються у вигляді об’єктів класу Report (Звіт), що містить атрибути назви (title), змісту (content) та дати створення (dateCreated). Звіти використовуються користувачем типу DecisionMaker (Особа, що приймає рішення) для ухвалення управлінських рішень (makeDecision()).

Клас RecommendationEngine відповідає за формування персоналізованих рекомендацій (generateRecommendation()) для користувачів на основі аналізу їхніх попередніх покупок (Purchase) та загальної активності. Зворотний зв’язок фіксується через клас Feedback (Зворотній зв'язок), що дозволяє враховувати суб’єктивну оцінку користувача (comment) стосовно запропонованих послуг або товарів.

Отже, діаграма класів демонструє цілісну модель системи, в якій усі компоненти взаємопов’язані: від реєстрації дій користувача — до аналітичної обробки інформації, створення звітів і підтримки управлінських рішень. Така структура забезпечує модульність, масштабованість та можливість подальшого розширення функціональності системи відповідно до нових вимог бізнес-середовища.

3 ПРОЕКТУВАННЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ АНАЛІЗУ ПОВЕДІНКИ КОРИСТУВАЧІВ

3.1 Діаграма розгортання

Діаграма розгортання (Deployment diagram) [13] належить до категорії структурних діаграм UML і використовується для відображення фізичної інфраструктури програмної системи. Вона моделює, як саме програмні компоненти (наприклад, веб-сервер, база даних, аналітичні модулі) розміщені на апаратних або віртуальних пристроях – вузлах (nodes), які можуть бути пов’язані каналами зв’язку.

Такий тип діаграми дозволяє візуалізувати мережеву архітектуру, логіку взаємодії між окремими елементами системи та забезпечує наочне представлення технічного середовища, в якому функціонує програмне забезпечення. Завдяки цьому можна своєчасно виявити потенційні «вузькі місця» у системі, оцінити розподіл навантаження, рівень відмовостійкості та забезпечити оптимальне масштабування.

Для забезпечення ефективної роботи розроблюваної інтелектуальної системи аналізу поведінки користувачів на комерційних платформах було спроєктовано архітектуру, що охоплює декілька взаємодіючих компонентів. Її головною метою є забезпечення надійного збору, зберігання, обробки та аналітичного аналізу великих обсягів даних про поведінку користувачів.

Архітектура системи передбачає чітке розділення ролей між джерелами даних, середовищем зберігання, аналітичними модулями та робочими станціями користувачів. Такий підхід дозволяє досягти масштабованості, високої продуктивності, а також спрощує технічне обслуговування окремих компонентів системи (рис. 3.1).

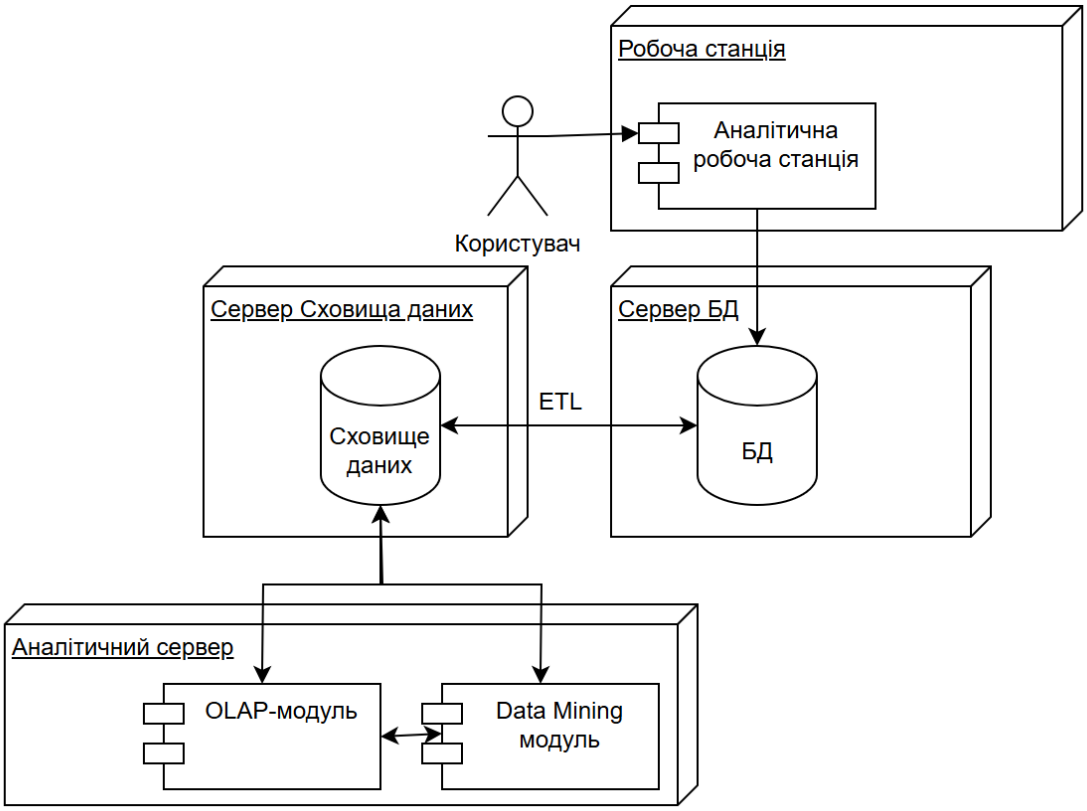

Рис. 3.1. Діаграма розгортання

Кінцевий користувач системи – це аналітик або маркетолог, який взаємодіє з аналітичною робочою станцією через спеціалізований інтерфейс. Користувачі мають змогу виконувати запити до OLAP-модулю, переглядати агреговані звіти, ініціювати запуск моделей Data Mining і вивчати закономірності в поведінці клієнтів на платформі.

Аналітична робоча станція – є спеціалізованим програмним середовищем, встановленим на локальному комп'ютері користувача. Вона забезпечує доступ до візуалізації даних, результатів класифікації, кластеризації або асоціативного аналізу. Саме через неї здійснюється інтерпретація результатів, створення звітів, виявлення інсайтів і прийняття рішень щодо стратегії розвитку платформи.

База даних містить первинні, оперативні дані про взаємодію користувачів з комерційною платформою: перегляди товарів, покупки, повернення тощо. Цей компонент функціонує як джерело «сирих» даних, які

ще не були агреговані або підготовлені для аналітики. БД оновлюється у в залежності наявності оновлених для завантаження даних самої платформи.

ETL (Extract, Transform, Load) – ключова ланка між операційною базою даних і сховищем даних. Вона дозволяє уникнути надлишкового навантаження на транзакційну базу даних і підвищити ефективність аналітики.

У процесі ETL [17]:

- Extract (витягування): відбувається регулярне вилучення інформації з основної БД;
- Transform (перетворення): дані проходять очищення, нормалізацію, структурування відповідно до аналітичної моделі;
- Load (завантаження): оброблені дані передаються у сховище для подальшого аналізу.

Сховище даних є централізованим сховищем, куди потрапляють структуровані й підготовлені для аналізу дані . Його структура реалізована за принципом багатовимірної моделі (зіркова), що дозволяє швидко здійснювати агрегацію, фільтрацію та сегментацію за різними вимірами: часом, категорією товару, статтю користувача, тощо. Сховище підтримує історичність даних та забезпечує узгоджене аналітичне середовище для всіх користувачів [14].

На **аналітичному сервері** розгорнуті два основні модулі:

- **OLAP-модуль** – виконує багатовимірний аналіз даних [15]. За його допомогою користувачі можуть «провалюватися» на нижчі рівні деталізації (drill-down), переглядати агреговані звіти за різними зрізами (slice-and-dice), виявляти сезонність та інші закономірності. OLAP-куб формується на основі даних зі сховища.
- **Data Mining-модуль** – відповідає за застосування алгоритмів інтелектуального аналізу даних: класифікація (наприклад, методом 1-Rule або Наївного Байєса), кластеризація (K-Means), пошук асоціативних правил (Apriori) [16]. Цей модуль дозволяє виявляти

неочевидні патерни у поведінці користувачів і формувати на основі цього рекомендації або персоналізовані пропозиції.

Між OLAP та Data Mining модулями існує взаємозв'язок – результати глибинного аналізу можуть бути використані для побудови нових вимірів або гіпотез, які потім візуалізуються в OLAP-звітах.

3.2 Опис джерела даних

З огляду на обмежений доступ до реальних даних електронної комерції, під час проєктування аналітичної системи було вирішено сформувати штучний датасет, заснований на поширених моделях поведінки користувачів в онлайн-магазинах. За основу було взято відкритий набір даних Amazon із платформи Kaggle [18], який було адаптовано, розширено та структуровано у такий спосіб, щоб він відображав типові взаємодії з комерційними платформами – від пошуку товарів і переглядів до покупок, повернень і взаємодії з рекомендаціями.

Створення синтетичного набору даних дозволило уникнути проблем із конфіденційністю та забезпечити достатню гнучкість для моделювання різних сценаріїв користувацької активності.

Набір даних охоплює такі основні інформаційні складові:

- дані про товари: назва продукту, категорія до якої відноситься, ідентифікатори (рис. 3.2);

	id_product	product_name	id_category
1	1	iPhone 15 Pro Max	1
2	2	Samsung Galaxy S23 Ultra	1
3	3	Google Pixel 8 Pro	1
4	4	OnePlus 11R	1
5	5	Xiaomi 13T Pro	1
6	6	MacBook Air M2	1
7	7	Dell XPS 13	1

Рис. 3.2. Таблиця «Товари»

- демографічні характеристики користувачів: ідентифікатор, ім’я, вік, стать (рис. 3.3);

	id_user	user_name	age	gender
1	1	Andrii	51	Male
2	2	Alina	48	Female
3	3	Andrii	19	Male
4	4	Oksana	29	Female
5	5	Yurii	58	Male
6	6	Yulia	51	Female
7	7	Yurii	41	Male
8	8	Anastasiia	30	Female

Рис. 3.3. Таблиця "Користувачі"

- часові параметри: дата, місяць, рік (рис. 3.4);

	id_date	year	month	day
1	1	2024	1	1
2	2	2024	1	2
3	3	2024	1	3
4	4	2024	1	4
5	5	2024	1	5
6	6	2024	1	6
7	7	2024	1	7
8	8	2024	1	8

Рис. 3.4. Таблиця "Дата"

- показники поведінки: кількість переглядів, кількість покупок, кількість повернень.

Ці параметри було спеціально підібрано для того, щоб забезпечити можливість проведення проведення багатовимірного аналізу через OLAP-інструменти, побудову гіперкубів та виконання операцій Drill-down, Slice and Dice, Roll-up тощо. Крім того, їх достатньо для реалізації завдань інтелектуального аналізу даних (Data Mining), зокрема класифікації, пошуку асоціативних правил і кластеризації.

Згенеровані дані було приведено до форми, що відповідає вимогам моделювання сховища даних. Завдяки узгодженій структурі та чітко визначеним зв’язкам між сутностями, цей набір даних ефективно використовується в рамках ETL-процесу (Extract, Transform, Load) для

заповнення сховища та подальшої аналітичної обробки. Така підготовка джерела даних забезпечує достовірність експериментів, дозволяє масштабувати дослідження та адаптувати методики до інших платформ.

У підсумку, з урахуванням усіх особливостей формування датасету, його можна було вважати репрезентативною основою для побудови інформаційно-аналітичної системи, орієнтованої на дослідження споживчої поведінки.

3.3 Опис сховища даних

Для ефективного зберігання, обробки та подальшого аналізу поведінкових даних користувачів на комерційній платформі було спроєктовано і реалізовано спеціалізоване сховище даних. Його архітектура побудована за принципом моделі «зірка» (star schema) (рис. 3.5), яка є однією з найпоширеніших і найбільш оптимізованих форм організації даних у сховищах для потреб аналітики та OLAP-запитів [19].

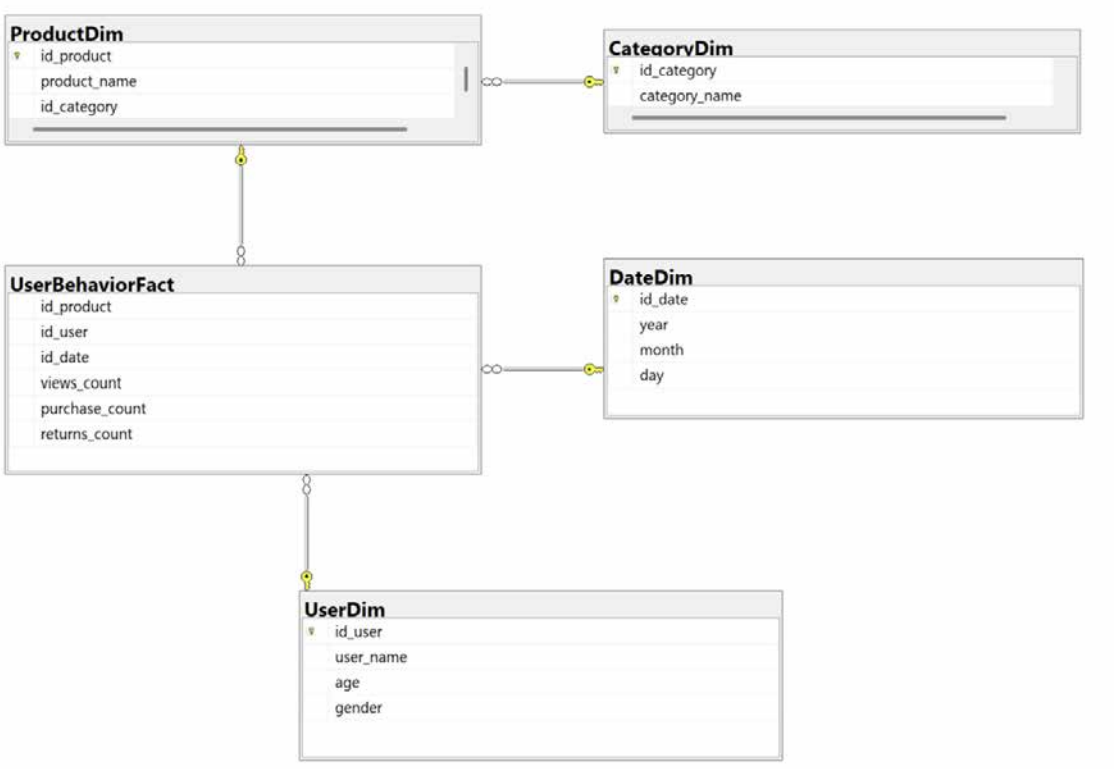

Рис. 3.5 Структура сховища даних

Центральне місце займає таблиця фактів UserBehaviorFact, яка акумулює ключові числові показники, що відображають активність користувачів: кількість переглядів товарів, кількість здійснених покупок і кількість повернень. Кожен запис у таблиці фактів представляє собою агрегований підсумок дій певного користувача щодо конкретного товару у визначену дату.

Для забезпечення повноти контексту та можливості гнучкої фільтрації, аналізу і агрегації, таблиця фактів з'єднана з чотирма вимірними таблицями: ProductDim, CategoryDim, UserDim та DateDim.

Таблиця вимірів ProductDim містить інформацію про найменування товару та ідентифікатор. Через поле id_category реалізовано зв'язок із таблицею CategoryDim, яка дозволяє класифікувати товари за категоріями, що у подальшому забезпечує аналіз популярності певних груп продуктів.

Таблиця UserDim містить атрибути користувача: унікальний ідентифікатор, ім'я, вік і стать, що дозволяє здійснювати демографічний аналіз поведінкових шаблонів.

Нарешті, DateDim дає змогу виконувати аналіз динаміки активності у часовому розрізі – за роками, місяцями, днями.

Така структура дозволяє ефективно реалізовувати OLAP-операції (drill-down, roll-up, slice, dice), будувати зручні дашборди та формувати індивідуальні звіти за гнучкими критеріями. Водночас вона є повністю сумісною з подальшим застосуванням методів інтелектуального аналізу даних (Data Mining), забезпечуючи основу для побудови моделей класифікації, виявлення асоціативних правил, сегментації користувачів тощо.

3.4 Розгортання гіперкубу в середовищі BI MS SQL Server

Одним із ключових етапів розробки інтелектуальної системи є створення багатовимірного куба, який забезпечує зручний доступ до даних для проведення багатовимірного аналізу. Для реалізації цього завдання

використовувався SQL Server Analysis Services (SSAS), що є потужним інструментом для побудови OLAP-моделей [20].

Перед початком роботи над кубом було виконано підключення до сховища даних: у середовищі Visual Studio створено новий проект Analysis Services Multidimensional and Data Mining Project. Було встановлено з'єднання із базою даних сховища (рис. 3.6) за допомогою модуля Data Source Wizard, яка містить таблиці фактів та вимірів, розроблені в попередніх етапах.

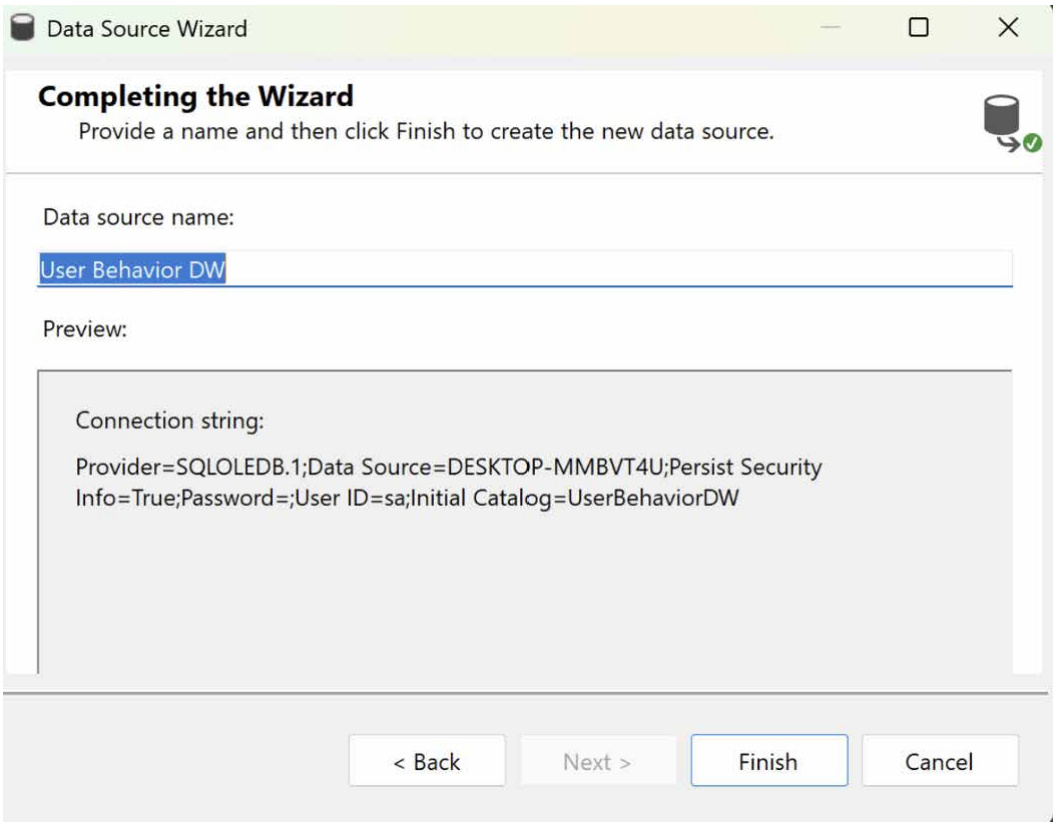

Рис. 3.6. Додавання джерела даних

Наступним етапом стало формування представлення джерела даних за допомогою майстра Data Source View (DSV). Дане представлення виступає логічною моделлю, яка ізолює користувача від фізичної структури джерела даних, надаючи можливість налаштовувати зв'язки між таблицями, об'єднувати поля або змінювати їхні назви для зручності подальшої роботи.

На рисунку 3.7 продемонстровано процес налаштування DSV, яке використовує таблиці з оперативної бази даних для аналізу.

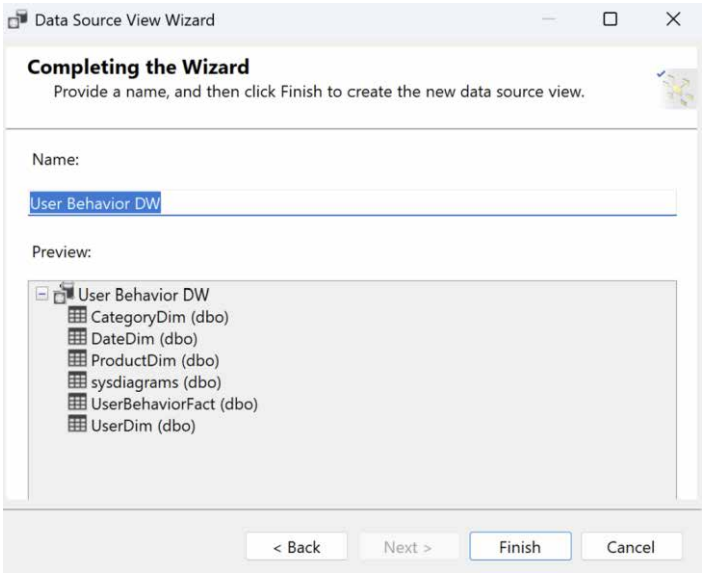

Рис. 3.7. Створення уявлення джерела даних

За допомогою Cube Wizard було створено багатовимірний куб для аналізу поведінки користувачів. У процесі налаштування було виконано наступні дії (рис. 3.8-3.10):

- обрано таблицю фактів UserBehaviorFact як основу для побудови куба;
- додано створені виміри (ProductDim, CategoryDim, UserDim, DateDim) до куба;
- визначено атрибути, які використовуватимуться у вимірах та для встановлення зв’язків між таблицями.

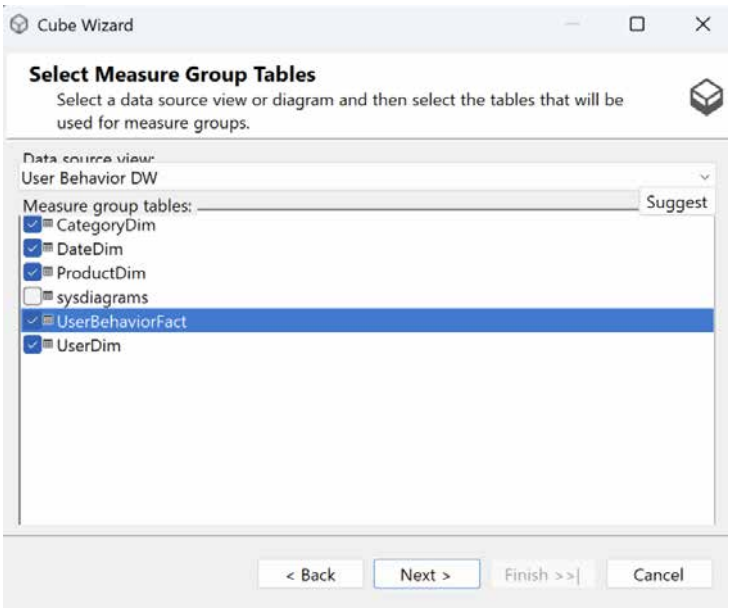

Рис. 3.8. Визначення таблиць вимірів

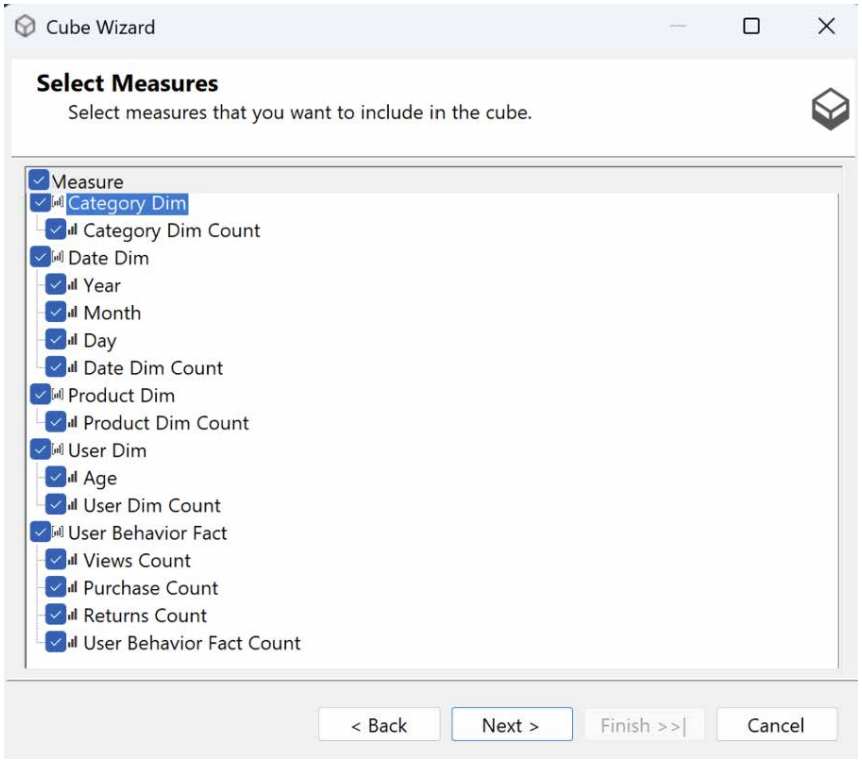

Рис. 3.9. Вибір атрибутів для вимірів

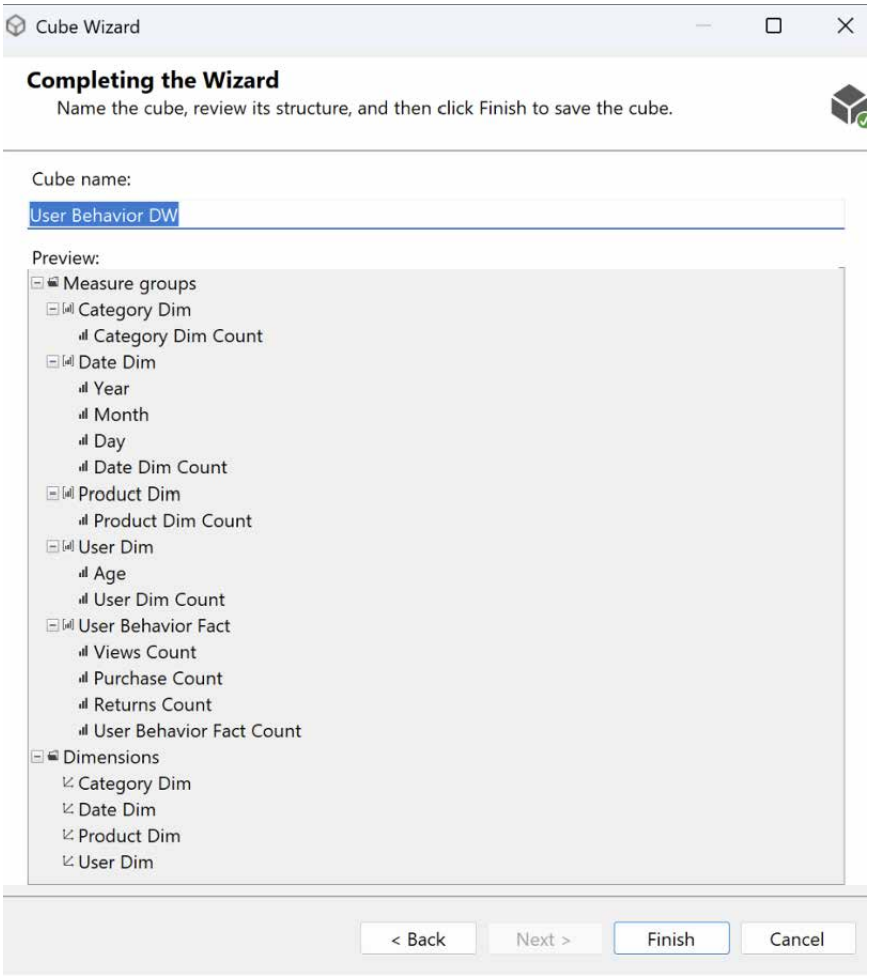

Рис. 3.10. Завершення налаштувань куба

Після завершення налаштувань виконано процесинг (обробку) куба (рис. 3.11) , який передбачає завантаження даних у куб для подальшого використання в аналітичних запитах:

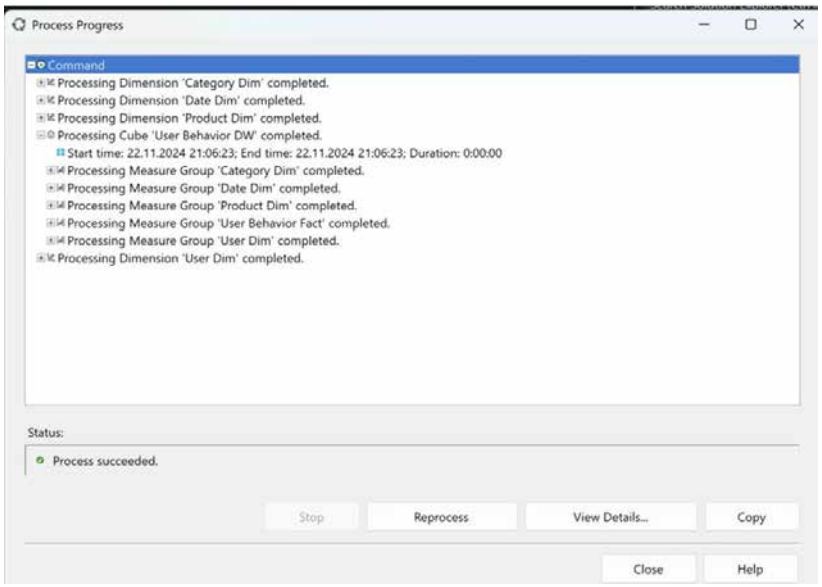

Рис. 3.11. Процес обробки куба

Результатом виконаних дій є розгорнутий багатовимірний куб (рис. 3.12), який дозволяє виконувати OLAP-запити для аналізу поведінки користувачів:

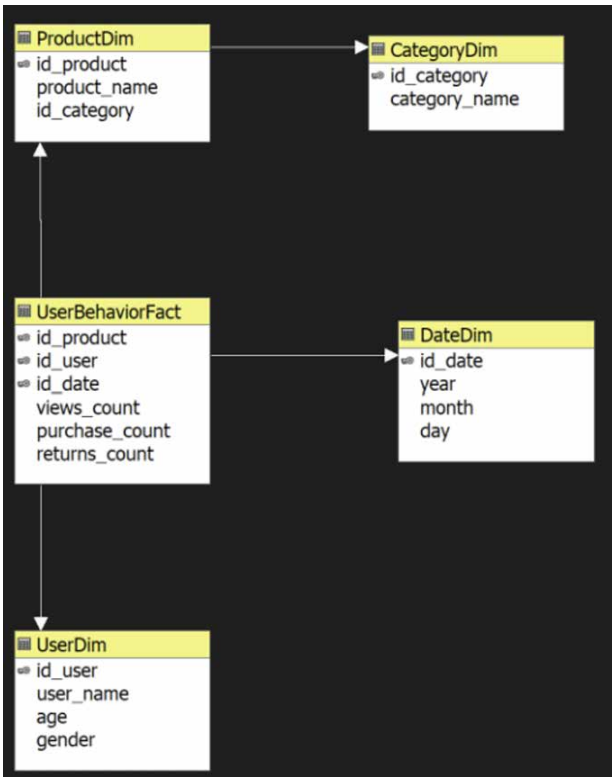

Рис. 3.12. Створений куб

3.5 Наповнення кубу даними

Реалізація процесу отримання, трансформації та завантаження даних у сховище виконується за допомогою модуля Data Flow у середовищі SQL Server Integration Services (SSIS) [21]. Цей процес забезпечує з'єднання з джерелами даних, їх обробку відповідно до вимог системи та завантаження у таблиці сховища даних.

Основні етапи реалізації Data Flow:

1. Створення нового пакета у SSIS: у середовищі Visual Studio створено новий проект Integration Services, у якому додано пакет для виконання ETL процесу. У пакеті реалізовано компонент Data Flow, який виконує всі необхідні дії для роботи з даними.
2. Налаштування джерела даних: для джерела даних обрано таблиці з оперативної бази даних (ОБД), які зберігають інформацію про активність користувачів, товари, категорії та час. Використано компонент OLE DB Source для встановлення з'єднання з базою даних та налаштування SQL-запитів для вибірки необхідних даних.
3. Завантаження даних у сховище: для завантаження даних у таблиці сховища даних використано компонент OLE DB Destination. Завантаження виконано в такі таблиці сховища: UserBehaviorFact, ProductDim, CategoryDim, UserDim та DateDim. Для забезпечення продуктивності та інтеграції даних було налаштовано механізм перевірки первинних та зовнішніх ключів.

На рисунках 3.13-3.15 представлено структуру Data Flow, що демонструє послідовність обробки даних: від джерела до завантаження даних для вимірів.

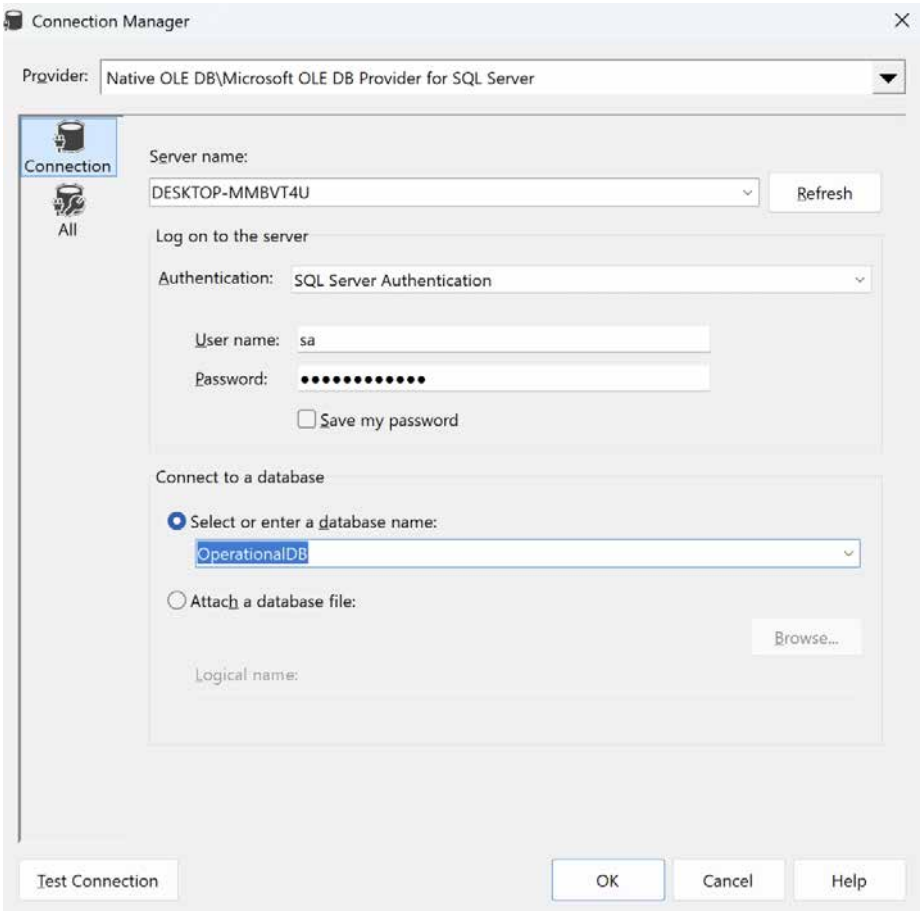

Рис. 3.13. Підключення оперативної БД

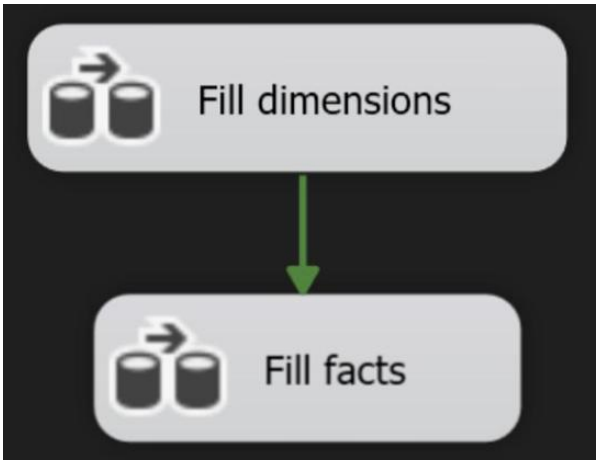

Рис. 3.14. Створення проєкту для наповнення даних

Рис. 3.15. Результат наповнення вимірів

Після завантаження даних у таблиці вимірів (ProductDim, CategoryDim, UserDim, DateDim) було виконано заповнення таблиці фактів UserBehaviorFact (рис. 3.16-3.18). Процес передбачав об'єднання зовнішніх ключів із відповідних таблиць вимірів та заповнення числових метрик, таких як views_count, purchase_count та returns_count, що характеризують активність користувачів, що дозволило створити завершену модель, придатну для багатовимірного аналізу.

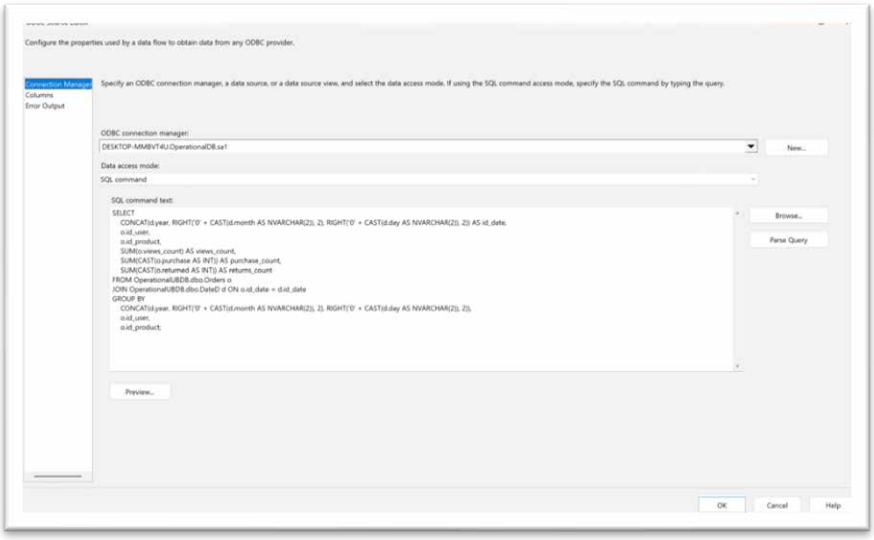

Рис. 3.16. Запит для наповнення таблиці фактів

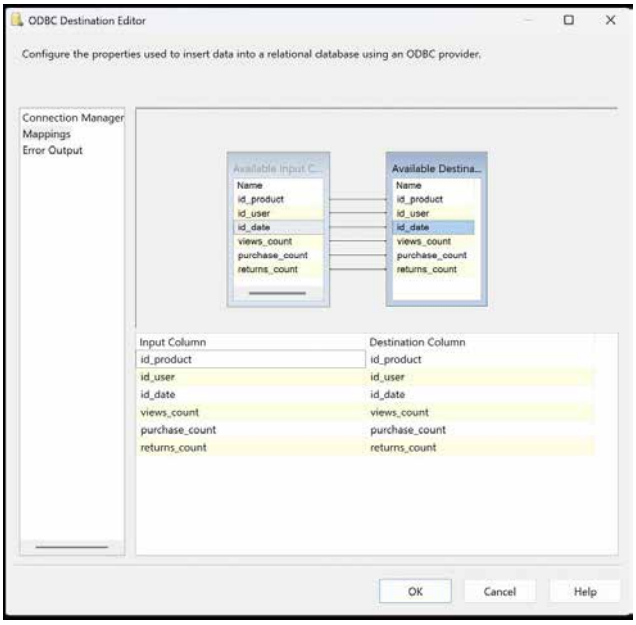

Рис. 3.17. Співставлення для перевірки

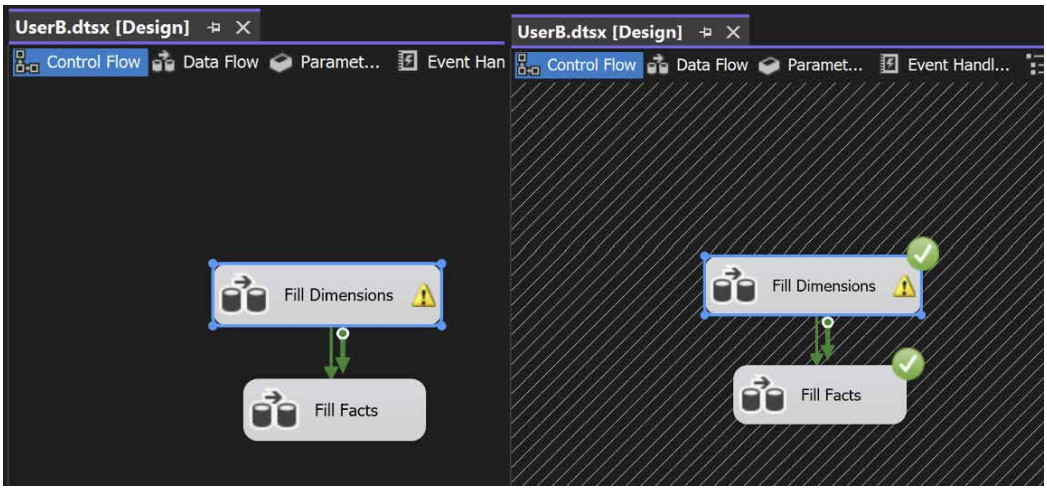

Рис. 3.18. Результат успішного наповнення факту та вимірів

У результаті реалізації ETL-процесу всі таблиці вимірів і фактів були успішно наповнені відповідними даними (рис. 3.19), що забезпечило готовність сховища до подальшого аналітичного опрацювання, побудови звітів і застосування методів інтелектуального аналізу.

```
USE UserBehaviorDW;
```

110 %

Results | Messages

	id_user	user_name	age	gender
1	1	Andrii	51	Male
2	2	Alina	48	Female
3	3	Andrii	19	Male
4	4	Oksana	29	Female
5	5	Yurii	58	Male
6	6	Yulia	51	Female
7	7	Yurii	41	Male
8	8	Anastasiia	30	Female
9	9	Yurii	18	Male
10	10	Oksana	48	Female
11	11	Andrii	36	Male
12	12	Natalia	36	Female
13	13	Maksym	30	Male
14	14	Olena	36	Female
15	15	Vladyslav	22	Male
16	16	Iryna	40	Female
17	17	Yurii	41	Male
18	18	Iryna	54	Female
19	19	Andrii	35	Male
20	20	Oksana	51	Female
21	21	Roman	41	Male
22	22	Nadiia	40	Female
23	23	Andrii	38	Male
24	24	Alina	44	Female
25	25	Oleksii	22	Male
26	26	Natalia	60	Female
27	27	Vladyslav	32	Male
28	28	Yulia	35	Female
29	29	Ihor	32	Male
30	30	Alina	37	Female
31	31	Maksym	40	Male
32	32	Olena	24	Female
33	33	Vladyslav	32	Male
34	34	Nadiia	58	Female
35	35	Taras	33	Male
36	36	Nadiia	40	Female
37	37	Vladyslav	36	Male
38	38	Natalia	34	Female
39	39	Vladyslav	52	Male
40	40	Alina	60	Female
41	41	Oleksii	51	Male
42	42	Oksana	58	Female
43	43	Oleksii	47	Male

Query executed successfully. DESKTOP-

Рис. 3.19. Приклад заповнення таблиці з даними про користувачів

4 АНАЛІЗ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ

4.1 Дослідження результатів розрахунку KPI

У сучасних інформаційних системах підтримки прийняття рішень ключову роль відіграє багатовимірний аналіз даних, який реалізується за допомогою технологій OLAP (Online Analytical Processing). Завдяки OLAP можливо досліджувати великі обсяги інформації з різних ракурсів (вимірів), швидко виявляючи закономірності та відхилення у динаміці показників. Такий підхід є особливо ефективним для обробки KPI (Key Performance Indicators) – ключових показників ефективності, які є основою оцінки бізнес-результатів та поведінки користувачів на платформі [22].

Вони дозволяють не лише оцінити загальний стан системи, а й деталізувати поведінкові шаблони користувачів, виявити вузькі місця в процесах взаємодії з товаром та надати ґрунтовані рекомендації для покращення бізнес-результатів. У OLAP-аналізі KPI виступають базовими агрегованими метриками, які візуалізуються через зрізи даних у кубі, що дозволяє швидко та гнучко отримувати аналітичну інформацію у багатовимірному просторі.

У межах даного дослідження було обрано та реалізовано кілька базових KPI (рис. 4.1), що дають уявлення про ефективність роботи окремих товарів та категорій (програмний код для процесу створення KPI можна переглянути в Додатку Б):

- **KPI_LowPurchaseRate (Низький рівень купівельної активності)**: цей показник базується на співвідношенні кількості переглядів до кількості фактичних покупок товару (рис. 4.1). Його мета – виявлення позицій, які активно розглядаються користувачами, але не викликають бажання придбати. Така ситуація може свідчити про недосконаліть в описі товару, неактуальну ціну або низьку

привабливість пропозиції. Аналіз цього KPI дозволяє маркетинговій команді оперативно реагувати – наприклад, оновлювати зображення, текстові описи, переглядати цінову політику чи запускати додаткові рекламні кампанії;

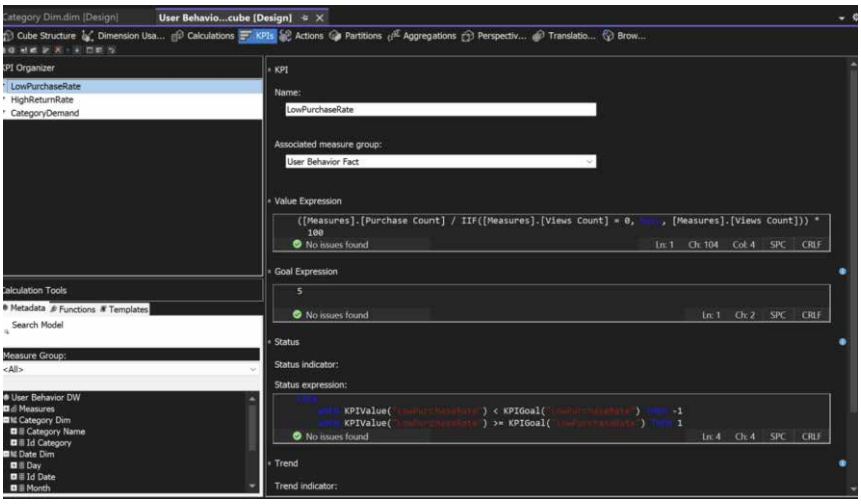

Рис. 4.1. KPI LowPurchaseRate

- **KPI_HighReturnRate (Високий відсоток повернень)**: формується на основі аналізу частки повернених товарів відносно загальної кількості продажів. Його підвищені значення можуть свідчити про низьку якість продукції, розбіжність між очікуваннями клієнтів і фактичними характеристиками товару, або про неналежне інформування споживача. KPI такого типу допомагає зосередити увагу на продуктах, що викликають незадоволення, і вжити заходів щодо зменшення частоти повернень (рис. 4.2);

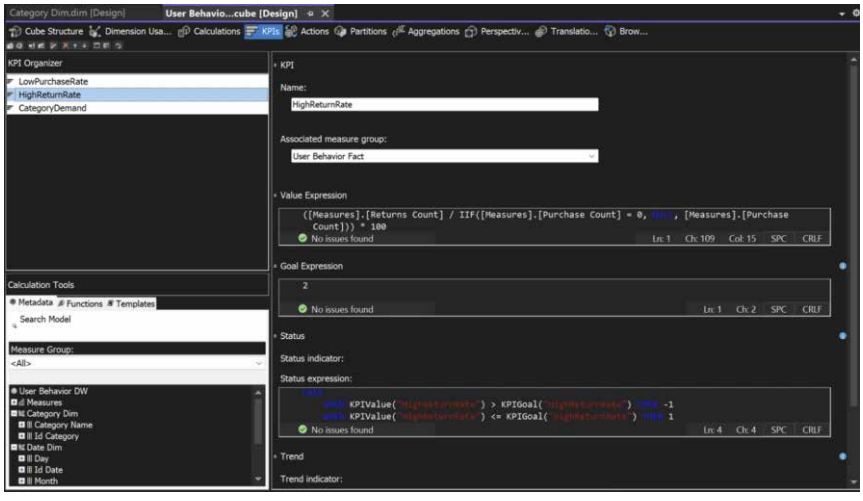

Рис. 4.2. KPI HighReturnRate

- **KPI_CategoryDemand (Попит на категорію товарів):** оцінює загальну кількість покупок у межах кожної категорії. Даний показник дозволяє визначити найбільш популярні категорії на платформі, порівняти їх динаміку між собою, а також ідентифікувати категорії з потенційним зростанням або ті, що втрачають актуальність. Такий аналіз є надзвичайно важливим для планування асортименту, логістики та ефективного розміщення акцентів у рекламних кампаніях (рис. 4.3).

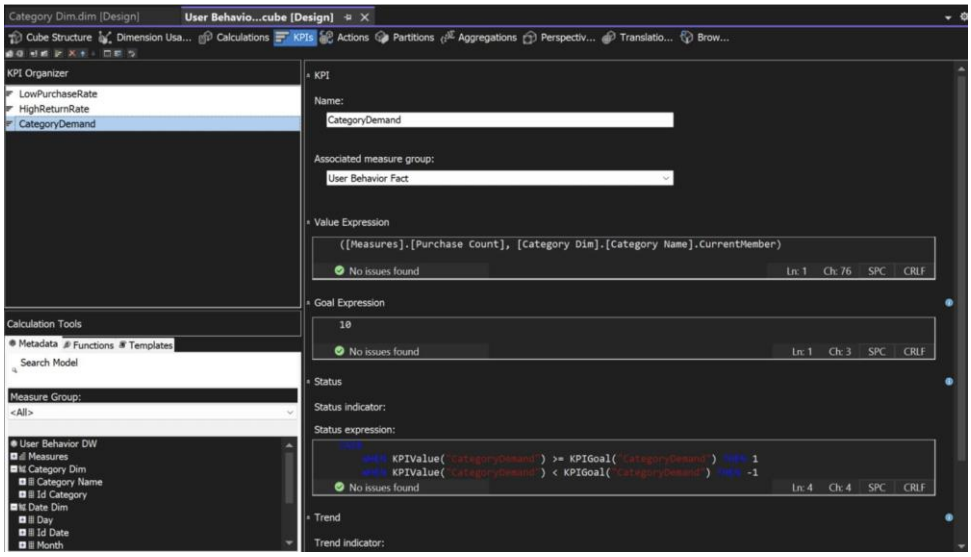

Рис. 4.3. KPI CategoryDemand

Візуалізація результатів дозволила підтвердити наявність типових закономірностей: товари з високою залученістю (переглядами), але низькими продажами, категорії з підвищеним рівнем повернень, а також лідери продажів серед категорій, які користуються найбільшим попитом (рис. 4.4).

Display Structure	Value	Goal	Status
CategoryDemand	2	10	
HighReturnRate	50	2	
LowPurchaseRate	2.63157894736842	5	

Рис. 4.4. Результати розрахунків KPI

Загалом застосування KPI у межах OLAP-аналізу дозволяє перетворити великі обсяги поведінкових даних на чіткі аналітичні висновки, що мають практичну цінність для вдосконалення роботи комерційної платформи.

4.2 Оцінка результатів OLAP-звітності

У межах цього дослідження для реалізації системи звітності було використано платформу SQL Server Reporting Services (SSRS) [23]. Даний інструмент забезпечив створення інтерактивних і наочних звітів на основі побудованого OLAP-куба, що дозволило виявити приховані закономірності у поведінці користувачів комерційної платформи.

SSRS надала можливість швидко формувати динамічні візуалізації з використанням різних рівнів агрегації даних, фільтрації та параметризації, що значно спростило аналіз багатовимірних даних, збережених у сховищі. У звітах було реалізовано стандартні OLAP-операції: drill-down (поглиблення до рівня окремих категорій та користувачів), roll-up (агрегація до категорій), slice (відбір за певним атрибутом, наприклад, стать) та dice (перехресна фільтрація по кількох вимірах одночасно).

Одним із прикладів реалізації є графік порівняння частоти переглядів і покупок у різних категоріях товарів (рис 4.5).

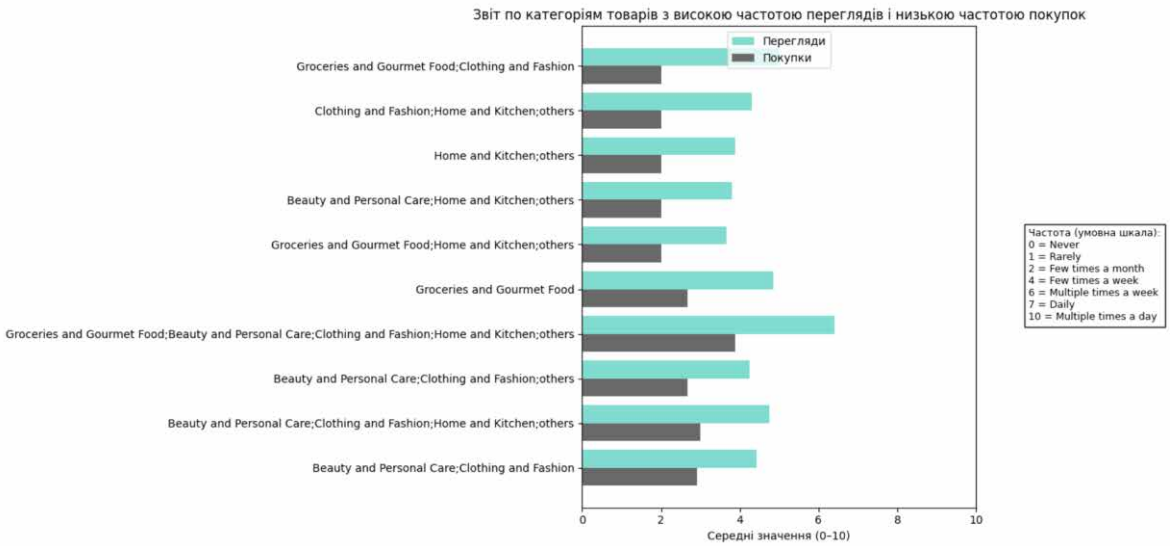

Рис. 4.5. Звіт по категоріям товарів з високою частотою переглядів і низькою частотою покупок

Цей звіт виявив категорії з високим рівнем зацікавленості, але з низькою конверсією, що сигналізує про можливі бар'єри для покупки. Такі товари

можуть потребувати оновлення візуального контенту, опису або перегляду цінової політики.

Крім цього, було створено два окремі кругові звіти для гендерного аналізу, які відображають найпопулярніші категорії серед чоловіків і жінок (рис. 4.6).

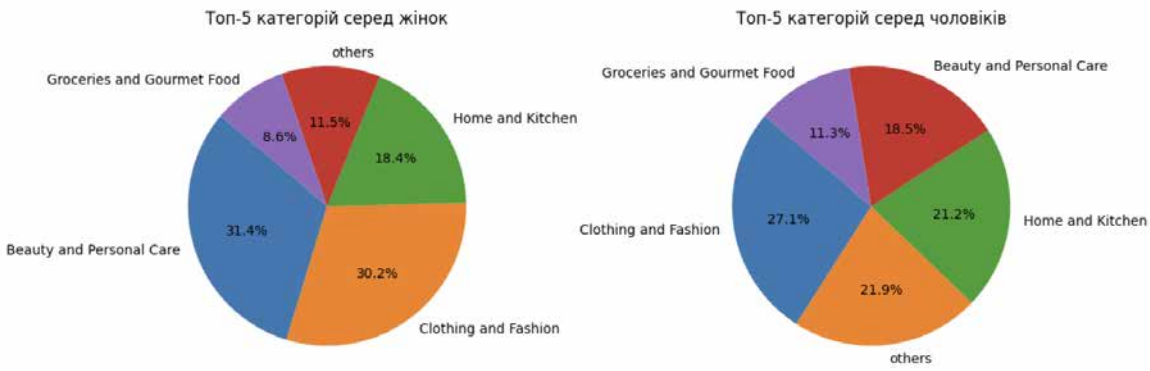

Рис. 4.6. Звіт по найпопулярнішим категоріям товарів серед жінок та чоловіків

Наприклад, для жінок найбільшим попитом користувались категорії Beauty and Personal Care та Clothing and Fashion, тоді як чоловіки віддавали перевагу Home and Kitchen та Electronics. Така деталізація є критично важливою для розробки персоналізованих стратегій просування та вдосконалення рекомендаційних механізмів.

Усі побудовані звіти інтегрувалися безпосередньо з джерелом даних, що дозволяло забезпечити їх актуальність, автоматичне оновлення і гнучкість у подальшому аналізі.

Підбиваючи підсумки, OLAP-звіти в SSRS стали невіддільним інструментом дослідження, дозволивши не лише візуалізувати ключові KPI, але й здійснити повноцінний аналіз поведінкових шаблонів користувачів у багатовимірному просторі.

4.3 Класифікація за методом 1-Rule та інтерпретація результатів

Класифікація – одна з ключових задач інтелектуального аналізу даних, що передбачає визначення належності об'єкта до одного з наперед визначених класів на основі його характеристик. У межах цього дослідження для класифікації активності користувачів було застосовано алгоритм 1-Rule (1R) – простий, але ефективний підхід, який формує класифікаційні правила на основі лише одного атрибуту [24].

Даний алгоритм працює за принципом побудови окремих правил для кожного значення обраного атрибуту, після чого обирається та змінна, яка забезпечує найменшу кількість класифікаційних помилок. Незважаючи на свою простоту, 1R часто демонструє хороші результати при попередньому аналізі даних і використовується як базовий підхід для порівняння зі складнішими моделями.

Для реалізації було створено окремий Windows Forms-додаток на мові C# у середовищі Visual Studio. Програма підключалась до сховища UserBehaviorDW, звідки виконувалося завантаження даних з таблиці фактів UserBehaviorFact та приєднання до неї таблиць вимірів: DateDim, UserDim, ProductDim, CategoryDim.

У результаті кожен запис, що аналізувався, містив такі параметри:

- кількість переглядів (views_count);
- стать користувача (gender);
- назву товару (product_name);
- категорію товару (category_name);
- дату дії (розбито на рік, місяць і день).

На основі кількості переглядів усі записи було розділено на три класи активності користувачів:

- Low – якщо views_count ≤ 30;

- Medium – якщо views_count ≤ 60;
- High – якщо views_count > 60.

Дане значення було обрано як залежну змінну, яку має передбачати алгоритм. У самій реалізації 1R алгоритм послідовно перевіряє кожен атрибут (незалежну змінну), групуючи дані за його значеннями. Для кожної групи визначається найпоширеніший клас (так званий MajorityClass), після чого рахується точність – частка правильно класифікованих записів.

У даній реалізації аналіз виконувався на основі таких атрибутів: назва товару (product_name), категорія товару(category_name), стать користувача (gender), місяць взаємодії (month).

Для кожного значення атрибутів програма формує правило класифікації. Наприклад:

- Якщо product_name = "Razer BlackWidow", то клас = Low (60%);
- Якщо month = 11, то клас = Low (65,1%).

Усі результати відображаються у таблиці dataGridView, яка показує значення атрибуту, передбачений клас, кількість правильних класифікацій і точність у відсотках (рис. 4.7-4.9).

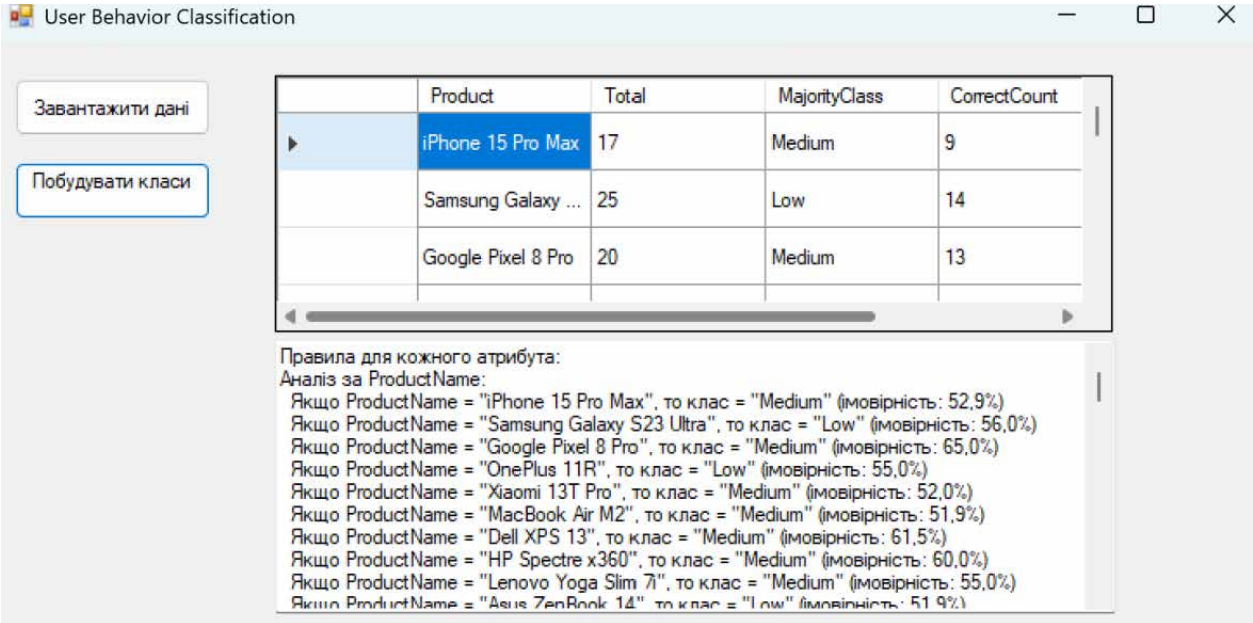

Рис. 4.7. Результати аналізу 1-Rule за назвою товару

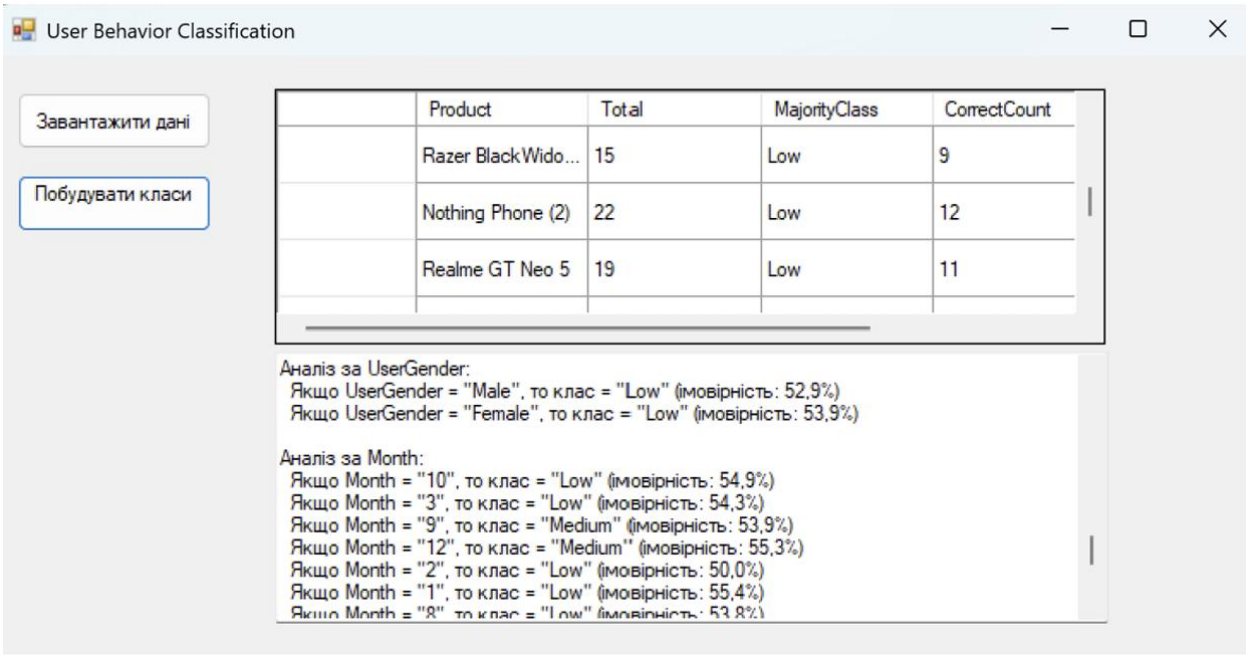

Рис. 4.8. Результати аналізу за атрибутами: стать та місяць

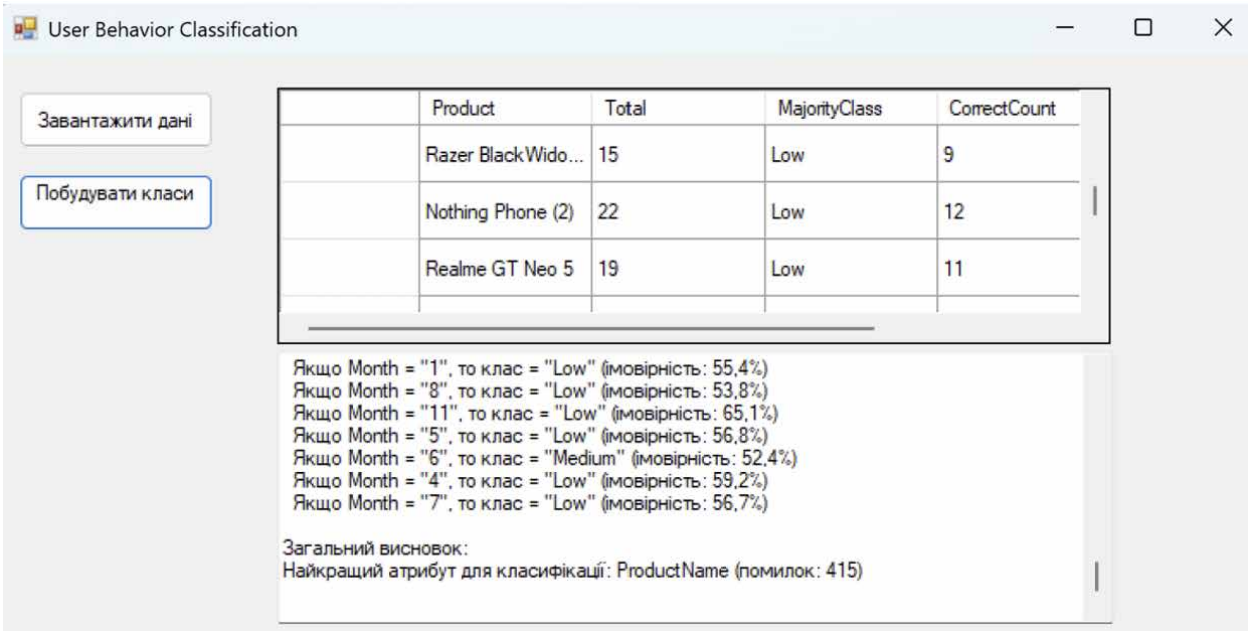

Рис. 4.9. Загальні висновки по класифікації

Такий підхід дає змогу швидко побачити, які характеристики є більш інформативними з точки зору класифікації.

За підсумками аналізу, найкращим атрибутом для класифікації виявилась назва товару (product_name), оскільки для цього параметра точність є найвищою серед усіх перевірених змінних.

4.4 Аналіз підсумків класифікації за методом наївного Байеса

Метод Наївного Байеса є ймовірнісним алгоритмом класифікації, що ґрунтується на теоремі Байєса та припущенні про незалежність ознак між собою, що означає, що кожен атрибут чи ознака впливає на результат незалежно від інших [25].

Попри свою «наївність», метод демонструє високу ефективність у багатьох практичних задачах – зокрема в текстовій аналітиці, фільтрації спаму, прогнозуванні поведінки користувачів тощо. Його основні переваги – швидкість, простота та інтерпретованість результатів.

Суть методу полягає в тому, що для кожного класу обчислюється: апріорна ймовірність класу (ймовірність появи класу в цілому) та умовна ймовірність кожного атрибута при умові, що об’єкт належить до цього класу. Далі ці ймовірності комбінуються, і клас із найбільшою підсумковою ймовірністю обирається як прогнозований (демонстрація програмного коду реалізації методу надано в Додатку В).

При виконанні роботи для реалізації методу були використані два підходи: через Python та десктопний застосунок на C# (Windows Forms). Обидва варіанти працювали з однаковими даними, отриманими з аналітичної бази UserBehaviorDW. Дані для аналізу формувались на основі таких параметрів:

- назва товару (product_name);
- категорія товару (category_name);
- стать користувача (gender);
- місяць, коли відбулася взаємодія (month);
- кількість переглядів (views_count), за якою визначався рівень активності.

Як і в попередній класифікації, рівень активності було поділено на три

класи:

- Low – до 10 переглядів;
- Medium – від 11 до 30;
- High – більше 30.

Даний показник став цільовим класом, який потрібно було передбачити. У Python реалізація базувалась на побудові частотної моделі: з кожної унікальної комбінації атрибутів (наприклад, "Ноутбук ASUS" – "Електроніка" – "Жіноча" – "Січень") формувалися списки класів активності, а далі обчислювались ймовірності належності до кожного з класів. Враховувались як апріорні ймовірності (частка кожного класу у загальній вибірці), так і умовні (наприклад, ймовірність того, що користувач-жінка, яка переглядає товар у січні, буде в класі Medium).

Кожна комбінація атрибутів отримувала прогнозований клас (PredictedClass) та відповідні ймовірності P(Low), P(Medium), P(High). Наприклад, для комбінації «iPhone 15 Pro Max – Smartphones – Male – 10 (November)» модель повертала: PredictedClass = Medium, P(Low) = 41.87%, P(Medium) = 58.13%, P(High) = 00.00%.

Такий підхід дозволив не тільки класифікувати, а й оцінити ступінь впевненості моделі у кожному рішенні відповідно до отриманих результатів (рис. 4.10) .

```
C:\Users\User\Desktop\Studies\NULES\2 семестр\Data mining\naive_bayes_user_behavior.py:30: UserWarning: pandas only supports SQL
Alchemy connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2 objects are not tested
. Please consider using SQLAlchemy.
  df = pd.read_sql(query, conn)
Витягнуті перші 5 записів з даних:
        product_name category_name  gender  month  views_count
0  iPhone 15 Pro Max   Smartphones    Male     10           16
1  iPhone 15 Pro Max   Smartphones  Female      3           14
2  iPhone 15 Pro Max   Smartphones  Female      9           15
3  iPhone 15 Pro Max   Smartphones  Female     12           10
4  iPhone 15 Pro Max   Smartphones  Female      2            3

 Результати класифікації:
          Product     Category  Gender Month PredictedClass P(Low) P(Medium) P(High)
0   iPhone 15 Pro Max  Smartphones    Male    10         Medium  41.87     58.13    0.00
1   iPhone 15 Pro Max  Smartphones    Male     3         Medium  41.35     58.65    0.00
2   iPhone 15 Pro Max  Smartphones    Male     9         Medium  33.59     66.41    0.00
3   iPhone 15 Pro Max  Smartphones    Male    12         Medium  32.34     67.66    0.00
4   iPhone 15 Pro Max  Smartphones    Male     2         Medium  37.20     62.80    0.00
5   iPhone 15 Pro Max  Smartphones    Male     1         Medium  42.41     57.59    0.00
6   iPhone 15 Pro Max  Smartphones    Male     8         Medium  40.86     59.14    0.00
7   iPhone 15 Pro Max  Smartphones    Male    11            Low  52.45     47.55    0.00
8   iPhone 15 Pro Max  Smartphones    Male     5         Medium  43.77     56.23    0.00
9   iPhone 15 Pro Max  Smartphones    Male     6         Medium  35.00     65.00    0.00
10  iPhone 15 Pro Max  Smartphones    Male     4         Medium  46.23     53.77    0.00
11  iPhone 15 Pro Max  Smartphones    Male     7         Medium  43.70     56.30    0.00
12  iPhone 15 Pro Max  Smartphones  Female    10         Medium  42.89     57.11    0.00
13  iPhone 15 Pro Max  Smartphones  Female     3         Medium  42.37     57.63    0.00
14  iPhone 15 Pro Max  Smartphones  Female     9         Medium  34.53     65.47    0.00
15  iPhone 15 Pro Max  Smartphones  Female    12         Medium  33.26     66.74    0.00
16  iPhone 15 Pro Max  Smartphones  Female     2         Medium  38.18     61.82    0.00
17  iPhone 15 Pro Max  Smartphones  Female     1         Medium  43.43     56.57    0.00
18  iPhone 15 Pro Max  Smartphones  Female     8         Medium  41.87     58.13    0.00
19  iPhone 15 Pro Max  Smartphones  Female    11            Low  53.48     46.52    0.00
20  iPhone 15 Pro Max  Smartphones  Female     5         Medium  44.80     55.20    0.00
21  iPhone 15 Pro Max  Smartphones  Female     6         Medium  35.95     64.05    0.00
22  iPhone 15 Pro Max  Smartphones  Female     4         Medium  47.27     52.73    0.00
```

Рис. 4.10. Результати отриманні через Python

Аналогічна логіка була реалізована і в C# з виведенням результатів у DataGridView. Перевага цього підходу – зручний графічний інтерфейс для перегляду результатів (рис. 4.11), можливість взаємодії з кнопками («Завантажити дані», «Класифікувати»), а також швидке тестування без запуску зовнішніх скриптів.

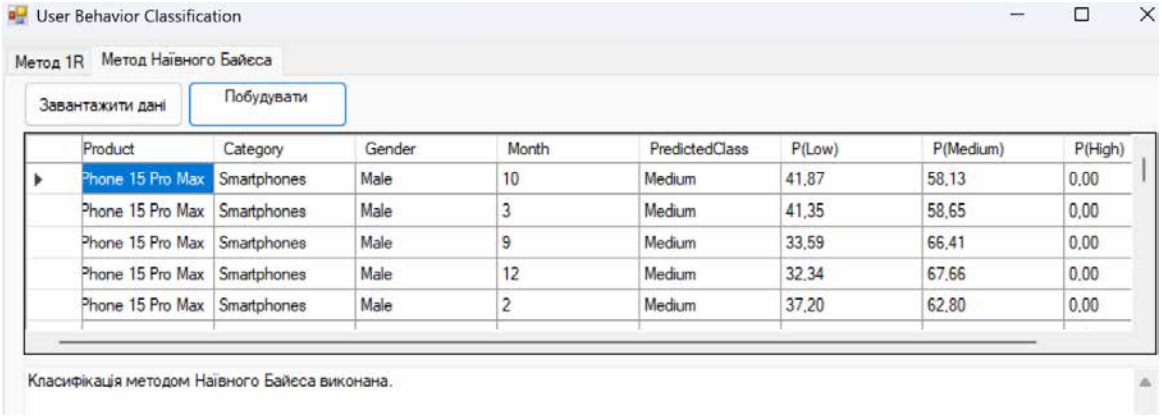

Рис. 4.11. Результати отримані через C#

Обидві реалізації надали однакові результати і показують стабільне визначення класу.

4.5 Аналіз асоціативних правил для виявлення поведінкових залежностей

Метод асоціативних правил є підхідом Data Mining для виявлення закономірностей типу $A \rightarrow B$ у великих масивах даних. Він відповідає на запитання «що часто відбувається разом із чим» (напр., після перегляду певної категорії користувачі часто переходять до пов’язаних товарів) [26].

Силу й корисність правил оцінюють трьома метриками: support (частота спільної появи A і B), confidence (ймовірність B за умови A) та lift (ступінь залежності; >1 – позитивний зв’язок).

У межах магістерської роботи метод застосовано практично: із сховища UserBehaviorDW (SQL Server) через pyodbc (ODBC Driver 17) сформовано вибірку, що з’єднує UserBehaviorFact з ProductDim, CategoryDim, UserDim,

DateDim; використано чотири ознаки – product_name, category_name, gender, month. Кожен рядок трактується як транзакція; дані перетворено у формат «присутність/відсутність» за допомогою TransactionEncoder (DataFrame). Далі в Python застосовано Apriori з mlxtend.frequent_patterns: у базовому режимі min_support = 0.10 (фокус на найпоширеніших комбінаціях), у розширеному – min_support = 0.05 (пошук нішевих зв’язків). Правила згенеровано функцією association_rules з метрикою confidence та порогом min_threshold 0.30→0.20; вивід – у читабельному форматі «[умова] → [наслідок], довірчість: X.XX» із зазначенням загальної кількості правил. Увесь процес автоматизовано, параметри легко змінюються без переписування логіки, що робить рішення гнучким і готовим до подальшої інтеграції в Data Mining-процес.

У результаті реалізації було сформовано набір закономірностей, які демонструють повторювані шаблони у даних, що описують поведінку користувачів. Отримані правила мають вигляд залежностей між такими атрибутами, як назва товару, категорія товару, стать користувача та місяць взаємодії.

Асоціативні правила генерувалися на основі попередньо побудованих частих наборів ознак (frequent itemsets), при цьому застосовувалися різні комбінації значень підтримки (support) та довірчості (confidence), що дозволило оцінити не лише загальні шаблони поведінки, але й виявити більш рідкісні та неочевидні зв’язки.

У першому варіанті реалізації використовувалися більш суворі обмеження: min_support = 0.1, confidence = 0.3. У результаті було сформовано невелику, але інформативну вибірку асоціативних правил (рис. 4.12).

```
C:\Users\User\Desktop\Studies\NULES\2 семестр\Data mining\lab5_apriori.py:31: UserWarning: pandas only supports SQLAlche
my connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2 objects are not tes
ted. Please consider using SQLAlchemy.
  df = pd.read_sql(query, conn)
Знайдені асоціативні правила (mlxtend):
Правило: [Gaming Accessories] →[Female], довірчість: 0.51
Правило: [Smartphones] →[Female], довірчість: 0.57
Правило: [Gaming Accessories] →[Male], довірчість: 0.49
Правило: [Laptops] →[Male], довірчість: 0.53
```

Рис. 4.12. Перша вибірка асоціативних правил

Типовими прикладами стали правила, які пов’язують загальні характеристики користувачів з популярними категоріями товарів. Завдяки високим значенням достовірності ці правила мають практичну цінність і можуть слугувати основою для побудови базових рекомендаційних механізмів або оглядової аналітики.

Другий варіант реалізації був спрямований на розширення глибини аналізу. Було знижено пороги: min_support = 0.05, confidence = 0.2, що дозволило виявити значно більшу кількість асоціативних зв’язків (рис. 4.13), у тому числі менш очевидні або специфічні комбінації.

```
C:\Users\User\Desktop\Studies\NULES\2 семестр\Data mining\lab5_apriori.py:31: UserWarning:
my connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Oth
ted. Please consider using SQLAlchemy.
  df = pd.read_sql(query, conn)
Знайдені асоціативні правила (mlxtend):
Правило: [02] →[Female], довірчість: 0.61
Правило: [03] →[Female], довірчість: 0.57
Правило: [06] →[Male], довірчість: 0.63
Правило: [12] →[Female], довірчість: 0.5
Правило: [12] →[Male], довірчість: 0.5
Правило: [Gaming Accessories] →[Female], довірчість: 0.51
Правило: [Female] →[Gaming Accessories], довірчість: 0.21
Правило: [Home Appliances] →[Female], довірчість: 0.5
Правило: [Laptops] →[Female], довірчість: 0.47
Правило: [Smartphones] →[Female], довірчість: 0.57
Правило: [Female] →[Smartphones], довірчість: 0.24
Правило: [Tablets] →[Female], довірчість: 0.51
Правило: [Gaming Accessories] →[Male], довірчість: 0.49
Правило: [Male] →[Gaming Accessories], довірчість: 0.21
Правило: [Home Appliances] →[Male], довірчість: 0.5
Правило: [Laptops] →[Male], довірчість: 0.53
Правило: [Male] →[Laptops], довірчість: 0.23
Правило: [Smartphones] →[Male], довірчість: 0.43
Правило: [Tablets] →[Male], довірчість: 0.49

Загальна кількість знайдених правил: 19
```

Рис. 4.13. Розширена версія вибірки асоціативних правил

Незважаючи на нижчі значення довірчості, отримані правила несуть додаткову аналітичну цінність і можуть бути корисними для сегментації аудиторії або глибокого аналізу нетипових сценаріїв.

Порівнюючи обидва підходи можна зробити кілька висновків:

- використання високих порогів підтримки й довірчості забезпечує стислий і точний набір правил, придатний для звітів або інтеграції в системи з обмеженим обсягом рекомендацій;
- зниження порогів дозволяє отримати ширший спектр знань, що особливо корисно на етапах дослідження або побудови більш

адаптивних моделей поведінки користувачів.

В свою чергу комбінація обох підходів дає змогу досягти балансу між точністю та повнотою, обираючи залежно від поставлених завдань найрелевантніші правила. Таким чином, проведений асоціативний аналіз продемонстрував потенціал методу Apriori у виявленні структурованих шаблонів у поведінці, а також підтвердив доцільність використання гнучких параметрів для адаптації глибини дослідження.

Надалі ці результати можуть бути використані для побудови систем рекомендацій, зокрема в комбінації з кластеризацією чи класифікаційними моделями.

4.6 Висновки кластерного аналізу

Кластерний аналіз є методом, який дозволяє автоматично розподіляти об’єкти на групи (кластери) за схожістю певних характеристик.

Головна мета – знайти всередині великого обсягу даних природні групи, де елементи кожної групи мають більше спільного між собою, ніж із рештою. При цьому наперед не потрібно знати, до яких саме груп належать об’єкти, оскільки система виявляє їх самостійно.

Кластеризація належить до методів неконтрольованого навчання, адже алгоритм не має заздалегідь заданих класів і формує їх на основі аналізу самих даних, що робить метод особливо цінним для дослідження нових наборів даних, виявлення прихованої структури, а також для підготовки вхідних даних для інших алгоритмів.

У межах цього етапу дослідження було реалізовано задачу кластерного аналізу користувачів на основі даних із сховища UserBehaviorDW.

Для цього було використано реалізацію на мові програмування Python з бібліотеками для аналізу даних та машинного навчання, такими як pandas, scikit-learn, matplotlib, seaborn та pyodbc.

На першому кроці з сховища даних за допомогою SQL-запиту було сформовано набір даних, який включав наступні ознаки:

- кількість переглядів товарів (views_count);
- кількість унікальних категорій, з якими взаємодіяв користувач (unique_categories);
- стать користувача (gender, у форматі 1 – жінка, 0 – чоловік);
- місяць активності (month).

Представлені дані були обрані як ключові характеристики, що потенційно можуть відображати поведінкову модель користувача в системі. Перед кластеризацією було виконано масштабування ознак (standardization), щоб жодна з них не домінувала над іншими через різний масштаб. Після цього для визначення оптимальної кількості кластерів було застосовано метод “лікоть”. Побудувавши графік (рис. 4.14) інерції залежно від кількості кластерів, було встановлено, що найкраще рішення забезпечує поділ на три кластери.

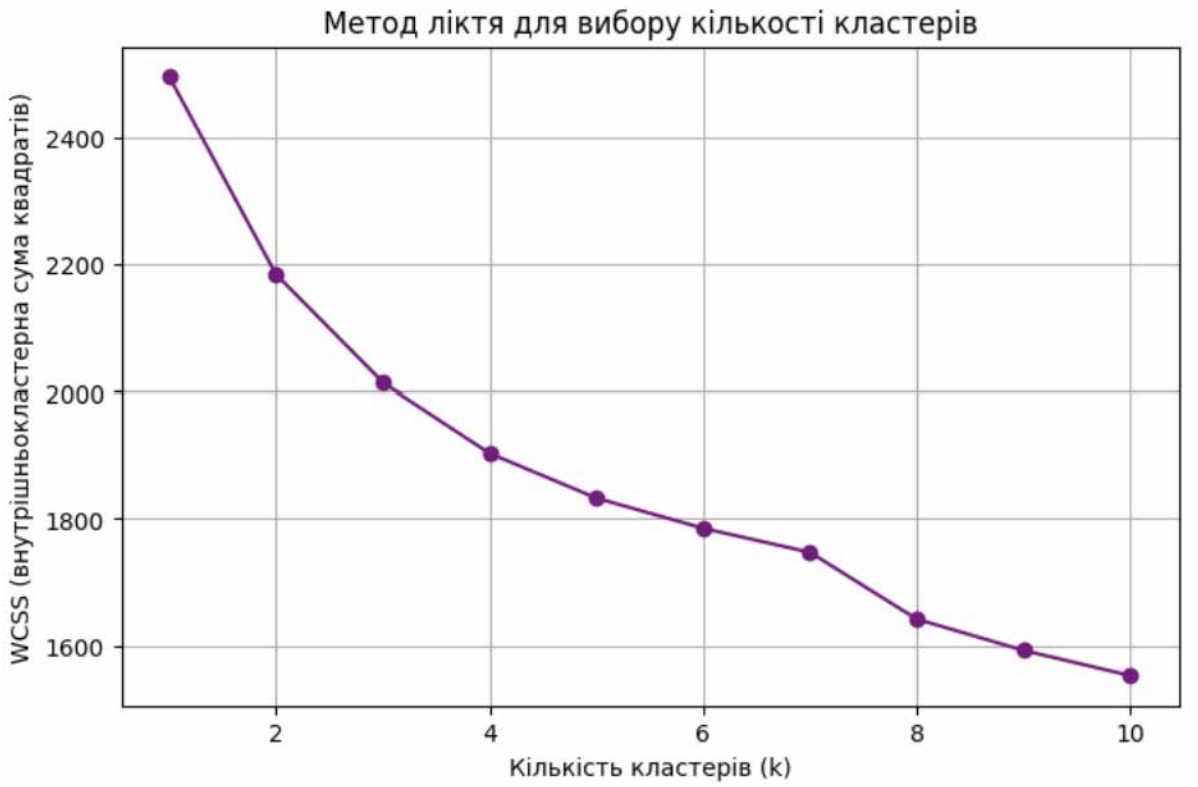

Рис. 4.14. Графік методу ліктя для вибору кількості кластерів

Для кластеризації застосовувався алгоритм K-Means, який автоматично розподілив усіх користувачів у три групи на основі обчисленої схожості за заданими параметрами.

Отримані кластери спочатку мали номери від 0 до 2, але з метою кращого розуміння результатів їм було присвоєно більш змістовні назви. Назви були визначені за допомогою аналізу центроїдів кластерів – тобто середніх значень ознак у кожній групі. Зокрема, було виокремлено такі групи:

- Активні користувачі: група включає користувачів, які активно взаємодіють із платформою;
- Пасивні користувачі: до цього кластеру віднесені користувачі, які є менш зацікавленими або випадковими відвідувачами;
- Потенційні користувачі: користувачі, які демонструють посередню активність. Вони періодично взаємодіють із функціональністю, але ще не сформували стійкої моделі поведінки.

Для кращої інтерпретації та візуалізації результатів кластеризації було застосовано PCA (метод головних компонент), котрий дозволив зменшити розмірність даних до двох вимірів і побудувати зрозумілий графік розподілу об’єктів на площині. Зменшення кількості вимірів при цьому не вплинуло суттєво на структуру даних, однак зробило її більш наочною та придатною для подальшого аналізу.

Завдяки такій трансформації з’явилася можливість візуально оцінити не лише ступінь відокремленості кластерів, а й щільність розташування об’єктів усередині кожної групи.

Побудований графік чітко демонструє, наскільки добре алгоритм K-середніх розділив користувачів за поведінковими характеристиками.

На графіку позначені точки, що представляють користувачів з різними кольорами відповідно до кластера, а також центри кластерів виділені

помаранчевими хрестиками (рис. 4.15). Такий підхід значно полегшує інтерпретацію кластеризації навіть у випадку багатовимірних вхідних даних.

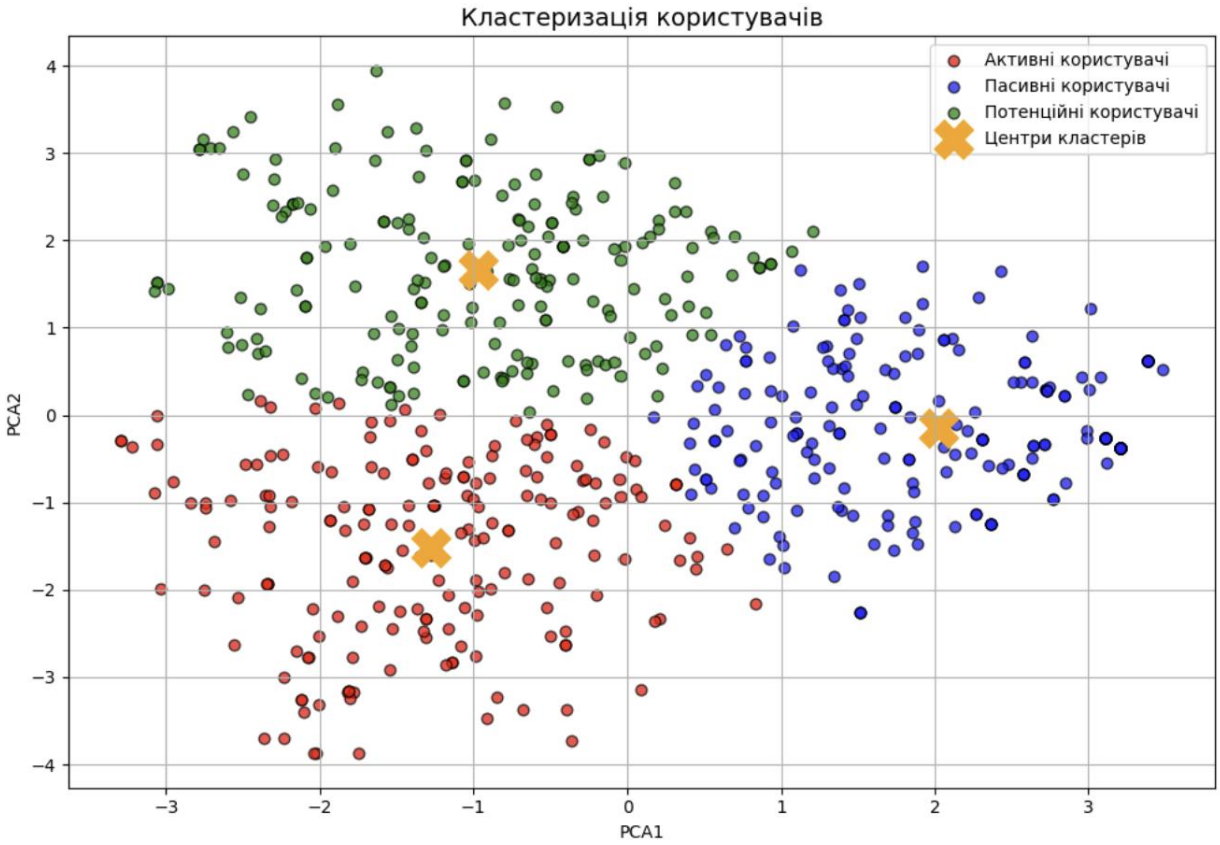

Рис. 4.15. Отримані результати кластеризації користувачів

На основі проведеного кластерного аналізу було виділено три основні групи користувачів комерційних платформ, які мають відмінні поведінкові характеристики:

- Кластер 0 – Активні користувачі (червоний кластер): велика кількість переглядів, інтерес до кількох категорій. Вони ключова активна аудиторія, яка найімовірніше буде готова до покупок або підписок. Варто зосередити на ній таргетовані акції та VIP-функціонал в реальних комерційних платформах;
- Кластер 1 – Пасивні користувачі (синій кластер): об’єднав малозалучених користувачів, які демонструють низьку активність. Для цієї групи доцільно було б застосовувати окремі мотиваційні

механізми: наприклад, пуш-нагадування або спрощені onboarding-сценарії;

- Кластер 2 – Потенційні користувачі (зелений кластер): демонструє користувачів з піковою активністю в певні місяці, що вказує на сезонну поведінку. Для цієї групи ефективними будуть тимчасові пропозиції, обмежені за часом розсилки та спецпропозиції до свят.

Така візуалізація дозволила наочно побачити чітке розділення користувачів за типами поведінки. Центри кластерів, позначені хрестиками, допомагають зрозуміти середні характеристики кожної групи. Таке групування є корисним для подальшої персоналізації рекомендацій, створення рекламних кампаній та загального покращення взаємодії користувачів із платформою.

ВИСНОВКИ

У результаті виконання дослідження в межах написання магістерської роботи було реалізовано повноцінну інтелектуальну систему для аналізу поведінки користувачів на комерційних платформах, яка охоплює повний цикл збору, обробки, зберігання, моделювання та візуалізації даних.

Система ґрунтується на сучасних підходах до створення сховищ даних, аналітики KPI та застосування методів Data Mining, що дозволяє трансформувати первинну інформацію про взаємодію користувачів із платформою у стратегічно цінні інсайти.

Побудоване сховище UserBehaviorDW забезпечило можливість централізованого зберігання структурованих даних про перегляди, покупки та повернення, товарні категорії, гендерні характеристики та часові параметри поведінки користувачів. Завдяки реалізованій зірковій схемі моделювання та узгодженим ключам між таблицями фактів і вимірів, досягнуто гнучкості в побудові аналітичних запитів та звітів. Внутрішня структура сховища дозволяє масштабування і доповнення новими ознаками у разі зміни потреб системи.

На етапі розрахунку ключових показників ефективності (KPI) було визначено низку метрик, що характеризують загальну активність користувачів. Побудовані ж звіти дозволили виокремити сегменти з низькою частотою покупок, проте при цьому достатньо великій кількості переглядів на платформі та виявити шаблонні зацікавленості в категоріях в залежності від того чи клієнт є чоловіком, чи жінкою. Такі дані можуть бути безпосередньо використані, наприклад для таргетованих маркетингових кампаній або можливого покращення рекомендаційної системи платформи.

Класифікація за алгоритмом 1-Rule, в свою чергу, виявила, що найінформативнішим атрибутом у контексті рівня активності користувача є конкретні товари. Наприклад, моделі iPhone 15 Pro Max найчастіше з’являлись у групі з високою активністю, що дозволяє зробити припущення про вплив

преміальних товарів на залучення аудиторії, що відкриває можливості для стратегічного просування конкретних товарів із високим потенціалом конверсії.

Модель наївного Байєса продемонструвала високий рівень точності при класифікації користувачів за рівнем активності (Low, Medium, High), з урахуванням поєднання атрибутів категорії, статі, місяця взаємодії та товару. Це дозволяє не лише поділяти існуючу базу за подібними класами, але й прогнозувати потенційну поведінку нових користувачів, що може бути інтегровано у рекомендательні системи чи email-стратегії.

Асоціативний аналіз з використанням алгоритму Apriori дозволив виявити часті поєднання характеристик, які найчастіше трапляються разом у поведінці користувачів. Наприклад, було встановлено, що чоловіки, які переглядали технічні категорії у зимові місяці, з високою ймовірністю виявляли інтерес до нових релізів смартфонів. Такі правила можуть використовуватись для побудови динамічних блоків «цікаве для вас» або налаштування банерної реклами в подальшому.

Завдяки ж кластерному аналізу на основі алгоритму K-середніх користувачі були поділені на три природні групи: активні, пасивні та потенційні. Розрахунок центроїдів підтвердив, що найбільш перспективною для впливу є третя група – потенційні користувачі, які демонструють посередню активність і можуть бути залучені до більш глибокої взаємодії за допомогою персоналізованих механік.

Кластеризація була візуалізована через зниження розмірності методом головних компонент (PCA), що дозволило аналітично оцінити щільність, віддаленість і перетин сегментів, а також сформувати гіпотези для подальшого A/B тестування.

Таким чином, побудована система довела свою здатність інтегрувати та інтерпретувати багатовимірні поведінкові дані, з високим потенціалом для практичного застосування у реальних умовах функціонування онлайн-платформ. Результати дослідження можуть бути основою для впровадження

аналітичного модуля в реальних системах електронної торгівлі, а також для навчання моделей прогнозування відтоку клієнтів, динаміки продажів або попиту на нові товари. Особливо важливою є можливість повторного використання та масштабування рішення для суміжних галузей – від освітніх платформ до стримінгових сервісів.

Запропонований підхід є прозорим і відтворюваним, що робить його цінним для менеджерів, які ухвалюють рішення на основі даних. Усі обчислення, класифікації та візуалізації виконано з дотриманням принципів пояснюваної аналітики, що забезпечує довіру до результатів. Тому зважаючи на стрімке зростання кількості комерційних онлайн-платформ, інтелектуальні системи аналізу поведінки користувачів мають стати обов’язковим елементом цифрової трансформації бізнесу. Представлена система не лише вирішує поставлену дослідницьку задачу, але й формує основу для подальших удосконалень у напрямі автоматизації рішень, інтеграції із CRM-системами та розробки рекомендаційних модулів на основі поведінкових шаблонів.

ДЖЕРЕЛА

1. Start learning about Google Analytics. URL: https://developers.google.com/analytics (дата звернення: 14.10.2025).
2. Analytics that drive decisions. URL: https://mixpanel.com/home/ (дата звернення: 01.08.2025).
3. The digital analytics platform for AI-guided growthtesting everythingnonstop optimizationgetting real answers. URL: https://amplitude.com/ (дата звернення: 02.08.2025).
4. Research on Purchasing Behavior Pattern of E-commerce Platform Consumers Based on Big Data Analysis URL: https://www.researchgate.net/publication/391517725_Research_on_Purchasing_Behavior_Pattern_of_E-commerce_Platform_Consumers_Based_on_Big_Data_Analysis (дата звернення: 04.08.2025).
5. "Exploring the Role of Personalization in E-commerce: Impacts on Consumer Trust and Purchase Intentions". URL: https://www.researchgate.net/publication/383603236_Exploring_the_Role_of_Personalization_in_E-commerce_Impacts_on_Consumer_Trust_and_Purchase_Intentions (дата звернення: 14.08.2025).
6. Personalization in personalized marketing: Trends and ways forward. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/mar.21670 (дата звернення: 20.09.2025).
7. Analysis of E-Commerce Purchase Patterns Using Big Data: An Integrative Approach to Understanding Consumer Behavior. URL: https://www.researchgate.net/publication/376577004_Analysis_of_E-Commerce_Purchase_Patterns_Using_Big_Data_An_Integrative_Approa

ch_to_Understanding_Consumer_Behavior (дата звернення: 24.09.2025).

8. Categorizing Online Shopping Behavior from Cosmetics to Electronics: An Analytical Framework. URL: https://www.researchgate.net/publication/344505790_Categorizing_Online_Shopping_Behavior_from_Cosmetics_to_Electronics_An_Analytical_Framework (дата звернення: 04.10.2025).
9. Для чого потрібні UML діаграми? URL: https://www.run-it.com.ua/dlia-choho-potribni-uml-diahramy/ (дата звернення: 05.10.2025).
10. Use Case Diagram - Unified Modeling Language (UML) URL: https://www.geeksforgeeks.org/system-design/use-case-diagram/ (дата звернення: 10.10.2025).
11. Як будувати UML-діаграми. Розбираємо три найпопулярніші варіанти. URL: https://dou.ua/forums/topic/40575/ (дата звернення: 11.10.2025).
12. Що таке діаграма класів UML і найкращий творець діаграм класів UML. URL: https://www.mindonmap.com/uk/blog/what-is-uml-class-diagram/ (дата звернення: 12.10.2025).
13. Створення схеми розгортання UML. URL: https://surl.lt/icnyll (дата звернення: 15.10.2025).
14. Що таке сховище даних і як його побудувати. URL: https://robotdreams.cc/uk/blog/48-chto-takoe-data-warehouse-i-kak-ego-postroit (дата звернення: 16.10.2025).
15. Що таке онлайн-аналітична обробка (OLAP)? URL: https://aws.amazon.com/what-is/olap/ (дата звернення: 16.10.2025).
16. What is data mining? URL: https://www.ibm.com/think/topics/data-mining (дата звернення: 16.10.2025).
17. What is ETL? URL: https://www.ibm.com/think/topics/etl (дата звернення: 17.10.2025).

18. Amazon consumer Behaviour Dataset. URL: https://www.kaggle.com/datasets/swathiunnikrishnan/amazon-consumer-behaviour-dataset (дата звернення: 19.10.2025).
19. Зіркова схема URL: https://data-life-ua.com/db/zirkova-skhema/ (дата звернення: 19.10.2025).
20. SQL Server Analysis Services overview. URL: https://learn.microsoft.com/en-us/analysis-services/ssas-overview?view=sql-analysis-services-2025 (дата звернення: 19.10.2025).
21. SQL Server Integration Services. URL: https://learn.microsoft.com/en-us/sql/integration-services/sql-server-integration-services?view=sql-server-ver17 (дата звернення: 22.10.2025).
22. Ключові показники ефективності (KPI) для торгівельного бізнесу. URL: https://keepincrm.com/kpi-calculation (дата звернення: 05.11.2025).
23. What is SQL Server Reporting Services (SSRS)? URL: https://learn.microsoft.com/en-us/sql/reporting-services/create-deploy-and-manage-mobile-and-paginated-reports?view=sql-server-ver17 (дата звернення: 06.11.2025).
24. Learn-One-Rule Algorithm. URL: https://www.geeksforgeeks.org/machine-learning/learn-one-rule-algorithm/ (дата звернення: 08.11.2025).
25. Введення в наївний алгоритм Байєса. URL: https://codelabsacademy.com/uk/blog/naive-bayes (дата звернення: 09.11.2025).
26. Асоціативні правила. URL: https://studfile.net/preview/4494757/page:2/ (дата звернення: 09.11.2025).

ДОДАТКИ

ДОДАТОК А

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

СЕРТИФІКАТ

учасника VII Всеукраїнської науково-практичної конференції студентів і аспірантів «Теоретичні та прикладні аспекти розробки комп'ютерних систем '2025»

виданий

Саяпіній Марії Сергіївні

студентці групи ІУСТ-24002м ОС «Магістр»

***за доповідь* «Інтелектуальна система аналізу поведінки користувачів на комерційних платформах»**

Завідувач кафедри комп'ютерних наук НУБіП України **Б.Л. Голуб**

24 квітня 2025 р.

ДОДАТОК Б

Процес створення KPI:

1. LowPurchaseRate:

Value Expression:

```
([Measures].[Purchase Count] / IIF([Measures].[Views Count] = 0, NULL,
[Measures].[Views Count])) * 100
```

Goal Expression:

```
5
```

Status Expression:

```
CASE
WHEN KPIValue("LowPurchaseRate") < KPIGoal("LowPurchaseRate")
THEN -1
WHEN KPIValue("LowPurchaseRate") >= KPIGoal("LowPurchaseRate")
THEN 1
END
```

2. HighReturnRate:

Value Expression:

```
([Measures].[Returns Count] / IIF([Measures].[Purchase Count] = 0, NULL,
[Measures].[Purchase Count])) * 100
```

Goal Expression:

```
2
```

Status Expression:

```
CASE
WHEN KPIValue("HighReturnRate") > KPIGoal("HighReturnRate") THEN
1
WHEN KPIValue("HighReturnRate") <= KPIGoal("HighReturnRate") THEN
1
END
```

3. CategoryDemand:

Value Expression:

```
([Measures].[Purchase Count], [Category Dim].[Category
Name].CurrentMember)
```

Goal Expression:

```
10
```

Status Expression:

```
CASE
WHEN KPIValue("CategoryDemand") >= KPIGoal("CategoryDemand")
THEN 1
WHEN KPIValue("CategoryDemand") < KPIGoal("CategoryDemand")
THEN -1
END
```

ДОДАТОК В

ПРОГРАМНИЙ КОД АЛГОРИТМУ NAIVE BAYES.

NaiveBayesClassifier.cs:

NaiveBayesClassifier.cs

```
using System;
using System.Collections.Generic;
using System.Data;
using System.Linq;

namespace UserBehavior1RL3
{
    public class NaiveBayesClassifier
    {
        public List<FactData> Data { get; set; }

        public DataTable Classify()
        {
            var totalCount = Data.Count;

            // Уникаємо null ключів — беремо лише ті записи, де ActivityLevel задано
            var classGroups = Data
                .Where(d => !string.IsNullOrEmpty(d.ActivityLevel))
                .GroupBy(d => d.ActivityLevel)
                .ToDictionary(g => g.Key, g => g.Count());

            var products = Data.Select(d => d.ProductName).Distinct();
            var categories = Data.Select(d => d.CategoryName).Distinct();
            var genders = Data.Select(d => d.UserGender).Distinct();
            var months = Data.Select(d => d.Month).Distinct();

            DataTable results = new DataTable();
            results.Columns.Add("Product");
            results.Columns.Add("Category");
            results.Columns.Add("Gender");
            results.Columns.Add("Month");
            results.Columns.Add("PredictedClass");
            results.Columns.Add("P(Low)");
            results.Columns.Add("P(Medium)");
            results.Columns.Add("P(High)");

            foreach (var prod in products)
            {
                foreach (var cat in categories)
                {
                    foreach (var gen in genders)
                    {
                        foreach (var mon in months)
                        {
                            var probs = new Dictionary<string, double>();

                            foreach (var cls in classGroups.Keys)
                            {
                                double prior = (double)classGroups[cls] / totalCount;
                                double p1 = Prob(cls, d => d.ProductName == prod);
                                double p2 = Prob(cls, d => d.CategoryName == cat);
                                double p3 = Prob(cls, d => d.UserGender == gen);
                                double p4 = Prob(cls, d => d.Month == mon);
                                probs[cls] = prior * p1 * p2 * p3 * p4;
                            }

                            double sum = probs.Values.Sum();
                            if (sum == 0) continue;
```

```
                            var normalized = probs.ToDictionary(kv => kv.Key, kv =>
kv.Value / sum);
                            string predicted = normalized.OrderByDescending(kv =>
kv.Value).First().Key;

                            results.Rows.Add(prod, cat, gen, mon, predicted,
                                $"{(normalized.ContainsKey("Low") ? normalized["Low"] :
0) * 100:F2}",
                                $"{(normalized.ContainsKey("Medium") ?
normalized["Medium"] : 0) * 100:F2}",
                                $"{(normalized.ContainsKey("High") ? normalized["High"]
: 0) * 100:F2}");
                        }
                    }
                }
            }

            return results;

            double Prob(string cls, Func<FactData, bool> predicate)
            {
                var classData = Data.Where(d => d.ActivityLevel == cls);
                int count = classData.Count();
                if (count == 0) return 0;
                return (double)classData.Count(predicate) / count;
            }
        }
    }
}
```