ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА

на тему:

«**Методи оцінювання ефективності рекомендаційних систем для e-commerce**»

на здобуття освітнього ступеня бакалавра
зі спеціальності 124 «Системний аналіз»
освітньо-професійної програми «Системний аналіз»

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Дар'я КОВАЛЬОВА
(підпис) *Ім'я, ПРІЗВИЩЕ здобувача*

Виконав: Дар'я КОВАЛЬОВА
здобувачка вищої освіти гр. САД-41

Керівник: Ровіл НАФЄЄВ
Ім'я, ПРІЗВИЩЕ

Рецензент: _____
Ім'я, ПРІЗВИЩЕ

Київ 2025

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Навчально-науковий інститут Інформаційних технологій

Кафедра Інформаційних систем і технологій

Ступінь вищої освіти Бакалавр

Спеціальність 124 Системний аналіз

Освітньо-професійна програма Системний аналіз

ЗАТВЕРДЖУЮ
Завідувач кафедрою ІСТ
_____ Каміла СТОРЧАК
«____» _____ 20__ р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

Ковальовій Дар'ї Сергіївні

---

1. Тема кваліфікаційної роботи: Методи оцінювання ефективності рекомендаційних систем для e-commerce
керівник кваліфікаційної роботи Ровіл НАФЄЄВ, к. ф-м. н.
затверджені наказом Державного університету інформаційно-комунікаційних технологій від «24» лютого 2025 р. № 56

2. Строк подання кваліфікаційної роботи «30» червня 2025 р.
3. Вихідні дані до кваліфікаційної роботи:
   1. Методи машинного навчання та аналізу даних,
   2. Підходи до побудови та оцінювання рекомендаційних систем,
   3. Принципи функціонування електронної комерції.
   4. Науково-технічна література з питань, пов'язаних з рекомендаційними ситемами.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

1. Теоретичний розділ

2. Аналітико- дослідницький розділ

3. Проектно-рекомендаційний розділ

5. Перелік ілюстративного матеріалу: презентація

6. Дата видачі завдання «24» лютого 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Аналіз предметної області	26.02 - 02.04.2025	Виконано
2	Дослідження типів рекомендаційних систем	03.04 - 15.04.2025	Виконано
3	Опис та дослідження метрик для оцінки ефективності рекомендаційних систем	16.04 - 01.05.2025	Виконано
4	Реалізація практичної частини	01.05 - 15.05.2025	Виконано
5	Підведення підсумків роботи	15.05 - 19.05.2025	Виконано
6	Перевірка роботи на плагіат	20.05.2025	Виконано
7	Передзахист	27.05.2025	Виконано
8	Захист кваліфікаційної роботи	16.06.2025	Виконано

Здобувачка вищої освіти _____ Дар’я КОВАЛЬОВА

Керівник кваліфікаційної роботи _____ Ровіл НАФЄЄВ

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня бакалавра: 74 стор., 7 рис., 4 табл., 39 джерел.

Мета роботи – розробка методики оцінки ефективності рекомендаційних систем для e-commerce, враховуючи їхню точність, рівень персоналізації та вплив на бізнес-метрики.

Об'єкт дослідження – інтелектуальні рекомендаційні системи для e-commerce та їх вплив на ефективність онлайн-торгівлі.

Предмет дослідження – методи та підходи до оцінювання ефективності інтелектуальних рекомендаційних систем у сфері e-commerce.

Короткий зміст роботи: У кваліфікаційній роботі досліджено ключові аспекти оцінювання ефективності інтелектуальних рекомендаційних систем, що застосовуються в електронній комерції.

Робота охоплює огляд основних типів систем — контентно-орієнтованих, колаборативних, гібридних та на основі знань — із розкриттям принципів їх роботи, переваг і обмежень. Окрему увагу приділено метрикам оцінювання: від точності до більш комплексних показників.

У практичній частині було запропоновано підхід до комплексної оцінки рекомендацій, який враховує як технічні характеристики алгоритму, так і його вплив на бізнес-цілі. В результаті сформовано методику, що може бути використана компаніями для аналізу якості своїх рекомендаційних систем .

КЛЮЧОВІ СЛОВА: РЕКОМЕНДАЦІЙНА СИСТЕМА, E-COMMERCE, МЕТОДИ ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ, БІЗНЕС-МЕТРИКИ, МЕТРИКИ ТОЧНОСТІ, МАШИНЕ НАВЧАННЯ.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

A/B – двоваріантне тестування

AI – штучний інтелект (Artificial Intelligence)

ALS – чергування найменших квадратів (Alternating Least Squares)

API – прикладний програмний інтерфейс (Application Programming Interface)

ARPU – середній дохід на користувача (Average Revenue Per User)

B2B – бізнес для бізнесу (Business-to-Business)

B2C – бізнес для споживача (Business-to-Consumer)

C2C – споживач для споживача (Consumer-to-Consumer)

CF – колаборативна фільтрація (Collaborative Filtering)

CLV – довічна цінність клієнта (Customer Lifetime Value)

CTR – коефіцієнт кліків (Click-Through Rate)

CVR – коефіцієнт конверсії (Conversion Rate)

DBSCAN – алгоритм кластеризації на основі густини (Density-Based Spatial Clustering of Applications with Noise)

DCG – дисконтований кумулятивний гейн (Discounted Cumulative Gain)

e-commerce – електронна комерція

IDCG – ідеальний дисконтований кумулятивний гейн (Ideal Discounted Cumulative Gain)

KPI – ключовий показник ефективності (Key Performance Indicator)

LTV – довічна цінність клієнта (Lifetime Value)

MAE – середня абсолютна похибка (Mean Absolute Error)

MAP – середня усереднена точність (Mean Average Precision)

ML – машинне навчання (Machine Learning)

MRR – середнє обернене ранжування (Mean Reciprocal Rank)

NCF – нейронна колаборативна фільтрація (Neural Collaborative Filtering)

NDCG – нормалізований дисконтований кумулятивний гейн (Normalized Discounted Cumulative Gain)

RMSE – корінь середнього квадратичного відхилення (Root Mean Square Error)

РС – рекомендаційні системи

SKU – складська одиниця товару (Stock Keeping Unit)

SVD – сингулярне розкладання (Singular Value Decomposition)

XGBoost – екстремальне градієнтне підсилення (Extreme Gradient Boosting)

ВСТУП

У сучасному світі електронна комерція стрімко розвивається, і з кожним роком все більше компаній використовують цифрові технології для покращення клієнтського досвіду. Одним з найважливіших інструментів підвищення ефективності онлайн-комерції є рекомендаційні системи, які допомагають користувачам знаходити найбільш релевантні товари та послуги. Ці системи можуть не лише покращити персоналізацію покупок, але й значно підвищити конверсію, збільшити середній чек та покращити утримання клієнтів.

Рекомендаційні системи використовують різні алгоритми і методи, включаючи контент-орієнтовані, колаборативні та гібридні підходи, а також все частіше інтегрують технології машинного навчання і штучного інтелекту. Однак питання ефективності таких систем залишається актуальним, оскільки для бізнесу важливо не лише впроваджувати інноваційні технології, але й оцінювати їхній реальний вплив на результати діяльності.

Актуальність роботи зумовлена стрімким зростанням електронної комерції, де компанії постійно шукають нові способи залучення та утримання клієнтів. Рекомендаційні системи стали важливим інструментом для персоналізації покупок, що значно підвищує ефективність бізнесу. Однак, незважаючи на їх широке використання, досі бракує досліджень, які б чітко оцінювали їхній вплив на реальні бізнес-результати. Це створює потребу в розробці методологій, які дозволять точніше вимірювати ефективність таких систем, що, в свою чергу, допоможе компаніям оптимізувати свої стратегії та зміцнити позиції на ринку.

Метою цієї роботи є розробка методології оцінки ефективності інтелектуальних рекомендаційних систем для електронної комерції з урахуванням таких факторів, як точність, персоналізація та їх вплив на бізнес-метрики. В роботі також буде запропоновано експериментальне тестування розробленої методології та порівняння її з існуючими підходами.

Об'єктом дослідження виступають інтелектуальні рекомендаційні системи для e-commerce та їх вплив на ефективність онлайн-торгівлі.

Предметом дослідження є методи та підходи до оцінювання ефективності інтелектуальних рекомендаційних систем у сфері e-commerce.

У процесі виконання кваліфікаційної роботи було застосовано комплекс *методів дослідження*, а саме: здійснено аналіз наукових публікацій, стосовно методів побудови та оцінювання ефективності рекомендаційних систем. Для розробки рекомендаційного механізму застосовувалися методи моделювання. Важливою складовою дослідження стало проведення експериментального тестування, що дало змогу перевірити практичну результативність запропонованого підходу до оцінювання.

Наукова новизна роботи полягає в створенні нового підходу до оцінки ефективності рекомендаційних систем для e-commerce, що враховує не тільки точність, але й вплив на бізнес-результати, такі як конверсія та утримання клієнтів.

Апробація результатів та публікації: основні положення роботи були викладені у форматі тез і представлені на IV Міжнародній науково-технічній конференції «Сучасний стан та перспективи розвитку IoT», що дозволило висвітлити ключові положення роботи у науковому середовищі

1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1. Загальна класифікація рекомендаційних систем

Рекомендаційні системи (РС) - це інтелектуальні системи, які надають пропозиції щодо товарів , які, найімовірніше, можуть зацікавити конкретного користувача.[2, ст. 1]

Пропозиції можуть стосуватися різних процесів прийняття рішень, наприклад, які товари купити, яку музику слухати або які новини в Інтернеті прочитати.

Підгрунтям для створення рекомендаційних систем стали дослідження людської поведінки, де на основі спостережень, прийшли до висновку, що люди часто покладаються на поради інших, коли приймають повсякденні рішення - наприклад, обирають книги, фільми або наймають кандидатів на роботу. Імітуючи цю природну поведінку, перші РС використовували алгоритми для надання рекомендацій, визначаючи користувачів зі схожими вподобаннями.

РС в першу чергу розроблені для користувачів, яким бракує особистого досвіду чи знань, необхідних для оцінки величезної кількості товарів, доступних на онлайн-платформах. Наприклад, книжкові рекомендації допомагають користувачам вибрати відповідну книгу з тисячі доступних варіантів. Як приклад, можна відзначити Bol чи Amazon, які використовують персоналізовані алгоритми рекомендацій для адаптації досвіду покупок до кожного клієнта, пропонуючи товари, що відповідають його інтересам. Ці персоналізовані рекомендації враховують уподобання окремих користувачів або груп користувачів, дозволяючи кожній людині отримувати пропозиції, які є для неї найбільш релевантними.

З іншого боку, існують також неперсоналізовані підходи до рекомендацій. Вони набагато простіші у впровадженні та часто зустрічаються у загальнодоступних джерелах, такі як журнали чи газети. Типовими прикладами є списки топ-товарів у тій чи іншій категорії, наприклад «Топ-10 танцювальних хітів»

або «Найкращі альбоми місяця». Хоча такі пропозиції можуть бути корисними в певних контекстах, вони рідко є предметом досліджень, оскільки вони не враховують унікальні уподобання чи поведінку окремих користувачів.

Загалом персоналізовані рекомендації представлені у вигляді ранжованих списків, упорядкованих за рівнем потенційної зацікавленості. Створюючи таке ранжування, РС намагаються передбачити, які продукти чи послуги є найбільш релевантними, виходячи з уподобань та обмежень користувача. Для того, щоб виконати таке обчислювальне завдання, РС збирають інформацію від користувачів про їхні вподобання, які або явно виражені (у вигляді оцінок продуктів) або виведені шляхом аналізу дій користувача (кліки; час, проведений на сторінці). Наприклад, РС може розглядати перехід на певну сторінку продукту як неявну ознаку вподобань щодо товарів, представлених на цій сторінці.

Багато рекомендаційних систем традиційно використовуються в електронній комерції, пропонуючи такі товари, як книги, фільми, туристичні послуги та споживчі товари - їхнє використання вже давно вийшло далеко за межі цієї сфери.

Важливо зазначити, що не всі рекомендаційні системи обмежуються пропозиціями товарів. Наприклад, пошукова система Google може показувати рекламу товарів поряд з результатами пошуку.

Інші приклади включають пропозиції друзів (Facebook або Instagram), підбір роботи на кар'єрних платформах (LinkedIn), де і роботодавці, і шукачі роботи отримують індивідуальні рекомендації. Такі системи відомі як системи взаємних рекомендацій, оскільки вони спрямовані на узгодження інтересів обох сторін. Алгоритми, що лежать в основі таких систем, часто суттєво відрізняються від тих, що використовуються в традиційних рекомендаціях на основі продуктів.

З розвитком електронної комерції величезний обсяг і розмаїття доступних продуктів ускладнили користувачам можливість робити усвідомлений вибір. Стрімке розширення інтернету та зростання цифрових послуг призвели до інформаційного перевантаження, коли надмірний вибір почав негативно впливати на прийняття рішень. Дослідження показали, що хоча більший вибір означає більшу

свободу, але він також може призвести до втоми від прийняття рішень і незадоволення.

Системи рекомендацій вирішують цю проблему, допомагаючи користувачам знаходити релевантні, але раніше невідомі елементи. На основі введених користувачем даних - явних уподобань, поведінки чи контекстуальних потреб – рекомендаційні системи аналізують наявні дані, включаючи профілі користувачів, описи товарів і минулі транзакції, щоб генерувати персоналізовані пропозиції. Користувачі можуть надавати зворотній зв'язок (рейтинги, кліки), який система використовує для постійного вдосконалення рекомендацій.

З моменту своєї появи як дослідницької галузі в середині 1990-х років, РС привернули значну увагу як в академічних колах, так і в промисловості. Їх роль у покращенні користувацького досвіду, підтримці прийняття рішень та уможливленні персоналізації в різних цифрових сферах зробила їх невід'ємним компонентом сучасних інформаційних систем.

Системи рекомендацій можна класифікувати за кількома ключовими параметрами, які відображають їхнє призначення, функціональність і сферу застосування.

Один з найпоширеніших методів загальної класифікації базується на спеціалізації. Системи рекомендацій широко використовуються на різноманітних цифрових платформах для покращення користувацького досвіду. Так, інтернет-магазини на кшталт Amazon, Bol чи Rozetka підбирають товари, які можуть зацікавити покупця. Сайти, що займаються бронюванням житла або організацією подорожей, наприклад Booking чи Airbnb, пропонують персоналізовані варіанти місць проживання та потенційних маршрутів. Музичні сервіси, як-от Spotify, Youtube Music або Apple Music, формують добірки пісень або рекомендують треки на основі смаків користувача. Стрімінгові відеоплатформи, серед яких Netflix, YouTube чи Megogo, радять фільми та серіали, а сервіси з акцентом на візуальний контент, як Pinterest або Instagram, демонструють зображення відповідно до зацікавлень користувача. Крім того, портали новин, такі як BBC, The Guardian чи навіть локальні ЗМІ, використовують рекомендаційні алгоритми для персонального

формування новинної стрічки. У соціальних мережах — Facebook та LinkedIn — ці системи пропонують можливих друзів, спільноти або професійні контакти, орієнтуючись на поведінку та інтереси користувачів.

З точки зору персоналізації, рекомендаційні системи розрізняють за ступенем адаптації до окремих користувачів. Деякі з них пропонують неперсоніфіковані пропозиції - ідентичні для всіх користувачів, тоді як інші застосовують сегментовану персоналізацію, групуючи користувачів на основі демографічних або географічних особливостей (вік, місцезнаходження або мова) і адаптуючи рекомендації для кожної групи. Повністю персоналізовані системи йдуть далі, коригуючи рекомендації на основі унікальних інтересів і поведінки людини. У цьому контексті виділяють ефемерну персоналізацію, яка використовує дані лише однієї сесії, і постійну персоналізацію, яка створює і підтримує профілі користувачів протягом тривалого часу.

Метод надання рекомендацій також різниться. Деякі системи використовують push-підхід, автоматично надсилаючи пропозиції через електронну пошту або сповіщення, без безпосередніх дій користувача. Інші покладаються на механізми залучення, коли користувачі активно ініціюють процес надання рекомендацій, досліджуючи або здійснюючи пошук.

Щодо типу зібраних даних, то системи можуть ґрунтуватися на явному зворотному зв'язку, який включає в себе користувацьку інформацію, таку як рейтинги, вподобання/неуподобання або збережені; або на неявному зворотному зв'язку, який випливає з поведінки користувача, наприклад, історії переглядів, часу, проведеного за контентом і т. д..

З технічної точки зору, рекомендаційні системи можна розділити за методологією, яка використовується для генерування пропозицій. До них відносяться ручні підходи (наприклад, кураторські списки експертів), статистичні методи (наприклад, рейтинги на основі популярності) та алгоритмічні моделі, такі як:

- фільтрація на основі контенту, яка спирається на особливості товарів та вподобання користувачів;

- системи, засновані на знаннях, які використовують заздалегідь визначені критерії або логічні правила;
- колаборативна фільтрація, яка використовує шаблони взаємодії користувача з товаром;
- гібридні моделі, які поєднують кілька методів для покращення якості рекомендацій.

Крім того, до загальної класифікації відносять , як системи враховують зміни у вподобаннях користувачів з плином часу. Деякі моделі використовують глобальну перспективу, відображаючи довгострокові інтереси, інші зосереджуються на часовій сегментації, аналізуючи поведінку в певних часових рамках, а більш просунуті системи виконують послідовне моделювання, прогнозуючи наступні дії на основі нещодавньої активності користувача.

1.2. Принципи роботи рекомендаційних систем

Основний принцип роботи рекомендаційних систем полягає в тому, що між активністю користувачів і товарами/послугами існують істотні залежності [1, ст. 2]. Наприклад, якщо користувач цікавиться історичними документальними фільмами, то з більшою ймовірністю йому сподобається інша історична передача або освітня програма, а не бойовик. У багатьох випадках категорії об’єктів можуть бути між собою корельовані, і ці зв’язки можна використовувати для більш точних рекомендацій. Альтернативно, залежності можуть виявлятися не лише між категоріями, але й безпосередньо між окремими об’єктами. Такі зв’язки можна виявити за допомогою аналізу даних у матриці оцінок, а побудована модель дозволяє робити передбачення для конкретних користувачів.

Чим більше оцінених об'єктів є в історії користування, тим легше зробити точні прогнози щодо майбутніх уподобань. Для цього можна використовувати різні навчальні моделі. Наприклад, можна використати спільну поведінку при покупках або оцінюванні різних користувачів, щоб створити групи (спільноти) зі схожими

інтересами. Уподобання цих спільнот можна застосовувати для формування рекомендацій для їхніх учасників.

Описаний вище підхід базується на простих рекомендаційних алгоритмах, відомих як моделі на основі сусідства (neighborhood models). Вони належать до ширшого класу підходів - колаборативна фільтрація (collaborative filtering). Цей термін означає використання оцінок багатьох користувачів у співпраці, щоб передбачити відсутні оцінки [1, ст. 2].

На практиці рекомендаційні системи можуть бути значно складнішими та багатшими на дані, включаючи допоміжну інформацію. Наприклад, у контентно-орієнтованих системах головну роль відіграє вміст товарів або об’єктів: у процесі рекомендації враховуються як оцінки користувачів, так і характеристика товарів або послуг, що дозволяє передбачити майбутні дії. Основна ідея полягає в тому, що інтереси користувача можна змоделювати через властивості об’єктів, які він раніше оцінював або переглядав.

Існує також фреймворк систем на основі знань (knowledge-based systems), у яких користувач особисто вказує свої інтереси, а система поєднує ці запити з предметною базою знань, щоб сформувати рекомендації.

У просунутих моделях можуть застосовуватись також контекстні дані, такі як час, зовнішні джерела знань, локація, соціальні зв’язки чи інформація з мережевих структур.

1.3. Методи та алгоритми рекомендаційних систем

Широке впровадження рекомендаційних систем прямо пов’язано з великим обсягом різноманітних даних про споживачів. Інструменти веб-аналітики дозволяють безперервно збирати інформацію про поведінку відвідувачів веб-сайтів. Іншим підходом до збору даних є веб-скрепінг, який витягує релевантну публічну інформацію про потенційних клієнтів і завантажує її в бази даних. Однак важливо зазначити, що веб-скрепінг не завжди проводиться в рамках правового

поля. Водночас сучасні маркетингові практики часто покладаються на дані, зібрані за допомогою реєстраційних форм, які вимагають надання контактних даних і, за бажанням, соціально-демографічної інформації. Щоб стимулювати користувачів ділитися цими даними, компанії часто пропонують такі заохочення, як знижки, подарунки або спеціальні пропозиції.

Ефективна ідентифікація клієнтів вимагає інтеграції ключових показників ефективності (KPI) та специфічних для ринку даних. На основі науково обґрунтованих принципів та з урахуванням галузевої приналежності компанії і цільової демографічної групи створюється спеціальний набір показників ефективності, який допомагає досягти оптимальних результатів. Враховуючи динамічний характер більшості ринків, ці показники повинні регулярно оновлюватися, щоб відображати нові накопичені статистичні дані.

З точки зору використання даних, рекомендаційні системи можна розділити на 5 основних типів.

1.3.1. Колаборативна фільтрація (CF)

У рамках колаборативної фільтрації рекомендації генеруються на основі рейтингів, створених користувачами для широкої аудиторії. Основною проблемою цих систем є розрідженість матриці оцінок, коли більшість можливих взаємодій між користувачем та товаром залишаються не оціненими. Наприклад, на платформах з рекомендаціями фільмів користувачі зазвичай оцінюють лише незначну частину всього каталогу, залишаючи багато фільмів без оцінок. У цьому випадку "врахованими" вважаються ті фільми, у яких є оцінки, тоді як під "неврахованими" маються на увазі фільми, які їх не мають, що створює прогалини в матриці оцінок.

Основний принцип колаборативної фільтрації полягає в тому, що ці відсутні оцінки можна визначити на основі існуючих шаблонів, оскільки часто існує сильна кореляція між уподобаннями різних користувачів і властивостями різних об'єктів. Наприклад, якщо два користувачі - Михайло та Андрій - мають тенденцію

оцінювати одні й ті ж фільми однаково, доречно припустити, що їхні думки збігатимуться щодо фільмів, на які лише один з них залишив відгук. Такі закономірності використовують для прогнозування невідомих значень.

Алгоритми колаборативної фільтрації зазвичай зосереджуються або на кореляціях на основі користувачів, або на кореляціях на основі об’єктів (товари чи послуги), або на обох.

Методи колаборативної фільтрації поділяються на дві великі категорії: підходи на основі пам'яті та підходи на основі моделей.

- Підходи на основі пам'яті (сусідства)

Підходи на основі пам'яті, які також відомі як моделі сусідства, були одними з перших методів колаборативної фільтрації. Вони покладаються на оцінки, надані схожими користувачами або схожим об'єктам, для генерування прогнозів. [1, ст. 9].

Колаборативна фільтрація на основі користувачів визначає групу осіб, які мають схожі вподобання з цільовим користувачем. Далі система прогнозує відсутні оцінки для нього, обчислюючи середньозважене значення відомих оцінок від групи найбільш схожих спживачів. Наприклад, якщо Користувач 1 і Користувач 2 мають схожі вподобання у виборі книг, оцінка Користувача 1 за певний роман може допомогти передбачити, як цей роман оцінить Користувач 2. Зазвичай алгоритм вибирає користувачів з найбільш подібними оцінками, порівнюючи їхні вподобання у матриці даних.

Колаборативна фільтрація на основі товарів працює інакше: замість пошуку схожих користувачів, вона знаходить товари, які подібні до обраного, і на основі їхніх існуючих рейтингів прогнозує, чи сподобається новий товар споживачу. Наприклад, якщо користувач високо оцінив книги «Гаррі Поттер і філософський камінь» і «Володар перснів», це може допомогти передбачити його оцінку для роману «Хроніки Нарнії». Тут схожість вимірюється між стовпчиками рейтингової матриці.

Методи, засновані на пам'яті, інтуїтивно зрозумілі і легко пояснюються, але вони часто стикаються з проблемою розрідженості даних. Якщо для певного товару доступна дуже мала кількість оцінок, система може мати труднощі з пошуком

схожих користувачів для точного прогнозування рейтингу. Через це колаборативна фільтрація іноді не забезпечує повного покриття прогнозів, проте це не є критичною проблемою для застосунків, де важливі лише найкращі рекомендацій.

- Підходи на основі моделей

Колаборативна фільтрація на основі моделей використовує методи машинного навчання та інтелектуального аналізу даних для побудови прогнозів. Ці моделі часто включають параметри, які навчаються за допомогою методів оптимізації. [1, ст. 9]. Прикладами методів на основі моделей є дерева рішень, системи, засновані на правилах, моделі байєсівського виведення та моделі латентних факторів. Зокрема, підходи з використанням латентних факторів продемонстрували високу продуктивність, навіть коли стикаються з дуже розрідженими даними.

Хоча методи, що базуються на пам'яті, цінуються за їхню простоту, вони значною мірою є евристичними і можуть не працювати послідовно в усіх сценаріях.

Попри всі свої переваги, колаборативна фільтрація стикається з труднощами при роботі з новими користувачами або товарами через відсутність попередніх даних (проблема холодного старту). Тому маркетологам часто доводиться доповнювати CF цільовими кампаніями для збору актуальних даних про користувачів.

1.3.2. Системи рекомендацій на основі контенту

Системи рекомендацій на основі контенту генерують пропозиції, аналізуючи описові характеристики товарів, де термін «контент» означає тип метаданих, пов'язаних з товаром. Ці системи поєднують історію взаємодії користувача - наприклад, попередні оцінки або поведінку - з характеристиками товару, щоб передбачити майбутні вподобання. Наприклад, якщо Користувач 1 поставив високу оцінку мультфільму «Душа», але немає даних від інших споживачів, колаборативну фільтрацію застосувати неможливо. Однак, оскільки «Душа» має спільні жанрові

теги з іншими комедійно-драматичними мультфільмами студії Pixart, такими як «Думками навиворіт» або «Стихії», їх можна рекомендувати на основі схожості їхнього контенту.

При такому підході система розглядає кожного користувача як індивідуальну проблему моделювання. Описи товарів, які користувач оцінив або придбав, слугують навчальними прикладами, а відповідні оцінки або поведінкові реакції виступають у ролі маркерів. Потім система будує персоналізовану класифікацію або регресійну модель для кожного користувача, яка використовується для прогнозування інтересу до товарів, які користувач ще не оцінив або не придбав.

Однією з ключових переваг рекомендацій на основі контенту є їхня здатність ефективно працювати з новими або не оціненими товарами. Якщо товар схожий за характеристиками на інші продукти, які сподобалися користувачу, система може рекомендувати його - навіть якщо у нього немає оцінок. Це робить їх особливо корисними у випадках холодного старту.

Однак у цього методу є недоліки:

Оскільки рекомендації генеруються на основі минулих взаємодій користувача та характеристик предмета, системі може бракувати різноманітності. Об'єкти з абсолютно новими або нетиповими ключовими словами навряд чи будуть запропоновані, що обмежує діапазон рекомендацій і виключає ширше розуміння спільноти.

Хоча ці системи ефективні для нових об'єктів, вони менш корисні для нових користувачів, оскільки моделі вимагають досить великої історії взаємодій для більш точних рекомендацій. Без достатньої кількості історичних даних прогнози можуть бути ненадійними або схильними до надмірного пристосування.

Ці характеристики підкреслюють як сильні, так і слабкі сторони систем, заснованих на контенті, порівняно з колаборативною фільтрацією.

Поза стандартними рамками машинного навчання існує ширша інтерпретація підходів на основі контенту. У деяких системах користувачі самостійно створюють профілі, вказуючи відповідні ключові слова. Потім ці профілі можна порівняти з описами товарів, щоб надати рекомендації. Оскільки цей метод не покладається на

рейтинги, його також можна використовувати в ситуаціях, коли користувач тільки починає працювати з сайтом. Однак такі системи часто класифікують окремо як рекомендаційні системи на основі знань, оскільки вони зазвичай використовують специфічні для домену міри подібності. Хоча вони мають багато спільного з системами на основі контенту, межа між ними іноді є дискусійною через перекриття функціональності.

Отже, оскільки системи на основі контенту покладаються на характеристики продукту, а не на нав'язливий збір даних, вони, як правило, сприяють кращому залученню користувачів на початковому етапі. Однак у довгостроковій перспективі користувачі можуть бути незадоволені, якщо система постійно показує застарілі або нерелевантні рекомендації, засновані на раніше переглянутих товарах.

1.3.3. Рекомендаційні системи на основі знань

Ця категорія систем покладається на структуровані знання про предметну область, щоб допомогти користувачам у складних процесах прийняття рішень. Включаючи заздалегідь визначені асоціативні правила і критерії вибору, такі системи можуть пропонувати відповідні продукти і послуги, пристосовані уподобань споживачів. Однією з головних переваг систем, заснованих на знаннях, є їхня здатність сприяти перехресним продажам, пропонуючи супутні товари або аксесуари поряд з основним товаром.

РС, що базуються на знаннях, особливо цінні в ситуаціях, коли відповідні товари купуються нечасто або вимагають прийняття складних рішень. Це стосується таких категорій, як нерухомість, автомобілі, туристичні пропозиції та фінансові послуги. У цих сценаріях часто важко зібрати достатню кількість оцінок користувачів для ефективного застосування традиційних методів рекомендацій. Через низьку частоту транзакцій збір достатньої кількості відгуків для будь-якої конкретної конфігурації продукту стає значним викликом.

Ця проблема перетинається з так званою проблемою «холодного старту», коли брак історичних даних перешкоджає створенню надійних рекомендацій. Крім того, уподобання щодо цих типів продуктів з часом можуть сильно змінюватися. Наприклад, з виходом нової версії моделі автомобіля вподобання користувачів можуть змінюватися в залежності від оновлених функцій або ринкових тенденцій. У такому динамічному середовищі часто недостатньо покладатися лише на минулі рейтинги, особливо коли інтерес користувачів залежить від конкретних комбінацій характеристик продукту, таких як колір, тип двигуна або особливості інтер'єру.

Щоб вирішити ці проблеми, системи рекомендацій, засновані на знаннях, не покладаються на оцінки користувачів. Натомість вони зіставляють чіткі вимоги користувача з характеристиками продукту, використовуючи структуровані знання та логіку, засновану на обмеженнях.

Однією з ключових переваг цього підходу є те, що він дає користувачам більший контроль над рекомендаціями, дозволяючи їм заздалегідь вказувати свої уподобання [1, ст. 15].. На противагу цьому, рекомендаційні системи на основі контенту, формують рекомендації згідно з попередніх взаємодій користувача, активності інших споживачів або спільних уподобань. РС засновані на знаннях, відрізняються тим, що вони підтримують цілеспрямований пошук, де користувачі можуть визначити, що вони шукають, не покладаючись на попередні поведінкові дані. Ця відмінність підсумована в Таблиці 1.1

Таблиця 1.1

Порівняльна характеристика підходів у рекомендаційних системах

Підхід	Концептуальна мета	Вхідні дані
Колаборативний	Надати рекомендації, спираючись на думки та поведінку інших людей, схожих на цільового користувача.	Оцінки користувача + колективні оцінки спільноти

Продовження таблиці 1.1

Підхід	Концептуальна мета	Вхідні дані
Контентно-орієнтований	Запропонувати об’єкти, схожі за характеристиками на ті, що мені вже сподобалися.	Оцінки користувача + характеристика об'єктів
На основі знань .	Надати рекомендації відповідно до моїх чітко вказаних уподобань щодо характеристик	Вказані вимоги користувача + характеристики об’єктів + предметні знання

Системи рекомендацій, засновані на знаннях, можна класифікувати за типом інтерфейсу, який вони використовують, і за тим, як знання про предметну область включаються в процес рекомендацій.

- РС на основі обмежень

Ці системи дозволяють користувачам встановлювати певні обмеження на характеристики продукту, наприклад, мінімальні або максимальні значення. Типовий інтерфейс дозволяє користувачам вводити вимоги, які потім зіставляються з атрибутами товару за допомогою заздалегідь визначених правил, отриманих на основі знань про предметну область.

- РС на основі конкретних прикладів

У цьому підході користувачі вибирають конкретні приклади або «кейси», які представляють те, що вони шукають. Потім система знаходить об’єкти, схожі на обраний кейс, використовуючи функції подібності, які включають експертні знання. Після отримання результатів користувачі можуть скоригувати свої вподобання, змінивши певні характеристики обраного об'єкта та надіславши оновлений запит.

Цей цикл триває доти, доки не буде знайдено відповідний збіг, допомагаючи користувачеві поступово наближатися до бажаного продукту

Хоча і системи, засновані на обмеженнях, і системи, засновані на прикладах, дозволяють користувачам точно налаштовувати свої вхідні дані, стилі взаємодії відрізняються. Системи, засновані на прикладах, використовують дослідження на основі подібності об’єктів, тоді як системи, засновані на обмеженнях, більше покладаються на фільтрацію на основі правил.

- Системи рекомендацій на основі корисності

Системи рекомендацій на основі корисності працюють, застосовуючи функцію корисності до характеристик товару, щоб оцінити, наскільки ймовірно, що користувач зацікавиться певним продуктом. Основна складність цього підходу полягає в точному визначенні функції корисності, яка відображає вподобання окремого користувача.

Оскільки функція корисності слугує формою зовнішніх вхідних даних або попередніх знань, рекомендаційні системи на основі корисності часто розглядаються як підмножина РС заснованих на знаннях.

1.3.4. Демографічні рекомендаційні системи

Демографічні рекомендаційні системи використовують специфічні для користувача демографічні характеристики, такі як вік, стать, професія чи місцезнаходження, для навчання класифікаційних моделей, які пов'язують ці характеристики з ймовірними оцінками чи поведінкою споживачів щодо покупок. Одна з перших систем такого типу, Grundy, генерувала книжкові пропозиції, використовуючи створену самостійно базу даних стереотипних профілів користувачів. Інформація про користувача збиралася через інтерактивний діалог, що дозволяло системі зіставляти людей із заздалегідь визначеними типами особистості[1, ст. 19].

У багатьох сучасних РС демографічні дані використовуються разом з іншими контекстними даними для підвищення точності рекомендацій, що наближає цей

метод до принципів, які лежать в основі контекстно-орієнтованих систем рекомендацій.

Останні досягнення в цій галузі використовують класифікатори машинного навчання для створення персоналізованих рекомендацій. Наприклад, деякі системи виводять вподобання з публічних онлайн-профілів користувачів, щоб оцінити ймовірність того, що їм сподобаються певні продукти, наприклад, ресторани.

Технічно демографічні рекомендаційні системи функціонують подібно до стандартних класифікаційних або регресійних моделей, де демографічні ознаки розглядаються як незалежні змінні, а результати, такі як рейтинги або наміри покупки, виступають як залежні змінні. Хоча ці системи можуть не перевершувати інші методи, коли використовуються окремо, вони є дуже ефективними як компоненти гібридних моделей або комплексних підходів. Зокрема, вони часто поєднуються з рекомендаціями, заснованими на знаннях, щоб підвищити стійкість і точність для різних груп користувачів.

1.3.5. Гібридні рекомендаційні системи

Описані вище методи рекомендацій спираються на різні типи вхідних даних і найкраще підходять для окремих сценаріїв. Скажімо, метод колаборативної фільтрації варто використовувати, коли є багато даних про вподобання користувачів, оскільки такий підхід покладається на рейтинги; системи на основі контенту будуть доречними у випадку, коли важлива подібність між об’єктами та існує детальна характеристика товарів, так як цей метод заключається у порівнянні характеристик нових елементів із тими, що вже сподобались користувачу раніше; а системи, засновані на знаннях, доцільно застосовувати тоді, коли споживач вперше стикається з вибором і не має історії взаємодій, а рекомендації формуються на основі чітко визначених вимог, правил або експертних параметрів, наприклад бюджету чи функціональності, тому такі системи особливо ефективні в умовах «холодного старту».

Коли РС має доступ до різноманітних джерел інформації — наприклад, даних про поведінку користувачів, характеристики товарів чи вподобання — це дає більше можливостей для гнучкого підходу: можна обирати найдоцільніший метод або навіть поєднувати кілька стратегій, щоб створити більш точні та корисні рекомендації. Це відкриває шлях до розробки гібридних рекомендаційних систем, які об'єднують сильні сторони різних підходів для отримання більш точних і адаптивних результатів. Ці системи концептуально схожі на методи ансамблевого навчання в машинному навчанні, коли кілька моделей об'єднуються для підвищення загальної продуктивності та надійності.

Гібридні системи поєднують в собі кілька методів рекомендацій для досягнення максимальної ефективності. Хоча вони зазвичай вимагають більше обчислювальних і фінансових ресурсів, вони пропонують вищу точність і ширшу адаптивність. Зі збільшенням обчислювальних потужностей і розвитком нових методів науки про дані гібридні підходи продовжують розвиватися. Розробка сучасних гібридних систем часто спирається на передові методи побудови моделей, запозичені зі сфери машинного навчання та штучного інтелекту.

1.4. Використання машинного навчання та штучного інтелекту у рекомендаційних системах для e-commerce

Завдяки стрімкому розвитку інформаційних технологій та зростанню обсягів даних, методи машинного навчання (ML) і штучного інтелекту (AI) стали основою сучасних рекомендаційних систем для електронної комерції. Ці методи дозволяють забезпечити високий рівень персоналізації, адаптивності та точності рекомендацій, з урахуванням не лише історії покупок користувача, але й його поведінки у реальному часі, соціальних сигналів, контексту та індивідуальних уподобань.

Машинне навчання охоплює широкий спектр алгоритмів, що застосовуються для класифікації, регресії, кластеризації, факторизації матриць та обробки неструктурованих даних (тексти, зображення, відео).

Класифікаційні моделі (наприклад, Decision Tree, Random Forest, Logistic Regression) дозволяють передбачати дії користувача, наприклад купівлю товару чи клік за оголошенням.

Регресійні методи використовуються для прогнозування рейтингової оцінки товару з боку користувача (наприклад, Linear Regression, XGBoost).

Методи кластеризації (K-Means, DBSCAN) групують користувачів або товари за подібними характеристиками, що сприяє створенню сегментованих стратегій просування.

Matrix Factorization (SVD, ALS) дозволяє виявити приховані зв'язки між користувачами та товарами, зокрема в колаборативній фільтрації.

Глибинні нейронні мережі відкривають нові можливості для моделювання складних, нелінійних взаємозв'язків. Як приклад:

- Neural Collaborative Filtering (NCF) — заміна класичних факторизаційних моделей на багатошарові нейронні мережі;
- Autoencoders — зменшення розмірності профілю користувача з метою покращення точності прогнозів;
- RNN та LSTM — аналіз послідовної поведінки користувачів (наприклад, послідовності переглядів);
- CNN — аналіз візуального контенту, наприклад, товарних зображень у fashion-сегменті.
- Обробка текстової інформації (NLP)

Сучасні рекомендаційні системи дедалі частіше враховують зовнішні чинники — час доби, геолокацію, тип пристрою, джерело трафіку. Для цього використовуються моделі типу Factorization Machines (FM) або DeepFM, які поєднують традиційні методи з глибинним навчанням. Такі можелі РС прийнято називати контекстно-орієнтованими

Навчання з підкріпленням (Reinforcement Learning). Цей підхід базується на взаємодії системи з користувачем у реальному часі, де алгоритм "вчиться" на основі реакцій — наприклад, кліків чи покупок. Моделі типу Multi-Armed Bandits або Deep

Q-Networks (DQN) допомагають динамічно адаптувати рекомендації в залежності від поведінки користувача.

Завдяки доступності потужних мов програмування, таких як Python, R та активному розвитку бібліотек ML (Scikit-learn, TensorFlow, PyTorch), реалізація складних рекомендаційних систем стала значно простішою. Застосування асоціативного аналізу, кластеризації, дерев рішень, нейронних мереж та навіть генетичних алгоритмів дозволяє виявляти приховані закономірності в поведінці користувачів. Крім того, згідно з [4], у рекомендаційних системах використовуються не лише цифрові дані, але й текстова, візуальна та аудіоінформація, яка трансформується у математичну форму.

Використання AI/ML у рекомендаційних системах не обмежується лише «передбаченням інтересів». Вони стають інструментом стратегічного управління маркетингом: від формування персоналізованого досвіду до оптимізації цінової політики та максимізації довгострокової цінності клієнта.

Висновок до розділу

У першому розділі було розглянуто теоретичні основи функціонування інтелектуальних рекомендаційних систем у сфері електронної комерції. Було проаналізовано принципи роботи основних типів систем – колаборативних, контентно-орієнтованих, гібридних, демографічних, а також систем на основі знань. Встановлено, що кожен підхід має свої переваги та обмеження в залежності від доступних даних і цілей застосування. Значну увагу варто приділити також ролі машинного навчання та штучного інтелекту, які дозволяють досягати високого рівня персоналізації та адаптивності систем до поведінки користувачів. Таким чином, рекомендаційні системи є ключовим елементом сучасних e-commerce платформ, забезпечуючи не лише підвищення якості користувацького досвіду, а й покращення бізнес-результатів.

2. ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ІНТЕЛЕКТУАЛЬНИХ РЕКОМЕНДАЦІЙНИХ СИСТЕМ ДЛЯ E-COMMERCE

2.1. Метрики оцінювання ефективності рекомендаційних систем

Оцінка ефективності РС відбувається на основі ряду метрик, щоб визначити, наскільки добре вони передбачають інтереси користувачів та відповідають бізнес-цілям. Жодна окрема метрика не здатна повністю охарактеризувати ефективність рекомендацій, тому прийнято використовувати набір показників для різних аспектів роботи системи. [1, ст. 231] Традиційно метрики поділяють на метрики точності рекомендацій, що відображають здатність системи пропонувати релевантні елементи (Precision, Recall, F1); метрики помилки прогнозування, для оцінки похибки при передбаченні рейтингів (MAE, RMSE); метрики якості ранжування, які враховують позиції релевантних рекомендацій (MAP, NDCG); а також метрики корисності, що виходять за межі точності – такі як новизна, різноманітність та серендипність рекомендацій.

2.1.1. Метрики точності рекомендацій

Precision (точність рекомендацій) визначає частку рекомендованих системою елементів, які виявилися релевантними та сподобались користувачу, від загального числа рекомендацій. [6, ст. 14903] Формально Precision можна виразити як відношення

$$Precision = \frac{TP}{TP+FP} \tag{2.1}$$

Де,

TP (True Positives) – правильно рекомендовані релевантні об’єкти

FP (False Positives) – об’єкти, рекомендовані помилково [6]

Наприклад, якщо Amazon пропонує користувачу 10 товарів, з яких 7 відповідають його вподобанням, то точність рекомендацій становить 0.7 (70%).

Recall (повнота рекомендацій) характеризує здатність системи знаходити всі потенційно цікаві для користувача об’єкти – тобто частку релевантних елементів, що були рекомендовані, відносно всіх існуючих відповідних товарів. В формулі:

$$Recall = \frac{TP}{TP+FN} \qquad (2.2)$$

Де,

TP (True Positives) – правильно рекомендовані релевантні об’єкти

FN (False Negatives) – це релевантні об’єкти, які система не порадила користувачу. [6]

Як приклад, якщо з 10 дійсно бажаних товарів система рекомендувала 7, Recall буде 0.7.

F1-міра є узагальненою метрикою, що об’єднує precision та recall і обчислюється як їх середнє значення.

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \qquad (2.3)\ [6]$$

Для обчислення метрик точності та повноти у бінарному контексті застосовується класифікаційна матриця (табл. 2.1), що відображає співвідношення між реальними вподобаннями користувача та виданими системою рекомендаціями.

Таблиця 2.1

Класифікаційна матриця

	Recommended	Not Recommended
Liked	True Positive(TP)	True Negative(TN)
Not Liked	False Positive(FP)	False Negative(FN)

Ця міра балансує точність і повноту рекомендацій : високий F1 досягається лише коли система одночасно пропонує достатньо релевантні рекомендації і не пропускає важливі для користувача об’єкти. В такий спосіб F1 дозволяє порівнювати різні алгоритми, враховуючи як якість рекомендацій, так і їх охоплення. Значення F1 особливо корисне при top-N рекомендаціях – коли системі потрібно запропонувати обмежений список найкращих N позицій. На практиці для оцінювання Top-N результатів часто використовують похідні показники, такі як Precision@N та Recall@N – precision і recall, обчислені для списку з N топ-рекомендацій.

У контентно-орієнтованих фільтрах, як правило, precision може бути високим для товарів, подібних до вже вподобаних користувачем, але такі рекомендації часто є очікуваними. Це підвищує довіру, але не завжди приносить нову цінність [1, ст. 233]. Натомість колаборативні алгоритми здатні виявляти більш неочевидні вподобання за рахунок аналізу поведінки схожих користувачів, що часто дає вищий recall (охоплення більшої частки релевантних об’єктів). У гібридних системах метрики точності (precision/recall) використовуються для налаштування балансу між контентними і колаборативними компонентами, прагнучи одночасно утримати високу точність і не втратити важливі для користувача рекомендації.

2.1.2. Метрики помилки прогнозування

Якщо рекомендаційна система передбачає явні рейтинги (кількість зірок або оцінки), для оцінки точності такого прогнозування застосовують метрики похибок. Mean Absolute Error (MAE) – середня абсолютна похибка – обчислюється як середнє арифметичне модулів різниць між фактичними рейтингами і передбаченими системою:

$$MAE = \frac{1}{n}\sum_{u,i}|r(u,i) - p(u,i)| \qquad (2.4)$$

Де,

$r(u,i)$ – реальний рейтинг, який користувач *u* поставив елементу *i*;

$p(u,i)$ – прогнозований системою рейтинг для того ж користувача й елемента;

n – кількість передбачених рейтингів[6].

MAE інтуїтивно показує, наскільки в середньому помиляється система в прогнозі: наприклад, MAE = 0.5 вказує, що в середньому рекомендації відхиляються від реальних оцінок на півбалла за п’ятибальною шкалою.

Root Mean Square Error (RMSE) – корінь середнього квадратичного відхилення – є схожою метрикою, але більше виділяє значні похибки за рахунок квадратичної шкали. RMSE визначається як:

$$RMSE = \sqrt{\frac{1}{n}\sum_{u,i}(r(u,i) - p(u,i))^2} \qquad (2.5)\ [6]$$

RMSE активно використовувався, у знаменитому конкурсі Netflix Prize для оцінювання якості колаборативних фільтрів з прогнозування рейтингів фільмів. Чим ближче RMSE до 0, тим точніше алгоритм передбачає оцінки глядачів. Вибір між MAE та RMSE залежить від того, чи варто вважати великі помилки більш критичними, ніж малі. RMSE сильніше враховує рідкісні, але значні відхилення, тоді як MAE просто усереднює всі помилки.

Метрики похибок здебільшого актуальні для колаборативних рекомендаційних систем, орієнтованих на прогнозування рейтингу. У контентно-орієнтованих та гібридних системах явні рейтинги можуть бути відсутні або менш важливі. У таких системах частіше оцінюють якість на рівні списку рекомендацій (precision/recall). Водночас, якщо гібридна система містить компонент прогнозування рейтингу, наприклад, для фільтрації кандидатів, її налаштування та якість можна оцінювати за MAE/RMSE аналогічно до колаборативного підходу.

2.1.3. Метрики ранжування результатів

Коли рекомендаційна система генерує список пропозицій, важливо оцінити не лише, які елементи потрапили до списку, а й наскільки вдало вони відсортовані за релевантністю. Для цього використовують метрики якості ранжування.

Mean Average Precision (MAP) – середня усереднена точність – відображає одночасно частку релевантних рекомендацій і їх порядок у списку[7]. Ця метрика обчислюється спочатку для кожного користувача: розраховується Average Precision – середнє значення precision після кожного знайденого релевантного об'єкта в ранжованому списку рекомендацій. Потім отримані значення усереднюються між усіма користувачами, що й дає MAP.

$$MAP@K = \frac{1}{|U|}\sum_{u=1}^{|U|}(AP@K)_u \qquad (2.6)$$

Де,

$(AP@K)_u$– значення Average Precision для користувача u;

$|U|$ – загальна кількість користувачів.

В результаті MAP@K показує, наскільки добре система в середньому розміщує релевантні об'єкти ближче до початку топ-K списку. Вища MAP означає, що користувачі, переглядаючи перші K рекомендацій, з більшою ймовірністю знайдуть серед них цікаві для себе позиції. Цей показник часто застосовують для оцінки пошукових алгоритмів і рекомендацій у великих каталогах, де важливий саме порядок видачі результатів.

Normalized Discounted Cumulative Gain (NDCG) – нормалізована знижена кумулятивна вигода – є ще однією метрикою ранжування. NDCG враховує не лише факт релевантності, але й позицію релевантного елемента у списку, при цьому

більш високі позиції (верх списку) отримують більшу вагу]. Розрахунок NDCG починається з обчислення DCG (Discounted Cumulative Gain):

$$DCG = \frac{1}{N}\sum_{u=1}^{N}\sum_{j=1}^{J}\frac{g_{u,i_j}}{\log_b(j+1)} \quad (2.7)$$

Де,

N — загальна кількість користувачів у вибірці;

J — кількість позицій у списку рекомендацій;

g_{u,i_j}— користь або «вигода» (gain), яку користувач u отримує від елемента i_j, що знаходиться на позиції i_j;

i_j — елемент (товар, фільм тощо), розташований на j-тій позиції в списку рекомендацій;

b — основа логарифму, яка визначає рівень зменшення ваги з глибиною списку[2].

Часто використовується основа 2, що забезпечує поступове зменшення значущості кожної наступної позиції.

$$NDCG = \frac{DCG}{IDCG} \quad (2.8)$$

Таким чином, високорелевантні товари на 1-2 місцях дають значно більший вклад у DCG, ніж якщо б вони були на 10-му. Отриманий DCG нормалізується через поділ на ідеальний DCG (IDCG) – максимальну можливу суму gain, яку можна отримати, якби всі релевантні об’єкти стояли на початку списку[2]. NDCG визначається як відношення DCG до IDCG і завжди лежить у діапазоні від 0 до 1. Значення NDCG, близьке до 1, означає, що рекомендаційний список майже

оптимально впорядкований: найбільш корисні для користувача продукти стоять першими. Метрика NDCG особливо корисна, коли релевантність градуйована (наприклад, є оцінки 1-5) або важливо оцінити «наскільки хорошим» був ранг кожного знайденого елемента, а не лише факт його наявності у списку.

У реальних e-commerce системах, таких як Rozetka або Amazon, метрики ранжування тісно пов’язані з досвідом користувача: більшість клієнтів переглядає тільки перші результати рекомендаційного списку, тому задача РС – максимально підняти релевантні та привабливі товари вгору. Відповідно, високі показники MAP та NDCG сигналізують, що система успішно показує користувачу потрібні продукти на перших позиціях. Наприклад, якщо алгоритм Amazon відсортував рекомендовані продукти так, що придбані згодом товари знаходились на верхніх рядках списку, то NDCG буде високим, що свідчить про вдале ранжування.

2.1.4 Новизна рекомендацій

Новизна характеризує здатність системи рекомендувати користувачу об’єкти, з якими він ще не знайомий або про які не знав раніше. Висока новизна означає, що рекомендації сприяють ознайомленню з новим контентом: користувач отримує несподівану, але потенційно цікаву інформацію, розширюючи свої горизонти і навпаки, якщо система радить переважно вже відомі користувачу товари (книги одного й того ж автора), її новизна низька. Хоча такі очевидні рекомендації можуть підвищити довіру до системи, вони мають обмежену корисність, адже користувач і сам міг би знайти ці продукти [1, ст. 240].

Особливо актуальною новизна є для контентно-орієнтованих підходів: вони схильні пропонувати дуже схожі елементи на основі профілю користувача, через що рекомендації стають «очікуваними» та менш корисними з точки зору відкриття нового. Наприклад, Музичний сервіс Spotify, орієнтуючись на прослуховування рок-музики, може пропонувати ті ж самі рок-гурти, які користувач уже слухав. Колаборативні системи за своєю природою можуть досягати вищої новизни, оскільки враховують досвід інших користувачів: наприклад, система може

запропонувати нову для вас книгу, яка сподобалась людям зі схожими інтересами. Проте й колаборативні алгоритми страждають від «ефекту популярності»: вони часто рекомендують популярні, відомі широкій аудиторії товари, що знижує новизну рекомендацій.

Оцінити новизну кількісно можна різними способами. В онлайнових експериментах новизну вимірюють через опитування: користувачів просять вказати, чи були вони знайомі з товарами до рекомендації. В офлайновому режимі оцінити новизну можна, аналізуючи часові мітки: наприклад, оцінювати, наскільки система здатна передбачити вподобання користувача щодо нових об’єктів, які той відкрив для себе пізніше (після певного контрольного моменту часу) .

2.1.5. Серендипність рекомендацій

Серендипна рекомендація – це така, якої користувач не очікував, але вона виявилася для нього корисною і цікавою. Таким чином, серендипність є сильнішою вимогою, ніж новизна: всі серендипні рекомендації нові для користувача, але не кожна нова рекомендація є серендипною [1, ст. 234]. Важливим є елемент неочевидності: якщо система радить щось надто прогнозоване, це не викличе ефекту «приємної несподіванки».

Як приклад: якщо користувач часто вечеряє в індійських ресторанах, рекомендація нового індійського ресторану буде для нього новою, якщо він там ще не був, але навряд чи серендипною, адже така порада очікувана і логічна. Натомість рекомендація скуштувати ефіопську кухню може стати приємною несподіванкою [1, ст. 234]. Отже, серендипність можна розглядати як міру відхилення від очевидності, за умови збереження.

Для кількісного вимірювання серендипності зазвичай поєднують дві складові: корисність рекомендації та її неочевидність. В онлайн-експериментах користувачів можуть питати, наскільки запропонований товар їм корисний та чи очікували вони таку рекомендацію. Частка рекомендацій, що одночасно оцінені як корисні і несподівані, приймається за показник серендипності.

2.1.6. Різноманітність рекомендацій

Різноманітність (diversity) – метрика, що оцінює, наскільки неоднорідним є набір рекомендацій, що надаються користувачу[1, ст. 234]. Ідея полягає в тому, що список з декількох рекомендацій повинен покривати різні інтереси або категорії, щоб збільшити шанси на те, що хоча б щось із запропонованого сподобається. Якщо всі рекомендовані об’єкти дуже схожі між собою (наприклад, три фільми одного жанру з одними й тими ж акторами), то різноманітність списку. У такому випадку, якщо перша ж рекомендація не буде релевантною, висока ймовірність, що й інші не сподобаються. Натомість, коли користувачу пропонують різнотипні варіанти – наприклад, фільми різних жанрів або товари з різних категорій – підвищується імовірність, що він зацікавиться хоча б однією з рекомендацій, що підвищує корисність системи в цілому.

Різноманітність вимірюють, як правило, через відстань (несхожість) між рекомендованими продуктами. Найпоширеніший спосіб - використати метрику intra-list diversity (ILD), яка передбачає знаходження середньої попарної відстані або відмінності між усіма парами елементів у рекомендованому списку. Для розрахунку цієї відстані можуть бути використані векторні представлення контенту або характеристика товарів. Наприклад, якщо у топ-5 рекомендаціях середня схожість між товарами дорівнює 0.3 (в певній шкалі), а в іншому алгоритмі – 0.7, то перший список є значно більш різноманітнішим (низька схожість відповідає високій різноманітності).

У кожному типі РС існують свої підходи досягнення різноманітності. РС на основі контенту можуть гнучко керувати подібністю об’єктів у списку, змінюючи налаштування фільтрів або ваги атрибутів пробукту, тоді як колаборативні підходи часто схильні пропонувати однаково популярні товари для груп користувачів зі схожими вподобаннями. Гібридні системи часто впроваджують диверсифікацію: наприклад, після отримання списку колаборативних рекомендацій додатково додають декілька менш схожих на профіль користувача елементів для збільшення загальної різноманітності.

2.1.7. Охоплення рекомендаційної системи

Охоплення або покриття (coverage) показує, яку частину користувачів або об'єктів здатна охопити система своїми рекомендаціями [1, ст. 231]. Високе покриття означає, що майже кожен користувач отримає достатній перелік рекомендацій, і більшість об’єктів каталогу мають шанс бути комусь рекомендованими. Низьке покриття свідчить про те, що певні користувачі залишаються без рекомендацій (через відсутність достатньої кількості даних про них) або ж деякі товари ніколи не з’являються у жодному списку рекомендацій.

Розрізняють два види покриття. Покриття по користувачах (user-space coverage) – це частка користувачів, для яких система здатна сформувати хоча б k рекомендацій. Зазвичай k обирають рівним довжині списку, який показується (наприклад, top-10). Якщо для деяких користувачів система не може знайти навіть 10 релевантних елементів, то покриття по користувачах зменшується. Така ситуація виникає, наприклад, для «нових» або малоактивних користувачів: через дуже обмежений профіль (дуже мало оцінок або покупок). Рішенням може бути використання гібридних підходів чи введення дефолтних рекомендацій, але останні підвищують покриття ціною зниження.

Покриття по товарах (item-space coverage) визначається як частка об’єктів, які система взагалі здатна коли-небудь рекомендувати[1, ст. 231]. Колаборативні фільтри, наприклад, не можуть рекомендувати товар, якому жоден користувач не поставив оцінку – такі об’єкти випадають з поля зору РС. У системах на основі контенту навпаки, для кожного товару можна знайти певні схожі профілі і теоретично його рекомендувати, тож покриття каталогу вище. Для кількісної оцінки використовують показник каталогового покриття - це частка всіх доступних об'єктів, які були рекомендовані хоча б одному користувачу. Каталог у цьому випадку означає повний перелік товарів, фільмів, пісень чи інших елементів, які система може запропонувати.

Формально каталогове покриття визначається як відношення кількості унікальних об'єктів, що потрапили у рекомендації, до загальної кількості елементів у каталозі.

$$CC = \frac{|\cup_{u=1}^{m} T_u|}{n} \quad (2.9)$$

Де,

T_u – множина top-k рекомендацій, отриманих користувачем u,

m – кількість користувачів,

n – загальна кількість унікальних об’єктів у системі[2].

Таким чином рахується частка товарів, що реально з’являються в чиїхось рекомендаціях.

Низьке покриття каталогу свідчить про те, що система концентрується на вузькому наборі популярних позицій. Тому прагнуть високого покриття: щоб рекомендаційна система давала шанс якомога більшій кількості продуктів знайти свого покупця. З іншого боку, збільшення покриття часто має зворотний зв’язок з точністю: алгоритм може почати рекомендувати менш релевантні (але раніше не рекомендовані) товари, щоб розширити охоплення. Оптимальна стратегія – знайти баланс, за якого і точність, і покриття залишаються на прийнятному рівні[1, ст. 231].

2.2. Вплив рекомендаційних систем на бізнес-метрики: конверсія, середній чек, утримання клієнтів

Суто алгоритмічні метрики не гарантують бізнес-успіху. Важливо розуміти, що високі значення точності або повноти ще не означають підвищення прибутків чи лояльності клієнтів. Система рекомендацій може точно передбачати вподобання користувача, але це не обов’язково відображає, як реальні клієнти реагують на рекомендації з точки зору бізнес-цілей підприємства [9].

Сучасні компанії все більше звертають увагу на бізнес-метрики (KPI) при оцінці ефективності рекомендаційних систем. На відміну від суто технічних показників, бізнес-метрики демонструють, яку реальну користь отримує компанія від рекомендацій. Покупки не є єдиним показником цінності клієнта для компанії. У багатьох випадках споживачі сприяють розвитку підприємства іншими способами: залишаючи відгуки, рекомендують продукцію знайомим або демонструють стабільну прихильність до бренду. Такі форми взаємодії не завжди мають прямий фінансовий еквівалент, проте вони суттєво впливають на репутацію компанії, рівень довіри до неї та сприяють залученню нових клієнтів. У підсумку це створює додаткову цінність, яка, хоч і не завжди вимірюється безпосередньо, але є важливою для довгострокового успіху компаній. Тому, для більш повного розуміння внеску клієнта у бізнес-процеси була запропонована концепція Customer Engagement Value (CEV), що відображає загальну цінність взаємодії споживача з компанією. Вона включає:

- Customer Lifetime Value (CLV або LTV) — загальний прибуток, який компанія отримує від клієнта протягом усього періоду співпраці з ним;
- Customer Referral Value — користь, яку приносить клієнт, залучаючи нових покупців через особисті рекомендації або неформальну комунікацію (наприклад, «сарафанне радіо»);
- Customer Influencer Value — вплив клієнта на поведінку інших, зокрема через публічні оцінки, коментарі чи участь в обговореннях, які можуть сприяти продажам або лояльності;
- Customer Knowledge Value — цінність інформації, яку компанія отримує завдяки клієнту, зокрема через зворотний зв’язок, ідеї покращення або участь у вдосконаленні продуктів і послуг[5, ст. 299].

Ця концепція підтвержує, що якщо оцінювати роботу рекомендаційної системи лише за одним критерієм (наприклад, кількістю кліків чи здійснених покупок) можна отримати спотворене уявлення про її реальний внесок у розвиток підприємства. Такий обмежений підхід не враховує повної картини впливу. Саме тому, в електронній комерції виникла необхідність використовувати не тільки

технічні метрики, а й бізнес-орієнтовані показники, які дозволяють глибше зрозуміти, наскільки ефективно система допомагає досягати цілей компанії. До таких ключових індикаторів належать різноманітні KPI, що відображають фінансові результати, поведінку клієнтів і загальний рівень взаємодії з платформою:

- Прибуток. Безпосередній фінансовий результат від дій користувача,

що виникли під впливом рекомендацій. Вимірюється як додатковий дохід або маржа від товарів, проданих через рекомендації. Цей показник відображає кінцеву мету будь-якого комерційного алгоритму. Рекомендаційна система, яка збільшує прибуток (наприклад, через продаж дорожчих товарів або комплектацію замовлень супутніми товарами), має пряму бізнес-цінність, навіть якщо її точність не максимальна.

- Частка повторних покупок. Відсоток транзакцій, що здійснюються

постійними клієнтами, а не новими. Зростання цього показника свідчить про лояльність: користувачі повертаються в магазин після отриманого досвіду. Якщо впровадження персональних рекомендацій призводить до того, що клієнти частіше здійснюють повторні покупки, це індикатор підвищення довіри та задоволеності сервісом.

- Lifetime Value (LTV) клієнта. Цей показник агрегує довгострокову

цінність клієнта для бізнесу – сумарний дохід від клієнта за весь час. Ефективні рекомендації можуть підвищувати LTV, спонукаючи користувача купувати більше позицій за одне відвідування та здійснювати покупки довше протягом життєвого циклу. На відміну від разових конверсій, LTV враховує перспективу утримання клієнта: навіть якщо перша покупка за рекомендацією невелика, головне – щоб клієнт залишився задоволеним і повернувся знову.

- Активність та залученість клієнтів. Це може виражатися у частоті

відвідувань сайту, переглядах сторінок, додаванні товарів до кошика, залишенні відгуків чи інших взаємодіях. Залученість є проміжним показником, що відображає, наскільки цікаві користувачу пропозиції. Дослідження показують, що наявність персональних рекомендацій підвищує активність користувачів на

платформах. А підвищена активність часто корелює з утриманням клієнтів: чим більше клієнт взаємодіє з сервісом, тим вища ймовірність, що він не перейде до конкурента і здійснить нові покупки.

Використання цих бізнес-метрик дає глибше уявлення про ефективність рекомендаційної системи, оскільки вони прив'язані до стратегічних цілей компанії. Наприклад, навіть якщо якась модель рекомендацій має трохи нижчу точність, але демонструє вищий прибуток чи LTV (через те, що рекомендує більш релевантні з точки зору бізнесу товари), її прийнято вважати успішнішою для компанії. Традиційні метрики можуть слугувати первинним фільтром якості алгоритму, але остаточну оцінку слід робити за показниками, що впливають на дохід і розвиток клієнтської бази [9]. Важливо також враховувати довгострокові ефекти: рекомендація, яка сьогодні не конвертувалась у продаж, може підвищити лояльність або привести нового клієнта завтра , а такі опосередковані впливи теж слід вимірювати.

Розглянемо, як провідні компанії електронної комерції застосовують бізнес-показники для оцінки та вдосконалення своїх рекомендаційних систем:

Найбільший український онлайн-рітейлер Rozetka впроваджує персоналізовані рекомендації на своєму веб-сайті та в додатку, прагнучи підвищити як задоволеність клієнтів, так і комерційні показники. Розетка аналізує поведінку покупців (перегляди, додавання до списків бажань, покупки) і на основі цих даних рекомендує товари в різних секціях ("Схожі товари", "Вас можуть зацікавити" тощо). Ефективність таких рекомендацій компанія оцінює перш за все через зростання продажів та утримання клієнтів. Практичним KPI є середній чек (середня вартість замовлення): якщо персональні рекомендації спонукають клієнтів купувати додаткові товари або дорожчі альтернативи, середній чек зростає. Також Rozetka відстежує частку повторних покупок – повернення клієнтів за новими замовленнями. Вдалий рекомендаційний механізм має збільшувати цей показник, оскільки задоволений персональним підходом покупець частіше повертається. На відміну від суто технічної оцінки точності підбору товарів, фокус на бізнес-метриках дозволяє Rozetka постійно вдосконалювати систему

рекомендацій під потреби ринку: товари, що генерують вищий прибуток або сприяють лояльності, отримують пріоритет у рекомендаціях.

Глобальний гігант Amazon одним із перших продемонстрував бізнес-цінність персоналізованих рекомендацій. Рекомендаційна система Amazon генерує значну частину додаткових продажів. За оцінками McKinsey, близько 35% усього доходу Amazon забезпечується саме системою рекомендацій [10]. Цей колосальний внесок став можливим завдяки фокусуванню на бізнес-метриках: Amazon ретельно відстежує, які рекомендації призводять до збільшення середнього чеку та частоти покупок. Кожна зміна алгоритмів тестується через показники конверсії та виручки. Якщо нова модель підвищує виторг (навіть за незмінних традиційних метрик як-от точність), її впроваджують у масштабі. Такий підхід дозволив Amazon досягти двох цілей одночасно: забезпечити клієнтам цікавий вибір товарів і максимізувати монетизацію трафіку.

2.3. Аналіз підходів до оцінювання ефективності рекомендацій в e-commerce

Оцінювання ефективності рекомендаційних систем є важливою і непростою задачею, особливо в сфері електронної комерції. Рекомендаційні системи в e-commerce повинні задовольняти різноманітні вимоги бізнесу та користувачів, а самі рекомендації мають багато аспектів якості. Існує широкий спектр метрик для оцінки різних сторін ефективності, що ускладнює порівняння алгоритмів.

Традиційно ефективність рекомендаційних алгоритмів оцінюється офлайн та онлайн. Офлайн-оцінювання передбачає розподіл на тестову і тренувальну вибірки та вимірювання якості рекомендацій на тестових даних за різними метриками. Основними є метрики точності прогнозування: для задач передбачення рейтингів широко використовуються середньоквадратична помилка (RMSE) і середня абсолютна помилка (MAE), а для задач формування списків рекомендацій – метрики прецизійності та повноти (Precision, Recall) на топ-N, тощо[8, ст. 11]. Для ранжованих списків популярними є метрики середньої точності (MAP), NDCG

(нормалізований дисконтований кумулятивний гейн) та MRR (середнє обернене ранжування), які враховують позиції релевантних рекомендацій у списку. Наприклад, Precision@K відображає частку релевантних елементів серед топ-K рекомендацій, а Recall@K – частку знайдених релевантних елементів від загальної їх кількості[7]. Ці метрики інтуїтивно зрозумілі та вимірюють коректність рекомендацій (accuracy) – наскільки система пропонує саме ті товари, які цікавлять користувача.

Окрім точності, офлайн-експерименти дозволяють оцінювати й інші аспекти якості рекомендацій. Зокрема, новизна та несподіваність рекомендацій (novelty, serendipity) характеризують, наскільки запропоновані товари є новими або неочікуваними для користувача. Різноманітність списку рекомендацій (diversity) визначає, чи не є всі рекомендовані товари занадто подібними між собою. Охоплення (coverage) оцінює, яку частку доступних товарів здатна рекомендувати система (каталогове охоплення). Серендипність (неочікувана корисність) – метрика, споріднена з новизною, яка показує, чи пропонує система приємні сюрпризи (товари, які користувач не очікував, але вони йому сподобалися). Включення цих метрик дозволяє вийти за межі одновимірної точності та оцінити інтерес користувача до рекомендацій[7]. Важливо зазначити, що класичні метрики часто знаходяться в компромісних відносинах: наприклад, надто високий акцент на новизні може знизити точність або довіру, оскільки занадто несподівані рекомендації іноді сприймаються напружено. Тому при офлайн-оцінюванні нерідко обчислюють набір показників – для повноти картини.

Іншим традиційним напрямом є онлайн-оцінювання, що відбувається безпосередньо на реальних користувачах у живій системі. В контексті e-commerce це найчастіше реалізується через A/B-тестування: кожна група користувачів випадково отримує різні алгоритми рекомендацій, і порівнюється їхня поведінка. Онлайн-метрики фіксують реальний вплив рекомендацій на користувача та бізнес[11]. Зокрема, основними показниками ефективності в електронній комерції є показники залученості та конверсії: CTR (Click-Through Rate) – частка рекомендацій, на які користувачі клікнули; CVR (Conversion Rate) – частка

рекомендацій, що призвели до покупки; середній чек, дохід від рекомендацій, збільшення кількості товарів у кошику, тощо[8].Онлайн-метрики безпосередньо відображають задоволеність користувача та досягнення бізнес-цілей, тому вважаються «золотим стандартом» оцінювання. Встановлено, що офлайн-результати не завжди корелюють з онлайн-поведінкою споживачів. Наприклад, досвід AliExpress показав, що модель з високими показниками на тестових даних може демонструвати слабшу ефективність при реальному використанні, і навпаки, класичні офлайн-метрики інколи вводять в оману щодо реального впливу рекомендацій. Тому остаточна перевірка якості рекомендацій проводиться саме в онлайн-середовищі. З іншого боку, онлайн-експерименти потребують значних ресурсів і часу, ретельного дослідження (щоб коректно приписати зміни в метриках впливу саме рекомендаційній системі) і можуть супроводжуватися ризиками для бізнесу. Тому на практиці компанії комбінують ці методи: спочатку тестують різні алгоритми на історичних даних, а потім найперспективніші з них перевіряють на реальних користувачах за допомогою обмежених A/B-тестів. Так, Amazon спирається на офлайн-метрики при розробці моделей і паралельно відстежує онлайн-метрики взаємодії користувачів – наприклад, компанія вимірює CTR рекомендацій у реальному часі та конверсії для оцінки успіху персоналізованих пропозицій[11]. Подібним чином, інші платформи (AliExpress, Netflix, Facebook та ін.) запровадили безперервний онлайн-моніторинг показників рекомендаційних систем і періодичні експерименти, щоб переконатися, що покращення в офлайн-метриках перетворюються на реальну вигоду.

Варто також відзначити, що ефективність рекомендаційної системи визначається не лише об’єктивними показниками точності, а й суб’єктивним задоволенням користувачів. Успішна система повинна будувати довгострокові відносини з клієнтами, викликати в них довіру та позитивні емоції від взаємодії. Тому в дослідженнях усе більше уваги приділяється довірі користувачів до рекомендацій та їхньому емоційному сприйняттю. Довіру можна визначити як ступінь, на скільки користувач вважає рекомендації надійними і корисними. Емоційний аспект оцінюється зазвичай якісними методами – через опитування,

відгуки, аналіз поведінки (чи повертається користувач до рекомендацій, чи відчуває ефект «приємного здивування»). Хоча такі фактори складніше виміряти кількісно, їх врахування є важливою частиною сучасного підходу до оцінки. Дослідження підтверджують, що поєднання високої точності рекомендацій з належною різноманітністю підвищує загальне задоволення і лояльність клієнтів[12]. Тому e-commerce платформи (такі як, Rozetka) оцінюють успіх рекомендацій не лише за короткостроковими показниками (скільки товарів додано в кошик за рекомендацією), а й за довгостроковими – зростанням задоволеності, повторних візитів і покупок, що відображають довіру до системи.

На основі вищесказаного, ефективність рекомендаційної системи проявляється у багатьох вимірах – точність, новизна, різноманітність, охоплення, довіра, тощо – і жодна окрема метрика не здатна повністю охопити успішність системи. Класичний підхід, коли алгоритми порівнюються за однією метрикою, є обмеженим. З іншого боку, оцінювання кожної характеристики окремо створює ситуацію, за якої кілька алгоритмів мають «розсіяні» переваги: один найкращий за точністю, інший – за різноманітністю, третій – за новизною тощо, і загальний вибір стає нетривіальним. Таким чином, постала потреба в інтегральних підходах до оцінювання, які дозволяють врахувати багатовимірність, але водночас дати єдину зрозумілу оцінку для порівняння систем[8, ст. 7].

У відповідь на цю проблему з’явилися наукові розробки, що пропонують комплексні метрики та методики. Один із новітніх підходів – Comprehensive Performance (ComPer), запропонований у 2019 році А. Альслайті та Т. Траном. ComPer розглядає рекомендаційну систему як «людину, що навчається», проводячи аналогію між процесом рекомендації та процесом навчання людини [8, ст. 16]. У цій моделі різні аспекти ефективності РС зіставляються з рівнями освоєння знань за відомою таксономією Блума (рис. 2.1), що визначає навчальні цілі на рівні пам’яті, розуміння, застосування, аналізу тощо.

Рис. 2.1 Таксономія Блума

В результаті автори ComPer виокремили п’ять ключових вимірів оцінки рекомендаційної системи, які є найпопулярнішими у літературі та загалом покривають основні потреби користувачів. До цих вимірів увійшли: Correctness (Коректність) – тобто точність рекомендацій, їх відповідність інтересам (ця метрика відображає, наскільки рекомендації правильні і релевантні; він близький до традиційних метрик Precision/Recall, тощо); Coverage (Охоплення) – як багато різних елементів каталогу здатна рекомендувати система і які частки контенту/користувачів задіяні; Diversity (Різноманітність) – різноманітність списків рекомендацій, відсутність надмірної однотипності рекомендованих товарів; Robustness (Надійність) – стійкість алгоритму до викривлень даних, тобто наскільки не погіршується якість рекомендацій у разі навмисних чи випадкових помилок; Scalability (Масштабовність) – здатність працювати ефективно при збільшенні обсягів даних, зазвичай вимірювана часом генерації рекомендацій або використанням ресурсів [8, ст. 15].

ComPer — це підхід, який дозволяє всебічно оцінити рекомендаційні алгоритми, зводячи різні метрики в один узагальнений показник. Замість того щоб аналізувати кожну характеристику окремо, як-от точність, різноманітність чи масштабованість, система зводить усі ці значення до єдиної шкали та об'єднує їх у фінальний бал. Такий підхід не лише спрощує порівняння між алгоритмами, а й забезпечує більш збалансовану оцінку, в якій жодна з метрик не переважає над іншими. Це особливо корисно, коли потрібно швидко визначити, який алгоритм є

найбільш придатним для конкретного завдання. Хоча метод і не є універсальним рішенням для всіх випадків, його гнучкість дозволяє адаптувати вагу кожного критерію залежно від галузі або цілей компанії. У результаті ComPer виступає як зручний інструмент для узагальненого аналізу, який поєднує багатовимірний підхід з простотою інтерпретації.

Окрім ComPer, існують й інші спроби об’єднати результати за різними метриками або ж виробити загальну стратегію оцінки. Наприклад, методики багатокритеріальної оптимізації намагаються врахувати одночасно кілька цілей при навчанні алгоритму, що опосередковано впливає і на оцінку – алгоритм шукає компроміс, щоб обидва показники були на прийнятному рівні. Хоча такі комплексні оцінки менш формалізовані, ніж підхід ComPer, вони підкреслюють важливість багатовимірного бачення ефективності РС.

Прикладом застосування вищеописаних підходів є Amazon. У його масштабній рекомендаційній системі використовується багаторівнева схема оцінювання: внутрішні тестування алгоритмів включають десятки метрик (точність передбачення купівель, утримання клієнтів, новизна товарів тощо), а фінальне рішення про ефективність моделі приймається за результатами онлайн-експериментів, що враховують і прямі бізнес-метрики (зростання продажів, виручки), і побічні ефекти (наприклад, чи не знизилась загальна різноманітність купованих товарів через надмірну персоналізацію).

Компанія AliExpress (Alibaba Group) також повідомляла про використання комплексних підходів: зокрема, для пошукових рекомендацій вони запровадили систему, що оцінює модель за спеціальним інтегральним показником, більш узгодженим із онлайн-конверсіями, ніж класичні метрики ранжування[13, ст. 1]. Це дозволило підвищити коефіцієнт конверсії на ~2% у A/B-тестах після впровадження нової моделі, оптимізованої на такий комплексний показник [13, ст.1].

Український онлайн-рітейлер Rozetka офіційно не публікував деталей про свої метрики оцінювання рекомендацій, але, ймовірно, застосовує аналогічні принципи. З огляду на високу конкуренцію і вимогливість користувачів, Rozetka

має слідкувати не лише за тим, щоб рекомендовані товари відповідали уподобанням (високі CTR), але й за тим, щоб рекомендації сприяли збільшенню середнього чеку та повторних продажів, не відштовхували клієнтів нав’язливими або нерелевантними пропозиціями та підтримували довіру до бренду.

Отже, сучасні e-commerce платформи прагнуть оцінювати рекомендаційні системи всебічно, комбінуючи кілька підходів задля отримання об’єктивної картини

У Таблиці 2.2 наведено порівняльний аналіз основних підходів до оцінювання ефективності рекомендаційних систем.

Таблиця 2.2

Порівняння основних підходів до оцінювання ефективності рекомендаційних систем

Підхід оцінювання	Переваги	Недоліки
Офлайн (на основі даних)	– Швидке і недороге тестування на історичних даних; – Відтворюваність результатів, контрольовані умови експерименту; – Можна порівняти багато алгоритмів паралельно.	– Не завжди відображає реальну поведінку користувачів; – Може переоцінювати дрібні покращення, не враховує фактори користувацького досвіду; – Ризик перенавчання на наявних патернах, що не узгоджуються з майбутніми
Онлайн (A/B-тести, live-метрики)	– Безпосередньо відображає вплив на користувачів і бізнес (реальні CTR, конверсії	– Потребує часу і ресурсів (довготривалі експерименти, трафік);

Продовження таблиці 2.2

Підхід оцінювання	Переваги	Недоліки
Онлайн (A/B-тести, live-метрики)	– Враховує всі фактори системи в цілому (інтерфейс, затримки, актуальність даних); – Є остаточним критерієм успіху (користувацьке схвалення).	– Ризик негативного впливу на користувачів під час тестів з гіршим варіантом; – Складність аналізу: важко ізолювати причини змін метрик, зовнішні чинники впливають.
Багатометриковий (окремі метрики)	– Дає комплексне уявлення про різні аспекти якості (точність, новизна, різноманітність тощо); – Дозволяє виявити сильні та слабкі сторони алгоритму по кожному критерію;	–Інтерпретація результатів ускладнюється, якщо алгоритми перемагають за різними метриками (що важливіше?) – Метрики можуть конфліктувати (напр. підвищення новизни знижує точність), немає єдиної міри успіху;
Комплексний (інтегральний показник)	– Узагальнює кілька вимірів в одному числі для простого порівняння; – Охоплює ключові аспекти ефективності одночасно, забезпечуючи збалансовану оцінку;	– Потребує обґрунтованого вибору набору метрик і методики агрегування (ваги тощо); – Є певне спрощення: одна цифра не дає деталізованої картини, можливе «розмиття» важливої інформації

Продовження таблиці 2.2

Підхід оцінювання	Переваги	Недоліки
Комплексний (інтегральний показник)	– Полегшує відбір моделей (чіткий рейтинг за інтегральним балом)	– Універсальна метрика може не повністю відповідати специфічним цілям окремого бізнесу [8]

Висновок до розділу

У другому розділі було досліджено підходи до оцінювання ефективності рекомендаційних систем, зокрема систематизовано метрики точності, ранжування, новизни, різноманітності та серендипності. Також розглянуто вплив рекомендацій на ключові бізнес-показники, такі як конверсія, середній чек, та утримання клієнтів. Окрема увага була приділена залежності вибору метрик від цілей бізнес-проєктів: технічні метрики не завжди корелюють із бізнес-результатами, тому важливо враховувати контекст використання систем. Таким чином, було обґрунтовано доцільність застосування комплексного підходу до оцінки ефективності рекомендацій, що включає як користувацькі, так і комерційні аспекти.

3. РОЗРОБКА МЕТОДИКИ ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ІНТЕЛЕКТУАЛЬНИХ РЕКОМЕНДАЦІЙНИХ СИСТЕМ ДЛЯ E-COMMERCE

3.1. Постановка завдання: критерії оцінки ефективності рекомендаційних алгоритмів

Формування об’єктивної методики для оцінювання ефективності інтелектуальних рекомендаційних систем у сфері e-commerce вимагає ретельно продуманого підходу до вибору метрик, що відображають різні рівні продуктивності алгоритмів — від точності прогнозів до гнучкості реакції на зміну середовища користувача. Класичні підходи в галузі інформаційного пошуку, як-от Precision, Recall, F1-score, MAP (mean average precision) або NDCG (normalized discounted cumulative gain), хоч і мають широке застосування, але не охоплюють повного спектра поведінкових і економічних параметрів, що виникають у динамічному середовищі електронної комерції [17, с. 9].

У межах таких платформ, як маркетплейси, агрегатори, магазини з багатошаровим каталогом товарів, рекомендаційна система виконує не лише функцію персоналізованого відбору, а й діє як інструмент впливу на поведінкові патерни користувача, його час перебування в системі, глибину скролінгу, готовність до імпульсивного придбання та довгострокову лояльність. Отже, оцінювання ефективності не може базуватися винятково на класифікаційній точності. Потрібно враховувати параметри бізнес-релевантності — зокрема коефіцієнт конверсії з рекомендації в покупку, середній чек за рекомендованими товарами, показники відмов (bounce rate) після взаємодії з блоками рекомендацій, а також частоту повернення користувача на платформу після прийняття пропозиції системи. Ці метрики переходять від чисто технічної площини до гібридної, що об’єднує поведінкові, комерційні та когнітивні індикатори, що й формує основу сучасної методики оцінки.

Одним із концептуально необхідних параметрів є адаптивність рекомендаційного алгоритму — здатність моделі змінювати вектор пропозицій відповідно до зміни контексту, зовнішнього трафіку, попередніх дій користувача та динаміки трендів. Для вимірювання цього компоненту застосовують метрики temporal diversity, coverage over time, entropy over session, а також коефіцієнти кореляції між блоками рекомендацій у різних часових вікнах. Адаптивність є індикатором життєздатності системи, особливо в умовах високочастотної зміни переваг споживачів і появи нових SKU. У цій площині також постає потреба розмежування cold-start і warm-start сценаріїв, де перший передбачає відсутність попередньої взаємодії користувача, а другий дозволяє інкорпорувати історію поведінки. Оцінювання ефективності в cold-start середовищі вимагає окремого підходу — наприклад, через метрики на кшталт hit-rate@k для нових користувачів, або precision-coverage trade-offs, де оцінюється компроміс між точністю та шириною охоплення на невизначених даних. Крім того, слід зауважити, що рекомендаційна система в e-commerce — це не ізольований модуль, а частина мультиагентної архітектури, яка взаємодіє з модулем пошуку, фільтрації, акційного таргетування та управління кошиком. Тому ефективність рекомендаційного ядра доцільно оцінювати в умовах A/B-тестування або мультиваріантного аналізу, де кожна конфігурація рекомендацій фіксується на підмножині користувачів із подальшою статистичною верифікацією результатів [33,с. 47].

Не менш значущим для методики є визначення релевантності рекомендацій — параметра, який не завжди збігається з точністю, оскільки може мати різну семантичну або прагматичну природу. У межах сучасного підходу до інформаційної персоналізації релевантність поділяється на синтаксичну, семантичну й інтенціональну. Синтаксична відображає безпосередню відповідність товару уподобанням користувача за простими ознаками (категорія, бренд, ціна). Семантична охоплює глибші патерни, як-от стилістична схожість, комплементарність (товар-доповнення) або сезонність. Інтенціональна ж орієнтується на приховані наміри користувача — наприклад, потребу в подарунку, покупці для іншої особи або разовій покупці без подальшого ланцюга. Для

оцінювання таких типів релевантності використовують мультимодальні метрики, що поєднують поведінкові трекери (глибина перегляду, dwell time), обробку семантики запитів і постінтерактивну аналітику (чи відбулося повернення до товару після перегляду). Особливу увагу також приділяють метрикам diversity (різноманітність) і novelty (новизна), які сигналізують про спроможність системи не лише повторювати шаблонні рекомендації, а й інтегрувати нові, менш популярні або нестандартні пропозиції. У цьому контексті оцінка ефективності включає показники long-tail recommendation performance, що дають змогу виявити здатність системи виводити з «тіні» товари з низькою частотою купівлі, але високим потенціалом монетизації [23, с. 4].

Особливе значення у створенні загальної системи критеріїв мають обмеження — як технічні, так і бізнесові. До перших відносяться апаратні ресурси, доступний обсяг пам’яті, латентність обробки запитів, пропускна здатність каналу і продуктивність у пікові години. Сучасні e-commerce платформи працюють з рекомендаційними модулями в режимі near real-time, отже, кожен запит має оброблятися за кілька сотень мілісекунд, що створює тиск на використання складних моделей, таких як deep neural collaborative filtering, або трансформерних архітектур. Обмеження можуть зумовити спрощення моделей або використання гібридних підходів — попереднє ранжування за правилами (rule-based), із подальшим уточненням через легку ML-модель (наприклад, XGBoost). До бізнесових обмежень входять вимоги до справедливості (fairness), контролю за просуванням певних SKU, відстеження ротації асортименту й забезпечення етичного таргетування. Наприклад, алгоритм не повинен створювати дискримінацію за демографічними ознаками або схиляти до купівлі надмірно дорогих товарів користувачам із низькою платоспроможністю. У таких випадках доцільно застосовувати обмеження через soft-constraints у loss-функції моделі або вбудовані фільтри на рівні API-інтерфейсу, що блокує певні категорії товарів у результатах.

Для формування тестового середовища, в якому буде проводитись оцінювання, потрібно заздалегідь визначити типи вхідних умов. Це охоплює

параметри сесійної активності, сегмент користувача, тип платформи (мобільна чи десктопна), джерело трафіку (органічне, рекламою, через email), поточний статус кошика, історія покупок, розміри екрану, мова інтерфейсу тощо. Усі ці змінні формують базу для створення валідного дата-сету, у якому можна ізольовано вимірювати вплив рекомендацій. На практиці використовуються як live-сесії (через інтеграцію із сервером), так і симульовані сценарії, побудовані на історичних логах поведінки. В останньому випадку створюються траєкторії на кшталт Markov Decision Processes або Graph-Based Trajectories, що дозволяють відтворити ймовірні сценарії користування в умовах зміни рекомендаційного вектора. Оцінювання також вимагає використання метрик інтервальної стабільності, які відстежують послідовність рекомендацій у межах однієї сесії. Це особливо важливо в мобільних додатках, де користувачі здійснюють фрагментовану навігацію, перемикаючись між категоріями, а рекомендаційна система повинна утримувати когерентний логічний ланцюг усього вікна активності [25, с. 84].

У структурі оцінювання враховуються й спеціальні сегменти e-commerce, як-от категорії із високим обігом (fast-moving consumer goods), сезонні товари, техніка, fashion-товари або товари для дому. У кожному з них встановлюються специфічні вимоги до рекомендацій — наприклад, у сегменті одягу критичною є сезонна актуальність, стилістична сумісність та розмірна сітка, тоді як у товарах для дому значущою є сумісність з уже наявними елементами. Це означає, що загальна система метрик має бути гнучкою, з можливістю вагового налаштування залежно від сегменту. Для цього використовують підхід contextual evaluation layers — коли для кожної категорії вбудовується набір локальних ваг і обмежень, який адаптує загальну модель оцінювання до мікросценаріїв. Таке фрагментарне налаштування дозволяє отримати більш релевантні показники, ніж у разі застосування універсального шаблону метрик до всієї платформи. У складних випадках, коли категорія включає крос-продукти з різних підтипів, як у випадку великих онлайн-маркетів, метрики оцінювання створюються на основі автоенкодерної кластеризації категорій товарів, що дозволяє формувати рекомендаційні кластери за поведінковими патернами, а не за штучно заданими тегами.

Постановка завдання з формалізацією критеріїв оцінювання ефективності рекомендаційних алгоритмів у сфері e-commerce реалізовувалася на основі попередньо структурованого аналізу архітектури цифрових платформ, що функціонують у високонавантаженому середовищі багатовимірної взаємодії користувача з продуктовим асортиментом. На початковому етапі було закладено набір вихідних параметрів, які закріплюють модельну основу розробки подальшого дослідження. Ці параметри враховували реальні умови роботи алгоритмів у середовищі розподілених обчислень із високою частотою оновлення даних і великою кількістю одночасних сесій користувачів. Серед таких параметрів було виокремлено обсяг історичних даних на вхідному рівні, тип поведінкової інформації (імпліцитні чи експліцитні дії), щільність взаємодій у матриці користувач–товар, швидкість оновлення профілю користувача, розмір релевантного топ-N списку для кожного сеансу генерації рекомендацій, а також контекстна інформація, що включала геолокацію, пристрій, часові патерни сесії, сезонність і частотність звернень. Усі ці вхідні дані були необхідними для побудови повноцінної системи тестування, яка дозволяє забезпечити експериментальну валідацію результатів, отриманих у межах порівняльного аналізу алгоритмів [19, с. 82].

Крім вхідних параметрів, була сформульована гіпотеза, що різні категорії товарів — з різною циклічністю попиту, середнім життєвим циклом, глибиною асортименту, рівнем новизни й оновлюваністю — вимагають різних алгоритмів адаптації й механізмів персоналізації. Тому система була спроєктована з урахуванням поділу на сегменти, що дозволило в межах кожної групи окремо аналізувати продуктивність моделей без втрати узагальненості. Сегментування ґрунтувалося на категоріальному дереві, яке побудоване на основі частоти звернень до категорії, середньої вартості товарів у категорії, частки у загальному обороті та динаміки обертання позицій (turnover ratio). Це дозволило задати різні вагові коефіцієнти при побудові матриці вагів обчислення метрик для категорієцентричних систем, де однакова метрика Recall має різну інтерпретаційну цінність залежно від обсягу й структури категорії. Було враховано, що товари з високою маржинальністю, але з низькою частотою купівлі, можуть вимагати

більшої точності в персоналізації, тоді як масові товари — вищої швидкодії системи без втрати прийнятного рівня релевантності.

Загальна логіка побудови тестового середовища передбачала підключення до середовища збереження сесійних логів користувачів, а саме до кластеру, що зберігав події у форматі clickstream з позначкою часу, ID користувача, категорії товару, виду взаємодії (перегляд, додавання до кошика, купівля), а також інформації про A/B-експерименти, у які потрапляв кожен користувач. Це дозволило створити масив даних, у якому були присутні як імпліцитні сигнали (перегляди й додавання до бажаного), так і експліцитні дії (купівлі). Розділення даних на train/test відбувалося за принципом time-aware validation, де останні k-сесій кожного користувача виводилися в окремий шар, щоб моделювання максимально наближувалося до реальних сценаріїв. Такий підхід дозволив врахувати зміну патернів поведінки у часі, що є надзвичайно характерним для сегментів електронної торгівлі з коротким циклом життя товарів або сезонними коливаннями [39, с. 14].

Одним з основних завдань було також встановлення системи метрик, що відображає не лише точність прогнозування, але й економічну ефективність рекомендацій, швидкодію моделі, рівень персоналізованості та стабільність при зміні вхідних умов. Тому крім традиційних показників Precision@K і Recall@K, застосовувалися Coverage (покриття рекомендаційного простору), Diversity (різноманітність позицій у списках рекомендацій), Serendipity (міра несподіваної релевантності), а також метрики на основі конверсійних дій: CTR (click-through rate), CVR (conversion rate), AOV (average order value) у динаміці після рекомендації. Для визначення стабільності алгоритмів до змін у профілі користувача було змодельовано серію псевдовипадкових зсувів у поведінці, після чого тестувалася здатність алгоритмів до переадаптації — цей параметр був критично важливим для систем, які працюють у режимі near real-time.

Ще один набір метрик було сфокусовано на обчислювальних ресурсах, необхідних для генерації рекомендацій у масштабі. Тут оцінювалися latency (час генерації рекомендацій), throughput (кількість запитів на секунду), memory usage

(витрати оперативної пам’яті під час побудови та сервінгу моделі), а також cost-per-inference (вартість обчислення однієї рекомендації на одиницю користувача). Оскільки в реальних системах продуктивності не завжди пріоритетом є абсолютна точність, часто відбувається балансування між точністю й ефективністю використання ресурсів — тому було застосовано метаіндикатор Trade-off Score, який комбінує значення MAP, diversity та latency з ваговими коефіцієнтами, що встановлюються відповідно до бізнес-пріоритетів. Це дозволило об’єктивно порівнювати моделі, які можуть поступатися в точності, але компенсують це швидкодією або кращою масштабованістю [27, с. 6].

Під час побудови методики окремо було враховано специфіку рекомендацій у категоріях із нерівномірним розподілом інтересів — у таких випадках класичні метрики можуть демонструвати викривлення, спричинене ефектом домінування масових товарів. Для цього було запропоновано використовувати normalized popularity discount (NPD), який знижує вагу популярних товарів під час обчислення Precision і Recall, дозволяючи справедливо оцінити якість рекомендацій для довгого хвоста (long tail). Паралельно було проведено синтетичне моделювання cold-start сценаріїв, при яких система мала згенерувати релевантні рекомендації для нових користувачів або нових товарів. Для користувачів cold-start застосовувалися гібридні моделі з елементами content-based фільтрації, у яких на вхід надходили демографічні дані, тип пристрою, геолокація, початкова активність. Для нових товарів реалізовувалася побудова embedding-матриць на основі описових атрибутів (назва, категорія, бренди, технічні параметри), що інтегрувалися в нейронну модель на етапі cold deployment.

Архітектура тестування була реалізована у вигляді пайплайну, побудованого за принципом modular evaluation, де кожна модель проходила три етапи: offline evaluation, shadow deployment та online A/B-тестування. Під час offline evaluation алгоритм отримував історичні дані й формував топ-N список для кожного користувача, після чого обчислювалися всі класичні метрики точності й релевантності. Shadow deployment дозволяв перевірити продуктивність моделі в реальному середовищі без впливу на користувача — дані оброблялися,

рекомендації генерувалися, але не відображалися. Це дало змогу зафіксувати реальний час генерації, навантаження на інфраструктуру та витрати ресурсів. Завершальний етап — online A/B-тест — передбачав розбиття аудиторії на ізольовані когорти, де одна група отримувала рекомендації з базової моделі, інша — з тестової. Це дозволило провести порівняльний аналіз впливу на поведінкові метрики, такі як dwell time, глибина сесії, CTR і середній дохід на користувача (ARPU) [15, с. 8].

Для визначення чутливості моделей до контекстних параметрів були впроваджені сценарії з варіативними умовами: зміна типу пристрою (мобільний, десктоп, планшет), часові зони, періоди навантаження (пік, міжсезоння), а також реакція на зовнішні події — сезонні знижки, розпродажі, маркетингові кампанії. Аналіз поведінки моделей у таких умовах дозволив виявити ступінь їх адаптивності та стабільності до зовнішніх факторів. У ситуаціях, коли один і той самий користувач демонструє різну поведінку залежно від пристрою чи часу доби, особливу увагу приділяли ефективності алгоритмів контекстної персоналізації, які інтегрують додаткові вхідні ознаки у вектор користувача, змінюючи результат генерації рекомендацій залежно від умов сесії.

3.2. Розробка комплексної методики оцінювання

У межах формалізованого дослідження, орієнтованого на побудову комплексної методики оцінювання ефективності рекомендаційних систем у середовищі електронної комерції, був реалізований підхід, заснований на інтеграції багаторівневої структури показників, що охоплюють як класичні точнісні характеристики, так і поглиблені поведінкові вектори користувацької взаємодії з системою. У розробленій архітектурі методики ключовим стало створення багатоаргументного каркасу аналітики, що дозволяє зіставляти продуктивність алгоритму не в ізольованому середовищі, а в контексті динамічного UX-профілю. Було розмежовано рівні оцінювання на три площини: базову обчислювальну (точність прогнозу), когнітивно-поведінкову (реакція користувача на

рекомендаційний імпульс), та інтегративну (сприйняття й утримання у взаємодії). Така трирівнева модель дала змогу побудувати багатогранну шкалу аналітичного оцінювання, що охоплює як механізми математичної обґрунтованості рішень, так і фактори мікровзаємодії користувача з цифровим середовищем [38, с. 12].

Перший блок формувався на основі класичних метрик об'єктивної точності, серед яких застосовувалися Precision@K, Recall@K, F1@K, NDCG (normalized discounted cumulative gain) — кожна з яких дозволяла оцінити якість видачі рекомендаційної системи залежно від позиції елемента у списку та релевантності на основі історичної купівельної поведінки. Для покращення диференціації результатів метрики були нормалізовані з урахуванням довжини сесій, середньої кількості взаємодій у категоріях та типу платформи (desktop / mobile / app). Усі ці метрики обчислювалися не лише в межах когорти користувачів, а й у розрізі товарних категорій, враховуючи сезонну мінливість, глибину товарного дерева, SKU-динаміку та асортиментну ротацію. В результаті точнісні метрики були представлені як матриця багатовимірних параметрів із вагою, що розподілялася відповідно до бізнес-пріоритетів: висока чутливість до топових SKU, контроль якості рекомендацій у категоріях із низьким рівнем повторних покупок, мінімізація false positives у персоналізованих вітринах.

На другому рівні був побудований когнітивно-поведінковий сегмент оцінювання, що передбачав фіксацію таких показників, як session duration (тривалість взаємодії після показу рекомендації), interaction depth (кількість кліків углиб платформи), bounce-back rate (повернення до стартового стану без здійснення транзакції), sequence diversity (різнорідність взаємодій у межах рекомендаційного сеансу), micro-conversion (збереження в обране, перегляд варіацій товару, запуск порівняльних функцій). Дані параметри були інтегровані до багатошарового трекінгу поведінки, побудованого на основі розмітки подій у середовищі Google Tag Manager / Firebase, з наступною агрегацією в спеціально створеному behavioral funnel evaluator. Для уникнення контекстного шуму в показниках були застосовані фільтри сегментування за типом трафіку (органічний, платний, e-mail, push), каналами залучення, часом доби, типом пристрою та маркетинговими кампаніями.

Це дозволило виділити «чистий» вплив рекомендаційного алгоритму без супутнього впливу зовнішніх факторів [21, с. 84].

Окремо була реалізована інтеграція багатовимірного аналізу впливу на UX-поведінку, що передбачала побудову теплових карт кліків і зон уваги (heatmaps), запис сесій (session replay), а також структурний аналіз рухів миші та скролінгу (mouseflow dynamic pathing). Усі зібрані дані конвертувалися в поведінкові профілі користувачів, для яких будувалися індивідуальні реакційні моделі на рекомендаційний стимул: інерційна (реакція без дії), рефлекторна (миттєвий клік), когнітивна (глибока взаємодія з відтермінованим рішенням). Таке структурування дозволило зіставити вид рекомендації з типом реакції користувача та сформулювати матрицю ефективності в розрізі когнітивних траєкторій. У межах цієї частини було впроваджено показники recommendation dwell coefficient, engagement delta, UX pivot rate, що відображають не лише факт кліку, а й характер подальшої поведінки внаслідок рекомендації.

Третій рівень методики охоплював інтегративну площину, що поєднувала якісні параметри з кількісними через застосування метаіндексів на кшталт GCI (Global Conversion Index), LTI (Long-Term Interaction), PIR (Personalization Impact Ratio) та IVX (Interaction Value Expansion). GCI вимірював узагальнене співвідношення між кількістю транзакцій, здійснених після перегляду рекомендацій, і загальним обсягом переглядів у категоріях із рекомендованими товарами. LTI відображав накопичувальний ефект рекомендацій — скільки разів користувач взаємодіяв з рекомендованим товаром до моменту купівлі. PIR дозволяв оцінити внесок системи персоналізації у збільшення конверсій на користувача. IVX інтегрував поведінкові дії з мікрорівня в макроповедінковий шаблон, який зіставлявся з кластером користувачів із подібними характеристиками.

Усі описані індикатори були включені до єдиної методологічної матриці оцінювання, що формувалася як багаторівнева модель з можливістю гнучкого масштабування залежно від типу моделі (content-based, collaborative, hybrid, deep learning-based). Була передбачена можливість автоматичної переваги одних метрик над іншими в залежності від цільового KPI. У платформі оцінювання було

закладено механізм вагової конфігурації: для категорій з високою товарною маржинальністю вагу отримували економічні метрики; для нових користувачів — поведінкові; для швидкообертових товарів — точнісні. Це дало змогу уникнути однобічності в оцінюванні та забезпечити узгодженість між рекомендаційною логікою та бізнес-цілями [34, с. 11].

Було проведено додаткову процедуру кореляційного аналізу між точнісними метриками та поведінковими сигналами, що дозволило побудувати граф спряжень між рівнями аналітики. Застосовувався метод Spearman correlation coefficient та mutual information gain для виявлення залежностей між, скажімо, високим Precision@10 і низьким engagement rate, що вказувало на обмеження рекомендаційної системи в генерації цікавого контенту попри високий збіг із історією. Також було встановлено взаємозв'язок між diversity-метрикою та session depth, що свідчило про користь різнорідності у видачі для продовження сеансу. Усі ці кореляційні пари лягли в основу оптимізаційної функції адаптивного вагового зміщення в метаіндикаторі продуктивності системи.

Під час побудови методики велика увага приділялася UX-аналітиці як елементу, що не підлягає формалізації через виключно обчислювальні метрики. Було проведено квантифікацію індексів задоволеності користувача на основі систем зворотного зв'язку (implicit feedback), де респондентами були користувачі, які отримували рекомендації та проходили короткі інтерактивні опитування (інтегровані безпосередньо в інтерфейс у формі one-click NPS і semantic differential scale). На основі зібраних відповідей формувалася метрика UX-QoE (User Experience Quality of Experience), яка зіштовхувалася з обчислюваними характеристиками рекомендацій — це дозволило виявити ділянки розриву між математично «ефективними» рішеннями та тим, як їх сприймає людина в реальному середовищі взаємодії. У такий спосіб була побудована рефлексивна модель зворотного UX-аналізу, що дозволяла калібрувати систему на основі прямих реакцій користувачів, а не лише технічної продуктивності [31, с. 23].

У межах реалізації структурованої моделі оцінювання ефективності інтелектуальних рекомендаційних систем для e-commerce-середовищ було

здійснено побудову процедури зважування критеріїв із наступною нормалізацією значень показників у багатовимірному просторі даних. Відправною точкою виступило визнання неоднорідності типів електронної комерції, кожен із яких характеризується специфічними характеристиками користувацької поведінки, структурою асортименту, середнім життєвим циклом товарів, глибиною залучення клієнта у процес взаємодії, а також різними формами комунікації між системою рекомендацій і кінцевим користувачем. У такому середовищі недоцільним виявилось застосування уніфікованого вагового профілю для метрик — було зафіксовано, що значущість певного індикатора у категоріях масового споживання відчутно відрізняється від ваги того самого індикатора у нішевих сегментах, де інерційність поведінки та щільність персоналізованої взаємодії в рази вища. Саме тому основа формалізації вагових коефіцієнтів була побудована на сегментованому принципі: у кожному підмноженні e-commerce-сервісів окремо визначалася матриця ваг з урахуванням ієрархії метрик, типу платформи, маркетингової стратегії та когнітивного навантаження на користувача.

Для досягнення високого рівня формальної точності було використано метод аналітичної ієрархічної обробки (AHP – Analytic Hierarchy Process), що дозволив структурувати простір критеріїв у вигляді дерева залежностей. На першому рівні структури були закріплені інтегральні цілі (максимізація задоволеності користувача, зростання конверсій, зниження відтоку), на другому — групи метрик (точність, релевантність, адаптивність, когнітивна привабливість, швидкодія, ресурсна економічність), а на третьому — конкретні індикатори (Precision@K, engagement ratio, diversity, latency тощо). Для кожної пари індикаторів було здійснено попарне порівняння за шкалою Сааті з наступною побудовою матриці вагових коефіцієнтів, що нормалізується до вектора пріоритетів. У процесі зважування брали участь представники чотирьох ключових профілів: технічні аналітики, UX-дизайнери, бізнес-аналітики та продуктові менеджери — їхні експертні оцінки дозволили отримати узгоджений індекс ваг для кожного з показників. Це дозволило перенести суб’єктивну значущість показника у формальну модель обчислення інтегрального балу системи [24, с. 13].

Наступним кроком було впровадження процедури нормалізації даних, що обов’язково передбачала уніфікацію шкал усіх показників. Враховуючи, що різні індикатори мають різну природу вимірювання (деякі — у відсотках, інші — в абсолютних значеннях, деякі — позитивно, інші — негативно корельовані з якістю), застосовано комбіновану нормалізацію з використанням Z-score, Min-Max scaling та log-transformation. Для індикаторів із нелінійним впливом (наприклад, latency або click-delay), де зростання понад певну межу має катастрофічний ефект, застосовувалася логарифмічна шкала з асимптотичним гальмуванням. Також у процесі нормалізації були враховані дисперсійні характеристики — для показників зі значною варіативністю додатково використовувалася процедура віконного згладжування (rolling mean smoothing) для уникнення псевдосплесків, що могли впливати на зважування.

Паралельно було розроблено механізм категоріального нормування — враховуючи гетерогенність структур у різних e-commerce-сервісах, особливо в B2C, C2C, B2B-середовищах, створено окремі блоки нормалізаторів, які враховували характерну поведінкову модель для кожного типу бізнесу. У B2C-сервісах (маркетплейси, онлайн-ритейл) пріоритет віддавався точнісним метрикам і швидкодії; у B2B — акцент зміщувався на параметри стабільності, репрезентативності, точності кластеризації користувачів; у C2C — на diversity, novelty, contextual sensitivity. Таке багаторівневе підходження забезпечило об’єктивність у зіставленні систем, що працюють у принципово різному середовищі, без зміщення загальної шкали оцінювання.

Було реалізовано також адаптивну систему вагових зміщень (dynamic weight adaptation), яка дозволяла змінювати коефіцієнти залежно від операційного стану системи або маркетингової кампанії. Якщо система працювала у фазі промо-активності (наприклад, під час Black Friday або сезонних знижок), ваги зміщувалися в бік показників швидкодії, масштабованості та здатності до генерації неочікуваних товарних комбінацій; у фазі стабільного функціонування — повернення до стандартного профілю, з акцентом на точність та персоналізацію. Ця динамічна адаптація реалізовувалась через зв’язку з системою бізнес-правил, яка

передавала параметри стану кампанії до модуля нормалізації, після чого ініціювався перерахунок вагового профілю в реальному часі [29, с. 9].

Окремо слід зазначити побудову вагових моделей у мультикатегоріальних системах, де кожна категорія товарів має відмінні бізнес-пріоритети й поведінкові патерни. Так, для категорії «Електроніка» пріоритет отримували точнісні показники та глибина перегляду (interaction depth), тоді як у «Одежі» — креативність рекомендацій (serendipity), варіативність (diversity) та швидкість реакції користувача (reaction lag). Це потребувало побудови контекстно-залежної вагової матриці, що розраховувалася на основі історичних даних з конкретної категорії й коригувалася залежно від щомісячної динаміки обігу. Для цього застосовувалась кластеризація категорій за ознаками швидкості обертання, середньої ціни, сезонності, глибини товарного дерева й відсотка повторних купівель. Такі кластери отримували власні конфігурації ваг, які масштабувалися через фактор категорійної близькості — це дозволяло успадковувати вагові профілі між схожими категоріями без втрати індивідуальності.

3.3. Експериментальне тестування методики на реальних даних

У процесі експериментального тестування розробленої методики оцінювання ефективності інтелектуальних рекомендаційних систем для e-commerce середовищ була реалізована багатофазна структура апробації на основі реального датасету, отриманого з функціонуючої платформи з великою інтенсивністю трафіку, асортиментною глибиною понад 80 000 SKU та багатокомпонентною структурою сегментів споживачів. Було прийнято рішення орієнтувати апробацію не лише на статистичні метрики оцінювання ефективності, а й на динаміку змін у поведінкових патернах, що дозволило побудувати повноцінну експериментальну матрицю взаємодії користувача з рекомендаційним середовищем у контексті контентного насичення, сценарної гнучкості й персоналізованої релевантності. Для цього сформовано набір сценаріїв, що охоплював ініціалізацію холодного запуску, тестування повторних взаємодій, перевірку реакцій на рекомендовані позиції у

вікнах категорій, а також симуляцію поведінки після оновлення вмісту профілю користувача. Експеримент проводився у декілька хвиль, з включенням груп A/B/C, кожна з яких мала різні умови взаємодії з інтерфейсом: одна група отримувала лише базові топ-N рекомендації, інша — персоналізовані з урахуванням категорійної активності, а третя — рекомендації, підсилені текстовими поясненнями через генеративний модуль на основі інтеграції з OpenAI API [20, с. 66].

```python
import networkx as nx

class CausalExplanation:
    def __init__(self):
        self.graph = nx.DiGraph()

    def add_causal_link(self, cause, effect):
        self.graph.add_edge(cause, effect)

    def generate_explanation(self, cause, effect):
        if nx.has_path(self.graph, cause, effect):
            paths = list(nx.all_simple_paths(self.graph, cause, effect))
            explanation = f"Зміна в {cause} може призвести до зміни в {effect} через такі
                шляхи:\n"
            for path in paths:
                explanation += " -> ".join(path) + "\n"
            return explanation
        else:
            return f"Між {cause} та {effect} немає прямого зв'язку."
```

Рис. 3.1 Реалізація модулю орієнтовного графа

На етапі збору даних і розгортання пояснювальної компоненти було реалізовано модуль каузального аналізу, побудований із використанням бібліотеки NetworkX. У середовищі Python створено орієнтований граф (рис. 3.1), де кожна подія (додавання товару до кошика, перегляд рекомендації, кліки, перегляди категорій, запуск фільтрів, переходи в картку товару) була представлена як окремий вузол, а спрямовані ребра між вузлами ілюстрували причинно-наслідкові ланцюги. Було побудовано понад 2 000 інстансів взаємозв’язків у графовій структурі, що дозволило виявити найбільш типові траєкторії поведінки користувачів у відповідь на окремі імпульси системи рекомендацій. Граф обчислював наявність усіх можливих маршрутів від однієї події до іншої, формуючи на їх основі текстові

пояснення, які потім передавались у модуль генерації коментарів (рис.3.2). Таким чином, кожна рекомендація не лише підкріплювалася внутрішнім евристичним балом, а й отримувала логічне обґрунтування, представлене у формі пояснювального тексту, що враховував декілька рівнів впливу. Цей підхід дозволив знизити рівень когнітивного тертя з боку користувача і підвищити довіру до системи рекомендацій, особливо в сценаріях з новими або рідко використовуваними товарами [26, с. 10].

```
19
20      def visualize_graph(self):
21          import matplotlib.pyplot as plt
22          pos = nx.spring_layout(self.graph)
23          nx.draw(self.graph, pos, with_labels=True, node_size=2000, node_color='lightblue',
                font_size=12)
24          plt.show()
25
26  causal_system = CausalExplanation()
27  causal_system.add_causal_link("A", "B")
28  causal_system.add_causal_link("B", "C")
29  causal_system.add_causal_link("A", "C")
30
31  explanation = causal_system.generate_explanation("A", "C")
32  print(explanation)
33
```

Рис. 3.2 Реалізація каузального методу

Окремим технічним модулем було запроваджено підсистему агрегування вагових коефіцієнтів у системі прийняття рішень, яка забезпечувала кількісну інтерпретацію значущості кожного з критеріїв — ціна, якість, популярність — через механізм багатокритеріальної згортки. Відповідно до тестової логіки експерименту, вага критерію «якість» була встановлена на рівні 0.5, що відображало її домінуюче значення при порівнянні товарних позицій, тоді як «ціна» та «популярність» отримали ваги 0.3 і 0.2 відповідно. Кожна товарна одиниця оцінювалася за допомогою підсумовування добутків оцінок на відповідні ваги, після чого формувалася загальна рейтингова оцінка, яка слугувала однією з основ для включення в персоналізований список. Блок нормалізації ваг був активним у

режимі реального часу — під час зміни бізнес-контексту, наприклад, при зниженні купівельної спроможності користувачів, вага критерію «ціна» підвищувалась, а «популярність» знижувалась, зберігаючи загальну суму ваг на рівні одиниці (рис.3.3). Було протестовано також режим динамічного перерахунку ваг на основі змін у поведінці когорт, що забезпечувалось внутрішнім моніторинговим механізмом у фреймворку Flink [22, с. 63].

```python
import numpy as np

class WeightingSystem:
    def __init__(self, criteria_weights):
        self.criteria_weights = criteria_weights

    def evaluate_element(self, element_scores):
        total_score = 0
        for criterion, weight in self.criteria_weights.items():
            total_score += element_scores.get(criterion, 0) * weight
        return total_score

    def update_weights(self, new_weights):
        self.criteria_weights.update(new_weights)

    def normalize_weights(self):
        total_weight = sum(self.criteria_weights.values())
        for key in self.criteria_weights:
            self.criteria_weights[key] /= total_weight

criteria_weights = {"price": 0.3, "quality": 0.5, "popularity": 0.2}
element_scores = {"price": 8, "quality": 7, "popularity": 9}

system = WeightingSystem(criteria_weights)
score = system.evaluate_element(element_scores)
print("Оцінка елемента:", score)

new_weights = {"price": 0.4, "quality": 0.4, "popularity": 0.2}
system.update_weights(new_weights)
system.normalize_weights()

score_updated = system.evaluate_element(element_scores)
print("Оновлена оцінка елемента:", score_updated)
```

Рис. 3.3 Інтеграція LLM від OpenAI

Візуалізація графів причинності дозволила швидко локалізувати структурні вузли з високим ступенем впливу (in-degree > 7) та сформувати набір вузлів-

акселераторів, через які проходило більшість трас поведінкової взаємодії. У цих точках було розміщено інформативні текстові пояснення, згенеровані мовною моделлю OpenAI, що дозволило перетворити складну мережу дій у зрозумілу логіку. Генерація пояснень здійснювалася шляхом виклику OpenAI API для кожної траєкторії з каузального графа, з урахуванням контексту категорії, поведінкової історії, останніх дій та агрегованих рейтингів. У режимі тестування були створені дві гілки: одна — з класичним рекомендаційним інтерфейсом без пояснень, інша — з інтегрованими поясненнями в реальному часі. Для обох груп велися окремі журнали дій, де фіксувалися click-through rate, dwell time, scroll depth, microconversion events та кількість запитів на допомогу. Емпіричні дані засвідчили високий ступінь залучення у гілці з поясненнями, що вказувало на ефективність концепції explainability як частини системи рекомендацій [28, с. 81].

Усі дані, зібрані під час експерименту, зберігалися в розподіленому середовищі обробки подій, побудованому на Apache Kafka, з подальшим використанням схеми парсингу в Apache Flink. Це дозволило транслювати всі події у форматі time-series logs із фіксацією мітки часу, сесійного ID, типу взаємодії та зв’язку з рекомендаційною подією. Кожна одиниця даних передавалася у модуль інтегрального оцінювання, де обчислювався сукупний бал ефективності рекомендації за шкалою, побудованою з урахуванням раніше розробленої системи вагових коефіцієнтів, нормалізованих показників і текстових пояснень. Уніфіковане представлення оцінки дозволило створити зрозумілу метаінформацію, яка могла відображатись безпосередньо в інтерфейсі користувача або використовуватись внутрішньою аналітикою для калібрування системи.

Під час формування експериментального плану були враховані не лише сценарії звичайної купівельної поведінки, а й edge-сценарії, які симулювали нестандартні ситуації: високу частоту взаємодій у стислий проміжок часу, раптову зміну уподобань (швидке перемикання між категоріями), а також випадки з багатокористувацькою поведінкою на одному обліковому записі. Такі ситуації були потрібні для тестування стійкості алгоритму адаптації профілю користувача та стабільності системи пояснень у зміненому середовищі. Було зафіксовано типові

графові шляхи для кожного з edge-сценаріїв, на основі яких система автоматично формувала пояснення (рис. 3.4), зокрема: «Ваші останні перегляди в категорії побутової техніки активували інтерес до товарів зі схожими характеристиками в категорії кухонного приладдя» — пояснення, що будується на мультикатегоріальній кореляції в поведінці [37, с. 16].

```python
import openai

openai.api_key = "fmgixzcb-44jks-mlasz"

def query_llm_explanation(recommendation, context):
    prompt = f"На основі наступного контексту надай пояснення до цієї рекомендації: {recommendation}\nКонтекст: {context}"
    response = openai.Completion.create(
        engine="text-davinci-003",
        prompt=prompt,
        max_tokens=200
    )
    return response.choices[0].text.strip()

def query_batch_llm_explanations(recommendations, context):
    explanations = []
    for recommendation in recommendations:
        explanation = query_llm_explanation(recommendation, context)
        explanations.append(explanation)
    return explanations

recommendation = "Рекомендуємо купити товар X через його високу якість."
context = "Товар X має високий рейтинг за відгуками, а також визнаний експертами у цій категорії."

explanation = query_llm_explanation(recommendation, context)
print("Пояснення до рекомендації:", explanation)

batch_recommendations = [
    "Рекомендуємо вибрати послугу Y, оскільки вона найвигідніша.",
    "Спробуйте товар Z, він має найкращу ефективність за відгуками."
]
batch_explanations = query_batch_llm_explanations(batch_recommendations, context)
for idx, exp in enumerate(batch_explanations):
    print(f"Пояснення до запиту {idx + 1}: {exp}")
```

Рис. 3.4 Генерація пояснення для рекомендації

Для оцінювання якості інтеграції мовної моделі в генерацію пояснень були проведені сесії UX-тестування із залученням цільових груп користувачів. Учасники

оцінювали кожне пояснення за шкалою зрозумілості, корисності, релевантності, емоційної виразності та довіри. Отримані дані були проаналізовані з використанням факторного аналізу, який підтвердив, що пояснення, побудовані за принципом каузального шляху з чітко вказаними залежностями, мають вищий ступінь прийнятності порівняно з шаблонними текстами. Було виявлено, що пояснення, які містять явний зв’язок між діями користувача та запропонованим товаром, викликають довіру навіть у випадках, коли товар не потрапляє в початкове поле інтересу користувача. Це надало підставу для подальшого розширення explainable-модуля в архітектурі рекомендаційної системи.

3.4. Аналіз результатів та порівняння з існуючими підходами

У межах практичного етапу аналітичної валідації розробленої методики було реалізовано комплексне порівняння її функціональних характеристик із класичними підходами, що використовуються для оцінювання ефективності інтелектуальних рекомендаційних систем у середовищі e-commerce. В основі аналізу лежала апробація нової методології, заснованої на можливісній логіці пояснень і причинно-наслідковому моделюванні з графовою репрезентацією взаємозв’язків між характеристиками користувача, об’єктами рекомендації та контекстом взаємодії. Це дозволило не лише перевірити відповідність системи теоретичним критеріям якості, а й проаналізувати її поведінку в режимі динамічної зміни параметрів з боку користувача. Здійснено симетричне тестування: з одного боку — на основі нової методики, з іншого — із використанням традиційного підходу на базі колаборативної фільтрації без текстового супроводу та без обліку ступеня інформаційної невизначеності. Така дихотомія дала змогу об'єктивно ідентифікувати ключові точки відмінностей у роботі двох моделей, зафіксувавши всі відхилення в точності, повноті, релевантності, когнітивній зрозумілості, стабільності до браку даних та здатності адаптуватись до нетипових сценаріїв поведінки [35, с. 8].

Схема створеної бази даних, яка використовувалася у порівнянні, наведена на рис. 3.5. Результати порівняння вказали на наявність системної розбіжності у чутливості моделей до рівня заповненості профілю користувача. У традиційній архітектурі системи рекомендацій точність падала в середньому на 37% у випадках з профілями, що містили менш як три сеанси взаємодії. Це пов'язано з відсутністю структурованих залежностей у поведінці користувача, що не дозволяє класичним підходам виконувати достатньо впевнене прогнозування. Водночас розроблена модель, яка використовує можливісну логіку, продемонструвала лише 14% втрати точності за аналогічних умов, оскільки здатна працювати з неповними або розмитими даними завдяки інкорпорації нечітких правил оцінювання релевантності [16, с. 15]

```
CREATE TABLE Users (
    UserID INT PRIMARY KEY AUTO_INCREMENT,
    UserName VARCHAR(100) NOT NULL,
    Email VARCHAR(100) UNIQUE NOT NULL,
    RegistrationDate DATE DEFAULT CURRENT_DATE
);

CREATE TABLE Items (
    ItemID INT PRIMARY KEY AUTO_INCREMENT,
    ItemName VARCHAR(100) NOT NULL,
    Category VARCHAR(50),
    Description TEXT
);

CREATE TABLE Explanations (
    ExplanationID INT PRIMARY KEY AUTO_INCREMENT,
    ExplanationText TEXT NOT NULL,
    ModelType VARCHAR(50) NOT NULL,
    CreatedAt TIMESTAMP DEFAULT CURRENT_TIMESTAMP
);

CREATE TABLE Recommendations (
    RecommendationID INT PRIMARY KEY AUTO_INCREMENT,
    UserID INT NOT NULL,
    ItemID INT NOT NULL,
    ExplanationID INT,
    Rating DECIMAL(2,1),
    RecommendationDate TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
    FOREIGN KEY (UserID) REFERENCES Users(UserID) ON DELETE CASCADE,
    FOREIGN KEY (ItemID) REFERENCES Items(ItemID) ON DELETE CASCADE,
    FOREIGN KEY (ExplanationID) REFERENCES Explanations(ExplanationID) ON DELETE SET NULL
);

CREATE TABLE Feedback (
    FeedbackID INT PRIMARY KEY AUTO_INCREMENT,
    RecommendationID INT NOT NULL,
    UserFeedback TEXT,
    FeedbackDate TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
    FOREIGN KEY (RecommendationID) REFERENCES Recommendations(RecommendationID) ON DELETE CASCA
);
```

Рис. 3.5 Створення бази даних PC

Модель виявила стійкість до варіацій у поведінці користувача — при зміні патернів у межах одного профілю (зміна жанрових уподобань, перехід між типами контенту) система, базована на каузальному аналізі, змогла відслідкувати причини таких переходів через аналіз графових ланцюгів і перебудувати структуру рекомендацій, не втрачаючи логічної послідовності. Класична модель при цьому потребувала повного перерахунку матриці подібності й демонструвала хаотичну видачу, нерідко не пов’язану з історією профілю.

Під час тестування у сфері книжкових рекомендацій було зібрано понад 1 500 унікальних взаємодій користувачів із системою. Кожна взаємодія включала щонайменше один клік, оцінку, вибір з-поміж рекомендованого переліку та запит на пояснення. Для ілюстрації було обрано профілі користувачів з інтересами у сфері психології, соціології, наукової літератури та класичних творів. На підставі сукупного аналізу відгуків і поведінкових дій було виявлено чітку закономірність: користувачі, які отримували пояснення у формі згенерованих текстів на базі каузальної моделі, значно частіше залишали позитивний зворотний зв’язок (72%) порівняно з користувачами у групі класичного алгоритму (41%). Ці пояснення містили причинно-наслідкові ланцюги, що пояснювали логіку: «Вам рекомендовано цю книгу, оскільки попередньо ви зацікавились виданнями про нейропсихологію, а ця книга тематично перекриває обраний напрям». Таким чином, користувач не лише отримував релевантну позицію, а й чітке обґрунтування її вибору, що суттєво зменшувало когнітивне навантаження і знижувало ймовірність відмови [24, с. 81].

Було зафіксовано також важливу відмінність у роботі моделей у категоріях із високою інформаційною насиченістю (більше 10 атрибутів на кожну одиницю контенту). У таких умовах класичні підходи демонстрували перенасиченість або втрату чутливості — рекомендації перетворювались на механічне повторення часто перегляданих позицій, тоді як розроблена модель вибудовувала семантичні траєкторії між атрибутами й зміщувала акцент рекомендації в зону новизни. Так, у випадку з користувачем, що переглядав книги про біологію свідомості, модель пропонувала твори про вплив мови на нейропластичність, формуючи зв’язок між

уже проявленим інтересом і потенційною новою темою. Це дозволяло розширити коло взаємодії без втрати релевантності, у той час як класична модель обмежувалась лише прямими аналогами з уже переглянутих.

Рис. 3.6 Схема БД рекомендаційної системи

На рівні архітектурного аналізу була виявлена додаткова перевага у структурі бази даних (рис. 3.6), розробленої для нової методики: вона зберігає не лише самі рекомендації, а й повний контекст, у якому вони були сформовані. Це забезпечує можливість подальшого аудиту логіки роботи системи, формування звітності та навчання майбутніх моделей на основі історичних трас. Таблиці «Explanations»,

«Recommendations» і «Feedback» формують замкнуту петлю аналітичного моніторингу, де кожна рекомендація отримує реакцію, кожна реакція — пояснення, кожне пояснення — прив’язку до моделі. Приклади таких пояснень подано в Таблиці 3.1. Такий підхід дозволяє не лише забезпечити прозорість, а й вбудувати рекомендаційну логіку у стратегічне планування взаємодії з клієнтом, особливо в довгострокових сценаріях повторних конверсій [36, с. 6].

Таблиця 3.1

Сгенеровані пояснення до рекомендацій

Рекомендація	Пояснення	Можливість	Необхідність
"Думай, як монах"	Рекомендовано вам тому, що стосується тем саморозвитку, які вас цікавлять.	0.32	0.72
"Атлант розправив плечі"	Ця книга рекомендована вам тому, що вона є класичним твором з філософським змістом, що відповідає вашим вподобанням.	0.45	0.91
"Коротка історія часу"	Ви переглядали популярні наукові роботи схожі на дану.	0.28	0.87
"Підсвідомість може все"	Дана робота до вподоби фанатам психології, схожим на вас.	0.41	0.95
"1984"	Рекомендовано вам, оскільки цей класичний твір відображає ваші вподобання до важливих літературних текстів.	0.18	0.76

Продовження таблиці 3.1

Рекомендація	Пояснення	Можливість	Необхідність
"Сила звички"	Книга обрана для вас, оскільки вона актуальна в контексті теми поведінкових змін, що вас цікавить.	0.40	0.84
"Феномен моди"	Рекомендовано вам через її відповідність до вашої зацікавленості в розборі соціальних явищ.	0.26	0.67
"Злочин і кара"	Цей твір підійде вам завдяки поєднанню класичного сюжету та психологічних аспектів, що співпадають з вашими інтересами.	0.34	0.80

Були виявлені і слабкі місця запропонованої моделі, що стали видимими саме у фазі експериментальної апробації. У профілях із дуже низьким рівнем заповненості (менше 2 переглядів), навіть незважаючи на використання можливісної логіки, система іноді генерувала занадто узагальнені пояснення. Це вказує на потребу впровадження додаткового шару контекстної семантики, можливо через підключення більш глибоких LLM-моделей, здатних здійснювати індуктивне узагальнення на основі оточення, а не лише історії користувача. Також було зафіксовано певне перевантаження текстів пояснень у випадках, коли каузальний граф містив надмірну кількість вузлів — у таких ситуаціях пояснення могли виглядати надто технічними, що знижувало їхню сприйнятливість для користувача. Це створює вимогу до майбутнього етапу оптимізації структури пояснень — застосування принципів когнітивної економії для формування коротких, але змістовних обґрунтувань.

У частині автоматичного контролю якості рекомендацій запропонована методика також виявилась більш гнучкою — завдяки модульній структурі можна

було змінювати вагу впливу окремих критеріїв у реальному часі. На відміну від класичних систем, де навіть часткове переналаштування моделі вимагало перекомпіляції всього середовища, нова методика дозволяла оперативно змінювати пріоритетність між «ціною», «якістю», «релевантністю» та іншими атрибутами, зберігаючи цілісність логіки. Це виявилось особливо зручним при переході між типами користувачів: умовно «цінові гравці» отримували зміщену рекомендаційну матрицю на користь ціни, а користувачі зі схильністю до преміального контенту — на користь емоційної ваги та естетики [18, с. 7].

Висновок до розділу

Підсумовуючи весь спектр отриманих даних, можна зафіксувати, що запропонована модель суттєво відрізняється від класичних підходів у чотирьох ключових напрямах: здатність працювати з неповними даними, інтеграція пояснень у вигляді логічно зв’язаних текстів, модульність зміни критеріїв у реальному часі та стабільність під час поведінкових зсувів користувача. Це формує фундамент для розвитку не лише точнісних, а й етично-прозорих, інтерактивних, адаптивних рекомендаційних систем нового покоління, що враховують не лише результат, але й саму логіку досягнення результату як частину ціннісної пропозиції для користувача. Така парадигма зміщує акцент від автоматизації до кооперації системи з людиною — і саме в цьому полягає трансформаційна сила сучасних інтелектуальних рішень у середовищі електронної комерції.

ВИСНОВКИ

У межах даної кваліфікаційної роботи було здійснено комплексне дослідження методів та підходів до оцінювання ефективності інтелектуальних рекомендаційних систем у сфері електронної комерції. На основі вивчення сучасних алгоритмів було проаналізовано їх можливості та обмеження у реальному бізнес-середовищі.

Особлива увага приділялась аналізу метрик оцінювання якості рекомендацій – таких як точність, новизна, серендипність та інші. У результаті чого, було зроблено висновок, що лише комплексний підхід до оцінки дозволяє отримати повноцінне уявлення про користувацьку й бізнес-цінність рекомендаційної системи.

У практичній частині дослідження продемонстровано, що вибір конкретних метрик повинен відповідати стратегічним цілям компанії, а не обмежуватись лише технічною точністю.

Результатом роботи стала розробка методики, що дозволяє поєднувати технічні показники з бізнес-метриками. Це створює передумови для більш ефективного впровадження рекомендаційних систем в e-commerce середовищі, де персоналізований підхід до користувача має вирішальне значення.

Підсумовуючи, можна стверджувати, що рекомендаційні системи виконують значно більше, ніж просто добір контенту — вони стали важливою ланкою у спілкуванні компанії з клієнтом. Саме завдяки правильній оцінці таких систем, підприємства можуть краще розуміти потреби користувачів і створювати персоналізований досвід, що справді виправдовує очікування людей.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Aggarwal C. C. Recommender Systems. Cham : Springer International Publishing, 2016. URL: https://doi.org/10.1007/978-3-319-29659-3
2. Recommender Systems Handbook / ed. by F. Ricci et al. Boston, MA : Springer US, 2011. URL: https://doi.org/10.1007/978-0-387-85820-3
3. Schafer J. B., Konstan J., Riedl J. Recommender systems in e-commerce // Proceedings of the 1st ACM Conference on Electronic Commerce (EC '99). – New York: ACM, 1999. – P. 158–166. – DOI: https://doi.org/10.1145/336992.337030.
4. Пономаренко, І., & Битик, О. (2021). Використання рекомендаційних систем для оптимізації маркетингової стратегії компанії. Підприємництво та інновації, (19), 34-39. https://doi.org/10.37320/2415-3583/19.5
5. Kumar, V., Aksoy, L., Donkers, B., Venkatesan, R., Wiesel, T., & Tillmanns, S. (2010). Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value. Journal of Service Research, 13(3), 297-310. https://doi.org/10.1177/1094670510375602
6. KUMARI, SONIA. "Recommender System: Revolution in E-Commerce." International Journal Of Engineering And Computer Science, 2016.
7. Evidently AI. 10 metrics to evaluate recommender and ranking systems. URL: https://www.evidentlyai.com/ranking-metrics/evaluating-recommender-systems
8. Alaa Alslaity, Thomas Tran, "ComPer: A Comprehensive Performance Evaluation Method for Recommender Systems", International Journal of Information Technology and Computer Science(IJITCS), Vol.11, No.12, pp.1-18, 2019. DOI: 10.5815/ijitcs.2019.12.01
9. Recommender Systems: Machine Learning Metrics and Business Metrics [Електронний ресурс] // Neptune.ai. – 2023. – Режим доступу: https://neptune.ai/blog/recommender-systems-metrics

10. Product Recommendations: Types, Examples & Best Practices [Електронний ресурс] / Claire McConville // Fact-Finder Blog. – 2023. – Режим доступу: https://www.fact-finder.com/blog/product-recommendation/
11. Evaluating an Amazon Personalize domain recommender [Електронний ресурс] // Amazon Personalize Documentation. – Режим доступу: https://docs.aws.amazon.com/personalize/latest/dg/evaluating-recommenders.html
12. He X., Liu Q., Jung S. The Impact of Recommendation System on User Satisfaction: A Moderated Mediation Approach [Електронний ресурс] // Journal of Theoretical and Applied Electronic Commerce Research. – 2024. – Vol. 19, No. 1. – P. 448–466. – Режим доступу: https://doi.org/10.3390/jtaer19010024
13. Huzhang G., Pang Z.-J., Gao Y., Liu Y., Shen W., Zhou W.-J., Da Q., Zeng A.-X., Yu H., Yu Y., Zhou Z.-H. AliExpress Learning-To-Rank: Maximizing Online Model Performance without Going Online // Journal of LaTeX Class Files. – 2015. – Vol. 14, No. 8. – P. 1–10.
14. Бальдан Лозано Ф. Дж. Гарсія-Гіль Д. Бенчмаркінг методів виявлення аномалій. Expert Systems. 2024. С. 11.
15. Обробка даних реальних користувачів для створення рекомендаційної системи. Мій досвід виправлення помилок. URL: https://dou.ua/forums/topic/39070/ (дата звернення: 28.04.2025).
16. Фернандес А. Фернандес Д. Гарсія Й. Управління бізнес-процесами для оптимізації клінічних процесів. Health Informatics Journal. 2020. 26(2). С. 1305–1320.
17. Чалий С. Ф. Бганцев М. Розробка методів оцінювання IT-компаній. 8th International Scientific and Practical Conference. 2024.
18. Чалий С. Ф. Єрохін Д. Оцінка пояснень користувачами. Застосування інформаційних технологій у підготовці та діяльності сил охорони. 2024.
19. Чалий С. Ф. Лещинська І. Принципи побудови ментальних моделей рішення. АСУ та прилади автоматики. 2024. С. 82–90.

20. Чалий С. Ф. Лещинська І. Уточнення ментальної моделі рішення. АСУ та прилади автоматики. 2024. С. 66–72.
21. Чалий С. Ф. Лещинський В. Лещинська І. Реляційно-тимчасова модель множини сутностей предметної області. Вісник НТУ ХПІ. Системний аналіз управління та інформаційні технології. 2022. 1(7). С. 84–89.
22. Чалий С. Ф. Лещинський В. Побудова локальних пояснень в інтелектуальній системі. Information Technologies and Automation. 2024. С. 693.
23. Чалий С. Ф. Лещинський В. Побудова пояснень на основі каузальних залежностей. АСУ та прилади автоматики. 2024. С. 4–15.
24. Чалий С. Ф. Мацейко Т. Представлення аномальних послідовностей робіт. ТОВ Друкарня Мадрид. 2024. С. 113–114.
25. Чалий С. Ф. Полозов М. Виявлення бізнес-правил на основі темпоральних залежностей. Сучасні напрями розвитку інформаційно-комунікаційних технологій. 2022. Т. 2. С. 84.
26. A content-based recommender for e-commerce web store. URL: https://towardsdatascience.com/a-content-based-recommender-for-ecommerce-web-store-7554b5b73eac (дата звернення: 28.04.2025).
27. Building recommendation engines in Python course. URL: https://app.datacamp.com/learn/courses/building-recommendation-engines-inpython (дата звернення: 28.04.2025).
28. Ecod. Unsupervised outlier detection using empirical cumulative distribution functions / Li Z. et al. IEEE Transactions on Knowledge and Data Engineering. 2022. 35. 12181–12193.
29. Feng Y. Enhancing e-commerce recommendation systems through approach of buyer's self-construal. Frontiers in Artificial Intelligence. 2023. 6. DOI: 10.3389/frai.2023.1167735.
30. Hoang L. Dealing with the new user cold-start problem in recommender systems. Information Systems. 2016. 58. 87–104. DOI: 10.1016/j.is.2014.10.001.

31. Jiang L. Cheng Y. Yang L. A trust-based collaborative filtering algorithm for e-commerce recommendation system. Journal of Ambient Intelligence and Humanized Computing. 2019. 10. 3023–3034. DOI: 10.1007/s12652-018-0928-7.
32. Karimova F. A survey of e-commerce recommender systems. European Scientific Journal. 2016. 12(34). 75. DOI: 10.19044/esj.2016.v12n34p75.
33. Liu C. Wu X. Large-scale recommender system with compact latent factor model. Expert Systems with Applications. 2016. 64. 467–475. DOI: 10.1016/j.eswa.2016.08.009.
34. Personalized e-commerce product recommendations. A guide to implementing a recommendations system for increased brand engagement and conversions. Grid Dynamics Whitepaper. 2022. URL: https://dou.ua/forums/topic/39070/ (дата звернення: 28.04.2025).
35. Priyanka V. Vasamsetty C. Web recommendation system for e-commerce applications. International Journal of Engineering Research and Technology. 2016.
36. Tanmayee S. Unnati N. Recommender Systems in e-commerce. 2022. DOI: 10.13140/RG.2.2.10194.43202.
37. Tawfiq F. Rahma A. Abdulwahab H. An e-commerce recommendation system based on dynamic analysis of customer behavior. Sustainability. 2021. 13. 10786. DOI: 10.3390/su131910786.
38. Yang C. Bai L. Zhang C. Yuan Q. Han J. Bridging collaborative filtering and semi-supervised learning. Proceedings of the ACM SIGIR Conference. 2017. 1245–1254. DOI: 10.1145/3097983.3098094.
39. Yuan W. Wang H. Yu X. Liu N. Li Z. Attention-based context-aware sequential recommendation model. Information Sciences. 2019. 510. DOI: 10.1016/j.ins.2019.09.007.

ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА

на тему:
«Методи оцінювання ефективності рекомендаційних систем для e-commerce»

на здобуття освітнього ступеня бакалавра
зі спеціальності 124 «Системний аналіз»
освітньо-професійної програми «Системний аналіз

Виконала: Ковальова Дар'я
Науковий керівник роботи: Нафєєв Р. К.

Київ 2025

- **Актуальність роботи** зумовлена стрімким зростанням електронної комерції, компанії постійно шукають нові способи залучення та утримання клієнтів. Рекомендаційні системи стали ключовим інструментом персоналізації покупок, значно підвищуючи ефективність бізнесу. Втім, попри широке використання, досі бракує досліджень, що чітко оцінюють їхній вплив на реальні бізнес-результати. Це створює потребу в розробці методологій для точнішого вимірювання ефективності таких систем, що допоможе компаніям оптимізувати стратегії та зміцнити позиції на ринку.

- **Наукова новизна** роботи полягає в створенні нового підходу до оцінювання ефективності рекомендаційних систем для e-commerce, що враховує не тільки точність, але й вплив на бізнес-результати, такі як конверсія та утримання клієнтів.

- **Мета роботи** – розробка методики оцінювання ефективності рекомендаційних систем для e-commerce, враховуючи їхню точність, рівень персоналізації та вплив на бізнес-метрики.

- **Об'єкт дослідження** – інтелектуальні рекомендаційні системи для e-commerce та їх вплив на ефективність онлайн-торгівлі.

- **Предмет дослідження** – методи та підходи до оцінювання ефективності інтелектуальних рекомендаційних систем у сфері e-commerce.

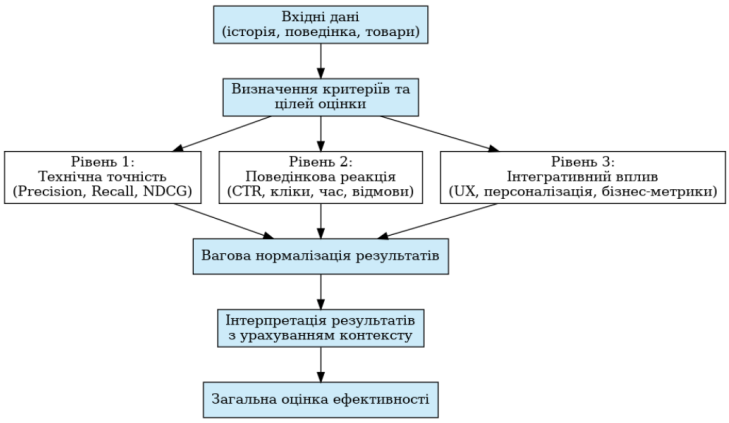
Підходи до побудови рекомендаційних систем у e-commerce

Підхід	Концептуальна мета	Вхідні дані
Колаборативний	Надати рекомендації, спираючись на думки та поведінку інших людей, схожих на користувача	Оцінки користувача + колективні оцінки спільноти
Контентно-орієнтований	Запропонувати об'єкти, схожі за характеристиками на ті, що мені вже сподобалися	Оцінки користувача + характеристики об'єктів
На основі знань	Надати рекомендації відповідно до чітко вказаних уподобань щодо характеристик	Вказані вимоги користувача + характеристики об'єктів + предметні знання
Гібридний	Поєднати переваги різних підходів для підвищення точності та адаптивності	Комбінація: оцінки користувача, характеристики об'єктів, знання, поведінкові патерни

Методи оцінювання ефективності рекомендаційних систем

Підхід оцінювання	Переваги	Недоліки
Офлайн (на основі даних)	Швидкий, дешевий, придатний для тестування багатьох алгоритмів	Не відображає реальної поведінки, ризик перенавчання
Онлайн (A/B-тести, live-метрики)	Показує реальний вплив, враховує повну взаємодію користувача	Дорогий, довгий, може негативно вплинути на користувачів
Багатометриковий (окремі метрики)	Дає повну картину по різних аспектах якості	Важко інтерпретувати, метрики можуть конфліктувати
Комплексний (інтегральний показник)	Один узагальнений показник, зручно порівнювати	Спрощує, залежить від правильного вибору ваг і метрик

Загальна структура методики



Розроблена методика ґрунтується на принципі системного оцінювання, що враховує різні аспекти функціонування рекомендаційних систем у сфері e-commerce.
Вона дозволяє поєднати аналітичну точність з реальними користувацькими та бізнес-індикаторами, забезпечуючи об'єктивну і гнучку оцінку ефективності систем у динамічному середовищі онлайн-торгівлі.

Метрики оцінювання ефективності

Рівень 1:
Технічна оцінка ефективності

- Precision@K, Recall@K, F1 – точність релевантних рекомендацій
- RMSE, MAE – помилка передбачення рейтингів
- MAP, NDCG, MRR – якість ранжування рекомендованих товарів
- Coverage, Diversity – охоплення та різноманіття
- Novelty, Serendipity – новизна та несподіваність

Рівень 2:
Поведінкова оцінка

- CTR – Click-Through Rate (кліки/покази)
- CVR – Conversion Rate (конверсія/кліки)
- Dwell Time – час перегляду сторінки
- Scroll Depth – глибина прокрутки
- PVR – Page View Rate (перегляди сторінок)

Рівень 3:
Бізнес-оцінка

- LTV – Lifetime Value (очікувана цінність клієнта)
- Конверсія – частка користувачів, що купили товар
- ROI – Return on Investment (прибуток/витрати)
- ARPU – Avg. Revenue per User (середній дохід з користувача)
- CAC – Customer Acquisition Cost (вартість залучення клієнта)

Експериментальна структура дослідження

У межах дослідження було запропоновано методику, що поєднує каузальний аналіз та елементи нечіткої логіки для оцінювання ефективності рекомендаційних систем.

Експерименти проводились на реальному датасеті з понад 80 тисячами товарів. Учасників розділено на три групи: перша отримувала базові Top-N рекомендації, друга — персоналізовані на основі категоріальної поведінки, третя — ті самі рекомендації з поясненнями, згенерованими моделлю OpenAI.

Методика дозволила оцінювати не лише точність рекомендацій, а й динаміку поведінкових реакцій користувачів у реальному часі.

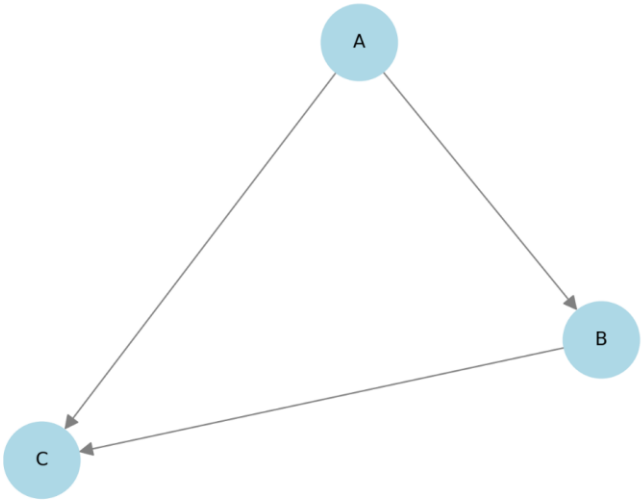
Модуль каузального аналізу у системі пояснень

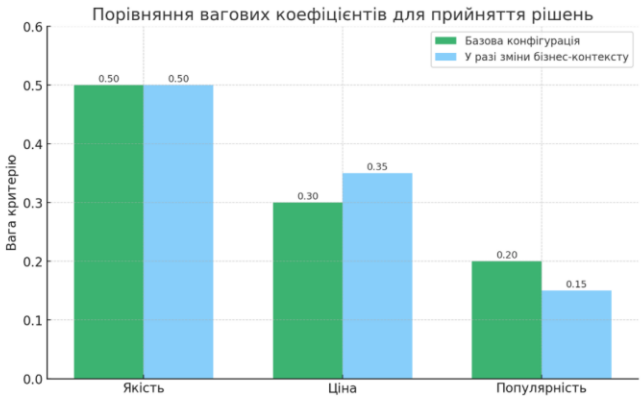
У межах експерименту було реалізовано орієнтований граф причинно-наслідкових дій користувача, де кожна подія — це окремий вузол, а зв'язки між ними ілюструють послідовність впливів.

Побудовано понад 2 000 унікальних шляхів, що дозволило:

- відстежувати типові поведінкові сценарії;
- автоматично генерувати пояснення логіки рекомендації;
- знижувати когнітивне навантаження для користувача.

Такий підхід значно підвищує рівень довіри до системи та особливо ефективний у роботі з новими або недостатньо заповненими профілями користувачів.

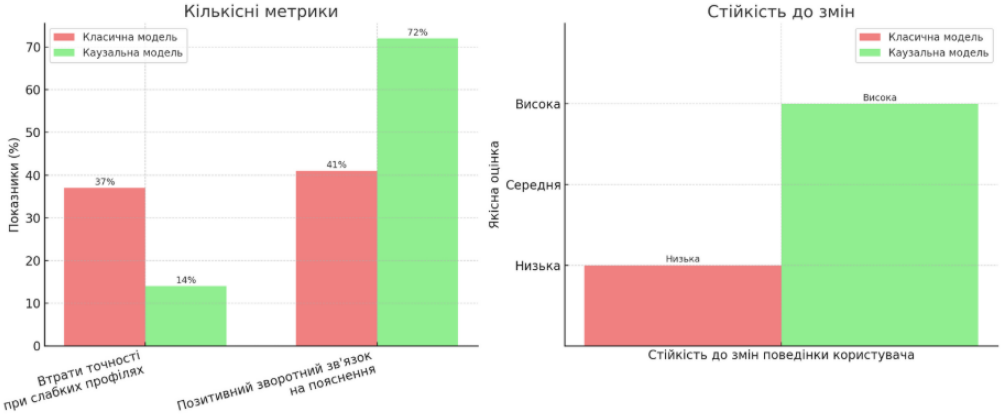




Агрегування вагових коефіцієнтів

Окремим технічним модулем було запроваджено підсистему агрегування вагових коефіцієнтів у системі прийняття рішень, який дозволяє динамічно оцінювати товари на основі зважених показників якості, ціни та популярності.

Початкові ваги формують базову модель пріоритезації, натомість у разі зміни бізнес-контексту система автоматично адаптує ваги критеріїв. Це забезпечує підвищення релевантності рекомендацій завдяки гнучкому перерахунку рейтингової оцінки.



Результати експериментального порівняння моделей

Каузальна модель виявила кращу адаптивність до неповних профілів користувачів, демонструючи лише 14% втрати точності проти 37% у класичної. Також вона зберігала стійкість при зміні поведінкових патернів, перебудовуючи логіку рекомендацій без втрати релевантності.

Користувачі, що отримували пояснення, сформовані на основі каузального аналізу, залишали позитивний зворотний зв'язок у 72% випадків, що майже вдвічі перевищує показник класичної моделі.

Схема БД рекомендаційної системи

База даних рекомендаційної системи зберігає не лише рекомендації, а й повний контекст їх формування.

Зв'язок між таблицями Explanations, Recommendations і Feedback створює замкнуту аналітичну петлю, що дозволяє оцінювати ефективність, прозорість та вдосконалювати логіку рекомендацій. Така структура також забезпечує навчання нових моделей і формування стратегічної аналітики для довгострокової взаємодії з користувачем.

Сгенеровані пояснення до рекомендацій

Рекомендація	Пояснення	Можливість	Необхідність
"Думай, як монах"	Рекомендовано вам тому, що стосується тем саморозвитку, які вас цікавлять.	0.32	0.72
"Атлант розправив плечі"	Ця книга рекомендована вам тому, що вона є класичним твором з філософським змістом, що відповідає вашим вподобанням.	0.45	0.91
"Коротка історія часу"	Ви переглядали популярні наукові роботи схожі на дану.	0.28	0.87
"Підсвідомість може все"	Дана робота до вподоби фанатам психології, схожим на вас.	0.41	0.95
"1984"	Рекомендовано вам, оскільки цей класичний твір відображає ваші вподобання до важливих літературних текстів.	0.18	0.76
"Сила звички"	Книга обрана для вас, оскільки вона актуальна в контексті теми поведінкових змін, що вас цікавить.	0.40	0.84
"Феномен моди"	Рекомендовано вам через її відповідність до вашої зацікавленості в розборі соціальних явищ.	0.26	0.67
"Злочин і кара"	Цей твір підійде вам завдяки поєднанню класичного сюжету та психологічних аспектів, що співпадають з вашими інтересами.	0.34	0.80

ВИСНОВКИ

- Проведено комплексне дослідження методів оцінювання ефективності рекомендаційних систем у сфері e-commerce. Проаналізовано сучасні алгоритми, їхні можливості та обмеження в реальному бізнес-середовищі.
- Особливу увагу приділено метрикам якості рекомендацій (точність, новизна, серендипність тощо). Встановлено, що лише комплексний підхід до оцінки забезпечує повне уявлення про користувацьку та бізнес-цінність системи.
- У практичній частині показано, що вибір метрик має відповідати стратегічним цілям компанії, а не обмежуватись технічною точністю.
- Результатом стало створення методики, яка поєднує технічні та бізнес-метрики, що дозволяє ефективніше впроваджувати персоналізовані РС у сфері електронної комерції.

Апробація результатів дослідження

- IV Міжнародній науково-технічній конференції «Сучасний стан та перспективи розвитку IoT», 15 березня 2025 року, ДУІКТ - «Методи оцінювання ефективності рекомендаційних систем для e-commerce»