

Київський національний університет імені Тараса Шевченка

Економічний факультет

Кафедра економічної кібернетики

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

**«Розробка рекомендаційної системи для підвищення ефективності
маркетплейсу»**

студентки 2 курсу магістратури

спеціальності 051 «Економіка»

ОНП «Економічна кібернетика»

денної форми навчання

Князевої Людмили Святославівни

Науковий керівник:

доктор економічних наук, професор

Затонацька Тетяна Георгіївна

Засвідчую, що в цій роботі немає запозичень із
праць інших авторів без відповідних посилань

Студент

(підпис)

Роботу допущено до захисту перед ЕК
рішенням кафедри економічної кібернетики
від 07 травня 2025 р., протокол № 12

Завідувач кафедри:

доктор економічних наук, професор

Ляшенко Олена Ігорівна

(підпис)

КИЇВ – 2025

РЕФЕРАТ

Кваліфікаційна робота магістра містить: 81 ст., 26 рис., 3 табл., 60 джерел, 4 додатки.

Ключові слова: електронна комерція, рекомендаційні системи, штучний інтелект, машинне навчання, маркетплейс, колаборативна фільтрація, фільтрація за вмістом, сентимент-аналіз.

Об'єкт дослідження: процеси персоналізації користувацького досвіду у маркетплейсах шляхом застосування рекомендаційних систем.

Мета дослідження: розробка рекомендаційної системи для підвищення ефективності маркетплейсу з використанням сучасних методів машинного навчання.

Методи дослідження: методи машинного навчання, зокрема класифікаційні алгоритми (логістична регресія, мультиноміальний наївний байєсівський класифікатор, класифікатор Ridge, пасивно-агресивний класифікатор та LightGBM), алгоритм кластеризації k-means, алгоритми колаборативної та контентної фільтрації, методи обробки природної мови (NLP) та аналіз сентиментів.

Наукова новизна, теоретична значимість дослідження: запропоновано рекомендаційну систему з модульною структурою, що дозволяє подолати проблеми масштабованості та «холодного старту», а також забезпечує гнучке та адаптивне поєднання різних алгоритмів залежно від завдань маркетплейсу та поведінки користувачів.

Практична цінність: створено універсальне рішення, придатне до впровадження на різних платформах електронної комерції, здатне працювати з реальними даними та враховувати емоційний контекст користувацьких відгуків для підвищення точності та релевантності рекомендацій.

RESUME

Taras Shevchenko National University of Kyiv,

Faculty of Economics, Department of Economic Cybernetics

Key words: e-commerce, recommendation systems, artificial intelligence, machine learning, marketplace, collaborative filtration, content-based filtration, sentiment analysis.

The graduation research of the student deals with development of a recommendation system for improving the marketplace efficiency. The work is interesting for introducing a modular recommendation system tailored to the dynamic nature of e-commerce marketplaces. By enabling flexible integration of components such as sentiment analysis, clustering, collaborative filtering, and content-based filtering, the system can adapt to different user behaviors, data structures, and business objectives. This adaptability makes the solution scalable and reusable across various platforms, while the use of real marketplace data and emotion-aware feedback analysis enhances recommendation relevance and boosts commercial effectiveness.

Pages 81, tables 3, bibliog. 60, append. 4.

ЗМІСТ

ВСТУП.....	6
РОЗДІЛ 1. РЕКОМЕНДАЦІЙНІ СИСТЕМИ ЯК НАПРЯМ РОЗВИТКУ ЕЛЕКТРОННОЇ КОМЕРЦІЇ ПІД ВПЛИВОМ ШТУЧНОГО ІНТЕЛЕКТУ	10
1.1. Стан та тенденції розвитку глобального ринку електронної комерції... 10	
1.2. Штучний інтелект як інструмент підвищення ефективності електронної комерції.....	21
1.3. Досвід використання рекомендаційних систем у світових компаніях... 29	
Висновки до розділу 1	37
РОЗДІЛ 2. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ РОЗРОБКИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ	38
2.1. Етапи рекомендаційного процесу.....	38
2.2. Методологічні підходи до розробки рекомендаційних систем.....	40
2.3. Метрики оцінювання рекомендаційних алгоритмів	46
Висновки до розділу 2	50
РОЗДІЛ 3. РОЗРОБКА РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ДЛЯ МАРКЕТПЛЕЙСУ	51
3.1. Інформаційне забезпечення: опис, структура та джерела даних	51
3.2. Аналіз сентиментів на основі класифікаційних моделей машинного навчання	54
3.3. Розробка рекомендаційної системи на основі алгоритмів колаборативної фільтрації.....	58
3.4. Розробка рекомендаційної системи за допомогою алгоритмів фільтрації на основі вмісту.....	64
Висновки до розділу 3	66
ВИСНОВКИ	68

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	71
ДОДАТКИ.....	78
<i>Додаток А</i>	78
<i>Додаток Б</i>	79
<i>Додаток В</i>	80
<i>Додаток Г</i>	81

ВСТУП

Актуальність теми дослідження. У сучасних умовах цифровізації економіки та стрімкого розвитку технологій штучного інтелекту електронна комерція зазнає істотних трансформацій. Зростання обсягів онлайн-продажів, зміна поведінки споживачів та зростаюча конкуренція серед торгівельних платформ вимагають запровадження інноваційних інструментів для персоналізації та підвищення ефективності взаємодії з клієнтами. Одним із найбільш перспективних напрямів є розробка рекомендаційних систем, які, використовуючи алгоритми штучного інтелекту та машинного навчання, дозволяють персоналізувати взаємодію з користувачами, підвищити рівень їх задоволеності та стимулювати повторні покупки.

Світовий досвід таких компаній, як Amazon, Netflix та інших демонструє, що впровадження рекомендаційних систем суттєво підвищує конкурентоспроможність бізнесу. Аналіз наукової літератури свідчить про наявність різних методологічних підходів (колаборативна фільтрація, фільтрація на основі вмісту, гібридні моделі), проте питання створення нових рішень залишається надзвичайно актуальним. Це пояснюється динамічністю інтересів користувачів, зміною середовищ застосування, а також постійним ускладненням структури даних. Багато сучасних систем виявляють недостатню гнучкість, стикаються з труднощами при роботі з новими користувачами або об'єктами та генерують упереджені рекомендації. У контексті швидкого розвитку таких технологій, як обробка природної мови, з'являється потенціал для суттєвого підвищення якості рекомендацій.

Об'єктом дослідження є процеси персоналізації користувацького досвіду у маркетплейсах шляхом застосування рекомендаційних систем. **Предметом дослідження** є алгоритми та методи побудови рекомендаційної системи для маркетплейсу, що дозволяють покращити користувацький досвід і підвищити ефективність продажів.

Метою дослідження є розробка рекомендаційної системи для підвищення ефективності маркетингу з використанням сучасних методів машинного навчання. Для досягнення даної мети було сформовано наступні **завдання**:

- вивчити сучасний стан і тенденції розвитку електронної комерції;
- дослідити можливості застосування ІІІ в електронній комерції;
- проаналізувати практики впровадження рекомендаційних систем у провідних світових компаніях;
- охарактеризувати основні етапи рекомендаційного процесу;
- систематизувати методологічні підходи до побудови рекомендаційних моделей;
- дослідити метрики оцінювання ефективності рекомендаційних алгоритмів;
- провести аналіз вхідних даних та джерел інформації для побудови рекомендаційної системи;
- реалізувати модель аналізу настроїв на основі класифікаційних алгоритмів;
- розробити рекомендаційну систему на основі алгоритмів колаборативної фільтрації;
- створити систему рекомендацій на основі фільтрації за змістом.

Методами дослідження є методи машинного навчання, зокрема класифікаційні алгоритми (логістична регресія, мультиноміальний наївний байєсівський класифікатор, класифікатор Ridge, пасивно-агресивний класифікатор та LightGBM), алгоритм кластеризації k-means, алгоритми колаборативної та контентної фільтрації, методи обробки природної мови (NLP) та аналіз настроїв. Для вирішення завдань розробки рекомендаційної системи підвищення ефективності маркетингу було використано мову програмування, обробки та аналізу даних Python.

Наукова новизна роботи полягає у розробці рекомендаційної системи з модульною структурою, де кожен функціональний компонент - класифікація для аналізу настрою, алгоритми кластеризації, колаборативної та контентної фільтрації - виконує роль окремого модуля, що може бути активований або змінений відповідно до конкретної задачі, структури даних або поведінкових патернів користувачів. Такий підхід удосконалює традиційні рекомендаційні системи, забезпечуючи масштабованість, гнучке поєднання та адаптацію різних моделей залежно від специфіки задач і характеристик маркетплейсу.

Практична новизна полягає у створенні універсальної системи, яка легко інтегрується в ІТ-інфраструктуру різних маркетплейсів, забезпечуючи релевантність рекомендацій та зростання показників ефективності платформи. Розроблений механізм дозволяє враховувати не лише історичну поведінку користувачів, але й емоційне забарвлення їхніх відгуків, що сприяє формуванню релевантних рекомендацій та підвищенню комерційної ефективності платформи.

Результати впровадження роботи. Розробки, які наведені у роботі, пройшли апробацію на Міжнародній науково-практичній конференції студентів, аспірантів та молодих вчених «Шевченківська весна 2025. Економіка України 2025: нові вектори розвитку в умовах глобальних трансформацій» за секцією «Моделювання та інформаційні технології в економіці: сучасні виклики та нові вектори розвитку» у тезах на тему «Реалізація рекомендаційного алгоритму на основі сегментації клієнтів та колаборативної фільтрації».

Інформаційну базу склали наукові праці вітчизняних і зарубіжних авторів; аналітичні звіти компаній McKinsey, Statista, та ін., а також дані з відкритих маркетплейсів.

Структура роботи. Кваліфікаційна робота складається зі вступу, трьох основних розділів, висновків, списку використаних джерел та додатків.

У першому розділі досліджено стан та тенденції подальшого розвитку ринку електронної комерції, роль штучного інтелекту та рекомендаційних систем, а також досвід провідних компаній.

У другому розділі проаналізовано теоретико-методологічні аспекти побудови рекомендаційних систем, класифікацію алгоритмів та метрики їх оцінювання.

Третій розділ присвячено практичній реалізації системи: аналізу структури даних, побудові моделей аналізу настроїв і розробці рекомендаційних систем на основі двох підходів - колаборативної фільтрації та фільтрації за вмістом.

РОЗДІЛ 1. РЕКОМЕНДАЦІЙНІ СИСТЕМИ ЯК НАПРЯМ РОЗВИТКУ ЕЛЕКТРОННОЇ КОМЕРЦІЇ ПІД ВПЛИВОМ ШТУЧНОГО ІНТЕЛЕКТУ

1.1. Стан та тенденції розвитку глобального ринку електронної комерції.

Термін «електронна комерція» є багатограним і охоплює різні аспекти комерційної діяльності, що здійснюється із використанням електронних засобів зв'язку. Упродовж останніх десятиліть науковці, міжнародні організації та практики запропонували низку підходів до визначення цього поняття, кожен з яких відображає окремі аспекти або функції електронної комерції. Зокрема, можна виділити економічний, інформаційно-технологічний, інституційно-правовий, маркетинговий та функціональний підходи.

Класичний економічний підхід фокусується на фінансово-торгівельній сутності явища, акцентуючи на транзакційній природі електронної комерції. Згідно з цим підходом, електронна комерція розглядається як форма комерційної діяльності, в якій торговельні операції здійснюються з використанням інформаційно-комунікаційних технологій, насамперед Інтернету. Так, наприклад, К. Лаудон та К. Тревер [1] визначають електронну комерцію як процес купівлі, продажу, передачі або обміну товарів, послуг та інформації за допомогою комп'ютерних мереж, зокрема Інтернету.

Інформаційно-технологічний підхід підкреслює технологічну складову електронної комерції. Електронна комерція розглядається як сукупність інструментів і платформ, що дозволяють здійснювати ділову активність в цифровому середовищі. Так, наприклад, Р. Калакота та Е. Уінстон [2] визначають електронну комерцію як використання електронних засобів і платформ (веб-сайтів, мобільних додатків, електронної пошти тощо) для ведення бізнесу. У цьому підході важлива роль відводиться таким технологіям, як онлайн-платформи, платіжні системи, логістичне ПЗ, CRM-системи тощо.

Згідно з інституційно-правовим підходом, електронна комерція розглядається як частина правовідносин, що регулюють купівлю-продаж через

електронні канали. Так, згідно з законом України «Про електронну комерцію» [3], електронна комерція – це відносини, спрямовані на отримання прибутку, які виникають під час здійснення правочинів щодо набуття, зміни або припинення цивільних прав та обов'язків, здійснені дистанційно з використанням інформаційно-комунікаційних систем, в результаті чого в учасників відносин виникають права та обов'язки майнового характеру.

У межах маркетингового підходу електронна комерція розглядається як інструмент електронного маркетингу й просування товарів. Основна увага приділяється взаємодії зі споживачами, формуванню лояльності та аналізу поведінки покупців у цифровому середовищі. Так, наприклад, Д. Чаффі [4] визначає електронну комерцію як використання електронних каналів для стимулювання продажів і побудови довгострокових відносин з клієнтами.

Функціональний підхід розглядає електронну комерцію як сукупність процесів, що забезпечують повний цикл здійснення торгової діяльності онлайн. Так, ОЕСР [5] визначає електронну комерцію як продаж або купівлю товарів чи послуг через комп'ютерні мережі методами, спеціально розробленими для отримання або розміщення замовлень, включаючи етапи замовлення, оплати, доставки та післяпродажного обслуговування. Таким чином, електронна комерція трактується як комплексна діяльність із залученням кількох ланок цифрового ланцюга поставок.

Підсумовуючи описані підходи, можна сформулювати комплексне визначення електронної комерції як форми комерційної діяльності, що здійснюється за допомогою електронних засобів (Інтернету, мобільних додатків, цифрових платформ) та охоплює обмін товарами, послугами та інформацією між суб'єктами ринку з урахуванням правових, логістичних та комунікаційних процесів.

Дослідженням стану та розвитку ринку електронної комерції присвячено багато робіт вітчизняних та закордонних вчених. Так, Кравченко Б. у своїй роботі

[6], аналізує динаміку розвитку українського ринку електронної комерції, фактори, які на неї впливали, визначаючи основні тенденції у споживчій поведінці та подальшому розвитку. Авторка відзначає велику перспективність даного ринку, проте також наголошує на викликах, з якими він стикається, у вигляді недосконалих інфраструктури та нормативно-правової бази, а також недовіри споживачів. У роботі Сантоса В., Аугусто Т. та ін. [7] визначено ключові поняття пов'язані з електронною комерцією, описано основні бізнес-моделі на ринку та технології, що на нього впливають. Автори підкреслюють зв'язки електронної комерції з глобалізацією, онлайн маркетплейсами і платформами та представляють ключові показники ефективності з точки зору продажів, маркетингу та обслуговування клієнтів, які учасники ринку електронної комерції повинні враховувати під час визначення власної стратегії електронної комерції. Щитов Б. та ін. у своїй статті [8] аналізують основні напрямки та прогнози розвитку е-комерції з точки зору технологічних та регуляторних змін, а також змін споживчої поведінки. Автори надають рекомендації для адаптації бізнесів у сфері електронної комерції до даних змін, зокрема, досліджуючи покращення доставки. Загалом, наявний широкий спектр досліджень, який розглядає проблематику даного ринку з різних кутів, однак положення наведені у даних роботах потребують перегляду з точки зору поточних змін, що відбуваються на ринку.

Сфера електронної комерції зародилася в США, після чого отримала розвиток в Європі і з кінця 90-х років XX століття стала активно розвиватися по всьому світу.

На сьогоднішній день 2,77 мільярдів осіб роблять покупки онлайн, що становить 33% населення світу - це на 2,2% більше, ніж минулого року. Прогнозується, що завдяки поширенню мережі Інтернет та зростаючій зручності платформ електронної комерції у 2026 році кількість осіб, що роблять покупки онлайн, зросте до 2,86 мільярдів. [9]

У 2025 році 21% роздрібних покупок відбуватиметься онлайн, що наразі є найвищим показником. Крім того, очікується, що до 2027 року 22,6% усіх роздрібних покупок буде здійснюватися в Інтернеті. З 2021 року частка роздрібних онлайн-покупок щорічно зростає в середньому на 0,32%. [10]

Очікується, що у 2025 році продажі електронної комерції в усьому світі сягнуть 6,88 трильйонів доларів, а до 2027 року – 8,03 трильйонів доларів. В такому разі, глобальні онлайн-продажі зростуть на 5,65 трильйонів доларів США за 10 років з 2017 року. [11]

Найбільший поштовх електронній комерції дала пандемія COVID-19 у 2020 році з рекордним зростанням у 897 мільярдів доларів лише за один рік. Після сплеску 2020 року у 2021 році темпи зростання сповільнилися до 17,42%, але електронна комерція все одно зросла майже на 740 мільярдів доларів порівняно з попереднім роком. Найповільніше галузь зростала у 2022 році додавши лише 323 мільярди доларів США (або 6,48%) до 5,31 трильйона доларів. (рис. 1.1.)

Очікується, що з 2024 до 2027 року глобальні продажі електронної комерції зростуть на 1,70 трильйона доларів. Загалом, глобальні продажі електронної комерції зростуть майже в шість разів з 1,34 трильйона доларів у 2014 році до прогнозованих 8,03 трильйона доларів у 2027 році. [11]



Рис. 1.1. Світові роздрібні продажі електронної комерції у 2015-2027 рр., трлн. дол. США

Джерело: побудовано автором на основі [10-11]

Азія є світовим лідером у сфері електронної комерції. У 2024 році галузь принесла даному регіону близько 1,97 трільйона доларів, що становить 47,90% від загального світового доходу від електронної комерції [12].

З відривом у 537 мільярдів доларів, друге місце посідають Північна і Південна Америка з доходами у 1,44 трлн доларів, що становить 34,86% світових доходів від електронної комерції за 2024 рік [12]. Разом Азія та Америка забезпечили понад 80% світового доходу від онлайн-покупок (рис. 1.2.).

Третю сходинку займає Європа, де дохід від роздрібної онлайн-торгівлі оцінюється у 632,7 мільярда доларів, що становить близько 15,37% від глобального обсягу [12].

Найменші частки світових доходів від електронної комерції (1,04% та 0,84%) належать Австралії та Океанії і Африці з доходами у 42,7 та 34,6 мільярдів доларів США відповідно [12].

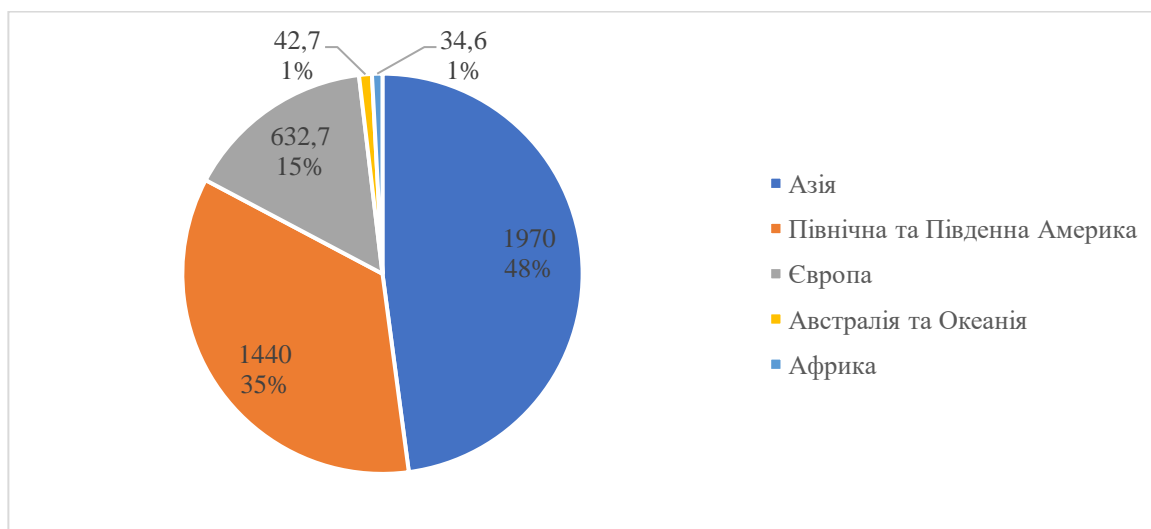


Рис. 1.2. Доходи від електронної комерції за регіонами у 2024 р., млрд. дол. США

Джерело: побудовано автором на основі [12]

Очікується [13], що дохід від онлайн-продажів продуктів харчування різко зросте з 230 мільярдів доларів у 2019 році до колосальних 1,23 трильйонів доларів у 2029 році, що зробить продовольчі товари найбільшим сегментом електронної комерції, випередивши одяг у 2027 році. Тим не менш, продажі одягу через онлайн-платформи також продовжують зростати і у 2029 році дохід від даного сегменту, за прогнозами, сягне 1,18 трильйонів доларів США, що на 127% більше порівняно з 520 мільярдами доларів у 2019 році (рис. 1.3.).

Медіа-електронна комерція, за оцінками, зросте на 129% з 280 мільярдів доларів у 2019 році до 640 мільярдів доларів у 2029 році, а дохід від онлайн-продажів електроніки - зі 190 до 400 мільярдів доларів [13].

За аналогічний період ринок іграшок і товарів для гобі зросте більш ніж удвічі, з 70 мільярдів доларів до 170 мільярдів доларів, а дохід від господарських товарів - з 220 до 720 мільярдів доларів [13].

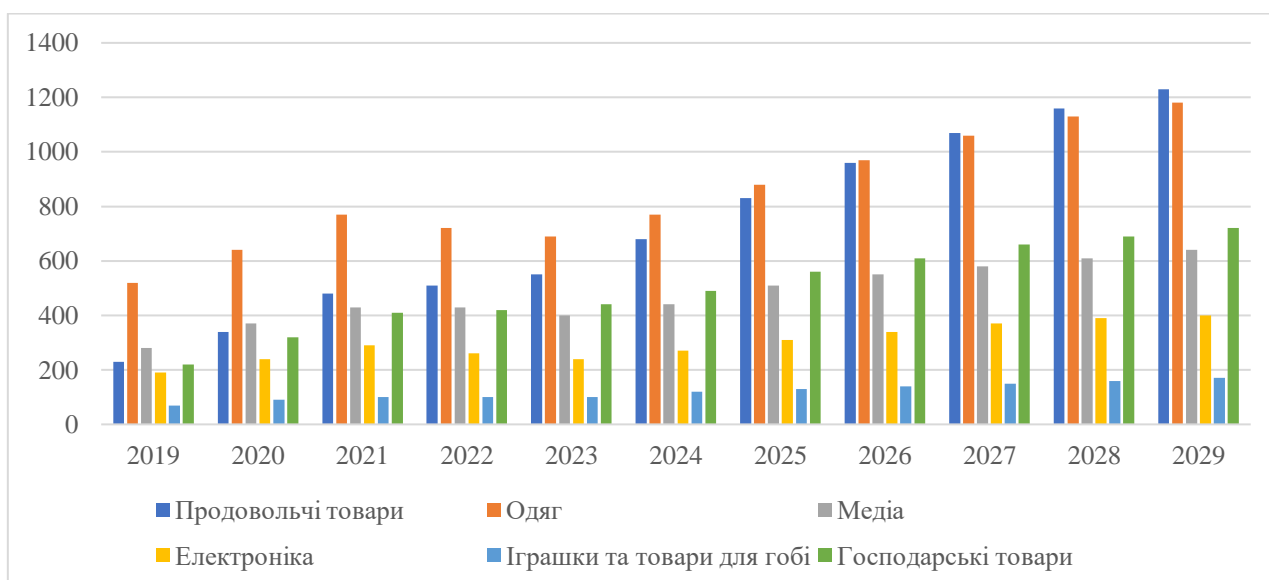


Рис. 1.3. Світові доходи від електронної комерції за сегментами у 2019-2029 рр., млрд. дол. США

Джерело: побудовано автором на основі [13]

Станом на 2025 рік у світі налічується 28 мільйонів сайтів електронної комерції, що на 2,9% перевищує показник минулого року [14]. Це означає, що приблизно 2162 веб-сайти електронної комерції запускалися щодня між 2024 і 2025 роками.

Половина світових сайтів електронної комерції базуються у США. Два з трьох найбільш відвідуваних сайтів були американськими компаніями – Amazon і Walmart. Однак найбільше зріс у популярності відносно минулого року китайський AliExpress, який показав величезне зростання на 44%. [15]

Що стосується залучення більшої кількості відвідувачів, то лідерами стали азійські компанії. Flipkart, інтернет-магазин з Індії, продемонстрував суттєве збільшення відвідуваності веб-сайту на 90%. Однак найвидатнішою компанією стала Lazada, кількість відвідувачів якої зросла майже на 200%, в основному завдяки її присутності на філіппінському ринку електронної комерції. [15]

Зростаючи, електронна комерція продовжує трансформуватись під впливом технологічних інновацій, змін у споживчій поведінці та глобальних економічних

факторів. Серед основних тенденцій, що визначають майбутнє глобальної електронної комерції, можна виділити наступні:

1. Забезпечення прозорості та безпеки завдяки блокчейн технологіям.

Оскільки електронна комерція продовжує зростати, зростають і занепокоєння щодо безпеки даних, шахрайства та довіри між покупцями та продавцями. Одним із інструментів, що сприяє вирішенню даних проблем, є блокчейн.

Блокчейн - це розподілений цифровий реєстр, який безпечно зберігає записи про транзакції в мережі комп'ютерів. Кожен блок у цьому ланцюзі містить криптографічний хеш попереднього блоку, мітку часу та дані транзакції. Це забезпечує те, що після запису дані не можуть бути змінені заднім числом без зміни всіх наступних блоків і досягнення згоди в мережі. [16]

Згідно з прогнозами [17], глобальний ринок блокчейн-технологій різко зросте з 17 мільярдів доларів США у 2023 році до понад 943 мільярдів доларів США до 2032 року, що свідчить про його швидке впровадження в усіх галузях, включаючи електронну комерцію. Забезпечуючи децентралізованість реєстрів та їх захищеність від несанкціонованого втручання, блокчейн підвищує безпеку транзакцій, запобігає шахрайству та покращує прозорість онлайн-покупок.

Бренди вже використовують блокчейн для забезпечення безпеки платежів, відстеження ланцюжків поставок і розумних контрактів, які автоматизують транзакції без посередників. Завдяки блокчейну покупці можуть перевіряти оригінальність продукту, відстежувати замовлення в режимі реального часу та захищати свої особисті дані, що дозволяє зробити їх досвід на безпечнішим і надійнішим.

Оскільки компанії віддають приділяють значну увагу конфіденційності та безпеці даних, блокчейн зіграє вирішальну роль у майбутньому електронної комерції, забезпечуючи прозорість транзакцій та впевненість споживачів під час здійснення покупок в Інтернеті.

2. Інтеграція тестування продуктів за допомогою доповненої реальності.

Однією з найбільших проблем онлайн-покупок є неможливість побачити та протестувати продукт особисто перед покупкою. Одним із шляхів подолання цієї проблеми є використання технологій доповненої реальності, що дозволяють інтегрувати цифрову інформацію з середовищем у якому перебуває користувач в режимі реального часу за допомогою цифрових пристроїв.

Очікується [18], що до кінця 2025 року кількість користувачів доповненої реальності в США перевищить 100 мільйонів, що становитиме 32% населення. Дана тенденція відобразиться і на електронній комерції, кардинально змінивши досвід онлайн-покупок, завдяки можливостям віртуальних примірок та інтерактивного 3D-перегляду продуктів, що дозволяє покупцям візуалізувати продукти в їхньому реальному середовищі, таким чином підвищуючи довіру покупців і знижуючи рівень повернення товарів.

3. Оптимізація для голосового пошуку.

Станом на 2025 рік 75% домогосподарств у США володіють розумними колонками [19], що породжує та доводить існуючу тенденцію на голосовий пошук, яка розвивається в сфері електронної комерції.

Голосові помічники, такі як Amazon Alexa та Google Assistant, змінили спосіб взаємодії споживачів із платформами електронної комерції, запропонувавши швидкий і зручний спосіб робити покупки. За допомогою простої голосової команди покупці можуть з легкістю шукати продукти, робити покупки та відстежувати замовлення. У відповідь багато компаній електронної комерції оптимізують свої сайти для голосового пошуку, забезпечуючи швидкі й точні відповіді на голосові запити. І оскільки голосова технологія продовжує розвиватися, її інтеграція в процес покупок, безсумнівно, підвищить зручність для клієнтів і сприятиме значному зростанню онлайн-продажів.

4. Використання мобільних пристроїв для здійснення покупок продовжує зростати.

Доступність та поширеність використання смартфонів призводить до того, що частка мобільної комерції в загальній електронній комерції постійно зростає, і очікується [20], що вона досягне 62% у 2027 році.

У відповідь на це бізнес часто фокусується на адаптації платформ електронної комерції для мобільних пристроїв і забезпечують безперебійну роботу на всіх пристроях, що означає оптимізацію контенту для менших екранів, спрощення замовлення в один клік і надання інших зручностей, які підтримують покупки на ходу.

5. Надання гнучких та різноманітних способів оплати.

Одним із ключових етапів, що формують досвід клієнта при здійсненні покупки є оформлення та оплата замовлення. Навіть найменші проблеми на цьому етапі можуть призвести до втрати клієнта. Однією з таких проблем може стати відсутність достатньої кількості варіантів оплати. Так, згідно з дослідженнями [21], 13% клієнтів покинуть свій кошик, якщо не побачать достатньо доступних способів оплати, у зв'язку з чим бізнеси мають приділяти значну увагу урізноманітненню варіантів оплати, таких як кредитні картки, платежі через цифрові гаманці, такі як Google Pay або Apple Wallet, або варіанти оплати в один клік, як, наприклад, PayPal.

6. Популяризація онлайн-покупок через соціальні мережі.

За останні декілька років Instagram, Facebook і TikTok стали не лише соціальними мережами, а і платформами електронної комерції. Так, станом на 2020 рік у США 110,4 мільйонів осіб здійснювали покупки через соціальні мережі [22].

Використання даних каналів стає все зручнішим для покупців, завдяки полегшеному пошуку та купівлі товарів, і в той же час дозволяє як великому, так і малому, бізнесу підвищувати впізнаваність бренду, розширювати свою аудиторію та пропонувати безперебійний досвід покупок.

7. Підвищення лояльності за рахунок моделі підписки.

За прогнозами The Washington Post [23], в 2025 році «економіка підписок» зросте до 1,5 трильйона доларів. Ця тенденція охопила в тому числі і широкий спектр галузей електронної комерції та продовжує розвиватися.

Для багатьох брендів електронної комерції ця бізнес-модель є привабливою, оскільки забезпечує не лише постійний дохід, але й підвищення лояльності клієнтів, наприклад, за рахунок надання покупцям доступу до ексклюзивних знижок і пропозицій, раннього доступу до нових продуктів і персоналізованої підтримки клієнтів.

8. Пріоритизація сталого розвитку.

Сучасні онлайн-покупці стають все більш обізнаними в питаннях екологічності. Згідно з дослідженням PwC [24], понад 70% покупців заявили, що готові були б платити більше за екологічно вироблені товари.

Особливо в індустрії моди та одягу молоді покупці віддають перевагу брендам, які керуються етичними і екологічними практикам. Згідно з опитуванням Drapers [25], 57% представників поколінь зумерів і міленіалів вважають, що екологічність важлива, коли йдеться про покупку одягу, аксесуарів або взуття - порівняно з 47% в опитуванні 2022 року. Ця зміна вподобань відображає зростаючу тенденцію, коли споживачі все більше враховують вплив своїх покупок на навколишнє середовище, що впливає на їх лояльність і загальне сприйняття бренду.

9. Використання даних клієнтів для персоналізації користувацького досвіду.

Як показують результати досліджень, персоналізація підвищує лояльність клієнтів. Так, наприклад, дослідження Google і Storyline Strategies [26] показало, що 72% споживачів з більшою ймовірністю будуть лояльними до бренду, якщо він пропонує персоналізований клієнтський досвід.

На сьогоднішній день персоналізація представляє собою створення безперебійного багатоканального досвіду, який залучає клієнтів у кожній точці взаємодії, завдяки налаштування кожного етапу покупки під вподобання клієнта за допомогою персоналізованих рекомендацій щодо продуктів, перенацілювання реклами, push-повідомлень і навіть динамічного ціноутворення.

Однак подібні маніпуляції з особистою інформацією можуть підривати довіру клієнтів до того, як компанії зберігають і використовують персональні дані, роздрібні торговці, надаючи клієнтам можливість «увімкнути» персоналізацію, надавши особисті дані. Це не тільки зміцнює довіру ваших клієнтів, але й полегшує організаціям прозорість щодо, використання даних клієнтів і дотримуватись законів про конфіденційність даних споживачів, включаючи європейський GDPR і Каліфорнійський закон про конфіденційність споживачів (CCPA).

10. Використання штучного інтелекту для вирішення проблем клієнтів та оптимізації бізнес-процесів.

За останні кілька років впровадження штучного інтелекту кардинально змінило галузь електронної комерції, дозволивши брендам автоматизувати кожен аспект свого бізнесу: від оптимізації обслуговування клієнтів за допомогою чат-ботів до персоналізації рекомендацій щодо продукту та оптимізації ланцюжка поставок. Зокрема, у Франції 84% опитаних бізнесів у галузі електронної комерції або вже активно впроваджують рішення на основі ШІ, або відзначають впровадження даних технологій як один з основних пріоритетів [27].

1.2. Штучний інтелект як інструмент підвищення ефективності електронної комерції

Зі зростанням конкуренції на ринку електронної комерції, одним з ключових інструментів для підвищення ефективності бізнесу у даній галузі стає впровадження технологій штучного інтелекту.

Поняття «штучний інтелект» виступає узагальнюючим терміном для сукупності методів і алгоритмів, які дозволяють комп'ютерним системам

навчатися на основі великих масивів даних, виявляти закономірності та приймати рішення або робити прогнози без явного програмування кожної дії. У цьому контексті ШІ реалізується здебільшого через машинне навчання - підгалузь, що створює моделі на основі статистичних підходів, зокрема нейронні мережі, дерева рішень, методи кластеризації та глибинного навчання.

Використання штучного інтелекту є поширеною темою для досліджень. Зокрема, Музиченко Т. та ін. у своїй роботі [28] аналізують практичне застосування ШІ у бізнес-процесах компаній, відзначаючи його позитивний вплив на обробку даних та якість обслуговування. Проте, автори також зазначають, що перспективність даного напрямку зумовлює важливість вирішення пов'язаних з ним проблем, зокрема, у аспектах регулювання та етичності. У статті Булаха О. [29] розвивається питання викликів для бізнесу електронної комерції пов'язаних зі штучним інтелектом та шляхів їх подолання. Махова Г. у своїй статті [30] розглядає основні підходи до визначення поняття «штучний інтелект» та узагальнює вплив ШІ на діяльність підприємницьких структур, підтверджуючи його наведеними результатами досліджень. Результати попередніх досліджень закладають фундамент для розуміння впливу штучного інтелекту, проте потребують уточнення з урахуванням поточних тенденцій та стану ринку електронної комерції, та розвитку алгоритмів ШІ.

Станом на 2025 рік електронна комерція є одним із лідерів у впровадженні штучного інтелекту, випереджаючи фінансові технології та онлайн-медіа. Використання штучного інтелекту в електронній комерції надзвичайно зросло за останні 5 років, і лише продовжить збільшуватися у найближчі роки. Очікується [31], що розмір глобального ринку електронної комерції з підтримкою штучного інтелекту сягне 9,01 мільярдів доларів США в 2025 році. Згідно з прогнозами [31], до 2034 року цей ринок зросте до 64,03 мільярдів доларів США, завдяки сукупному річному темпу зростання у розмірі 24,34% з 2024 по 2034 рік. (рис. 1.4.)

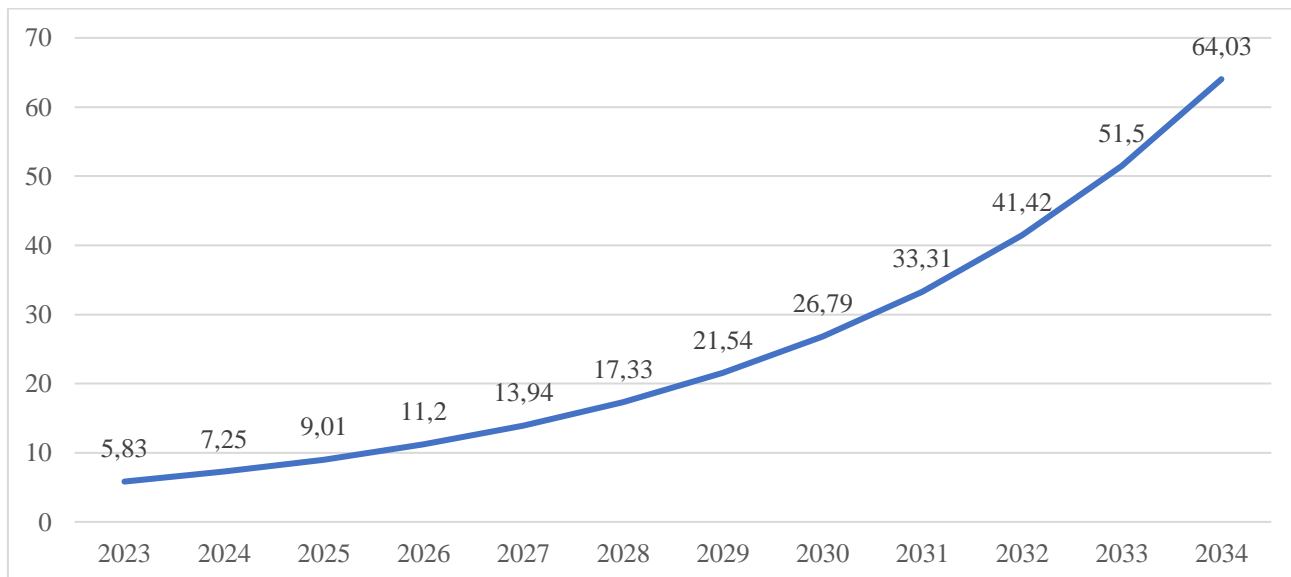


Рис. 1.4. Розмір глобального ринку штучного інтелекту у електронній комерції у 2023-2034 рр., млрд. дол. США

Джерело: побудовано автором на основі [31]

У 2023 році лідером на ринку штучного інтелекту в електронній комерції була Північна Америка і очікується, що регіон збереже свої позиції протягом прогнозованого періоду завдяки високим показникам проникнення Інтернету та витрат на рішення штучного інтелекту в державному секторі (рис. 1.5.) . Проте, за прогнозами, частка Азіатсько-Тихоокеанського регіону значно зросте протягом прогнозованого періоду в результаті економічних і технічних удосконалень регіону, що сприятимуть розширенню використання рішень на основі штучного інтелекту. Також в наступні роки очікується суттєве збільшення ринків Близького Сходу, Африки та Латинської Америки внаслідок зростання молодого населення та швидкого розвитку сектора інтернет-покупок. [31]

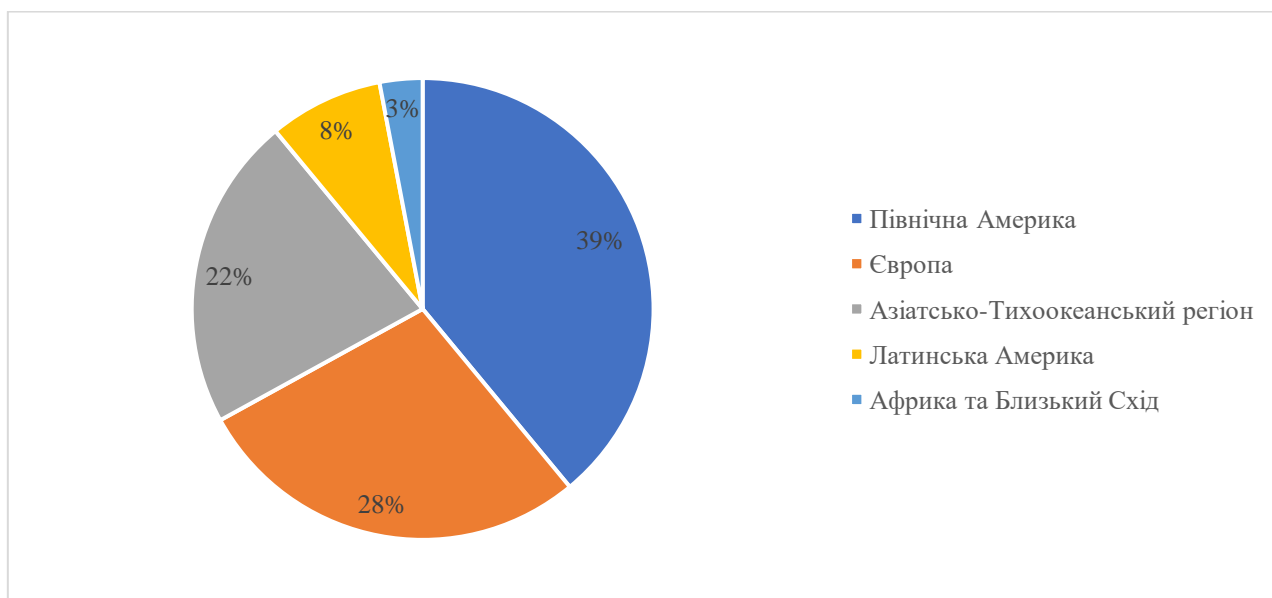


Рис. 1.5. Частка регіонів у глобальному ринку штучного інтелекту в електронній комерції у 2023 р., %

Джерело: побудовано автором на основі [31]

Серед основних інструментів та моделей штучного інтелекту, що використовуються в електронній комерції варто виділити наступні:

- Обробка природної мови (NLP) – область штучного інтелекту зосереджена на тому, щоб комп’ютерні системи могли інтерпретувати та створювати природну людську мову.
- Машинне навчання (ML) – підгалузь штучного інтелекту, що використовує статистичні методи, що дозволяють комп’ютерним системам навчатися на основі даних і робити прогнози чи рішення без явного програмування.
- Комп’ютерний зір (CV): Комп’ютерний зір - це сфера штучного інтелекту, яка дозволяє комп’ютерам інтерпретувати інформацію із зображень і відео.
- Інтелектуальний аналіз даних - це процес виявлення корисних, неочевидних, раніше невідомих і потенційно значущих закономірностей, структур, залежностей або трендів у великих обсягах даних за допомогою методів статистики, машинного навчання, штучного інтелекту та баз даних.

Дані алгоритми можуть використовуватись у будь-яких процесах та операціях. Найпоширенішими способами використання штучного інтелекту можна вважати наступні:

1. Персоналізація користувацького досвіду

Персоналізація є одним із найважливіших напрямів застосування ШІ в електронній комерції. Системи на основі машинного навчання аналізують дані про поведінку користувачів, такі як історія переглядів, попередні покупки, частота відвідування сайту, взаємодія з рекламними матеріалами та пошукові запити. На основі цієї інформації формуються персоналізовані рекомендації товарів, індивідуальні маркетингові повідомлення, а також адаптується вміст сторінки під конкретного користувача. Це дозволяє створити унікальний і релевантний досвід для кожного покупця, що, у свою чергу, підвищує ймовірність здійснення покупки, збільшує середній чек і покращує рівень задоволеності клієнтів. Персоналізація також сприяє зменшенню відтоку клієнтів, оскільки формує відчуття індивідуального підходу з боку бренду.

Згідно з дослідженням Netcore [32], 70% ритейлерів, які інвестували в персоналізацію клієнтського досвіду, звітують про рентабельність інвестицій, у обсязі більше ніж 400%. При цьому, 71% споживачів стверджують, робили б робили покупки частіше у компаній, що пропонують персоналізований досвід [33].

2. Чат-боти та віртуальні асистенти

Інтеграція інтелектуальних чат-ботів у сервіси електронної комерції дозволяє автоматизувати обслуговування клієнтів. Завдяки використанню технологій обробки природної мови (NLP), такі боти здатні розпізнавати інтенцію запитів користувачів, аналізувати контекст розмови та формувати релевантні відповіді. Чат-боти можуть виконувати низку функцій: надавати інформацію про товари, допомагати з навігацією сайтом, інформувати про статус замовлення, обробляти повернення та навіть виконувати роль консультантів. Перевагою цих систем є здатність працювати 24/7, зменшуючи навантаження на операторів контакт-центрів та прискорюючи процес обслуговування.

Згідно з результатами опитувань Intercom [34], 41% підприємств, що використовують чат-боти для продажів, спостерігали зростання продажів у середньому на 67%.

3. Прогнозування попиту та управління запасами

ШІ значно покращує процес прогнозування попиту, дозволяючи більш точно оцінити потреби споживачів на основі історичних даних, сезонних коливань, маркетингової активності та зовнішніх факторів (економічні тенденції, погодні умови, соціальні події). У результаті компанії можуть адаптувати свої складські стратегії, зменшуючи витрати на зберігання та запобігаючи дефіциту або надлишку продукції. Управління запасами за допомогою ШІ також включає автоматичне оновлення замовлень на поповнення, динамічне планування постачань та оптимізацію рівня товарних залишків на різних складах. Інтелектуальні системи здатні аналізувати дані в реальному часі та оперативно реагувати на зміни у поведінці споживачів чи на ринку загалом. Це особливо важливо в умовах швидкої ротації асортименту або в сегментах, де попит суттєво коливається.

За даними досліджень McKinsey [35], завдяки плануванню ланцюга постачань з використанням штучного інтелекту компанії мають змогу отримати зростання доходу до 4%, зменшення запасів до 20% і падіння витрат в ланцюзі постачання до 10%.

4. Динамічне ціноутворення

Алгоритми динамічного ціноутворення використовують машинне навчання для формування оптимальної ціни товару з урахуванням багатьох факторів, таких як попит, залишки на складі, поведінка користувача, активність конкурентів, день тижня чи час доби. Такий підхід дозволяє збільшити продажі під час високого попиту або стимулювати збут у періоди низької активності споживачів. ШІ-системи можуть оновлювати ціни в реальному часі, забезпечуючи максимальну адаптивність бізнесу до змін у зовнішньому середовищі. У результаті підвищується загальна прибутковість компанії та ефективність цінової стратегії.

За даними Boston Consulting Group [36], компанії, які використовували динамічне ціноутворення в електронній комерції, мали на 70% більше шансів досягти своїх стратегічних цілей у висококонкурентному середовищі. Загалом учасники дослідження отримали 100 мільйонів доларів додаткового прибутку завдяки інтеграції даної технології.

5. Виявлення шахрайства та забезпечення безпеки

ШІ значною мірою підвищує рівень безпеки електронної комерції. Використовуючи алгоритми виявлення аномалій, системи можуть ідентифікувати підозрілі транзакції, спроби несанкціонованого доступу або інші нетипові дії користувачів. Наприклад, якщо користувач здійснює покупку з нової геолокації або вводить невірні платіжні дані кілька разів поспіль, система може автоматично заблокувати транзакцію або вимагати додаткову автентифікацію. Такі рішення не лише зменшують ризики фінансових втрат, а й формують довіру клієнтів до платформи, що є критично важливим фактором у цифровому середовищі.

За результатами досліджень [37], організації, які використовують ШІ для виявлення шахрайства, повідомляють про 98% успіху у виявленні шахрайських дій, демонструючи ефективність технологій ШІ в цій сфері.

Таким чином, використання перелічених вище та інших технологій штучного інтелекту дозволяє компаніям електронної комерції підвищувати продуктивність та безпеку бізнесу, зменшувати операційні витрати, підвищувати рівень продажів за рахунок тагретованості маркетингових операцій та забезпечувати кращий досвід для клієнтів за рахунок дослідження їх поведінки та персоналізації.

Однак впровадження даних інструментів також супроводжуються рядом викликів, серед яких, зокрема:

1. Проблеми з конфіденційністю та захистом даних. Для ефективної роботи ШІ необхідна велика кількість персоналізованих даних, що створює ризики витоку інформації та викликає занепокоєння щодо дотримання законодавства про захист персональних даних.

2. Труднощі інтеграції в існуючі системи. Впровадження ШІ-технологій потребує суттєвої трансформації ІТ-інфраструктури, що може бути витратним та складним процесом, особливо для малого та середнього бізнесу.

3. Висока вартість впровадження. Розробка та підтримка ШІ-рішень потребує значних фінансових інвестицій, зокрема в інфраструктуру, програмне забезпечення та кваліфікований персонал.

4. Проблеми з інтерпретованістю моделей. Багато алгоритмів машинного навчання мають характер «чорної скриньки», що ускладнює розуміння механізмів прийняття рішень та може знижувати довіру з боку користувачів і менеджменту.

5. Етичні та соціальні питання. Автоматизація процесів може призводити до зменшення кількості робочих місць, що викликає соціальні ризики, а також породжує етичні дилеми щодо ролі машин у прийнятті рішень.

Так за результатами опитувань ритейлерів у США [38], 44 % генеральних директорів вважають, що безпека даних і конфіденційність є найбільшими проблемами при впровадженні рішень на основі штучного інтелекту в їхніх компаніях. Менеджери та співробітники були ще більше стурбовані цим аспектом (53 відсотки). Однак ще 43 відсотки співробітників назвали головною проблемою відсутність обізнаності, розуміння та/або досвіду використання інструментів ШІ. (рис. 1.6.)



Рис. 1.6. Результати опитування ритейлерів у США з приводу основних проблем у впровадженні штучного інтелекту у 2023 р.

Джерело: побудовано автором на основі [38]

1.3. Досвід використання рекомендаційних систем у світових компаніях.

В контексті змін споживчої поведінки та постійного зростання інформаційного навантаження все більшої ролі та поширення набуває використання рекомендаційних систем на основі інструментів штучного інтелекту для персоналізації користувацького досвіду та підвищення продажів за рахунок надання релевантних пропозицій користувачам. Такі технології вже широко та успішно використовуються великими компаніями.

Питання використання рекомендаційних систем, як data science технології, у компаніях піднімається у роботі Затоначької Т. та Фаренюк Я. [39], проте потребує подальшої деталізації та глибокого аналізу з точки зору використовуваних алгоритмів.

Однією з перших компаній, які звернули загальну увагу до цінності ШІ-інструментів у роздрібній торгівлі, стала Amazon, запорукою популярності якої великою мірою стали рекомендаційні системи. За даними McKinsey & Company [40], 35% доходу Amazon.com приносить його система рекомендацій.

На початку 2000-их компанія впровадила алгоритм спільної фільтрації «item-to-item» [41]. На відміну від традиційних на той момент методів, які зіставляли клієнтів із подібною історією покупок, цей підхід зосереджувався на виявленні кореляції між продуктами. Наприклад, якщо клієнти, які купили певну книгу, також часто купували певний набір предметів, ці товари будуть рекомендовані майбутнім клієнтам, які купили ту саму книгу.

Використання даної методики дозволило Amazon не лише надавати більш точні рекомендації, але й значно скоротити забезпечило значні обчислювальні переваги [41]. Так, щоб знайти групу клієнтів, історія покупок яких найбільше схожа на історію покупок певного відвідувача, необхідно порівняти історії покупок у всій базі даних клієнтів Amazon, що є занадто трудомістким та часозатратним процесом. Однак у середньому певний продукт, який продається в магазині Amazon, купує лише невелика частина клієнтів сайту. Це означає, що для перевірки історії останніх покупок кожного, хто купив певний товар, потрібно набагато менше пошуків, ніж для ідентифікації клієнтів, які найбільше схожі на даного відвідувача сайту.

Однак алгоритм мав і певні недоліки, зокрема, підхід не враховував того факту, що покупці, які купують більше, можуть спотворити рекомендації. Так, оскільки клієнт, який купує багато продуктів, ймовірніше купить А, ніж клієнт, який купує мало продуктів, покупці А в середньому купують більше ніж середньостатистичний клієнт Amazon. Але оскільки вони купують більше, вони також надзвичайно ймовірно купують В. Таким чином, дослідники Amazon прийшли до висновку, що недостатньо оцінити збільшену ймовірність покупки продукту В за умови покупки продукту А і натомість необхідно враховувати підвищену ймовірність покупки продукту В з будь-якою покупкою [41]. Тобто, вони дисконтували підвищену ймовірність купівлі В у великих покупців відповідно до ваги їхньої покупки.

Рекомендаційні системи також інтегровані і у стримінговий сервіс Amazon – Amazon Prime. У 2014 році команда Amazon Prime Video стала розробляти новий рекомендаційний алгоритм, експериментуючи з глибокими нейронними

мережами, зокрема автокодерами, щоб передбачити, які фільми користувач може переглянути наступним [41]. Спочатку ці моделі були менш ефективними порівняно з більш простими методами, такими як колаборативна фільтрація та списки бестселерів. Прорив стався, коли дослідники виявили, що в будь-який момент більшість кіноглядачів, ймовірно, віддадуть перевагу останнім релізам, а не забутій класиці у своїх улюблених жанрах [41]. Щоб врахувати даний патерн поведінки і таким чином підвищити точність рекомендацій, команда Amazon змінила підхід до проблеми: вони все ще використовували автокодер, але навчали його на даних перегляду фільмів, які були відсортовані в хронологічному порядку та з певної дати відсіювались [41].

Однією з найбільш масштабних і найдосконаліших промислових систем рекомендацій є рекомендаційна система Youtube, яка надає більш ніж мільярд користувачів персоналізований контент із постійно зростаючого масиву відео.

Система керується проектом глибокого навчання штучного інтелекту Google Brain та складається з двох нейронних мереж: одна для формування кандидатів, а інша для рейтингу [42, с. 191].

Мережа створення кандидатів використовує події з історії активності користувача на YouTube як вхідні дані та отримує невелику підмножину (сотні) відео з великого корпусу. Дана мережа забезпечує лише широку персоналізацію за допомогою колаборативної фільтрації, де подібність між користувачами виражається в таких загальних характеристиках, як ідентифікатори переглядів відео, маркери пошукових запитів і демографічні дані. [42, с. 191]

Представлення кількох «найкращих» рекомендацій у списку вимагає детального представлення, щоб виділити відносну важливість серед кандидатів із високим показником повноти. Мережа рейтингу виконує це завдання, призначаючи оцінку кожному відео відповідно до бажаної цільової функції, використовуючи багатий набір характеристик, що описують відео та користувача. Відео з найвищими балами представлені користувачеві, упорядковані за їхніми балами. [42, с. 191]

Двоетапний підхід до рекомендацій дозволяє Youtube давати рекомендації з дуже великого корпусу (мільйони) відео, залишаючись впевненими, що невелика кількість відео, які з'являються на пристрої, персоналізовані та цікаві для користувача.

Іншим прикладом використання рекомендаційних систем є Netflix. Після зміни бізнес-моделі на стрімінг компанія стала накопичувати великі масиви даних про клієнтів, що дозволило розробляти нові шляхи покращення продукту, зокрема, полегшуючи клієнтам пошук контенту. За даними McKinsey [40], 75% контенту, який користувачі переглядають на платформі, вони знаходять завдяки рекомендаціям. Рекомендаційна система Netflix включає в себе ряд алгоритмів, серед яких, зокрема [43, с. 2-5]:

1. Персоналізований рейтинг відео (PVR).

Зазвичай на кожній домашній сторінці є близько 40 рядків і до 75 відео на рядок; ці цифри дещо відрізняються на різних пристроях через апаратне забезпечення та міркування щодо взаємодії з користувачем. Відео в певному рядку зазвичай надходять за одним алгоритмом. Рядки жанрів керуються нашим алгоритмом персоналізованого рейтингу відео (PVR), що упорядковує весь каталог відео (або підмножин, вибраних за жанром або іншим фільтром) для кожного профілю користувача персоналізованим способом.

2. Рейтинг Top N відео.

Даний алгоритм створює рекомендації в рядку Top Picks і полягає в тому, щоб знайти кілька найкращих персоналізованих рекомендацій у всьому каталозі для кожного учасника. Рейтинг Top N і PVR мають схожі характеристики, наприклад, поєднання персоналізації з популярністю, а також визначення та включення тенденцій перегляду за різні часові вікна від дня до року. Проте, на відміну від PVR, що ранжує окремі підмножини каталогу, рейтинг Top N оптимізовано та оцінено за допомогою показників і алгоритмів, що спираються на верхні позиції усього каталогу.

3. «Популярні зараз».

Іншим важливим предиктором того, які відео переглядатимуть підписки сервісу є короткотермінові часові тенденції, що варіюються від кількох хвилин до, кількох днів, що в поєднанні з персоналізацією створює рейтинг тенденцій, який використовується для керування рядком «Популярні зараз». Цей рейтинг добре визначає два типи тенденцій: ті, що повторюються циклічно (наприклад, щороку), але мають короткостроковий ефект, коли виникають, наприклад, зростання перегляду романтичних відео під час Дня святого Валентина в Північній Америці, та одноразові, короткочасні події, наприклад, великий ураган, який наближається до густонаселеного району, висвітлюється багатьма ЗМІ, що викликає підвищений короткочасний інтерес до документальних фільмів і фільмів про урагани та інші стихійні лиха.

4. «Продовжити перегляд».

Враховуючи важливість епізодичного вмісту, який переглядається протягом кількох сеансів, а також свободу перегляду несерійного вмісту невеликими фрагментами, ще одним важливим алгоритмом ранжирування відео є упорядкування відео в рядку «Продовжити перегляд». Рейтинг продовження перегляду сортує підмножину нещодавно переглянутих елементів на основі оцінки того, чи має підписник намір продовжити перегляд чи переглянути відео повторно. Сигнали, які використовуються для цього, включають час, що минув з моменту перегляду, момент припинення (початок, середина чи кінець програми), чи переглядався після цього інший контент та використовувані пристрої.

5. «Тому що ви дивилися»

Іншим типом категоризації є рядки «Тому що ви дивилися», які прив'язують свої рекомендації до одного відео, яке переглянув підписник. Даними рядками керує неперсоналізований алгоритм подібності відео-відео, який обчислює ранжований список схожих відео для кожного відео в каталозі.

6. Генерація сторінки рекомендацій з відповідних рядків.

Відео, вибрані для кожного рядка, представляють оцінку найкращого вибору відео для показу конкретному користувачеві. Але більшість учасників мають різні настрої від сеансу до сеансу, і багато облікових записів використовуються кількома членами родини. Пропонуючи різноманітний вибір рядків, Netflix полегшує підписникам вибір відео, які підходять для часу, події чи члена сім'ї.

Алгоритм генерації сторінки використовує результати всіх уже описаних алгоритмів для створення кожної окремої сторінки рекомендацій, беручи до уваги релевантність кожного рядка для учасника, а також різноманітність сторінки. Типовий учасник має десятки тисяч рядків, які можуть розміститися на домашній сторінці, що ускладнює керування обчисленнями, необхідними для їх оцінки. У зв'язку з цим Netflix використовує повністю персоналізований математичний алгоритм, який може вибирати та впорядковувати рядки з великої групи кандидатів, щоб створити порядок, оптимізований з точки зору релевантності та різноманітності.

Даний набір алгоритмів дозволяє розширити розмір ефективного каталогу, підвищити кількість переглядів контенту та скоротити кількість клієнтів, що втрачаються через відписку від платформи. Загалом, ефект від персоналізації та рекомендацій допомагає компанії заощаджувати близько мільярда доларів на рік. [43, с. 6]

Іншою компанією, запорукою популярності якої стали рекомендаційні системи, є Spotify. Платформа збирає дані з кількох джерел: дані про взаємодії користувачів (відтворення, пропуски, плейлисти, пошукові запити, тощо); дані про контент, що включають детальну інформацію про кожну пісню: від жанру й темпу до тексту; контекстуальні дані, що описують де та коли відтворюється контент. Цей комплексний набір даних формує основу для ансамблю різноманітних методів і алгоритмів машинного навчання, що використовує компанія, серед яких, зокрема [44]:

- Колаборативна фільтрація.

Spotify використовує спільну фільтрацію як на основі користувачів, так і на основі елементів, пропонуючи як контент, який подається схожим користувачам, так і контент, який фігурує разом у плейлистах.

- Фільтрування на основі вмісту

Spotify використовує моделі глибокого навчання, такі як згорткові нейронні мережі (CNN) та рекурентні нейронні мережі (RNN), для ідентифікації шаблонів та особливостей (ритм, мелодія, тощо) у музиці на основі аналізу спектрограм та необробленого аудіо з метою розуміння унікального «відбитку» кожної пісні.

- Обробка природної мови (NLP)

Методи NLP застосовуються для аналізу назв плейлистів, текстів пісень і, потенційно, створеного користувачами контенту. Це може включати вкладання слів (наприклад, Word2Vec або GloVe) для фіксації семантичних зв'язків між текстовими дескрипторами.

- Глибоке навчання

Крім аудіоаналізу, моделі глибокого навчання можуть використовуватися для інтеграції різних типів даних (поведінка користувачів, аудіофункції, текстові дані) в уніфіковану модель рекомендацій. Тут можуть бути використані такі методи, як вкладання, механізми уваги та нейронні мережі графів.

- Навчання з підкріпленням

Щоб оптимізувати довгострокове залучення користувачів, Spotify використовує методи навчання підкріплення, з метою балансування рекомендацій нового різноманітного контенту та контенту подібного до того, що вже подобається користувачеві.

- Моделі з урахуванням часу

Враховуючи динамічний характер музичних уподобань, Spotify включає часову динаміку у свої моделі. Це може включати фактори спаду в часі у

колаборативній фільтрації або більш складні підходи, як-от періодичні мережі рекомендацій.

Система рекомендацій Spotify проявляється в кількох ключових функціях, кожна з яких унікально використовує вищезгадані технології [44]:

1. Тижнева добірка - персоналізований плейлист, що використовує комбінацію колаборативної фільтрації та підходів на основі вмісту, поєднуючи історію прослуховування користувача разом із світовими музичними трендами.

2. Release Radar – функція, що балансує уподобаннями користувача та актуальність завдяки окремій моделі, навченій прогнозувати прихильність користувача до нових релізів на основі історії його взаємодій з подібними виконавцями та жанрами.

3. Мікс дня – окремі плейлисти, що створюються завдяки сегментації смаків користувачів за допомогою таких алгоритмів, як кластеризація k-середніх, або більш просунутих методи, як моделі суміші Гауса, застосовані до шаблонів прослуховування користувача.

4. Радіо - функція рекомендацій на основі сеансу, що використовує послідовні моделі, або LSTM, щоб підтримувати контекст і узгодженість протягом сеансу прослуховування.

5. «Фанам також подобається» - функція рекомендації виконавця, шляхом застосування колаборативної фільтрації за елементами, можливо, розширене за допомогою алгоритмів на основі графів для фіксації складних взаємозв'язків у мережі виконавець-слухач.

Для технічної реалізації системи рекомендацій Spotify використовує лямбда-архітектуру, що дозволяє як пакетну обробку історичних даних, так і потокову обробку взаємодії користувачів, що дає їм змогу збалансувати потребу у всебічному аналізі з потребою в негайних, оперативних рекомендаціях. Щоб керувати величезною кількістю даних і надавати швидкі відповіді, Spotify використовує комбінацію баз даних у пам'яті та розподілених кешів. Це дозволяє їм попередньо обчислювати певні рекомендації та швидко їх отримувати,

водночас допускаючи коригування в реальному часі на основі негайних дій користувача. [44]

Висновки до розділу 1

Проведений аналіз динаміки розвитку глобального ринку електронної комерції засвідчив його стабільне зростання, цифрову трансформацію та зростаючі вимоги споживачів до якості сервісу. В умовах високої конкуренції компанії змушені впроваджувати інноваційні інструменти для утримання клієнтів і підвищення ефективності продажів. Одним з таких інструментів виступають технології штучного інтелекту, що забезпечують інтелектуальну обробку великих масивів даних, автоматизацію бізнес-процесів і, зокрема, побудову ефективних рекомендаційних систем. Проаналізований успішний досвід використання рекомендаційних систем у провідних світових компаніях демонструє їхню здатність підвищувати рівень персоналізації, позиттєвої цінності та задоволеності клієнтів і, як наслідок, доходів компанії.

Отримані результати свідчать про стратегічну важливість рекомендаційних систем для маркетплейсів як одного з найдинамічніших сегментів електронної комерції. З урахуванням цього, вибір теми дослідження є обґрунтованим, актуальним і відповідає потребам сучасного цифрового бізнесу. Необхідність поглибленого дослідження в напрямку розробки рекомендаційних систем полягає у потребі створення більш точних, адаптивних та контекстно чутливих алгоритмів, здатних ефективно працювати в умовах обмежених або неоднорідних даних, характерних для багатьох маркетплейсів. Це відкриває перспективи для подальших наукових розробок, спрямованих на підвищення якості персоналізованих сервісів з урахуванням реальних викликів ринку.

РОЗДІЛ 2. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ РОЗРОБКИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ

2.1. Етапи рекомендаційного процесу.

В узагальненому вигляді життєвий цикл рекомендації можна представити у вигляді петлі зворотного зв'язку між трьома ключовими компонентами: користувачем, даними та моделлю. Як показано на рис. 2.1. , цикл зворотного зв'язку складається з трьох етапів:

1. Збір даних (Користувач→Дані).

На цьому етапі система акумулює дані про взаємодію користувачів із платформою. Основним джерелом таких даних є користувацький фідбек, який поділяється на три основні типи:

- Явний зворотній зв'язок - це прямі оцінки користувачів, які свідомо виражають свою думку щодо контенту. Прикладами є виставлення зіркових оцінок (напр., 1-5 зірок), вподобання, написання відгуків або відповідей на опитування. Явний фідбек є високоточним, однак обмежений за обсягом, оскільки потребує активної участі користувача.

- Неявний зворотній зв'язок - включає інформацію, яку система отримує шляхом спостереження за поведінкою користувача без прямої взаємодії. Це може бути час перегляду контенту, частота кліків, прокрутка сторінки, завантаження файлів, додавання товару в кошик або відмова від покупки. Неявний фідбек характеризується великою кількістю даних, однак містить вищий рівень шуму і неоднозначності, адже не завжди чітко відображає вподобання користувача.

- Гібридний зворотній зв'язок - поєднує елементи як явного, так і неявного фідбеку, дозволяючи компенсувати слабкі сторони кожного окремо. Наприклад, система може враховувати як оцінку фільму, так і тривалість його перегляду, або об'єднувати реакцію користувача (лайк/дизлайк) із даними про

частоту взаємодії з подібним контентом. Гібридний підхід підвищує надійність профілю користувача, зменшуючи залежність від обмежень одного типу фідбеку.

Окрім користувацького фідбеку, збирається також мета-інформація про контент (жанр, теги, авторство, тематика) та контекстуальні фактори (час доби, пристрій, геолокація), що дозволяє сформувати більш повну картину взаємодії користувача з системою.

2. Навчання (Дані→Система)

Зібрані дані використовуються для побудови або оновлення моделі рекомендацій. Цей етап може охоплювати різноманітні підходи: від класичних методів, як-от колаборативна фільтрація або контентно-орієнтовані методи, до сучасних гібридних чи глибоких нейронних мереж. Навчання моделі здійснюється з метою виявлення прихованих патернів у поведінці користувачів і встановлення залежностей між користувачами та об'єктами. У процесі навчання модель оптимізується за допомогою функцій втрат і методів регуляризації, щоб забезпечити узагальнення на нові, ще не бачені приклади.

3. Подача рекомендацій (Система→Користувач)

На завершальному етапі система генерує список персоналізованих рекомендацій для кожного користувача на основі натренованої моделі. Це може відбуватись у режимі реального часу або періодично. Зокрема рекомендації може бути подано у вигляді товарів на головній сторінці («Рекомендовано для вас»), персоналізованих результатів пошуку, динамічного контенту на сторінці товару («Клієнти також купували»), та електронних листів або push-сповіщень з персональними добірками. Ефективність подачі залежить від швидкодії системи, можливостей масштабування та точності ранжування об'єктів. У деяких випадках рекомендації додатково адаптуються до поточного контексту користувача, враховуючи його поточну сесію, місце перебування чи попередні дії. Результати цього етапу знову потрапляють у систему як нові дані, завершуючи цикл зворотного зв'язку.

Через цей цикл користувачі та рекомендаційна система перебувають у процесі взаємної динамічної еволюції, де особисті інтереси та поведінка користувачів оновлюються завдяки рекомендаціям, а рекомендаційна система може прийти до самопідсилення моделі, використовуючи оновлені дані.

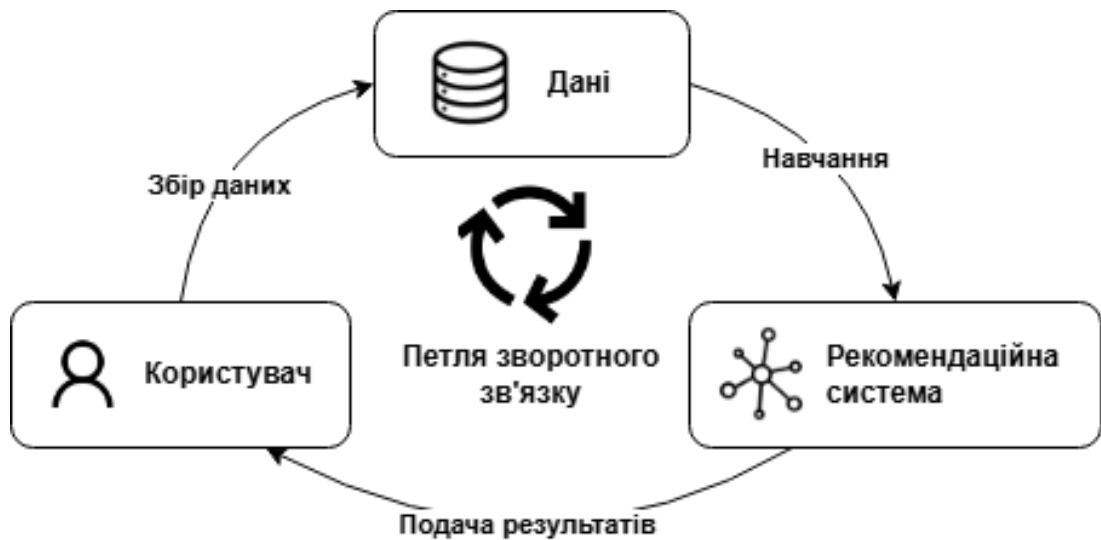


Рис. 2.1. Петля зворотного зв'язку у рекомендаційних системах.

2.2. Методологічні підходи до розробки рекомендаційних систем.

Підходи до побудови рекомендаційних систем є темою багатьох досліджень. Зокрема, огляди існуючих методик запропоновані у роботах Ісінкайє Ф., Фоладжмі Є. та Оджох Б. [45]; Рагуванші С. та Патерія Р. [46]; Фаяз З., Ебрахіміан М. та ін [47]. Проте, погляди науковців на виділення основних методик та аспектів розробки рекомендаційних систем відрізняються та потребують подальшого вивчення, узагальнення та систематизації.

В основі рекомендаційних систем лежать рекомендаційні алгоритми, які прийнято поділяти на три основних типи (рис. 2.2.): алгоритми на основі контенту, колаборативна фільтрація та гібридні підходи.

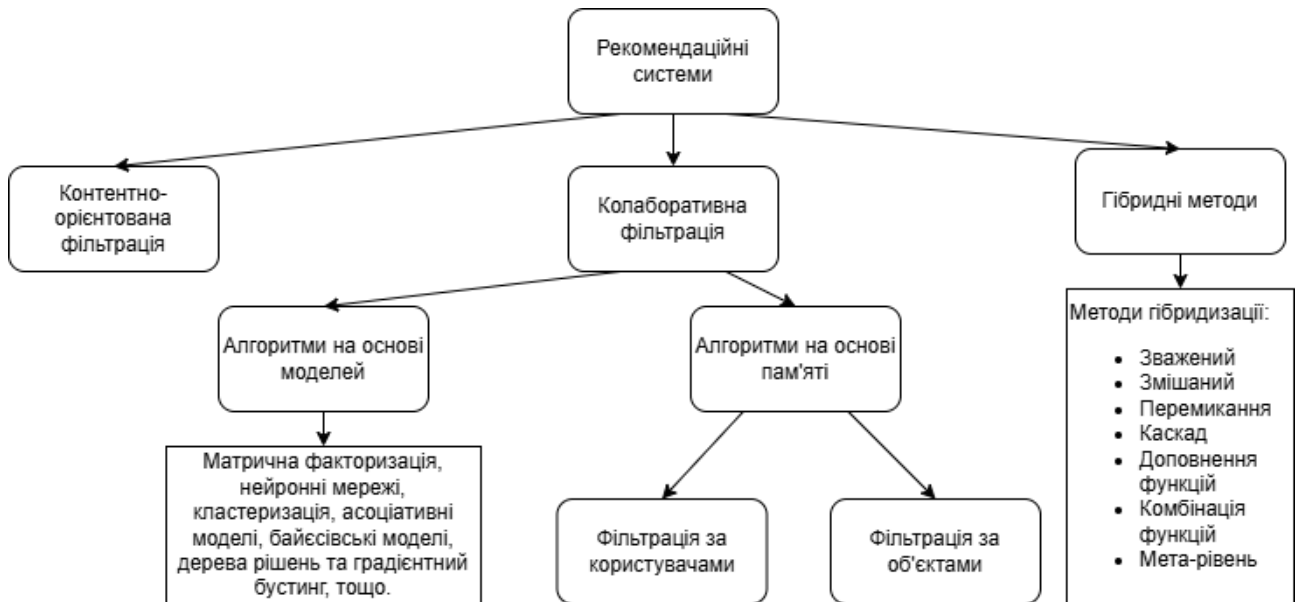


Рис. 2.2. Класифікація рекомендаційних алгоритмів

Алгоритми на основі контенту фокусуються на атрибутах елементів для побудови прогнозів вподобань споживачів. Основна ідея даних алгоритмів полягає в тому, що користувачеві слід рекомендувати об'єкти, схожі на ті, які він оцінював позитивно. Для цього система створює профіль користувача, що включає опис його уподобань на основі атрибутів товарів з якими він взаємодівав, і підбирає об'єкти, схожі за характеристиками до тих, які йому вже сподобались. Даний підхід є найбільш ефективним у ситуаціях, коли немає змоги авторизувати користувача, щоб отримати інформацію про нього, адже дозволяє сформувати тимчасовий профіль. Для того, щоб витягнути з описів товарів їх ключові атрибути з відповідною їм вагою використовуються алгоритми представлення елементів. Одним з найпоширеніших серед них є TF-IDF.

Нехай, N – загальна кількість об'єктів, які можуть бути рекомендовані користувачу, ключове слово k_j фігурує в n_i об'єктах, а $f_{i,j}$ – кількість входжень цього слова в об'єкт d_j . Тоді $TF_{i,j}$ (term frequency) – це відношення числа входжень тега до загальної кількості тегів об'єкта, тобто [48, с. 16]:

$$TF_{i,j} = \frac{f_{i,j}}{\max_z f_{z,j}} \quad (2.1)$$

Проте враховуючи виключно частоту входження тега може виникнути ситуація, коли у більшості об'єктів максимальна вага буде у найбільш поширених тегів, що може створювати викривлення інформації про вподобання. Для уникнення даної ситуації використовується IDF_i (inverse document frequency) – величина, зворотна частоті входження тега в об'єкт колекції, що визначається як [48, с. 16]:

$$IDF_i = \log \frac{N}{n_i} \quad (2.2)$$

Сукупність об'єктів, які користувач оцінив раніше, утворює профіль користувача ContentBasedProfile (u), що представляє собою вектор ваг ($w_{u,1}, \dots, w_{u,k}$), де кожна вага $w_{u,i}$ визначає важливість тега k_i для користувача u . Отже, ContentBasedProfile (u) і Content (s) можна уявити як TF-IDF вектори \vec{w}_u і \vec{w}_s , при цьому функція задоволеності користувача $h(u, s)$ може бути представлена як косинусний коефіцієнт векторів \vec{w}_u і \vec{w}_s , де K – спільна кількість тегів в системі [48, с. 17]:

$$h(u, s) = \cos(\vec{w}_u, \vec{w}_s) = \frac{\vec{w}_u \vec{w}_s}{\|\vec{w}_u\| \|\vec{w}_s\|} = \frac{\sum_{i=1}^K w_{i,u} w_{i,s}}{\sqrt{\sum_{i=1}^K w_{i,u}^2} \sqrt{\sum_{i=1}^K w_{i,s}^2}} \quad (2.3)$$

Окрім методів інформаційного пошуку, для фільтрації на основі контенту також можуть використовуватися складніші методи, в тому числі на основі машинного навчання, такі як байєсівські класифікатори, кластерний аналіз, дерева рішень та нейронні мережі.

Перевагами фільтрації на основі контенту є незалежність від профілів інших користувачів та можливість налаштування рекомендацій на основі даних за короткий період часу, що дозволяє надавати рекомендації навіть новим користувачам. Проте, дані алгоритми можуть бути певною мірою обмежені у відкритті нових патернів, а також потребують глибоких знань про об'єкти у вигляді розгорнутих описів характеристик товару.

Іншим підходом до розробки рекомендаційних систем є колаборативна фільтрація, яка ґрунтується на аналізі взаємодії багатьох користувачів із системою. Ідея полягає в тому, що користувачам, які мали схожі оцінки в минулому, ймовірно, сподобаються подібні об'єкти в майбутньому. Колаборативна фільтрація має кілька важливих переваг порівняно з фільтрацією на основі контенту, оскільки вона може працювати в ситуаціях відсутності достатньої кількості інформації про об'єкти, а також може виявляти неочевидні патерни, що не відображені у профілі користувача, і можуть бути пов'язані, наприклад, з його світоглядом. Алгоритми колаборативної фільтрації можна поділити на дві основні групи: алгоритми на основі моделей та алгоритми на основі пам'яті.

Алгоритми на основі пам'яті напряду використовують базу оцінок або взаємодій між користувачами й об'єктами для побудови рекомендацій. Даний вид алгоритмів поділяється на дві групи:

1. Фільтрація за користувачами, в межах якої схожість між користувачами оцінюється шляхом порівняння їхніх оцінок тих самих об'єктів, після чого обчислюється прогнозована оцінка об'єкта цільовим користувачем як середня оцінка об'єкта користувачами, подібними до цільового користувача, зважена на подібність цих користувачів із цільовим. Загальна формула даного алгоритму має наступний вигляд:

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{v \in N_u(i)} w(u,v) \cdot (r_{v,i} - \bar{r}_v)}{\sum_{v \in N_u(i)} |w(u,v)|} \quad (2.4)$$

де $\hat{r}_{u,i}$ - прогнозована оцінка цільового користувача для об'єкта i ;

$r_{v,i}$ - оцінка користувача v для об'єкта i ;

\bar{r} - середня оцінка користувача;

$w(u,v)$ - подібність між користувачами u і v ;

$N_u(i)$ - множина користувачів, подібних до u , які оцінили об'єкт i .

2. Фільтрація за об'єктами, в межах якої прогнози формуються на основі подібності між елементами, а не подібності між користувачами. Фільтрація за об'єктами передбачає побудову моделі, що на основі минулих взаємодій клієнтів ідентифікує схожість об'єктів, після чого для набору об'єктів з якими взаємодіяв цільовий користувач підбирає набір подібних товарів. Загальна формула даного алгоритму має наступний вигляд:

$$\hat{r}_{u,i} = \frac{\sum_{j \in N_i(u)} w(i, j) \cdot r_{u,j}}{\sum_{j \in N_i(u)} |w(i, j)|} \quad (2.5)$$

де $r_{u,j}$ - оцінка цільового користувача для подібного об'єкта j ;

$w(i, j)$ - подібність між об'єктами i та j ;

$N_i(u)$ - множина об'єктів, подібних до i , які користувач u вже оцінив.

Найбільш популярними мірами схожості у даних алгоритмах є кореляція Пірсона та косинусна подібність. Кореляція Пірсона вимірює лінійну залежність між двома векторами (наприклад, оцінки двох користувачів) і обчислюється за формулою:

$$w(u, v) = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{v,i} - \bar{r}_v)^2}} \quad (2.6)$$

де u, v - два користувачі;

$r_{u,i}$ - оцінка користувача u для елемента i ;

\bar{r} - середнє значення оцінок користувача u ;

I_{uv} - множина елементів, які оцінили обидва користувачі.

В той час як косинусна подібність вимірює косинус кута між двома векторами оцінок і обчислюється як:

$$w(u, v) = \frac{\sum_{i \in I_{uv}} r_{u,i} \cdot r_{v,i}}{\sqrt{\sum_{i \in I_{uv}} r_{u,i}^2} \sqrt{\sum_{i \in I_{uv}} r_{v,i}^2}} \quad (2.7)$$

Проте, варто зазначити, що з алгоритмами колаборативної фільтрації також пов'язний ряд проблем, серед яких:

1. Проблема холодного старту, яка полягає в тому, що через настану даних про нових користувачів, або продукти, такі системи не можуть робити точні рекомендації з їх приводу.

2. Проблема розрідженості даних, яка виникає через те, що велика кількість користувачів оцінює лише невелику кількість товарів, у зв'язку з чим навіть найпопулярніші товари мають відносно невелику кількість оцінок і матриця користувач-продукт є дуже розрідженою, що знижує якість рекомендацій.

3. Проблема масштабованості, яка пов'язана з тим, що при великій кількості користувачів і об'єктів обчислювальні витрати зростають.

Алгоритми на основі моделей передбачають прогнозування вподобань користувачів за допомогою моделей навчених на минулій історії оцінок. Моделі, що можуть використовуватися для колаборативної фільтрації на основі моделей, можуть включати моделі матричної факторизації, байєсівські моделі, нейронні мережі, моделі кластеризації, асоціативні моделі, дерева рішень та градієнтний бустинг, тощо.

Найпоширенішим типом рекомендаційних алгоритмів є гібридні алгоритми, що комбінують кілька підходів (наприклад, поєднання колаборативної фільтрації та фільтрації на основі вмісту) для усунення обмежень окремих методів. Наприклад, система може спочатку оцінити подібність об'єктів за вмістом, а потім скоригувати рекомендації на основі схожості з діями інших користувачів. Існують різні стратегії гібридизації, зокрема:

1. Зважена гібридизація, що передбачає паралельну реалізацію кількох рекомендаційних алгоритмів, кожен з яких генерує оцінки або рейтинги для

елементів, а потім результати об'єднуються у фінальний список рекомендацій із використанням вагових коефіцієнтів.

2. Змішана гібридизація, при якій рекомендації з кількох джерел об'єднуються в один список без обчислення середніх оцінок.

3. Метод перемикавання, який полягає у тому, що система вибирає один метод з декількох доступних для генерації рекомендацій у конкретній ситуації. Вибір залежить від контексту, даних, користувача чи якості інформації. Так, наприклад, для нових користувачів може використовуватися фільтрація на основі вмісту, а для давно активних – колаборативна фільтрація.

4. Каскадна гібридизація, за якої алгоритми працюють послідовно: один формує список кандидатів, а наступний - їх ранжує, фільтрує або уточнює.

5. Доповнення функцій, яке передбачає, що один алгоритм створює додаткові характеристики, які потім використовуються іншим методом як вхідні дані.

6. Комбінація функцій, коли всі дані з різних джерел об'єднуються в один вектор ознак, який подається до одного навчального алгоритму (наприклад, дерева рішень, нейромережі).

7. Мета-рівень, за якого один алгоритм використовується для побудови моделі (знань), яка потім застосовується іншим алгоритмом. Це глибша інтеграція, ніж просто доповнення ознак.

2.3. Метрики оцінювання рекомендаційних алгоритмів

Оцінювання якості рекомендаційних систем є критичним етапом їх розробки, що дозволяє визначити ефективність алгоритмів у передбаченні вподобань користувачів. Для цього використовується низка метрик, кожна з яких вимірює різні аспекти продуктивності рекомендацій. Основні метрики оцінювання якості рекомендаційних систем можна умовно поділити на три категорії (рис. 2.3.): метрики точності; метрики покриття, різноманітності та новизни; а також метрики задоволеності користувачів.



Рис. 2.3. Класифікація метрик оцінювання рекомендаційних алгоритмів

Метрики точності оцінюють, наскільки добре система передбачає вподобання користувача та включають в себе метрики статистичної точності, точності підтримки прийняття рішень та точність ранжування.

Статистичні показники точності оцінюють точність рекомендаційного алгоритму шляхом порівняння прогнозованих оцінок безпосередньо з фактичною оцінкою користувача та включають, зокрема такі показники середня абсолютна похибка (MAE) та середньоквадратична похибка (RMSE). MAE - це міра відхилення рекомендації від значення, заданого користувачем, яка обчислюється наступним чином:

$$MAE = \frac{1}{N} \sum_{u,i} |p_{u,i} - r_{u,i}| \quad (2.8)$$

Тут $p_{u,i}$ - це прогнозований рейтинг користувача u для елемента i , $r_{u,i}$ - фактичний рейтинг, а N - загальна кількість оцінок для набору елементів. Чим нижчий показник MAE, тим точніше система рекомендацій прогнозує оцінки користувачів. Середньоквадратична помилка (RMSE) обчислюється як:

$$RMSE = \sqrt{\frac{1}{N} \sum_{u,i} (p_{u,i} - r_{u,i})^2} \quad (2.9)$$

RMSE приділяє більше уваги великим похибкам, і, аналогічно до MAE, чим нижче RMSE, тим вища точність рекомендацій.

Метрики точності підтримки прийняття рішень застосовуються для бінарної оцінки релевантності рекомендацій. Precision визначає частку релевантних об'єктів серед рекомендованих, тоді як Recall - частку рекомендованих об'єктів серед усіх релевантних. F1-score за своєю суттю є комбінацією цим двох показників, що обчислюється як їх гармонічне середнє.

При оцінюванні точності, вкрай важливим є розміщення найбільш релевантних об'єктів у верхніх позиціях списку рекомендацій, з чим пов'язана наявність ряду метрик для вимірювання того, чи справді релевантні рекомендації знаходяться на початку списку. Однією з таких метрик є середня точність (Mean Average Precision, MAP), яка є узагальненням точності для випадків, коли кожному користувачу надається список з декількох рекомендованих об'єктів. Вона обчислюється як середнє значення Average Precision (AP) для всіх користувачів, де AP - це середнє значення точності на кожній позиції, де з'являється релевантний об'єкт у списку:

$$MAP = \frac{1}{|U|} \sum_{u \in U} \frac{1}{|Q|} \sum_{k=1}^K P @ k \cdot rel_u(k) \quad (2.10)$$

де $P @ k$ - Precision на позиції k ;

$rel_u(k) = 1$, якщо елемент на позиції k релевантний, інакше 0;

U – множина користувачів;

Q - множина дійсно релевантних елементів для u .

Середнє рангове місце (Mean Reciprocal Rank, MRR) використовується в задачах, де важливо, на якому місці серед списку рекомендацій знаходиться перший релевантний об'єкт. Міра визначається як середнє значення оберненого рангу першого релевантного елемента:

$$MRR = \frac{1}{|U|} \sum_{u \in U} \frac{1}{rank_u} \quad (2.11)$$

де $rank_u$ - позиція першого релевантного об'єкта у списку рекомендацій для користувача u .

Нормалізований дисконтований кумулятивний прибуток (NDCG) враховує не лише релевантність, а й позицію елементів у списку. Чим вище релевантний об'єкт розташовано в списку, тим більше його внесок у загальну метрику:

$$NDCG @ K = \frac{DCG @ K}{IDCG @ K}, \quad DCG @ K = \sum_{i=1}^K \frac{rel_i}{\log_2(i+1)} \quad (2.12)$$

де $DCG@K$ – дисконтований кумулятивний прибуток на позиції K ;

$IDCG@K$ – DCG для ідеального (відсортованого за релевантністю) списку;

rel_i – релевантність об'єкту на позиції i .

Окрім точності та ранжування, важливими аспектами є широта покриття (coverage), різноманіття рекомендацій (diversity), новизна (novelty) та випадковість (serendipity). Coverage визначає, яка частка доступних елементів у системі може бути рекомендована принаймні одному користувачеві (item coverage) або яка частка користувачів отримує хоча б одну рекомендацію (user coverage). Diversity вимірює відмінність між рекомендованими елементами; чим менш подібні вони між собою, тим вища різноманітність. Novelty показує, наскільки рекомендовані об'єкти є новими або несподіваними для користувача, тобто такими, які він ще не бачив. Serendipity оцінює здатність системи пропонувати корисні, але неочікувані рекомендації, які користувач інакше не знайшов би.

Також про якість рекомендаційних систем можуть свідчити показники задоволеності клієнтів такі як click-through rate (частота кліків на рекомендовані елементи), conversion rate (частка рекомендацій, що призвели до цільової дії, або результати опитувань користувачів).

Висновки до розділу 2

У даному розділі було визначено основні етапи рекомендаційного процесу, які включають збір даних, навчання та подачу рекомендацій; проаналізовано методологічні підходи до створення рекомендаційних систем, такі як колаборативна фільтрація, фільтрація за вмістом та гібридні моделі, які забезпечують гнучкість та адаптивність алгоритмів до конкретних задач; а також розглянуто ключові метрики оцінювання ефективності рекомендаційних алгоритмів, такі як RMSE, MAE, точність, повнота, F1-міра та NDCG, що є основою для порівняльного аналізу моделей.

Отримані теоретичні засади стали фундаментом для розробки комплексу прикладних моделей в наступному розділі.

РОЗДІЛ 3. РОЗРОБКА РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ДЛЯ МАРКЕТПЛЕЙСУ

3.1. Інформаційне забезпечення: опис, структура та джерела даних

Дана робота базується на базі даних отриманих з сайту Sephora [49]. Sephora - це французька багатонаціональна мережа роздрібної торгівлі засобами особистої гігієни та косметики, що була заснована у 1969 році. На сьогодні асортимент товарів даної компанії налічує близько 340 брендів та більш ніж 45 000 продуктів. За грудень 2024 року сайт компанії було відвідано більше ніж 63 мільйони разів, а прибутки від продажів через онлайн платформу склали більше 87 мільйонів доларів з показником конверсії 1,5 – 2,0% і середньою вартістю замовлення 75-100\$ [50].

База даних використана для побудови рекомендаційного алгоритму містить близько мільйону відгуків про 2000 продуктів, та такою інформацією як текст відгуку, рейтинг, назва продукту, ідентифікатор користувача та ін. Повна інформація про базу даних наведена у табл. 3.1.

Таблиця 3.1

Структура та опис даних

Колонка	Опис
author_id	Унікальний ідентифікатор автора відгуку
Rating	Оцінка товару по п'ятибальній шкалі
is_recommended	Чи рекомендує користувач даний продукт?
Helpfulness	Відношення позитивних оцінок відгуку до усіх оцінок відгуку
submission_time	Дата публікації відгуку
review_text	Текст відгуку
review_title	Заголовок відгуку
skin_tone	Тон шкіри автора

Колонка	Опис
eye_color	Колір очей автора
skin_type	Тип шкіри автора
hair_color	Колір волосся автора
product_id	Ідентифікатор продукту
product_name	Назва продукту
brand_name	Назва бренду
price_usd	Ціна (доларів США)
Highlights	Список тегів, що описують основні атрибути товару
primary_category	Первинна категорія товару
secondary_category	Вторинна категорія товару
tertiary_category	Третинна категорія товару

При дослідженні вихідних даних варто звернути увагу на те, що у даних, які лежать в основі рекомендаційних систем, може виникати упередження вибору. Упередження вибору виникає внаслідок того, що користувачі вільно обирають, які товари оцінювати, а отже їх оцінки не репрезентують усі оцінки, оскільки частина даних відсутня не випадковим чином [51, с. 6]. Наявність даного явища у досліджуваному датасеті підтверджується загальним розподілом оцінок та розподілом середніх оцінок користувачів, що представлені на рис. 3.1. та рис. 3.2. відповідно. Так на основі даних розподілів можна стверджувати, що користувачі зазвичай схильні надавати фідбек на товари, які їм сподобались. У той самий час може також мати місце ситуація, коли користувач через власні принципи не оцінює максимальною оцінкою навіть ідеальний товар.

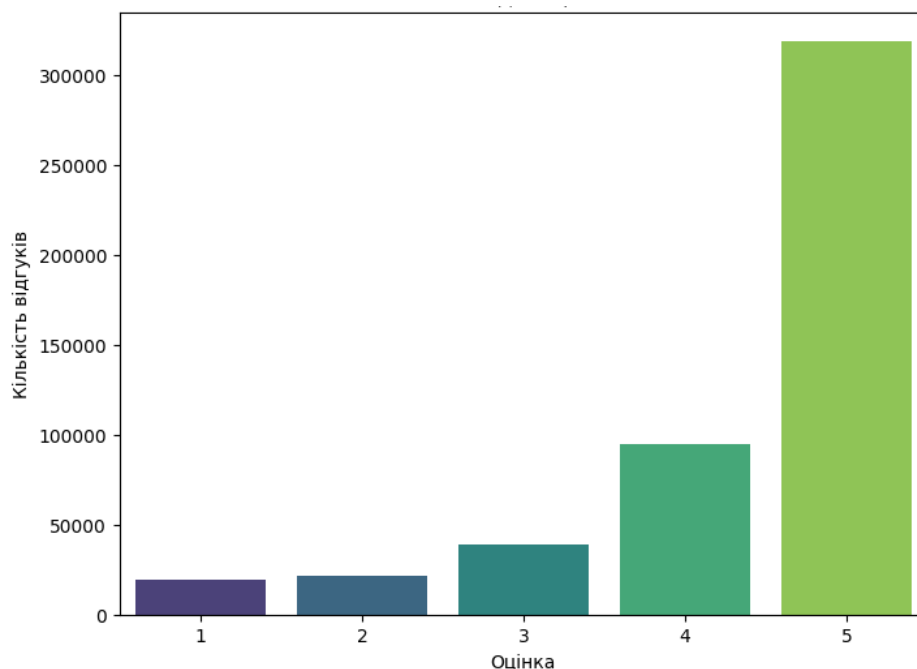


Рис. 3.1 . Розподіл оцінок товарів

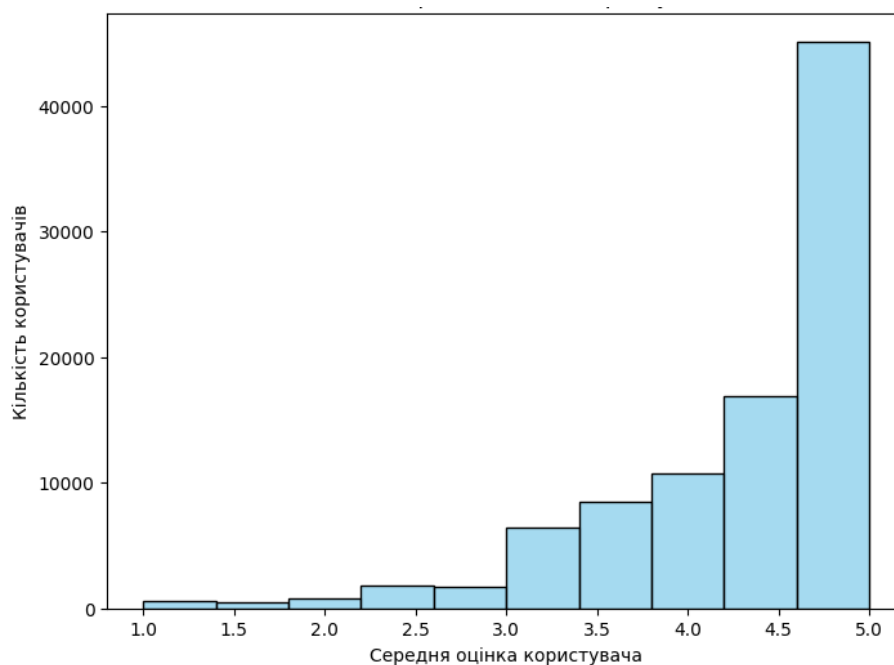


Рис. 3.2. Розподіл середніх оцінок користувачів

Іншим типом упереджень в даних, що може виникати, є упередження популярності, яке виникає через те, що на невелику частку популярних елементів припадає велика кількість взаємодій користувачів (рис. 3.3.). При навчанні на таких даних модель зазвичай дає вищі оцінки популярним елементам, ніж їхні ідеальні значення, тоді як непопулярним елементам просто пророкує негативні

оцінки. В результаті, популярні продукти рекомендуються навіть частіше, ніж їхня початкова популярність, відображена в наборі даних [51, с. 10].

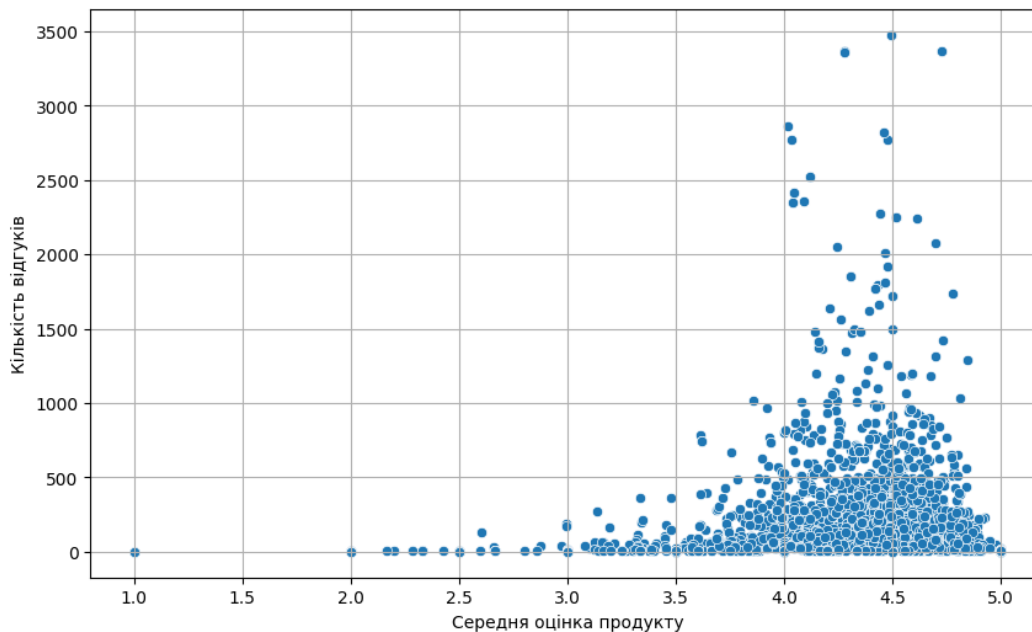


Рис. 3.3. Кількість відгуків та середня оцінка на продукти

Виходячи з зазначеного вище при подальшій реалізації рекомендаційного алгоритму є необхідність у нормалізації оцінок.

3.2. Аналіз настроїв на основі класифікаційних моделей машинного навчання

Одним із важливих попередніх кроків у розробці ефективної рекомендаційної системи є аналіз настроїв, що дозволяє глибше зрозуміти емоційне ставлення користувачів до певного контенту, товарів або послуг. На відміну від традиційних підходів, які базуються переважно на кількісних оцінках (наприклад, вподобаннях чи числових рейтингах), аналіз настроїв дає змогу витягувати корисну інформацію з текстових відгуків, коментарів і публікацій у соціальних мережах.

Проведення аналізу настроїв як попереднього етапу забезпечує кілька ключових переваг. По-перше, це сприяє збагаченню вхідних даних для моделі рекомендації, що підвищує її точність. Наприклад, два товари можуть мати однаковий середній рейтинг, але суттєво відрізнитися за якістю відгуків: один

може викликати переважно позитивні емоції, тоді як інший - нейтральні або негативні. По-друге, сентимент-аналіз дає змогу виявити латентні характеристики продукту або контенту, які не завжди відображаються в рейтингах, проте мають суттєвий вплив на вибір користувача. По-третє, врахування емоційної компоненти сприяє персоналізації рекомендацій: система може адаптувати пропозиції не лише на основі взаємодій, але й з урахуванням індивідуального емоційного профілю користувача.

В рамках даної роботи, основною метою проведення аналізу сентиментів є створити модель, яка на основі тексту відгуку передбачає, чи є відгук позитивним, нейтральним чи негативним. Для цього відгуки з рейтингом 4-5 приймаються як позитивні, з рейтингом 3 – як нейтральні і з рейтингом 2 та нижче – як негативні.

Першим кроком проведення аналізу є попередня обробка тексту, що включає об'єднання заголовку та тексту кожного відгуку у єдиний текстовий блок та очистку даних для зменшення «шуму» у тексті. Для очищення даних текстові блоки переводяться у нижній регістр, після чого з них видаляються спецсимволи, числа, пунктуація та стоп-слова. Стоп-слова - це загальноживані слова, які зазвичай несуть мало смислового навантаження (наприклад, артиклі та сполучники). Після цього проводиться лематизація – процес зведення слів до їх базової словникової форми (леми). Дані маніпуляції дозволяють скоротити обсяг даних для аналізу та зосередитись на більш значущих словах, що підвищує точність та ефективність обробки тексту.

```
def clean_text(text):
    text = str(text).lower()
    text = re.sub(r'^a-z\s', '', text)
    words = text.split()
    words = [lemmatizer.lemmatize(word) for word in words if word not in stop_words]
    return ' '.join(words)
```

Рис. 3.4. Функція для програмної реалізації очистки тексту

Отримані текстові блоки виступають у якості вхідних даних моделей, а вихідними даними є мітки, що відображають емоційне забарвлення відгуку. Ці дані розбиваються на навчальну та тестову вибірку у відношення 4:1 зі збереженням часток кожного класу. Для подання отриманих вибірок даних у моделі машинного навчання їх необхідно попередньо перевести у числові вектори. Для цього можна скористатися описаним у минулому розділі алгоритмом TF-IDF, який обчислює важливість слова у відгуку, з урахуванням того, як часто це слово з'являється в усіх блоках тексту, в результаті чого кожен текст перетворюється на вектор із чисел (в нашому випадку розміром до 10 тис.).

```
from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer(max_features=10000)
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)
```

Рис. 3.5. Програмна реалізація перетворення текстових блоків на числові вектори методом TF-IDF

В межах даної роботи для класифікації використовувались такі моделі як логістична регресія, мультиноміальний наївний байєсівський класифікатор, класифікатор Ridge, пасивно-агресивний класифікатор та LightGBM. Основним критерієм вибору методів класифікації була їх масштабованість. Програмну реалізацію методів класифікації подано в Додатку А.

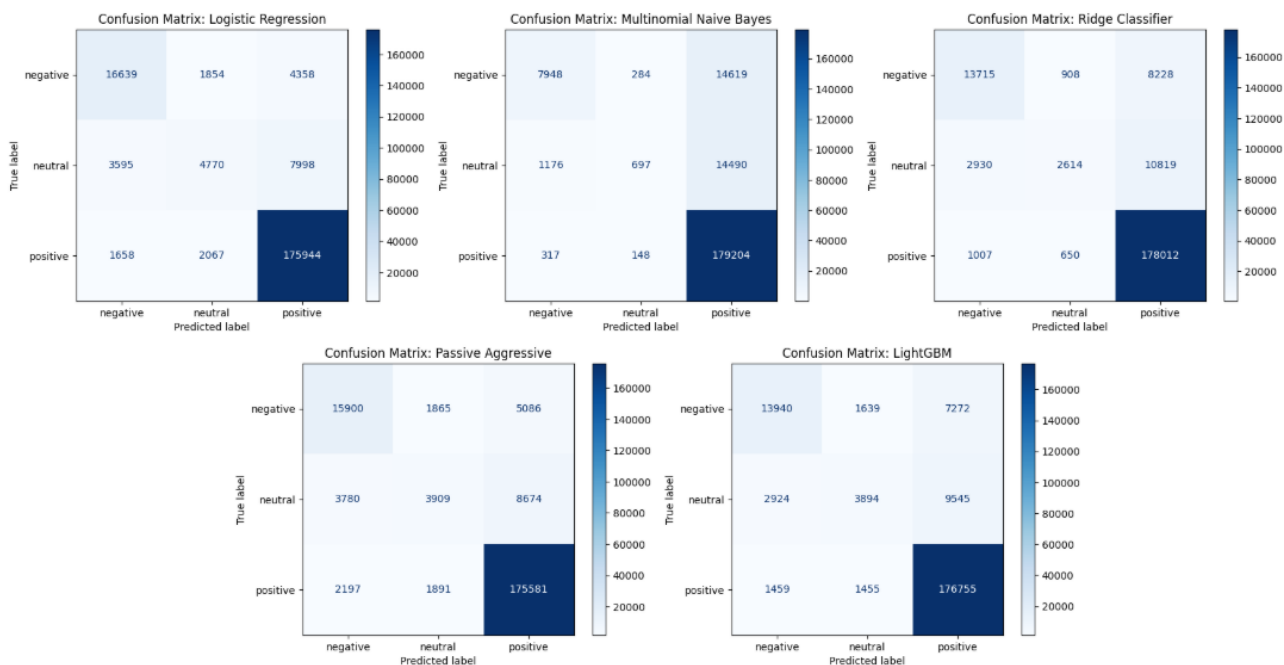


Рис. 3.6. Матриці похибок класифікаційних моделей

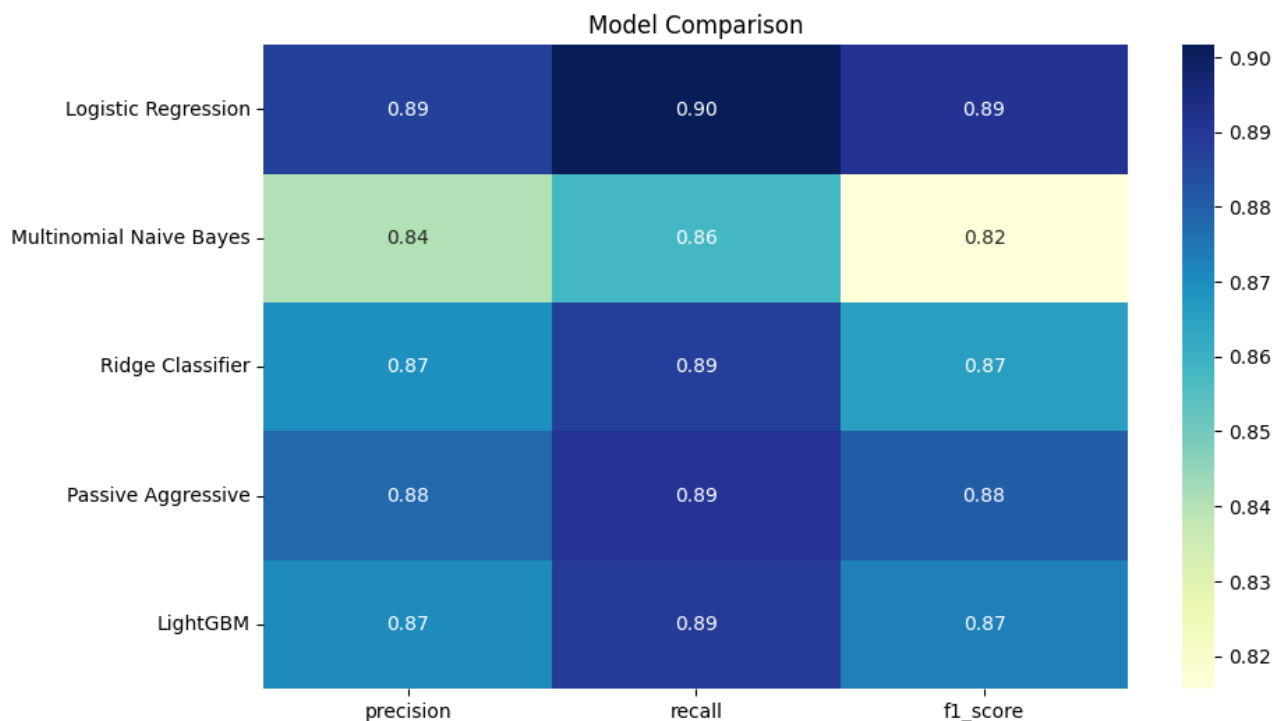


Рис. 3.7. Порівняння класифікаційних моделей

За результатами порівняння моделей, що представлені на рис. 3.6. та рис. 3.7., найкращою моделлю за всіма метриками є логістична регресія. Найвища точність (0,887), повнота (0,902) і F1-оцінка (0,891) свідчать про те, що модель добре балансує між правильністю передбачень і здатністю знаходити всі

релевантні приклади. Модель універсальна та стійка до різних типів текстових задач.

Дещо нижче, проте дуже близько за якістю йде пасивно-агресивний класифікатор, з показником F1-оцінки (0,880). Даний метод добре працює на великих потоках даних і ідеально підходить для онлайн-навчання, коли дані постійно прибувають.

Достатньо сильні результати показує LightGBM. Ця модель має добрий recall (0,889), що вказує на її гарну здатність прогнозувати класи. Проте, його precision (0,871) і F1-score (0,873) дещо поступаються Logistic Regression і Passive Aggressive.

Ridge класифікатор виступає на рівні з LightGBM, але трохи гірше у балансі precision/recall ($f1_score = 0,865$). Даний метод може бути хорошою альтернативою, якщо важливі регуляризаційний контроль та стійкість до переобучення.

Найгірші результати серед усіх моделей дає мультиноміальний наївний байєс. Хоча ця модель працює швидко і може бути хорошою для великих датасетів, її precision (0,840), recall (0,858) та F1-score (0,816) значно поступаються іншим моделям. Це свідчить про те, що модель має більше помилок при передбаченні, що, ймовірно, не дасть ідеальних результатів на цьому типі даних.

3.3. Розробка рекомендаційної системи на основі алгоритмів колаборативної фільтрації.

Як зазначалось у другому розділі, існує два підходи до колаборативної фільтрації на основі пам'яті. Серед них більш персоналізованим вважається колаборативна фільтрація на основі користувачів, за якої рекомендації формуються шляхом виявлення користувачів зі схожими вподобаннями. Проте, зі зростанням кількості користувачів у системі спостерігається значне зростання обчислювальної складності, пов'язане з необхідністю обчислення схожості між

кожною парою користувачів. Це призводить до проблеми масштабованості, яка стає критичною у великих системах.

Одним із підходів до вирішення цієї проблеми є попередня сегментація користувачів на основі певних ознак (наприклад, демографічних характеристик, поведінкових патернів або контексту використання). Така сегментація дозволяє обмежити пошук схожих користувачів лише в межах релевантного сегмента, що істотно знижує кількість необхідних обчислень та прискорює процес формування рекомендацій. Окрім того, сегментація може підвищити якість рекомендацій завдяки урахуванню додаткової інформації про користувачів, що не завжди присутня у стандартних матрицях рейтингів.

Оскільки на вибір косметичних засобів мають значний вплив зовнішні риси користувачів є сенс у інтеграції сегментації клієнтів за даними ознаками у загальний процес персоналізації. В рамках даної роботи для цього було використано алгоритм кластеризації k-means, який є оптимальним у даній ситуації через те, що ми працюємо з великим набором даних. Програмну реалізацію кластеризації подано в Додатку Б. Відповідно до графіка (рис. 3.8.), який демонструє залежність між кількістю кластерів та загальною сумою квадратів похибок в кластері, оскільки крива перегинається на позначці у $k=20$ кластерів, то саме на таку кількість кластерів будуть поділені користувачі.

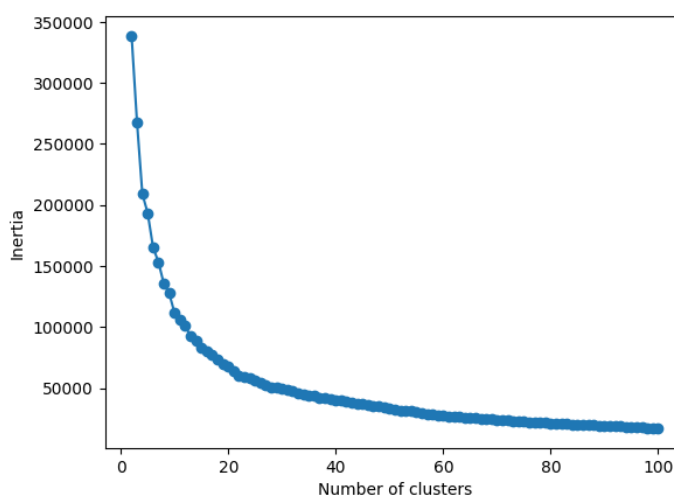


Рис. 3.8. Візуалізація методу ліктя для визначення кількості кластерів

В результаті отримано 20 кластерів, до кожного з яких належить від 10358 до 119112 користувачів. Візуалізація отриманих кластерів наведена на рис. 3.9. Силуетна міра для даної моделі становить 0,37, індекс Калінського-Харабаша – 47371,5, індекс Девіса-Болдіна – 1,143, на основі чого можна стверджувати що кластеризаційна модель має корисну структуру, адже відокремленість кластерів є задовільною, можливе перекриття не критичним, а структура кластерів стійкою.

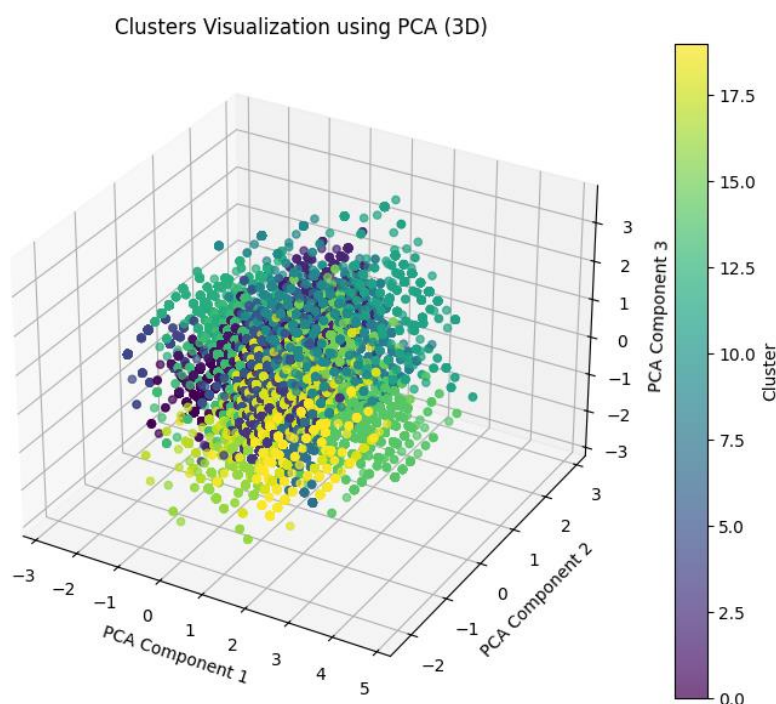


Рис. 3.9 . Візуалізація результатів кластеризації

Окрім, проблеми масштабованості кластеризація також може допомогти у вирішенні проблеми холодного старту. Наприклад, новим користувачам без попередньої історії покупок може бути запропоновано короткий опитувальник, за результатами якого користувача буде віднесено до одного з кластерів, після чого рекомендовано товари популярні серед користувачів даного кластеру.

<p>Кластер 0:</p> <ul style="list-style-type: none"> Cicapair Tiger Grass Color Correcting Treatment SPF 30 (Dr. Jart+) - 140 відгуків Mini Cicapair Tiger Grass Color Correcting Treatment SPF 30 (Dr. Jart+) - 137 відгуків Mini Superfood Antioxidant Cleanser (Youth To The People) - 111 відгуків <p>Кластер 1:</p> <ul style="list-style-type: none"> Daily Microfoliant Exfoliator (Dermalogica) - 747 відгуків Mini Daily Microfoliant Exfoliator (Dermalogica) - 747 відгуків Vitamin Enriched Face Base Priming Moisturizer (Bobbi Brown) - 551 відгуків <p>Кластер 2:</p> <ul style="list-style-type: none"> Ultra Repair Cream Intense Hydration (First Aid Beauty) - 585 відгуків Mini Take The Day Off Cleansing Balm Makeup Remover (CLINIQUE) - 442 відгуків Take The Day Off Cleansing Balm Makeup Remover (CLINIQUE) - 442 відгуків <p>Кластер 3:</p> <ul style="list-style-type: none"> Mini Superfood Antioxidant Cleanser (Youth To The People) - 255 відгуків Superfood Antioxidant Cleanser (Youth To The People) - 255 відгуків Daily Microfoliant Exfoliator (Dermalogica) - 236 відгуків 	<p>Кластер 4:</p> <ul style="list-style-type: none"> Superfood Antioxidant Cleanser (Youth To The People) - 241 відгуків Mini Superfood Antioxidant Cleanser (Youth To The People) - 240 відгуків Green Clean Makeup Meltaway Cleansing Balm Limited Edition Jumbo (Farmacy) - 227 відгуків <p>Кластер 5:</p> <ul style="list-style-type: none"> Daily Microfoliant Exfoliator (Dermalogica) - 278 відгуків Mini Daily Microfoliant Exfoliator (Dermalogica) - 278 відгуків Vitamin Enriched Face Base Priming Moisturizer (Bobbi Brown) - 197 відгуків <p>Кластер 6:</p> <ul style="list-style-type: none"> Green Clean Makeup Meltaway Cleansing Balm Limited Edition Jumbo (Farmacy) - 396 відгуків Green Clean Makeup Removing Cleansing Balm (Farmacy) - 396 відгуків Cicapair Tiger Grass Color Correcting Treatment SPF 30 (Dr. Jart+) - 367 відгуків <p>Кластер 7:</p> <ul style="list-style-type: none"> Green Clean Makeup Meltaway Cleansing Balm Limited Edition Jumbo (Farmacy) - 269 відгуків Green Clean Makeup Removing Cleansing Balm (Farmacy) - 269 відгуків Cicapair Tiger Grass Color Correcting Treatment SPF 30 (Dr. Jart+) - 223 відгуків
--	---

Рис. 3.11. Приклад виводу найпопулярніших товарів для кожного кластеру

Отримані кластери лягають в основу майбутнього рекомендаційного алгоритму на основі колаборативної фільтрації за користувачами: при пошуку користувачів зі схожими вподобаннями до цільового користувача спочатку відфільтровуються користувачі, що належать до того самого кластеру. Після цього для користувачів кластеру формується матриця користувач-продукт, де рядки відповідають користувачам, стовпці – продуктам, а значення дорівнюють середньому рейтингу.

Як було раніше зазначено, через упередження в даних, в рамках даної роботи існує необхідність нормалізації оцінок, як по користувачам так і по продуктам. Виходячи з цього доцільно скористатися методом подвійної стандартизації, що здійснюється шляхом віднімання середнього значення оцінок кожного користувача та ділення на стандартне відхилення для кожного користувача, після чого повторення даних операцій для кожного продукту. На основі отриманих значень розраховується схожість користувачів коефіцієнтом кореляції Пірсона.

Результатом реалізації моделі є надання користувачу рекомендацій, які б ґрунтувались на його схожості з іншими користувачами та середніми оцінками товарів, які ще не придбав цільовий користувач. Програмну реалізацію даного алгоритму подано в Додатку В.

```

Топ-5 рекомендованих продуктів:
Vinopure Pore Purifying Gel Cleanser (Caudalie) – передбачена оцінка: 5.00
Green Clean Makeup Meltaway Cleansing Balm Limited Edition Jumbo (Farmacy) – передбачена оцінка: 5.00
JuJu Exfoliating Bar (Drunk Elephant) – передбачена оцінка: 5.00
Grape Water Moisturizing Face Mist (Caudalie) – передбачена оцінка: 5.00
Pekee Cleansing Bar (Drunk Elephant) – передбачена оцінка: 5.00

```

Рис. 3.12. Вивід товарів рекомендованих методом колаборативної фільтрації за користувачами за прикладі користувача 21929544613

Для оцінки побудованої моделі було використано метрики статистичної точності та метрики точності прийняття рішень. Результати наведено у табл 3.2.

Таблиця 3.2

Метрики якості для рекомендаційного алгоритму на основі колаборативної фільтрації за користувачами

Метрика	RMSE	MAE	Precision	Recall	F1-score
Значення	0,8483	0,5101	0,7801	0,7579	0,7688

Альтернативно, також може бути використаний алгоритм колаборативної фільтрації за об'єктами, який хоч і вважається менш персоналізованим, але не має таких проблем з масштабованістю як фільтрація за користувачами.

Програмна реалізація даного алгоритму подана в Додатку Г і загалом є аналогічною до програмної реалізації фільтрації за користувачами, але схожість на основі матриці користувач-продукт розраховується не між користувачами, а між об'єктами, і обчислюється на основі косинусної подібності, а не коефіцієнту кореляції Пірсона. Оцінки моделі наведено у табл. 3.3 .

```

Введіть author_id: 21929544613

Рекомендовані товари:
- P500613: Banana Bright+ Vitamin C Eye Crème (OLEHENRIKSEN) – прогнозована оцінка: 5
- P393076: Ultra Repair Hydra-Firm Night Cream (First Aid Beauty) – прогнозована оцінка: 5
- P481972: La Mousse OFF/ON Foaming Face Cleanser (Dior) – прогнозована оцінка: 5
- P500240: Daily Milkfoliant Exfoliator (Dermalogica) – прогнозована оцінка: 5
- P442741: Peptide 21 Wrinkle Resist Serum (Peter Thomas Roth) – прогнозована оцінка: 5

```

Рис. 3.13. Приклад виводу товарів рекомендованих методом колаборативної фільтрації за об'єктами

Таблиця 3.3

Метрики якості для рекомендаційного алгоритму на основі колаборативної фільтрації за об'єктами

Метрика	RMSE	MAE	Precision	Recall	F1-score
Значення	1,0107	0,6390	0,6904	0,6847	0,6791

Як можна побачити, рекомендації отримані в результаті реалізації двох алгоритмів значно відрізняються. Згідно з отриманими метриками, модель фільтрації за користувачами значно переважає модель фільтрації за об'єктами за точністю рекомендацій. Це свідчить про те, що у контексті наявних даних фільтрація за користувачами краще відображає індивідуальні вподобання та поведінкові патерни.

Разом із тим, з алгоритмічної точки зору, підхід, що базується на користувачах, має низку обмежень. Зокрема, він гірше масштабується при зростанні кількості користувачів, є більш чутливим до змін їх поведінки, а також схильний до проблеми холодного старту для нових користувачів, оскільки вимагає наявності історії активності. Натомість, модель фільтрації за об'єктами, хоча й демонструє нижчу точність, має певні структурні переваги: вона краще працює при розріджених даних, простіше масштабується при великій кількості користувачів і забезпечує стабільніші рекомендації, оскільки подібність між об'єктами менш схильна до змін з часом. Крім того, підхід, що базується на об'єктах, легше інтерпретувати: система рекомендує об'єкти, схожі на ті, які користувач уже оцінив або переглянув.

Таким чином, вибір між фільтрацією за користувачами та об'єктами повинен залежати від контексту застосування системи рекомендацій. Якщо головним пріоритетом є висока точність персоналізації для користувачів з багатою історією взаємодій, підхід на основі користувачів є більш доречним. Водночас, у випадках із великими масивами даних, частими новими користувачами або обмеженими

відомостями про поведінку, модель фільтрації за об'єктами може бути більш надійним та практичним вибором. Найбільш ефективним підходом може стати гібридна модель, яка поєднує сильні сторони обох методів.

3.4. Розробка рекомендаційної системи за допомогою алгоритмів фільтрації на основі вмісту.

Запропоновані у попередньому пункті моделі на основі колаборативної фільтрації можуть бути недостатньо ефективні при роботі з новими користувачами або продуктами, через нестачу даних про пов'язані з ними взаємодії. Для вирішення цієї проблеми можна використовувати фільтрацію на основі вмісту, яка дозволяє забезпечити краще покриття для об'єктів або користувачів з малою кількістю взаємодій, за рахунок врахування характеристик об'єктів та вподобань користувачів щодо певних характеристик. Більше того, використання даного методу покращує стабільність рекомендацій, забезпечуючи базову надійність, та враховує особисті вподобання користувача щодо характеристик об'єктів, а не лише вподобання інших користувачів, що допомагає краще моделювати смаки користувачів з унікальними або вузькоспеціалізованими інтересами.

Першим кроком при розробці даної моделі є попередня обробка атрибутів товарів, з метою їх об'єднання у зручному текстовому представленні. Для цього у якості атрибутів було використано ключові теги, назву бренду, а також первинну, вторинну та третинну категорії, до яких належить товар.

```
def preprocess_product_features(df):
    df['highlights_str'] = df['highlights'].apply(lambda x: ' '.join(eval(x)) if isinstance(x, str) else '')
    df['combined_features'] = (
        df['brand_name'].fillna('') + ' ' +
        df['highlights_str'].fillna('') + ' ' +
        df['primary_category'].fillna('') + ' ' +
        df['secondary_category'].fillna('') + ' ' +
        df['tertiary_category'].fillna('')
    )
    return df
```

Рис. 3.14. Функція для програмної реалізації попередньої обробки атрибутів

Після цього отримані текстові блоки перетворюються у числові вектори з врахуванням, наскільки важливе слово в контексті певного документа порівняно з іншими, за допомогою методу TF-IDF.

```
df = preprocess_product_features(df)
tfidf = TfidfVectorizer()
tfidf_matrix = tfidf.fit_transform(df['combined_features'])
```

Рис. 3.15. Програмна реалізація перетворення текстових блоків на числові вектори методом TF-IDF

Наступним етапом є створення профілю користувача для того, щоб розуміти, що саме подобається користувачу, на основі товарів, які він позитивно оцінив. Для цього обчислюється середній вектор на основі TF-IDF векторів усіх товарів, які користувач рекомендував.

```
def build_user_profile(author_id, df, tfidf_matrix):
    user_reviews = df[df['author_id'] == author_id]
    recommended = user_reviews[user_reviews['is_recommended'] == True]

    if recommended.empty:
        return None

    product_indices = recommended.index.tolist()
    user_profile_vector = tfidf_matrix[product_indices].mean(axis=0)

    # Перетворення у numpy масив (1D)
    return np.asarray(user_profile_vector).reshape(1, -1)
```

Рис. 3.16. Функція для програмної реалізації створення профілю користувача

Після цього обчислюється схожість між вектором профілю користувача і всіма товарами за допомогою косинусної подібності. Отриманий список товарів фільтрується на предмет дублікатів і товарів, з якими користувач вже взаємодіяв, та ранжується за схожістю.

```

# Вилучаємо дублікати та продукти, які вже були оцінені користувачем
seen_products = df[df['author_id'] == author_id]['product_id'].unique()
df['similarity'] = cosine_similarities
recommendations = df[~df['product_id'].isin(seen_products)].sort_values(by='similarity', ascending=False)
recommendations = (
df[~df['product_id'].isin(seen_products)]
.sort_values(by='similarity', ascending=False)
.drop_duplicates(subset='product_id')
)

```

Рис. 3.17. Програмна реалізація фільтрації та ранжування списку рекомендованих продуктів

Результат реалізації алгоритму представляє собою список рекомендованих товарів, найбільш подібних до вподобаних користувачем з погляду їх контентних характеристик.

```

Топ рекомендованих продуктів для користувача 21929544613:
Продукт: Superberry Hydrate + Glow Dream Night Mask with Vitamin C (Бренд: Youth To The People). Рекомендаційний бал: 0.611
Продукт: Superclay Purify + Clear Power Mask with Niacinamide (Бренд: Youth To The People). Рекомендаційний бал: 0.597
Продукт: Superberry Dream Cleansing Balm (Бренд: Youth To The People). Рекомендаційний бал: 0.555
Продукт: 100% Plant-Derived Squalane (Бренд: The Ordinary). Рекомендаційний бал: 0.553
Продукт: Mini Superberry Hydrate + Glow Dream Mask (Бренд: Youth To The People). Рекомендаційний бал: 0.545

```

Рис. 3.18. Приклад виводу рекомендованих продуктів в результаті реалізації алгоритму на основі фільтрації за вмістом

Висновки до розділу 3

У результаті проведеної роботи було розроблено комплекс моделей, що можуть бути інтегровані у функціонал маркетплейсу для підвищення його ефективності.

В рамках запропонованого комплексу моделей, здійснено попередню обробку та аналіз текстових відгуків шляхом сентимент-аналізу, що надало додаткові можливості для оцінки користувацького досвіду за емоційним забарвленням. Для цього були використані класифікаційні моделі машинного навчання, які забезпечили досить високу точність у визначенні тональності.

Також було реалізовано рекомендаційні системи на основі колаборативної фільтрації, які показали хороші результати при наявності історичних даних про взаємодії користувачів із товарами. Окрім того, було запропоновано використання каскадного гібридного поєднання кластеризації для сегментування

користувачів з колаборативною фільтрацією за користувачами з метою подолання проблеми квадратичної складності даного алгоритму.

Паралельно, з метою забезпечення ефективних рекомендації у ситуаціях, коли інформація про попередні взаємодії є обмеженою, було створено модель рекомендаційної системи за допомогою фільтрації на основі вмісту, яка формує рекомендації, орієнтуючись на характеристики самих товарів.

Розроблений комплекс моделей, завдяки своїй модульній структурі та широкому функціоналу, має високий потенціал до гібридизації та адаптації під різноманітні завдання в електронній комерції - від класичних рекомендацій товарів до інтелектуального аналізу поведінки користувачів і контенту.

ВИСНОВКИ

На основі проведеного аналізу доведено, що ринок електронної комерції перебуває у фазі динамічного зростання і супроводжується цифровізацією торгівлі та зростанням потреби у персоналізованому сервісі. Також, було обґрунтовано доцільність застосування технологій штучного інтелекту в електронній комерції, зокрема в сегменті формування рекомендацій. Алгоритми штучного інтелекту забезпечують обробку великих масивів даних, моделювання поведінки користувачів і створення персоналізованого досвіду, що, в свою чергу, веде до підвищення рівня задоволеності клієнтів і зростання прибутковості бізнесу. На основі аналізу досвіду провідних світових компаній показано ефективність використання рекомендаційних систем у реальних умовах та описано основні підходи до впровадження рекомендаційних систем. Зокрема, проаналізовано, як компанії Amazon, Netflix, Youtube та Spotify інтегрують алгоритми персоналізації у свої сервіси. Виявлено, що використовуються як алгоритми колаборативної фільтрації, так і методи, засновані на вмісті та поведінкових патернах користувачів. Зазначено особливості реалізації, такі як, наприклад, застосування гібридних моделей.

У процесі дослідження охарактеризовано основні етапи рекомендаційного процесу, зокрема: збір даних, навчання та подання результатів. Також було систематизовано методологічні підходи до побудови рекомендаційних моделей, включаючи колаборативну фільтрацію, фільтрацію за вмістом і гібридні системи. Показано переваги та обмеження кожного підходу, що дозволяє обґрунтовано вибирати алгоритм відповідно до специфіки завдання й доступних даних. На основі аналізу метрик оцінювання показано, що для коректного порівняння моделей доцільно використовувати комплексну оцінку, яка включає показники точності, повноти, F1-міри, середньої абсолютної та середньоквадратичної помилки, показники точності ранжування та задоволеності користувачів. Це дозволяє об'єктивно оцінити якість сформованих рекомендацій.

У рамках побудови рекомендаційної системи проаналізовано вхідні дані та джерела інформації, необхідні для функціонування рекомендаційної системи. На основі отриманої бази даних, реалізовано модель аналізу настроїв на основі класифікаційних алгоритмів машинного навчання, серед яких логістична регресія, мультиноміальний наївний байєсівський класифікатор, класифікатор Ridge, пасивно-агресивний класифікатор та LightGBM. На основі тестування моделей показано, що найкращу якість класифікації (F1-оцінка - 0,891) забезпечує модель логістичної регресії. Це підтверджує доцільність включення етапу настроїв-аналізу в загальну структуру рекомендаційної системи. Розроблено рекомендаційну систему на основі алгоритмів колаборативної фільтрації, зокрема використано метод фільтрації за користувачами та об'єктами. Також, було запропоновано поєднання моделей кластеризації та фільтрації за користувачами, яке дозволило би подолати проблеми масштабованості. Результати тестування засвідчили, що кращу якість рекомендацій забезпечує алгоритм фільтрації за користувачами, з показниками RMSE – 0,8483 та F1-оцінки - 0,7688. Доведено ефективність такого підходу для користувачів із багатим історичним профілем взаємодії. Також було створено рекомендаційну систему на основі фільтрації за вмістом, яка використовує текстові та категорійні характеристики товарів. Показано, що такий підхід дозволяє ефективно генерувати персоналізовані рекомендації для нових користувачів, що є його важливою перевагою.

Комплекс побудованих моделей відзначається високою гнучкістю та функціональною розширюваністю, що створює підґрунтя для їх подальшої гібридизації під широкий спектр завдань у сфері електронної комерції. Завдяки модульному підходу, система може бути масштабована та доповнена іншими алгоритмами рекомендацій, адаптована до різних категорій товарів, типів користувачів і цілей бізнесу, що значно підвищує її прикладну цінність.

Більше того, оскільки всі моделі навчені та перевірені на реальних даних, можна стверджувати про широкі можливості застосування даної методики на практиці.

Подальші дослідження можуть бути спрямовані на більш детальне вивчення стратегій гібридизації для поєднання розроблених моделей під конкретні задачі та типи подачі рекомендацій, інтеграцію контекстуальних факторів (час, локація, сезонність) у процес рекомендацій, а також залучення більшого спектру алгоритмів на основі моделей, з відповідним дослідженням особливостей кожного з них.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Laudon K. C., Traver C. G. E-commerce 2019: Business, technology, society. Pearson, 2020.
2. Kalakota R., Whinston A. B. Frontiers of electronic commerce. Addison Wesley Longman Publishing Co., Inc., 1996.
3. Про електронну комерцію: Закон України від 03.09.2015 р. № 675-VIII. URL: <https://zakon.rada.gov.ua/laws/show/675-19#Text>
4. Chaffey D. E-business and E-commerce Management: Strategy, Implementation and Practice. Pearson Education, 2007.
5. Unpacking E-Commerce: Business Models, Trends and Policies. OECD Publishing, Paris, 2019. URL: <https://doi.org/10.1787/23561431-en>.
6. Кравченко Б. О. Розвиток електронної комерції в Україні: перспективи, виклики та тенденції розвитку у 2023 році. *Цифрові трансформації та інноваційні технології в економіці: виклики, реалії, стратегії*. Збірник матеріалів II Міжнародної науково-практичної конференції (17-19 травня 2023 р.). Сумський державний університет, 2023.
7. Santos V. et al. E-commerce: issues, opportunities, challenges, and trends. Promoting organizational performance through 5G and agile marketing. 2023. P. 224-244. doi: 10.4018/978-1-6684-5523-4.ch012
8. Щитов Д. М., Жадько К. С., Мормуль М. Ф. Тенденції розвитку ринку електронної комерції у світі та в Україні. Наукові перспективи. 2025. № 7(49). URL: <http://biblio.umsf.dp.ua/jspui/handle/123456789/6992>
9. 51 eCommerce Statistics In 2025 (Global and U.S. Data). SellersCommerce. 2024. URL: <https://www.sellerscommerce.com/blog/e-commerce-statistics/>
10. Revenue share of the e-commerce market worldwide from 2019 to 2029, by sales channel. Statista. 2025. URL: <https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide/>

11. Retail e-commerce sales worldwide from 2022 to 2028. Statista. 2025. URL: <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/>
12. Total retail e-commerce revenue worldwide in 2024, by region. Statista. 2024. URL: <https://www.statista.com/forecasts/1117851/worldwide-e-commerce-revenue-by-region>
13. Retail e-commerce revenue worldwide from 2019 to 2029, by segment. Statista. 2025. URL: <https://www.statista.com/forecasts/1223973/e-commerce-revenue-worldwide-by-segment>
14. Ecommerce usage distribution on the entire Internet. BuiltWith. 2025. URL: <https://trends.builtwith.com/shop/traffic/Entire-Internet>
15. Semrush. URL: <https://www.semrush.com/>
16. Ross R. et al. Developing cyber resilient systems: a systems security engineering approach. National Institute of Standards and Technology, 2019. №. NIST Special Publication (SP) 800-160 Vol. 2 (Draft). URL: <https://doi.org/10.6028/NIST.SP.800-160v2r1>
17. Blockchain technology cloud market size worldwide in 2021, with a forecast for 2030. Statista. 2023. URL: <https://www.statista.com/statistics/1319369/global-blockchain-technology-market-size/>
18. US Augmented and Virtual Reality Users Forecast 2022. EMarketer. 2022. URL: <https://www.emarketer.com/content/us-augmented-virtual-reality-users-forecast-2022>
19. Smart speaker household penetration rate in the United States from 2014 to 2025. Statista. 2019. URL: <https://www.statista.com/statistics/1022847/united-states-smart-speaker-household-penetration/>
20. Global Mobile E-Commerce Worth \$2.2 Trillion in 2023. Statista. 2023. URL: <https://www.statista.com/chart/13139/estimated-worldwide-mobile-e-commerce-sales/>
21. 49 Cart Abandonment Rate Statistics 2025. Baymard Institute. 2025. URL: <https://baymard.com/lists/cart-abandonment-rate>

22. Number of social commerce buyers in the United States from 2023 to 2028. Statista. 2024. URL: <https://www.statista.com/statistics/1120128/number-social-buyers-united-states/>
23. Long H., Van Dam A. Everything's becoming a subscription, and the pandemic is partly to blame. *The Washington Post*. 2021. URL: <https://www.washingtonpost.com/business/2021/06/01/subscription-boom-pandemic/>
24. Decision points: Sharpening the pre-purchase consumer experience. PwC. 2023. URL: <https://www.pwc.com/gx/en/industries/consumer-markets/consumer-insights-survey.html>
25. Gen Z and Millennials 2023: Understanding the next generation of shoppers. Drapers. 2023. URL: <https://www.drapersonline.com/guides/gen-z-and-millennials-2023>
26. Global Consumer Report: Current and Future Shopping Trends. BigCommerce. 2022. URL: https://www.ctidigital.com/hubfs/bigcommerce_global_consumer_trends_report.pdf
27. Is the deployment of AI in your e-commerce solutions an important subject in your business?. Statista. 2019. URL: <https://www.statista.com/statistics/1096189/artificial-intelligence-ecommerce-adoption-france/>
28. Музиченко Т. О., Скорба О. А., Шевчук А. А. Штучний інтелект як засіб оптимізації бізнес-процесів в електронній комерції. *Академічні візії*. 2023. №. 25. URL: <https://doi.org/10.5281/zenodo.10081884>
29. Булах О. В. Глобальний вплив штучного інтелекту та машинного навчання на ефективність електронної комерції. *Бізнес Інформ*, (8). 2023. С. 114-121. URL: <https://doi.org/10.32983/2222-4459-2023-8-114-121>
30. Махова Г. В. Штучний інтелект в підприємстві: можливості та перспективи використання. *Економіка та підприємництво : зб. наук. пр.* Київ: КНЕУ, 2022. Вип. 49. С. 75–89. URL: <https://ir.kneu.edu.ua/handle/2010/38603>

31. Artificial Intelligence in E-commerce Market Size, Share, and Trends 2024 to 2034. Precedence Research. 2024. URL: <https://www.precedenceresearch.com/artificial-intelligence-in-e-commerce-market>
32. Ecommerce Personalization Benchmark Report. Netcore. 2021. URL: <https://netcorecloud.com/wp-content/uploads/2021/04/Ecommerce-Personalization-Benchmark-Report-2021.pdf>
33. Retail Personalization Index 2022: Balancing Trust, Data Collection and Privacy. Sailthru. 2022. URL: <https://go.sailthru.com/2022-retail-personalization-index-retailer-consumer-survey/>
34. Yin S. Where chatbots are headed. Intercom. 2019. URL: <https://www.intercom.com/blog/the-state-of-chatbots/>
35. Aliche K., K Ganesh, Ganguly S., Shinghal S. Autonomous supply chain planning for consumer goods companies. McKinsey. 2022. URL: <https://www.mckinsey.com/capabilities/operations/our-insights/autonomous-supply-chain-planning-for-consumer-goods-companies>
36. Hazan J., Brégé C., Verwaerde J., Bassoulet A. Why AI Transformations Should Start with Pricing. Boston Consulting Group. 2021. URL: <https://www.bcg.com/publications/2021/ai-pricing-tranformations>
37. Global AI In Fraud Detection Market Report. Market.us. 2024. URL: <https://market.us/report/ai-in-fraud-detection-market/>
38. Main challenges in implementing AI-based solutions in the retail industry in the U.S. Statista. 2024. URL: <https://www.statista.com/statistics/1463707/ai-challenges-united-states/>
39. Затонацька Т., Фаренюк Я. Використання Data Science технологій в е-комерції: успішні кейси. Міжнародний форум EFBM 3.0 «Економіка. Фінанси. Бізнес. Управління. Від відновлення до зростання» 2024. С. 26-30.
40. MacKenzie I., Meyer C., Noble S. How retailers can keep up with consumers. McKinsey. 2013. URL: <https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers#/>

41. Hardesty L. The history of Amazon's recommendation algorithm. Amazon Science. 2019. URL: <https://www.amazon.science/the-history-of-amazons-recommendation-algorithm>.
42. Covington P., Adams J., Sargin E. Deep neural networks for youtube recommendations. *Proceedings of the 10th ACM conference on recommender systems*. 2016. P. 191-198. doi: <http://dx.doi.org/10.1145/2959100.2959190>
43. Gomez-Uribe C. A., Hunt N. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*. 2015. 6(4). P. 1-19. doi: <http://dx.doi.org/10.1145/2843948>
44. Mudaliyar M. Behind the Playlist: Analyzing Spotify's Recommendation System. *Medium*. 2024. URL: <https://medium.com/%40myliemudaliyar/behind-the-playlist-analyzing-spotifys-recommendation-system-5044a13f5ccf>
45. Isinkaye F. O., Folajimi Y. O., Ojokoh B. A. Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal*. 2015. 16(3). P. 261-273. doi:[10.1016/j.eij.2015.06.005](https://doi.org/10.1016/j.eij.2015.06.005)
46. Raghuwanshi S. K., Pateriya R. K. Recommendation systems: techniques, challenges, application, and evaluation. *Soft Computing for Problem Solving: SocProS 2017, Volume 2*. Springer Singapore, 2019. P. 151-164. doi: https://doi.org/10.1007/978-981-13-1595-4_12
47. Fayyaz Z. et al. Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *Applied sciences*. 2020. 10(21). doi: <https://doi.org/10.3390/app10217748>
48. Щербаков О. Е. Дослідження методів оцінки рекомендаційних систем. М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. Харків, 2019. 77 с. URL: <http://openarchive.nure.ua/handle/document/10953>
49. Sephora. URL: <https://www.sephora.com/>
50. Sephora.com ecommerce revenue, sales and traffic. Grips. URL: <https://gripsintelligence.com/insights/retailers/sephora.com>

51. Chen J. et al. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*. 2023. 41(3). P. 1-39. doi: <https://doi.org/10.1145/1122445.1122456>
52. Про деякі правові аспекти інформаційних послуг, зокрема, електронної комерції, на внутрішньому ринку ("Директива про електронну комерцію"): Директива 2000/31/ЄС Європейського парламенту та Ради від 8 червня 2000 року. URL: https://zakon.rada.gov.ua/laws/show/994_224#Text
53. Греков О. О., Головянко М. В. Сучасні методи побудови рекомендаційних систем. *Радіоелектроніка та молодь у XXI столітті*. Матеріали 29-го Міжнародного молодіжного форуму (16 – 19 квітня 2025 року). Харків, 2025. С. 18-20.
54. Mykhalchuk, T., Zatonatska, T., Dluhopolskyi, O., Zhukovska, A., Dluhopolska, T., & Liakhovych, L. Development of recommendation system in e-commerce using emotional analysis and machine learning methods. *11th IEEE international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS)*. 2021. 1. P. 527-535. doi: [10.1109/IDAACS53288.2021.9660854](https://doi.org/10.1109/IDAACS53288.2021.9660854)
55. Loukili M., Messaoudi F., El Ghazi M. Sentiment analysis of product reviews for e-commerce recommendation based on machine learning. *International Journal of Advances in Soft Computing & Its Applications*. 2023. 15(1). doi: 10.15849/IJASCA.230320.01
56. Patil, P., Kadam, S. U., Aruna, E. R., More, A., Balajee, R. M., & Rao, B. N. K. Recommendation system for e-commerce using collaborative filtering. *Journal Européen des Systèmes Automatisés*. 2024. 57(4). P. 1145-1153. doi: <https://doi.org/10.18280/jesa.570421>
57. Permana A. H. J. P. J., Wibowo A. T. Movie Recommendation System Based on Synopsis Using Content-Based Filtering with TF-IDF and Cosine Similarity. *International Journal on Information and Communication Technology (IJoICT)*. 2023. 9(2). P. 1-14. doi: <https://doi.org/10.21108/ijoiect.v9i2.747>

58. Abdul Hussien F. T., Rahma A. M. S., Abdulwahab H. B. An e-commerce recommendation system based on dynamic analysis of customer behavior. *Sustainability*. 2021. 13(19). doi: <https://doi.org/10.3390/su131910786>
59. Xu, K., Zhou, H., Zheng, H., Zhu, M., & Xin, Q. Intelligent classification and personalized recommendation of E-commerce products based on machine learning. *Applied and Computational Engineering*. 2024. 64. P. 147-153. doi: <https://doi.org/10.54254/2755-2721/64/20241365>
60. Sun A. Beyond Collaborative Filtering: A Relook at Task Formulation in Recommender Systems. *ACM SIGWEB Newsletter*. 2024. P. 1-11. doi: <https://doi.org/10.1145/3663752.3663756>

ДОДАТКИ

Додаток А

Програмна реалізація класифікаційних моделей для аналізу сентиментів

```
from sklearn.linear_model import LogisticRegression, RidgeClassifier, PassiveAggressiveClassifier
from sklearn.naive_bayes import MultinomialNB
import lightgbm as lgb
from sklearn.metrics import accuracy_score, f1_score
import pandas as pd

models = {
    'Logistic Regression': LogisticRegression(max_iter=500, solver='saga'),
    'Multinomial Naive Bayes': MultinomialNB(),
    'Ridge Classifier': RidgeClassifier(),
    'Passive Aggressive': PassiveAggressiveClassifier(max_iter=1000),
    'LightGBM': lgb.LGBMClassifier()
}

results = {}

for name, model in models.items():
    model.fit(X_train_tfidf, y_train)
    y_pred = model.predict(X_test_tfidf)
    acc = accuracy_score(y_test, y_pred)
    f1 = f1_score(y_test, y_pred, average='weighted')
    results[name] = {'accuracy': acc, 'f1_score': f1}

results_df = pd.DataFrame(results).T
print(results_df.sort_values(by='f1_score', ascending=False))
```

Програмна реалізація кластеризації методом k-means

```
import pandas as pd
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
import matplotlib.pyplot as plt

# Вибір необхідних колонок та унікальних користувачів
df = combined_reviews_df[['author_id', 'skin_tone', 'eye_color', 'skin_type', 'hair_color']].drop_duplicates()

# Обробка пропущених значень
df.fillna('Unknown', inplace=True)

# Перетворення категоріальних змінних у числові
encoder = LabelEncoder()
for col in ['skin_tone', 'eye_color', 'skin_type', 'hair_color']:
    df[col] = encoder.fit_transform(df[col])

# Масштабування даних
scaler = StandardScaler()
features = ['skin_tone', 'eye_color', 'skin_type', 'hair_color']
X = scaler.fit_transform(df[features])

# Визначення оптимальної кількості кластерів за методом "лікоть"
inertia = []
range_n_clusters = range(2, 101)
for n_clusters in range_n_clusters:
    kmeans = KMeans(n_clusters=n_clusters, random_state=42)
    kmeans.fit(X)
    inertia.append(kmeans.inertia_)

# Візуалізація методу "лікоть"
plt.plot(range_n_clusters, inertia, marker='o')
plt.title("Elbow Method")
plt.xlabel("Number of clusters")
plt.ylabel("Inertia")
plt.show()

# Вибір кількості кластерів
optimal_clusters = 28
kmeans = KMeans(n_clusters=optimal_clusters, random_state=42, n_init=10)
df['cluster'] = kmeans.fit_predict(X)

# Додавання результатів до початкового датафрейму
final_df = combined_reviews_df.merge(df[['author_id', 'cluster']], on='author_id', how='left')

# Аналіз результатів
print(final_df['cluster'].value_counts())
```

Програмна реалізація колаборативної фільтрації за користувачами

```

from surprise import KNNBaseline

# Налаштування параметрів baseline estimates
bsl_options = {
    'method': 'als',
    'n_epochs': 10,
    'reg_u': 15,
    'reg_i': 10
}

# Налаштування параметрів схожості
sim_options = {
    'name': 'pearson_baseline',
    'user_based': True
}

# Ініціалізація моделі
algo = KNNBaseline(bsl_options=bsl_options, sim_options=sim_options)

# Тренування моделі
algo.fit(trainset)

# Отримання списку всіх продуктів
all_items = set(trainset.all_items())
all_items_raw = [trainset.to_raw_iid(iid) for iid in all_items]

# Отримання списку продуктів, які користувач вже оцінював
user Rated_items = set([j for (j, _) in trainset.ur[trainset.to_inner_uid(target_author_id)]]
user Rated_items_raw = [trainset.to_raw_iid(iid) for iid in user Rated_items]

# Визначення продуктів, які користувач ще не оцінював
items_to_predict = set(all_items_raw) - set(user Rated_items_raw)

# Передбачення оцінок для неоцінених продуктів
predictions = []
for item_id in items_to_predict:
    pred = algo.predict(target_author_id, item_id)
    predictions.append((item_id, pred.est))

# Сортування прогнозів за спаданням оцінки
predictions.sort(key=lambda x: x[1], reverse=True)

# Виведення топ-5 рекомендацій
print("Топ-5 рекомендованих продуктів:")
for item_id, rating in predictions[:5]:
    # Фільтрація записів з відповідним product_id
    product_info = combined_reviews_df[combined_reviews_df['product_id'] == item_id]

    if not product_info.empty:
        # Отримання першого співпадіння
        product_name = product_info['product_name'].values[0]
        brand_name = product_info['brand_name'].values[0]
        print(f"{product_name} ({brand_name}) – передбачена оцінка: {rating:.2f}")
    else:
        print(f"Продукт з ID {item_id} – передбачена оцінка: {rating:.2f}")

```

Програмна реалізація колаборативної фільтрації за об'єктами

```

import pandas as pd
from surprise import Dataset, Reader, KNMBaseline

# Створимо мапу product_id -> (product_name, brand_name)
product_info_map = combined_reviews_df.drop_duplicates(subset=['product_id'])\
    .set_index('product_id')[['product_name', 'brand_name']].to_dict(orient='index')

# Витягаємо необхідні колонки
df = combined_reviews_df[['author_id', 'product_id', 'rating']]

# --- Підготовка Surprise ---
reader = Reader(rating_scale=(df['rating'].min(), df['rating'].max()))
data = Dataset.load_from_df(df[['author_id', 'product_id', 'rating']], reader)
trainset = data.build_full_trainset()

# --- Item-based collaborative filtering ---
sim_options = {
    'name': 'cosine',
    'user_based': False # item-based
}

bsl_options = {
    'method': 'sgd',
    'learning_rate': 0.005,
    'n_epochs': 20
}

algo = KNMBaseline(sim_options=sim_options, bsl_options=bsl_options)
algo.fit(trainset)

# --- Рекомендаційна функція ---
def get_top_n_recommendations(user_id, trainset, algo, n=5):
    all_items = trainset.all_items()
    all_item_raw_ids = [trainset.to_raw_iid(iid) for iid in all_items]

    try:
        user_inner_id = trainset.to_inner_uid(user_id)
        rated_items_inner = [iid for (iid, _) in trainset.ur[user_inner_id]]
        rated_items_raw = set(trainset.to_raw_iid(iid) for iid in rated_items_inner)
    except ValueError:
        return [] # Користувача немає в тренувальному наборі

    predictions = []
    for item_id in all_item_raw_ids:
        if item_id not in rated_items_raw:
            pred = algo.predict(user_id, item_id)
            predictions.append((item_id, pred.est))

    predictions.sort(key=lambda x: x[1], reverse=True)
    top_n = predictions[:n]

    # Повертаємо як список словників з повною інформацією
    result = []
    for pid, est_rating in top_n:
        info = product_info_map.get(pid, {})
        result.append({
            'product_id': pid,
            'product_name': info.get('product_name', 'Невідомо'),
            'brand_name': info.get('brand_name', 'Невідомо'),
            'predicted_rating': round(est_rating, 2)
        })
    return result

--- Головна частина ---
if __name__ == "__main__":
    input_id = input("Введіть author_id: ")
    if input_id in df['author_id'].unique():
        recommendations = get_top_n_recommendations(input_id, trainset, algo)
        if recommendations:
            print("\nРекомендовані товари:")
            for rec in recommendations:
                print(f"- {rec['product_id']}: {rec['product_name']} ({rec['brand_name']}) - прогнозована оцінка: {rec['predicted_rating']}")
        else:
            print("Для цього користувача немає рекомендацій.")
    else:
        print("Користувача не знайдено.")

```