

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

**Економічний факультет
Кафедра економічної кібернетики**

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

«Розробка гібридної рекомендаційної моделі у електронній комерції»

студентки 4 курсу
спеціальності 051 «Економіка»
ОПП «Економічна кібернетика»
денної форми навчання
Усенко Дарини Олександрівни

Науковий керівник:
доктор економічних наук,
професор
Чорноус Галина Олександрівна

Засвідчую, що у цій дипломній
роботі немає запозичень із
праць інших авторів без
відповідних посилань
Студент _____
(підпис)

Роботу допущено до захисту перед ЕК
рішенням кафедри економічної кібернетики
від 12 червня 2023 р., протокол № 17
Завідувач кафедри:
доктор економічних наук, професор
Ляшенко Олена Ігорівна

(підпис)

КИЇВ – 2023

РЕФЕРАТ

Кваліфікаційна робота бакалавра містить: 48 ст., 14 рис., 54 джерела, додатки

Ключові слова: електронна комерція, гібридна рекомендаційна система, контентна фільтрація, колаборативна фільтрація на основі користувачів, колаборативна фільтрація на основі об'єктів

Об'єкт дослідження: рекомендаційні системи в сфері електронної комерції.

Мета дослідження: розробка алгоритму реалізації гібридної рекомендаційної системи на основі демографічної фільтрації з метою створення індивідуального списку товарів для користувача в умовах обмеженої кількості вхідних даних..

Методи дослідження: методи наукової абстракції, порівнянь, індукції та дедукції, методи аналізу та синтезу, методи фільтрації контенту, широкий спектр методів математичного моделювання, лінійної алгебри та дослідження операцій (техніки колаборативної фільтрації) та методи програмної реалізації математичного моделювання в програмному середовищі Python.

Наукова новизна, теоретична значимість дослідження: полягає в систематизації теоретичних напрацювань та практики використання рекомендаційних систем в електронній комерції, розробці концепції побудови рекомендаційної системи на основі демографічної фільтрації, а саме місцерозташування користувача, а також її практичної реалізації.

Практична цінність: полягає у створенні гібридної рекомендаційної системи, яку можна застосовувати бізнесам у електронній комерції, зокрема тим, яким характерна велика залежність від місцерозташування користувача

ЗМІСТ

ВСТУП.....	4
РОЗДІЛ 1. ЕЛЕКТРОННА КОМЕРЦІЯ ТА МІСЦЕ РЕКОМЕНДАЦІЙНИХ СИСТЕМ В ЕЛЕКТРОННІЙ КОМЕРЦІЇ	7
1.1 Розвиток електронної комерції та її сучасний стан в світі та Україні	7
1.2 Рекомендаційні системи та їх використання в електронній комерції	13
Висновки до Розділу 1.....	20
РОЗДІЛ 2. ГІБРИДНІ РЕКОМЕНДАЦІЙНІ МОДЕЛІ	21
2.1 Види рекомендаційних моделей та методологія їх побудови.....	21
2.2 Підходи до побудови гібридних рекомендаційних систем.....	29
Висновки до Розділу 2.....	31
РОЗДІЛ 3. ПОБУДОВА ГІБРИДНОЇ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ	33
3.1 Концепція побудови гібридної рекомендаційної системи з використанням демографічної фільтрації	33
3.2 Розробка гібридної рекомендаційної системи на основі демографічної фільтрації	41
ВИСНОВКИ	49
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	51
ДОДАТКИ.....	57

ВСТУП

Актуальність дослідження. За останні десять років електронна комерція набула величезного значення в економічній сфері. Частка продажів електронної комерції збільшується у всьому світі, та становить від 11% до 22% від загального обсягу продажів. [1,2], з прогнозованим щорічним збільшенням щонайменше на 1-2 відсоткових пункти. Саме тому в умовах посилення конкуренції в цій сфері бізнесам слід докладати більше зусиль для залучення нових клієнтів та підтримки їх задоволеності сервісом, що у свою чергу, збільшує можливе використання спеціальних засобів.

Одним із таких інструментів, представлених у 1992 році, є рекомендаційні системи. Основна мета цього інструменту — створити список рекомендацій для об'єктів, які можуть бути цікавими для користувача, виходячи з його особистих характеристик або історії взаємодії з об'єктами (оцінка/перегляд/додавання в кошик). Рекомендаційні системи можуть вплинути на продаж, правильно просуваючи продукт і задовольняючи попит користувачів. Зокрема, полягає в тому, що розробка, використання та підтримка гібридних рекомендаційних систем навіть на малих обсягах доступних даних дозволяє надавати користувачам рекомендації, що можуть підвищити продажі та інші показники бізнесу, зокрема й малого.

Об'єктом дослідження є рекомендаційні системи в сфері електронної комерції.

Предметом дослідження є математичні моделі та технології гібридизації на основі технік колаборативної фільтрації для побудови рекомендаційної системи в сфері електронної комерції.

Метою дослідження розробка алгоритму реалізації гібридної рекомендаційної системи на основі демографічної фільтрації з метою створення індивідуального списку товарів для користувача. В даному випадку під алгоритмом розуміється набір кроків, виконання яких призводить до побудови найбільш ефективної рекомендаційної системи.

Система рекомендацій призначена для реалізації у вигляді скрипту

Python, що можна достатньо легко інтегрувати в додатки або веб-сторінки з платформами бізнесів.

Для досягнення цієї мети необхідно виконати наступні завдання дослідження:

- Узагальнити інформацію щодо динаміку розвитку та поточного стану електронної комерції в Україні та світі
- Дослідити сучасні практики використання рекомендаційних систем в електронній комерції
- Систематизувати знання щодо типів рекомендаційних систем та їх методології
- Проаналізувати існуючі підходи до побудови гібридних рекомендаційних систем, їх переваги та недоліки, а також методи, що можуть використовуватись
- Описати концепцію побудови рекомендаційної системи на основі демографічної фільтрації
- Запропонувати гібридну рекомендаційну систему на основі демографічної фільтрації

Методологічний інструментарій дослідження складають методи наукової абстракції, порівнянь, індукції та дедукції, методи аналізу та синтезу, методи фільтрації контенту, широкий спектр методів математичного моделювання, лінійної алгебри та дослідження операцій (техніки колаборативної фільтрації) та методи програмної реалізації математичного моделювання в програмному середовищі Python.

Наукова новизна: полягає в систематизації теоретичних напрацювань та практики використання рекомендаційних систем в електронній комерції, розробці концепції побудови рекомендаційної системи на основі демографічної фільтрації, а саме місцезоташування користувача, а також її практичної реалізації.

Практична значимість полягає у створенні гібридної рекомендаційної системи, яку можна застосовувати бізнесам у електронній комерції, зокрема тим, яким характерна велика залежність від місцерозташування користувача.

Інформаційною базою дослідження є монографії, наукові статті вітчизняних та зарубіжних вчених, фахові періодичні видання. Зокрема над дослідженнями рекомендаційних систем та їх використанням в електронній комерції працювали К. Єгуда, Б. Сарвар, Дж. Л. Герлокер, Г. Лінден, Ф. Річчі, С. Дж. Парк, Дж. А. Констан, П. Рєзнік, І. Чжан, Е. Кано, Т. Затонацька, Г. Черноус, О. Виноградова та інші.

Структура дослідження. Робота структурно складається зі вступу, трьох розділів, висновків і списку використаних джерел зі 54 найменувань. Повний обсяг роботи становить 48 сторінок, містить 14 рисунків та додатки.

РОЗДІЛ 1. ЕЛЕКТРОННА КОМЕРЦІЯ ТА МІСЦЕ РЕКОМЕНДАЦІЙНИХ СИСТЕМ В ЕЛЕКТРОННІЙ КОМЕРЦІЇ

1.1 Розвиток електронної комерції та її сучасний стан в світі та Україні

Електронна комерція (e-commerce) — це процес купівлі та продажу товарів і послуг через Інтернет або електронні мережі. Вона передбачає здійснення бізнес-операцій, включаючи покупку, продаж, рекламу, маркетинг та постачання товарів і послуг за допомогою електронних засобів зв'язку. Електронна комерція стала широко поширеною завдяки зростанню доступності Інтернету та зростанню електронних платежів. Вона включає в себе різні типи бізнесу, від онлайн-магазинів та маркетплейсів до електронного банківського обслуговування, електронних аукціонів та електронної пошти маркетингу.

Електронна комерція надає ряд переваг як для покупців, так і для продавців. Для покупців вона забезпечує зручний та доступний спосіб здійснення покупок з будь-якого місця за умови наявності підключення до Інтернету. Вони можуть швидко знаходити необхідні товари, порівнювати ціни та отримувати зручні способи доставки. Для продавців електронна комерція відкриває безмежні можливості досягнення глобальної аудиторії і збільшення обсягів продажів. Вони можуть знизити витрати на магазинну оренду та фізичну інфраструктуру, а також ефективно управляти запасами та логістикою. Електронна комерція також дає можливість збирати та аналізувати дані про покупців, що дозволяє створювати персоналізовані пропозиції та покращувати маркетингові стратегії.

У світі електронної комерції існують різні моделі бізнесу, включаючи B2C, B2B, C2C та інші. Крім того, розвиток електронної комерції супроводжується впровадженням нових технологій, таких як мобільна комерція, інтернет речей (IoT), штучний інтелект (AI) та блокчейн, що забезпечують ще більше можливостей для розвитку цього сегмента бізнесу.

Загалом, електронна комерція стала невід'ємною частиною сучасного бізнесу і надає безліч можливостей для здійснення ефективних бізнес-операцій

та залучення клієнтів. Вона продовжує розвиватися та інновувати, що робить її однією з найдинамічніших галузей сучасного світового ринку.

Електронна торгівля в Європі стрімко прогресує й набуває все більшої популярності. Згідно з інформацією від TradingPlatforms, на початку 2020 року в Європі було понад 480.9 млн користувачів онлайн-магазинів, що становило приблизно 64% загального населення. Перша половина 2020 року стала періодом надзвичайної несподіваності для розвитку електронної комерції по всьому світу і в даний момент пройшла одна з найбільш вражаючих трансформацій. За доволі короткий проміжок часу всі усталені уявлення про промисловість були перевернуті — змінилися звички покупців, методи й місця покупки, середній чек та інші фактори, що визначають ринок. Найбільший вплив на це мала пандемія Covid-19, зокрема:

- Пандемія прискорила цифрові трансформації, і цифрові рішення все більше потрібні для продовження економічної та соціальної діяльності віддалено або в змішаному форматі [6]
- Пандемія призвела до розширення електронної комерції на нові фірми, клієнтів і типи продуктів. Це забезпечило клієнтам доступ до значної різноманітності продуктів із зручності та безпеки їхніх домівок і дозволило компаніям продовжувати роботу, незважаючи на обмеження контактів та інші заходи обмеження. [7]
- Трансакції електронної комерції в багатьох країнах частково перейшли від розкоші та послуг до повсякденних потреб, актуальних для великої кількості осіб, що призвело до розширення потенційної бази користувачів в електронній комерції [7]
- Пандемія змусила багато малих підприємств переглянути традиційні бізнес-моделі або закритися назавжди. Нові та існуючі технології висуваються на передній план кожного бізнес-інструменту, і перспективні компанії вирішують питання щодо талантів, які виникають у зв'язку з цими новими наборами навичок цифрового бізнесу. [8]

Загалом пандемія COVID-19 мала як позитивний, так і негативний вплив на електронну комерцію, а також прискорила впровадження цифрових рішень у різних секторах. У 2021 році приблизно 2,14 мільярда людей – понад чверть населення світу – купували товари та послуги онлайн. [9]. Обсяги онлайн-торгівлі у 2021 році досягли понад 850 млрд доларів (рис 1.1), а на 2023 рік очікується зростання до 1.014,53 млрд доларів. Ці дані свідчать про постійне збільшення впливу електронної комерції на європейську економіку й надають нові можливості як покупцям, так і продавцям в цьому секторі ринку. [10]



Рис. 1.1. Обсяги онлайн торгівлі у 2012-2021 роках, млрд доларів.

Джерело: <https://www.cbre.com/insights/articles/omnichannel-what-is-the-share-of-e-commerce-in-overall-retail-sales>

Окрім того, діджиталізація (цифровізація) як глобальний процес використання цифрових технологій для зміни бізнес-моделі та надання нових можливостей для отримання прибутку та збільшення вартості призводить до накопичення великої кількості інформації. За оцінкою Петрока Тейлора загальний обсяг світових даних, які були створені, отримані, скопійовані і спожиті до 2025 року сягатиме понад 180 зетабайт (рис 1.2). Однак ці прогнози вимагають перегляду та оновлення через те, що у 2020 році кількість створених і використаних даних досягла нового рекорду. Зростання було вище, ніж

очікувалося раніше, через збільшений попит, спричинений пандемією COVID-19. [11]

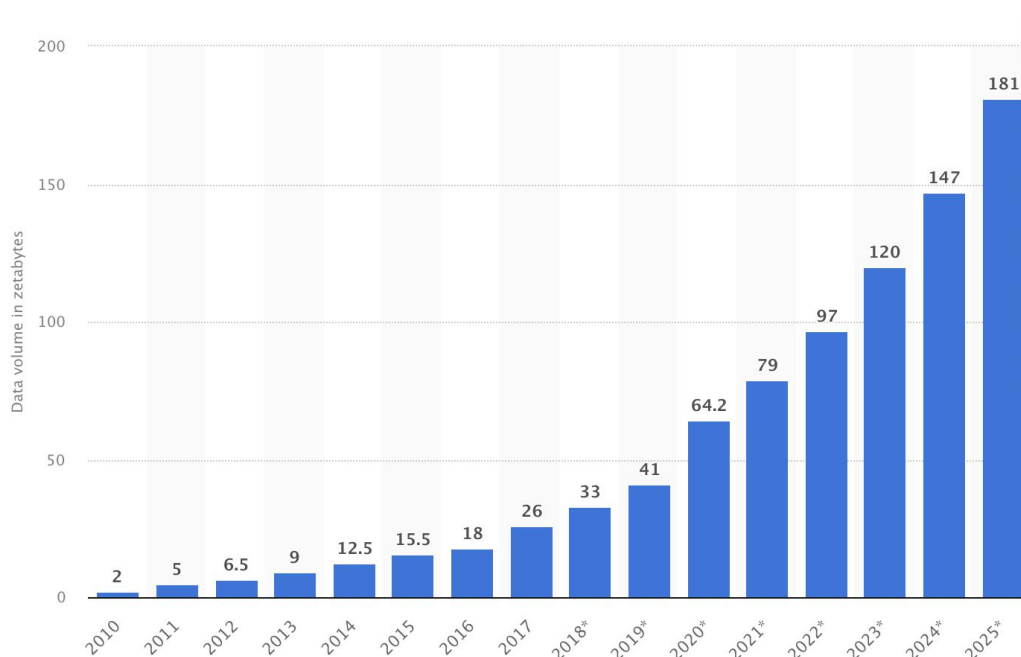


Рис. 1.2. Обсяг даних/інформації, створених, зібраних, скопійованих і використаних у всьому світі з 2010 по 2020 рік, з прогнозами з 2021 по 2025 рік
Джерело: <https://www.statista.com/statistics/871513/worldwide-data-created>

Розвиток електронної комерції в Україні протягом останніх десяти років відбувається дещо інтенсивніше завдяки приєднанню до Зони вільної торгівлі з ЄС та доступу до внутрішніх ринків Європейського Союзу з початку 2016 року. Це відкрило широкі можливості для європейських інвестицій та сприяло зростанню електронної комерції в країні.

Згідно з дослідженнями EcommerceDB [12], український ринок електронної комерції у 2020 році досяг прибутку в розмірі 838 мільйонів доларів, що розмістило Україну на 66 місці у світовому рейтингу. За статистикою, на початок 2021 року близько 68% населення України, що становить близько 30 млн громадян, мають доступ до Інтернету. Дев'ять із десяти користувачів Інтернету здійснюють покупки онлайн, що свідчить про популярність електронної комерції серед українських споживачів.

Україна також має кілька популярних онлайн-магазинів, таких як Rozetka, Epicentr та Allo, які залучають велику кількість відвідувачів. Багато українців

також здійснюють покупки на міжнародних платформах, таких як AliExpress, Amazon та eBay.

Найпопулярнішими категоріями товарів у 2020 році були електроніка, одяг та взуття, а також меблі та техніка. Ці категорії займають значну частку в загальному обсязі продажів у галузі електронної комерції. [13]

При цьому повномасштабне вторгнення також значно вплинуло на стан електронної комерції в Україні в 2022 році. Зокрема дослідження присвячене аналізу впливу війни на інтернет-торгівлю засвідчує, що під час воєнних дій багато мереж продажів зазнали значних збитків. У березні онлайн-продажі майже зникли, але вже в квітні електронна комерція почала оживати та показувати зростання. У липні спостерігалось відновлення долі онлайн-продажів порівняно з офлайн-продажами до передвоєнних показників. Війна наклала (і досі накладає) певні обмеження на розвиток електронного бізнесу, зокрема неможливість працювати бізнесам, що знаходяться на тимчасово окупованих територіях. [14]

Розширення електронної комерції відкриває безліч можливостей для бізнесу. Впровадження сучасних технологій дозволяє підприємствам краще розуміти потреби своїх клієнтів і забезпечувати їм персоналізовані послуги. Різноманітні маркетингові інструменти допомагають збільшити взаємодію зі споживачами та ефективно просувати товари та послуги.

Прогнозується, що продажі через Інтернет у світі до 2024 року будуть становити 21,8% від загальної кількості покупок населення. Цей ріст свідчить про те, що електронна комерція є дедалі більш популярною та затребуваною формою споживання товарів і послуг. Щодо прогнозування перспектив розвитку електронної комерції впродовж наступних 10 років – це є складним завданням, оскільки ця галузь динамічно розвивається та залежить від багатьох факторів. Однак, на основі поточних тенденцій та досліджень можна сформулювати кілька ключових напрямків, які ймовірно будуть визначати майбутність електронної комерції:

- Зростання мобільної комерції: За останні роки спостерігається стрімкий розвиток мобільних пристроїв та їх використання для покупок онлайн. Очікується, що цей тренд продовжиться, і мобільна комерція стане ще більш популярною. Зростання швидкості Інтернету, впровадження технологій 5G та зручність мобільних додатків сприятимуть збільшенню обсягу мобільних продажів.
- Збільшення впровадження штучного інтелекту та аналітики даних: Штучний інтелект та аналітика даних вже сьогодні використовуються для персоналізованих рекомендацій, прогнозування попиту та оптимізації процесів електронної комерції. Протягом наступних 10 років ці технології продовжать розвиватися, допомагаючи покращувати користувацький досвід, ефективність та прибутковість електронних магазинів.
- Розширення використання віртуальної та доповненої реальності: Віртуальна та доповнена реальність вже знаходять застосування в електронній комерції, дозволяючи клієнтам переглядати товари перед покупкою, випробувати їх або створювати віртуальний образ себе у новому вбранні. З очікуваним розвитком цих технологій в наступному десятилітті, вони стануть ще більш доступними та популярними серед споживачів.
- Зростання ролі соціальних медіа: Соціальні медіа вже сьогодні використовуються для просування товарів та залучення клієнтів, але їх вплив на електронну комерцію буде зростати. Відомості про споживачів, отримані з соціальних медіа, допоможуть бізнесам створювати більш цільовану рекламу та залучати більше клієнтів.
- Зміни в останній милі: Останній миль - етап доставки товару до споживача - є критичним для задоволення покупців. У майбутньому, з розвитком дронів, автономних автомобілів та інших інноваційних технологій, очікується поліпшення швидкості та якості доставки, зменшення витрат та покращення досвіду клієнтів. [15,16,17]

1.2 Рекомендаційні системи та їх використання в електронній комерції

Стрімке зростання електронної комерції за останні десять років призвело до ряду проблем, що стосуються як бізнесу так і споживача, серед яких можна виокремити проблеми пов'язані з:

- Забезпеченням безпеки і захисту персональних даних покупців, оскільки існує ризик витоку конфіденційної інформації, крадіжки особистих даних та шахрайства в Інтернеті.
- Відсутністю фізичного контакту і особистої взаємодії, оскільки електронна комерція позбавляє споживачів можливості особистого огляду товару, примірки або торгівлі
- Доставкою товарів, що є важливою складовою електронної комерції. Обмеження щодо розміру товарів, термінів та умов доставки, інших параметрів товару.
- Конкуренцією і перенасиченістю ринку. Електронна комерція є дуже конкурентною галуззю, і існує велика кількість онлайн-магазинів, які пропонують схожі або ідентичні товари і послуги, наприклад сервіси прокату фільмів чи купівлі музики. Це створює проблему перенасиченості ринку, де важко виділитися і залучити увагу споживачів.

Остання з перелічених проблем також включає в себе перенасичення нерелевантною інформацією, що і покликані вирішити рекомендаційні системи. Різні вчені по-різному визначали системи рекомендацій. Зокрема у праці науковців F.O. Isinkaye, Y.O. Folajimi та V.A. Ojokoh (2015) системи рекомендацій були визначені як «засіб допомоги та розширення соціального процесу використання рекомендацій інших для прийняття вибору за відсутності попередніх знань чи досвіду» [18]. У енциклопедії з машинного навчання (Encyclopedia of Machine Learning) зазначено, що «метою системи рекомендацій є створення значущих рекомендацій для колекції користувачів щодо предметів або продуктів, які можуть їх зацікавити» [19]. «Рекомендаційні системи – це окремий клас веб-додатків для прогнозування вибору користувача.» [20]. Серед інших визначень також можна знайти формулювання пов'язані з «передбаченням

оцінки, яку користувач може поставити продукту» [21] чи «ранжуванням продуктів чи користувачів» [22]. Загалом системи рекомендацій діють як інструменти фільтрації інформації, пропонуючи користувачам відповідний персоналізований вміст або інформацію. Ці системи передбачають найімовірніший продукт, який користувачі, найімовірніше, придбають і який представляє інтерес.

Використання систем рекомендацій в електронній комерції з часом досліджували різні вчені та дослідники. Використання комп'ютера для рекомендації найкращого предмету (перша реалізація концепції рекомендаційних систем) з'явилася в 1979 році в системі під назвою Grundy, що було електронним бібліотекарем, який надавав користувачам пропозиції щодо того, які книги читати. Пік досліджень у сфері рекомендаційних систем відбувся наприкінці 1990-х, коли Amazon запустили свій метод Collaborative Filtering (CF), який успішно підвищив лояльність їх клієнтів. [23].

Пізніше було розроблено велику кількість підходів для неперсоналізованих і персоналізованих рекомендаційних систем, зокрема:

- підхід асоціації продуктів (Poriya та ін., 2014) [24]
- контент-фільтрування інформації (Hooda та ін., 2014; Woldan та ін., 2020) [25,26]
- спільна фільтрація (на основі користувача (Resnik та ін., 1994; Li & Li, 2020) та на основі елементів (Sarwar та ін., 2001; Ujkani та ін., 2020; Ajaegbu, 2021) [27,28]
- багатокритеріальні системи рекомендацій (Shambour, 2021; Smirnov & Popomarev, 2020) тощо. [29].

Рекомендаційні системи є важливими інструментами в електронній комерції, соціальних мережах та стрімінгових сервісах. Вони допомагають передбачити та пропонувати користувачам товари, контент або послуги, які найбільш відповідають їхнім потребам і інтересам. Ці системи можуть бути розроблені, щоб рекомендувати продукти на основі різних факторів, таких як

найбільші продавці на веб-сайті або найбільш вірогідний продукт, який користувач зацікавлений у покупці, на основі різноманітних факторів та атрибутів користувача (-ів).

Використання рекомендаційних систем має ряд переваг як для бізнесу так і для покупців. Однією з переваг рекомендаційних систем для підприємців є збільшення продажів. Ці системи допомагають збільшити конверсію користувачів, оскільки рекомендації спонукають їх купувати більше товарів. Крім того, рекомендаційні системи дозволяють підприємцям продавати різні товари, включаючи ті, які користувачі можуть пропустити або не помітити без таких рекомендацій. Також системи рекомендацій допомагають збільшити кількість товарів для крос-продажу, пропонуючи аксесуари або додаткові товари, пов'язані з покупкою. Рекомендаційні системи також сприяють підвищенню лояльності клієнтів до магазину. Зручний сервіс з персоналізованими рекомендаціями задовольняє потреби користувачів, що часто призводить до повторних покупок і збільшення їхньої відданості певному магазину або бренду. Крім того, рекомендаційні системи допомагають підприємцям краще розуміти споживчий попит. Аналізуючи дані про вибір користувачів і їхні уподобання, підприємці можуть управляти запасами і придбати товари, які мають попит, забезпечуючи ефективну торговельну політику. Для споживачів головна цінність рекомендаційних систем полягає в тому, що ці системи можуть виступати в ролі особистих покупців і торгових помічників, допомагаючи клієнтам знаходити товари, які вони шукають, не витрачаючи час на безрезультатні поїздки до торгового центру. [30]

Досить актуальним є зараз дослідження впливу використання рекомендаційних систем на покупців та їх поведінку. Зокрема Silvana Dakduk, Enrique ter Horst, Zuleyma Santalla, German Molina and José Malavé (2017) у своїй роботі досліджуючи ринок електронної комерції в Колумбії дійшли висновку, що споживачі вважають електронну комерцію легкою та вважають, що використання Інтернету для покупок є вигідним, покращуючи результат покупки, роблячи процес купівлі більш ефективним та швидким. Обидва аспекти сприяють

збільшенню ймовірності позитивної оцінки інтернет-покупок, що, у свою чергу, збільшує ймовірність онлайн-покупки в найближчому майбутньому. Результати цього дослідження, показують, що ставлення залишається основною змінною у прогнозуванні появи онлайн-купівельної поведінки. Складовими ставлення є переконання, що стосуються атрибутів онлайн-покупок, наявність попередніх сприятливих результатів покупок, легкості та корисності використання, та їх оцінки іншими особами, які мають відношення до предмета. [31] Що в свою чергу вказує на необхідність впровадження рекомендаційних систем для покращення ставлення покупців до бізнесу/платформи.

Зараз системи рекомендацій в електронній комерції перетворюються з простих інновацій кількох веб-магазинів на серйозні бізнес-інструменти, які перебудовують увесь ландшафт електронної комерції. Багато гігантів електронної комерції використовують системи рекомендацій, щоб допомогти своїм клієнтам, щоб вони могли легко купувати продукти та збільшувати продажі.

Відомими компаніями, що використовують рекомендаційні системи в електронній комерції є [32]:

- Amazon, що використовують системи рекомендацій у книжковому розділі Amazon.com, що містить два окремі списки рекомендацій. Перший рекомендує книги, які часто купують клієнти, які придбали вибрану книгу. Другий рекомендує авторів, книги яких часто купують клієнти, які придбали твори автора вибраної книги. Також використання цих рекомендаційних систем доповнюється періодичними розсилками електронною поштою, щоб сповістити передплатників про їхні останні рекомендації в категоріях, на які ви підписалися. Рекомендації також здійснюються на основі особистих оцінок книг користувачем та його коментарів [33]
- CDNOW з їх функцією Album Advisor. Клієнти вводили від одного до трьох виконавців. У свою чергу, система рекомендувала 10 альбомів, пов'язаних із виконавцями, про яких йдеться. Також клієнти вказували, які альбоми їм належать і які виконавці є їхніми улюбленими. Система автоматично додавала

рекомендовані альбоми до покупок користувача. При цьому користувач міг оцінити чи подобається йому підбірка та окремі елементи в ній. Коли клієнт запитував рекомендації, система передбачала 6 альбомів, які можуть сподобатися клієнту, на основі того, що вже є у власності. Альбоми змінювались на основі відгуків користувачів. [34]

- Профіль зворотного зв'язку eBay. Функція профілю зворотного зв'язку дозволяє як покупцям, так і продавцям вносити свій внесок у профілі відгуків інших клієнтів, з якими вони співпрацювали. Зворотній зв'язок використовується для забезпечення системи рекомендацій для покупців, які можуть переглядати профіль продавців. Цей профіль складається з таблиці кількості кожного рейтингу за останні 7 днів, минулого місяця та останніх 6 місяців, а також загального підсумку (наприклад, 867 позитивних оцінок від 776 унікальних клієнтів). За додатковим запитом клієнти можуть переглядати індивідуальні оцінки та коментарі продавців. [35]

- Компанія Levi Strauss, що пропонує функцію Style Finder, яка дозволяє клієнтам веб-сайту отримувати рекомендації щодо предметів одягу Levi's. Клієнтам пропонується пройти опитувальник щодо їх вподобань в певних категоріях за 7-бальною шкалою. Після введення мінімальної кількості оцінок клієнти можуть вибрати «отримати рекомендації». Тут їм надано мініатюри 6 предметів рекомендованого одягу. Клієнти можуть надати відгук, що дозволяє їм вводити оцінку думки для рекомендованого предмета одягу. Відгуки можуть змінити один або всі шість рекомендованих пунктів.

- Moviefinder.com використовує функцію Match Maker, що дозволяє клієнтам знаходити фільми з подібним «настроєм, темою, жанром або акторським складом» до певного фільму. На інформаційній сторінці відповідного фільму клієнти натискають піктограму Match Maker і отримують список рекомендованих фільмів, а також посилання на інші фільми режисера оригінального фільму та ключових акторів. Друга функція - We Predict рекомендує фільми клієнтам на основі їхніх раніше зазначених інтересів. Клієнти оцінюють фільми, які вони переглянули, за 5-бальною шкалою від А до F. Як різновид цього, клієнти можуть

використовувати Powerfind для пошуку найкращих варіантів на основі синтаксичних критеріїв, таких як жанр, режисери чи актори, і вибрати, щоб вони сортувалися за їхнім персоналізованим

Серед українських бізнесів, що використовують рекомендаційні системи найбільшими є:

- Rozetka [36]. Для неавторизованого користувача на цьому сайті доступні рекомендації в різних категоріях, таких як акційні пропозиції, гарячі новини, схожі товари від інших продавців та популярні товари в даний момент. Для авторизованого користувача, окрім вищезазначеного, пропонуються додаткові категорії, такі як останні переглянуті товари, бестселери в категорії (нещодавно переглянуті), інші товари на вибір та рекомендації, які враховують особисті погляди користувача. Додані категорії аналізують пряму поведінку користувача, такі як конкретні продукти, схожі продукти або категорії, які вони переглядають.
- Епіцентр [37]. Для неавторизованого користувача цей сайт пропонує рекомендації в таких категоріях, як товар дня, популярні категорії, акції та найцікавіший. Ті ж самі категорії доступні і для авторизованого користувача. Епіцентр має додаткові рекомендації для категорій, які співпадають для авторизованих і неавторизованих користувачів. Користувачі позитивно сприймають факт, що в загальних категоріях вибираються рекомендації спеціально для них. Аналіз показує, що Епіцентр ймовірно використовує лише пряму інформацію від користувачів та їх географічне положення.
- Алло [38]. Для неавторизованого користувача на цьому сайті доступні такі категорії рекомендацій: продажі-лідери, новини, акції та найкраща ціна. Для авторизованого користувача додатково доступна категорія "Ви дивилися". Решта категорій товарів для цього сайту залишаються без змін.

Однак з іншого боку використання рекомендаційних систем в усіх бізнесах електронної комерції неможливе через ряд причин. Найголовніша, малі бізнеси не мають достатньо ресурсу, аби розробляти та/чи забезпечувати постійне функціонування потужної рекомендаційної системи, а також її оновлення та

оптимізацію. Також проблемою рекомендаційних систем є потенційне пряме чи опосередковане втручання третіх осіб в їх роботу через маніпуляцію даними на основі яких здійснюється фільтрація товарів для рекомендації. Зокрема, це може проявлятися через навмисне «завищення/заниження» оцінок. Що означає або неможливість коректного використання рекомендаційних систем на цій платформі, або необхідність оптимізації методів, що використовуються в рекомендаційній системі, або використання додаткових ресурсів компанії на створення запобіжних заходів щодо такого виду шахрайства.

Зокрема, проблеми використання рекомендаційних систем досліджували Shah Khusro, Zafar Ali та Irfan Ullah (2016). Вони виокремили наступні проблеми:

- Проблема холодного запуску – додання нових користувачів чи товарів в систему
- Синонімія - предмет представлено двома або більше різними назвами або записами, що мають подібні значення
- Shilling Attacks – навмисне надання неправдивих оцінок
- Конфіденційність - передача особистої інформації системам рекомендацій призводить до кращих рекомендаційних послуг, але може призвести до проблем конфіденційності та безпеки даних.
- Обмежений аналіз вмісту та надмірна спеціалізація - обмежений аналіз вмісту призводить до надмірної спеціалізації, у якій рекомендують елементи, тісно пов'язані з профілем користувача, але не пропонують нові предмети.
- Сіра вівця - зустрічається в чистих системах Collaborative Filtering, де думки користувача не збігаються не збігаються з жодною групою, і тому він не може отримати переваги від рекомендацій.
- Розрідженість - наявність величезного обсягу даних про товари в каталозі та небажання користувачів оцінювати товари роблять матрицю розпорошеного профілю, що призводить до меншої точності.
- Масштабованість - швидкість зростання алгоритмів найближчого сусіда демонструє лінійну залежність від кількості елементів і кількості користувачів. Типовій рекомендаційній системі стає важко обробляти такі масштабні дані.

- Проблема затримки – рекомендаційні системи на основі Collaborative Filtering стикаються з проблемою затримки, коли нові елементи додаються частіше до бази даних, що призводить до того, що протягом певного часу пропонуються лише вже оцінені елементи оскільки нещодавно додані елементи ще не оцінені та інші. [39]

Багато з перелічених проблем є властивими тому чи іншому методу, що використовується рекомендаційною системою. Зараз більшість рекомендаційних систем в електронній комерції використовує лише один тип даних, які клієнти створюють, коли здійснюють покупки (наприклад, прями оцінки, коментарі або історичні дані). При цьому поєднання кількох методів дозволяє усунути або зменшити вплив цих проблем на роботу системи. Саме тому сьогодні найбільшої актуальності набуває питання розробки та використання гібридних рекомендаційних систем, а також їх вдосконалення.

Висновки до Розділу 1

У цьому розділі окреслюється поточний стан та динаміка розвитку електронної комерції, а також актуальність використання рекомендаційних систем для підприємств електронної комерції та аналізується низка існуючих досліджень на цю тему.

Ключовою частиною є розгляд та вивчення переваг та недоліків використання рекомендаційних систем у електронній комерції. У ході роботи були проаналізовані основні проблеми, які можуть супроводжувати використання рекомендаційних систем.

У результаті дослідження було виявлено, що саме гібридні рекомендаційні системи дозволяють нівелювати або зменшити більшість негативних факторів, що супроводжують роботу рекомендаційної системи.

РОЗДІЛ 2. ГІБРИДНІ РЕКОМЕНДАЦІЙНІ МОДЕЛІ

2.1 Види рекомендаційних моделей та методологія їх побудови

Перш за все, рекомендаційна система збирає інформацію для полегшення процесу прийняття рішень щодо показу та рекомендацій вибору товарів, вмісту або навіть осіб у соціальних мережах або пропозиціях від друзів в соціальних мережах. Ця зібрана інформація відноситься до трьох складових:

- об'єктів, які рекомендуються
- користувача, на якого впливають ці рекомендації
- та інших користувачів, які вже взаємодіяли з платформою.

Інформацію про користувача поділяють на явну і неявну. Явна інформація складається з даних, які надають користувачі у відповідь на запитання або запити, сюди відносяться заповнення профілів, опитувальників інтересів, коментарі користувачів та їх оцінки. Неявна інформація створюється користувачами спонтанно та пов'язана з їхньою поведінкою під час користування платформою/застосунком, як от час проведений на сторінці, навігація, пошукові запити, кліки на товари чи інформаційні блоки тощо.

Рекомендаційні системи можна поділити на персоналізовані та неперсоналізовані.

Системи рекомендацій, засновані на неперсоналізованому підході, в основному рекомендують те саме всім користувачам, а це означає, що система рекламує тільки певні товари, а користувач отримує не завжди цікаві для нього рекомендації. Незначною перевагою є мінімальна взаємодія користувача з системою, що дозволяє останній обслуговувати нових користувачів відразу при їх першому відвідуванні програми, не чекаючи від них будь-яких дій.

Неперсоналізований підхід може бути реалізований багатьма методами, серед яких можна виділити два основні:

1. Метод статистичних даних. Створення рекомендаційного списку на основі статистичних даних працює за принципом: «Якщо всім щось подобається, це сподобається і вам». Наприклад, можна використовувати середню оцінку продукту серед користувачів.

2. Спосіб просування товару за атрибутами (наприклад дата додавання товарів у додаток, популярність товарів серед користувачів). Коли товари рекламуються за атрибутами, список рекомендацій може складатися з «новинок» або товарів, на які продавці вручну пропонує користувачам звернути увагу.

Персоналізовані рекомендаційних системи в свою чергу також можна поділити на два типи: колаборативна фільтрація та контентна фільтрація.

У системах рекомендацій на основі вмісту (або контентної фільтрації) всі елементи даних збираються в різні профілі предметів на основі їх опису або характеристик. Наприклад, у випадку з книгою, функціями будуть автор, видавець тощо. У випадку фільму функціями будуть режисер, актор і т. д. Коли користувач дає позитивну оцінку предмету, то інші елементи, присутні в цьому профілі елемента, об'єднуються разом для створення профілю користувача. Профіль користувача об'єднує всі профілі предметів, товари які оцінені користувачем позитивно. Елементи, присутні в цьому профілі користувача, потім рекомендуються користувачеві.

Одним із недоліків цього підходу є те, що він вимагає глибоких знань про предмет та його функції для точної рекомендації. Ці знання чи інформація можуть не бути завжди наявності для всіх товарів. Крім того, цей підхід має обмежені можливості для розширення існуючого вибору або інтересів користувачів (проблема спеціалізації). Однак такий підхід має багато переваг. Оскільки уподобання користувача, як правило, змінюються з часом, цей підхід має швидку здатність динамічно адаптуватися до мінливих уподобань користувача. Оскільки один профіль користувача є специфічним лише для цього користувача, цей алгоритм не потребує даних профілю будь-яких інших користувачів оскільки вони не впливають на процес рекомендацій. Це забезпечує безпеку та конфіденційність даних користувачів. Якщо нові елементи мають достатньо опису, на основі вмісту методи можуть подолати проблему холодного запуску, тобто ця техніка може рекомендувати продукт, навіть якщо цей елемент раніше не оцінював жоден користувач. Підходи фільтрації на основі вмісту найчастіше використовуються для рекомендацій новин, публікацій, веб-сторінок

тощо. Отже, контентна фільтрація використовує параметри контенту та користувачів для знаходження схожості і рекомендацій.

Метод колаборативної фільтрації для створення рекомендацій заснований на порівнянні користувачів з подібними інтересами та уподобаннями. Він враховує схожість між їхніми профілями, що базується на попередній поведінці користувача. Цей метод може бути побудований на основі лише поведінки даного користувача або залучати інших користувачів з подібними характеристиками. При виборі рекомендацій цей метод використовує рейтинги, які користувачі виставляють, а також їх перегляди та дії. Формалізовано, система рекомендацій має вигляд матриці $m \times n$ $P = \{r\}$, (2.1) що репрезентує оцінки користувачів $U = \{u_1, \dots\}$ (рядок відповідає користувачу) певним товарам, $I = \{i_1, \dots\}$ (по стовпчиках). Два основних завдання рекомендаційної системи – спрогнозувати рейтинги для всіх об'єктів або створити список найкращих N рекомендацій.

Техніка колаборативної фільтрації починається з пошуку групи або колекції користувачів X , чий вподобання, і антипатії подібні до користувача A . X називають сусідами A . Ті елементи, які подобаються більшості користувачів у X , потім рекомендуються користувачеві A . Ефективність алгоритму залежить від того, наскільки точно алгоритм може знайти околиці цільового користувача. Традиційно рекомендаційні системи, засновані на колаборативній фільтрації, страждають від проблеми холодного запуску та конфіденційності, оскільки існує потреба спільно використовувати дані користувачів. Однак для підходів колаборативного фільтрування не потрібні знання про функції елемента для створення рекомендації. Крім того, такий підхід може допомогти розширити на існуючі інтереси користувача, відкриваючи нові елементи.

Існує два основних методи CF: метод найближчого сусіда (the nearest neighbour, NN) і метод латентного фактору. Метод NN заснований на принципі, що споживачі, які віддали перевагу подібним об'єктам у минулому схильні віддавати перевагу подібним об'єктам у майбутньому.

Визначення «подібності» користувачів може бути реалізоване завдяки кільком методам. За умови що існує k товарів, і оцінки кожного користувача

можна представити у вигляді вектора у k-мірному просторі можна порівняти вектори оцінок для знаходження «подібних» користувачів обчисливши один з наступних показників, які можна використовувати як функції подібності:

- косинус подібності – обчислення косинусу кута між двома векторами. Найбільш подібні вектори мають значення косинуса 1, а найменш подібні – 0.

$$sim = \cos(\theta) = \frac{A * B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}, \quad (2.2)$$

де A_i та B_i є координатами векторів А та В

- коефіцієнт кореляції Пірсона – результат кореляції слугує показником «подібності» користувачів. Якщо кореляція дорівнює -1, то користувачі абсолютно не схожі, а якщо 1 – то максимально схожі між собою

$$r_{xy} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^m (y_i - \bar{y})^2}} = \frac{\cos(x, y)}{\sqrt{s_x^2 s_y^2}}, \quad (2.3)$$

де \bar{x} , \bar{y} - вибіркові середні,

x^m, y^m, s_x^2, s_y^2 - вибіркові дисперсії,

$$r_{xy} \in [-1; 1]$$

- Коефіцієнт Танімото – набуває значення від 0 до 1, найдоцільніше використовувати для бінарних оцінок (наприклад «подобається/не подобається»)

$$T(A, B) = \frac{N_c}{N_a + N_b - N_c}, \quad (2.4)$$

де N_a - кількість оцінок для товарів у множині А,

N_b - кількість оцінок для товарів у множині В,

N_c - кількість спільних для множин А і В оцінок для товарів.

- Манхеттенська відстань

$$d(A, B) = \sum_{i=1}^n |A_i - B_i|, \quad (2.5)$$

де A, B – користувачі та їх оцінки,

n – кількість предметів у матриці

- Евклідова відстань

$$d(A, B) = \sqrt{\sum_{i=1}^n (|A_i - B_i|)^2}, \quad (2.6)$$

де A, B – користувачі та їх оцінки,

n – кількість предметів у матриці

Колаборативні підходи також поділяються на два типи: підходи на основі пам'яті та підходи на основі моделі. Підходи, що базуються на пам'яті, рекомендують нові предмети, беручи до уваги переваги свого оточення. Вони безпосередньо використовують матрицю корисностей для передбачення. У цьому підході першим кроком є побудова моделі. Модель дорівнює функції, яка приймає матрицю корисності як вхідні дані. $Model = f(utility\ matrix)$, (2.7)

Потім даються рекомендації на основі функції, яка використовує модель і профіль користувача як вхід. Тут можемо давати рекомендації лише користувачам, чий профіль користувача належить до матриці корисностей. Тому, щоб надати рекомендації для нового користувача, профіль користувача має бути доданий до матриці корисності, а матриця подібності повинна бути обчислена повторно, що робить обчислення цієї техніки важкими.

$$Recommendation = f(defned\ model, user\ profle), \quad (2.8)$$

де $user\ profle \in utility\ matrix$

Підходи до колаборативної фільтрації на основі пам'яті знову поділяються на два типи: колаборативна фільтрація на основі користувачів (user-based CF) і колаборативна фільтрація на основі елементів (item-based CF). У підході, орієнтованому на користувача, оцінка користувача нового елемента обчислюється шляхом пошуку інших користувачів із оточення користувачів, які

раніше оцінювали той самий елемент. Якщо новий предмет отримує позитивні оцінки від «подібних» користувачів, новий продукт рекомендовано користувачеві А. В колаборативній фільтрації на основі користувачів дуже важливим елементом є саме вибір функції подібності, що були згадані вище. [40] Зокрема науковці Spertus E, Sahami M., and Buyukkokten O. провели масштабне дослідження, щоб оцінити шість різних функцій подібності в контексті соціальної мережі Orkut. І хоча їхні результати можуть бути упередженими через особливі умови проведення експерименту, варто відзначити, що найкращою реакцією на рекомендації були ті, які були сформовані за допомогою косинусної подібності. [41] З іншого боку, американські науковці Lathia N., Hailes S., та Carra L. також провели дослідження, в ході якого вони дійшли висновку, що в загальному випадку на точність передбачення рекомендаційної системи не впливає вибір міри подібності. В контексті їхньої роботи використання випадкової міри подібності іноді давало кращі результати, ніж використання будь-якого з добре відомих підходів. [42]. При цьому використання рекомендаційної системи колаборативної фільтрації на основі користувачів стикається з проблемами під час великого масштабування кількості користувачів.

У підході, заснованому на елементах, формування пропозицій засноване на зв'язку між об'єктами, який виведено з рейтингової матриці. Припущення яке лежить в основі цього підходу полягає в тому, що користувачі віддадуть перевагу об'єктам, схожим на інші об'єкти, які їм подобаються. Будується так звана «околиця» предметів, що складається з усіх подібних елементів, які користувач оцінив раніше. Тоді оцінка цього користувача для іншого нового елемента прогнозується шляхом обчислення середньозваженого значення всіх оцінок, присутніх у околиці елемента. IBCF знаходить найбільш схожі об'єкти (використовується k-NN) для створення прогнозів і рекомендацій. Перевагою методу є можливість заздалегідь розрахувати матрицю подібності.

Системи на основі моделей використовують різноманітні алгоритми інтелектуального аналізу даних і машинного навчання для розробки моделей прогнозування оцінки користувача для товару без оцінки. Під час обчислення

рекомендацій вони не покладаються на повний набір даних, а витягують функції з набору даних для обчислення моделі. Звідси й назва — модельна техніка. Ці методи також вимагають двох кроків для прогнозування: перший крок полягає в створенні моделі, а другий крок полягає в прогнозуванні рейтингів за допомогою функції (f), яка приймає модель, визначену на першому кроці, і профіль користувача як вхідні дані. Методи на основі моделі не вимагають додавання профілю нового користувача до матриці корисності, перш ніж робити прогнози, що дозволяє давати рекомендації навіть користувачам, які відсутні в моделі. Системи на основі моделі більш ефективні для групових рекомендацій. Вони можуть швидко рекомендувати групу предметів за допомогою попередньо навченої моделі. Точність цієї техніки значною мірою залежить від ефективності основного алгоритму навчання, який використовується для створення моделі. Методи, засновані на моделях, здатні вирішити деякі традиційні проблеми систем рекомендацій, такі як розрідженість і масштабованість за допомогою методів зменшення розмірності та навчання моделі. [21].

Алгоритми колаборативної та контентної фільтрації є достатньо адаптивними, адже вони використовують тільки базову інформація про елемент. При цьому такі рекомендаційні системи надають рекомендації низького ступеня участі.

Методи латентного семантичного аналізу. Такий спосіб аналізу зазвичай використовується для класифікації текстів, статей, відгуків та ін. Метод полягає у тому, що спочатку з тексту виключаються слова, що не мають сенсового навантаження (наприклад займенники чи сполучники) та є у будь-яких текстах (найпоширеніші нерелевантні для певної галузі слова). Після чого, в залежності від мови тексту може виникнути потреба в стеммінгу – «відкидання» закінчень (як от часових чи особових, зокрема характерно для української мови), початкової форми слова. Стемінг - це підхід, заснований на правилах. що полягає у пошуку мовної форми слова (кореня) шляхом відсікання його початку та кінця за заданими правилами. Правила обрізки визначені і зазвичай складаються з набору регулярних виразів. Основним недоліком цього підходу

є можлива втрата інформації під час обрізання частин слова. Іншим способом є лематизація – це словниковий підхід. Суть підходу полягає в морфологічному аналізі слів за допомогою словників, в які машина може зазирнути, щоб порівняти поточну форму слова з його словниковою формою (лемою).

Після цього виключаються слова, що зустрічаються менше 2 разів. Після первинної обробки даних проводиться сингулярне розкладання

$$M = U * W * V_t, \quad (2.9)$$

де, U , V_t – ортогональні матриці,

W – діагональна матриця

В результаті аналізу отриманих можна ідентифікувати перелік ключових слів, на основі яких виконуватиметься групування контенту чи товарів за їх тематикою. Такі рекомендаційні системи покликані надавати рекомендації, що знаходяться у сфері інтересів користувача. Однак це може поступово призвести до проблеми спеціалізації.

Також можна виокремити демографічну фільтрацію, яка класифікує користувачів у відповідні групи на основі їхньої демографічної інформації. Демографічна фільтрація, як і колаборативна фільтрація, приймає рішення на основі подібності користувачів, але використовує демографічні дані про користувачів, а не дані про поведінку.

Демографічна фільтрація створює групи користувачів на основі демографічних даних, а потім створює список рекомендацій, відстежуючи загальну купівельну поведінку користувачів у цих групах. Системи демографічної фільтрації використовують демографічну інформацію, таку як вік, стать та освіту, щоб зіставити користувача з групою. Новому користувачеві спочатку рекомендуються об'єкти, цікаві для групи, до якої його відніс алгоритм, проте при взаємодії із системою він також починає впливати на список рекомендацій своєї групи.

Прикладом демографічної фільтрації є показ оголошення про продаж певної нерухомості у тому самому місті, де користувач авторизувався в додатку. Такі системи рекомендацій, засновані на знаннях, рекомендують об'єкти на

основі явних користувацьких переваг, наприклад, за допомогою анкети.

2.2 Підходи до побудови гібридних рекомендаційних систем

Гібридна фільтрація (гібридна техніка фільтрації) – це сукупність двох або більше технік, які використовуються разом для мінімізації чи нівелювання обмежень окремих методів рекомендацій. Поєднання різноманітних технік можна виконувати різними способами. Гібридний алгоритм може включати результати, отримані за допомогою окремих методів, або використовувати фільтрацію на основі вмісту у колаборативному методі або використовувати техніку спільного фільтрування в контенто-орієнтованому методі. Це гібридне включення різних технік, як правило, призводить до збільшення продуктивності та підвищеної точності у багатьох рекомендаційних системах.

Загальна концепція побудови гібридної рекомендаційної системи включає в себе наступні завдання:

- Визначення проблеми, яку потрібно вирішити, наприклад, рекомендувати продукти користувачам на основі їхніх уподобань.
- Зібрати дані про поведінку користувачів, як-от покупки, оцінки та відгуки, а також атрибути товару, як-от ціна, категорія та опис. Ці дані можуть бути отримані з різних джерел, включаючи внутрішню базу даних, журнали веб-серверів та соціальні медіа.
- Попередня обробка даних для видалення шуму, обробки відсутніх значень і нормалізації даних.
- Вибір підходів, які будуть використовуватись у гібридній системі, як-от фільтрація на основі вмісту, колаборативна фільтрація або демографічна фільтрація.
- Запровадження підходів та навчання моделі на попередньо оброблених даних.
- Комбінування підходів, щоб надати кращі рекомендації користувачам. Кожна модель може мати свої переваги та обмеження, тому гібридна система комбінує їх для отримання кращих результатів. Це можна зробити, зробивши

передбачення на основі вмісту та на основі співпраці окремо, а потім поєднавши їх, додавши можливості на основі вмісту до підходу на основі співпраці (і навпаки), або об'єднавши підходи в одну модель.

- Оцінка продуктивності гібридної системи рекомендацій за допомогою різних показників, таких як точність, повнота.
- Удосконалення гібридної системи рекомендацій на основі результатів оцінки та відгуків користувачів. [43]

Поєднання кількох рекомендаційних систем в гібридну може бути досягнуто за допомогою різних методів, зокрема:

- Ансамблювання моделей. Використання ансамблю моделей дозволяє комбінувати прогнози з різних рекомендаційних систем для отримання остаточного рекомендаційного списку. Наприклад, можна застосувати ансамбль з використанням алгоритму голосування більшості, стекінгу моделей або використанням ансамблю алгоритмів. До компаній, що використовують такий метод належать Netflix та TripAdvisor [44]
- Взважене поєднання - використання вагових коефіцієнтів для різних рекомендаційних систем залежно від їхньої точності або надійності. Кожна система може мати свою вагу, і рекомендації обчислюються шляхом комбінування результатів з урахуванням цих ваг. Зокрема таку техніку використовують Amazon та YouTube [45]
- Гібридний підхід з розділеними моделями полягає у використанні окремих моделей для рекомендацій на основі різних факторів, які потім поєднуються для формування остаточного списку рекомендацій. Кожна модель може бути відповідальною за свою власну область рекомендацій. Наприклад, компанії Spotify та Airbnb. [46]

Ансамблювання моделей є ефективним підходом до поєднання прогнозів різних моделей з метою отримання більш точного та стійкого прогнозу. Існують декілька методологій, які можна використовувати для ансамблювання моделей. Найпопулярнішими є:

- Середнє (Average) ансамблювання:
- Зважене середнє (Weighted Average) ансамблювання:
- Багатокласове голосування (Majority Voting): Кожна модель надає свій прогноз, а потім обирається найчастіший прогноз серед усіх моделей.
- Багатошарове голосування (Stacking): Використовуються два рівні моделей. Перший рівень моделей навчається на вхідних даних і генерує прогнози. Прогнози переходять до другого рівня моделі, яка навчається використовувати ці прогнози для отримання остаточного прогнозу. [47]

Варто зауважити, що методологія та формули ансамблювання можуть варіюватися залежно від конкретних вимог та контексту задачі. Вибір підходу до ансамблювання моделей залежить від типу даних, кількості моделей, їхньої ефективності та різноманітності.

Окрім просто поєднання є також комплексніші підходи, що використовуються в гібридних моделях:

- Мета-рівень. Попередньо вивчена модель використовується як вхідні дані для іншої системи рекомендацій
- Комбінація функцій. Функції однієї рекомендаційної системи впроваджуються в іншу
- Розширення функцій. Результат однієї моделі застосовується як вхідні дані для іншої
- Каскадна гібридизація. Одна система імпровізує результат іншої
- Перемикання гібридизації. Дозволяє вибирати одну рекомендовану модель на основі поточних вимог. [21]

Висновки до Розділу 2

Ключовою частиною глави є розгляд та вивчення підходів до побудови рекомендаційних систем. У ході роботи були розглянуті такі підходи, як контентна фільтрація, колаборативна фільтрація, демографічна фільтрація, рекомендаційна система, що базується на знаннях, гібридний та неперсоналізований підхід.

В результаті дослідження було проведено порівняльний аналіз переваг та недоліків побудови та використання рекомендаційних систем, на основі якого було обрано гібридний підхід, що поєднує підходи колаборативної фільтрації на основі користувачів та на основі об'єктів, який включає подібність об'єктів за оцінками від користувачів та подібність об'єктів за їх атрибутивними характеристиками.

РОЗДІЛ 3. ПОБУДОВА ГІБРИДНОЇ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ

3.1 Концепція побудови гібридної рекомендаційної системи з використанням демографічної фільтрації

Як було зазначено в розділі 1, електронна комерція поступово набуває дедалі більшого поширення у світі. У 2020 році онлайн-продавці отримали значно більше нових клієнтів, а також збільшили продажі наявних клієнтів. Це призвело до дуже вдалого року для продавців: за оцінками eMarketer, у 2020 році продажі електронної комерції досягли колосальних 3,914 трильйона доларів. При цьому користувачі можуть поводитися зовсім по-різному в залежності від регіону, країни і іноді навіть окремої територіальної одиниці. Врахування географічного місця розташування користувача є важливим аспектом при розробці рекомендаційної системи. Існує кілька причин, чому географічний контекст має велике значення:

- **Регіональні вподобання:** Користувачі з різних географічних областей мають різні вподобання та смаки. Врахування географічного місця розташування дозволяє пропонувати товари та послуги, які відповідають вподобанням конкретного регіону.
- **Культурні та соціальні відмінності:** Розташування користувача може відображати його культурні та соціальні відмінності. Врахування цих аспектів допомагає покращити рекомендації, забезпечуючи відповідність до особливостей конкретної культури або спільноти.
- **Географічні обмеження:** Деякі товари або послуги можуть бути доступні тільки в певних географічних областях. Врахування місця розташування користувача дозволяє обмежити рекомендації до доступних в їх регіоні товарів або послуг.
- **Місцеві події та акції:** Географічне місце розташування може впливати на наявність місцевих подій, акцій та пропозицій. Врахування цього контексту дозволяє пропонувати рекомендації, пов'язані зі спеціальними подіями або акціями, що відбуваються в конкретному регіоні.

- Купівельна спроможність та рівень цін: Навіть країни Європи можуть дуже відрізнятися обсягами мінімального та середнього доходу на душу населення.

Зокрема, це питання досліджувалося у дослідженні 2020-2021 Global Ecommerce Consumer Behavior Report що проводилося компанією ClearSale [48] зазначено, що користувацька поведінка та досвід користувача на продукті залежить від його країни розташування. Наприклад, на рисунку 3.1 показаний розподіл відповідей користувачів, що прийняли участь в опитуванні щодо їх досвіду та поведінки при онлайн-купівлях. Дослідження показує, що вдвічі більша частка жителів США купують товари онлайн один-два рази на тиждень порівняно з Канадою чи Австралією.

How often do you shop online?

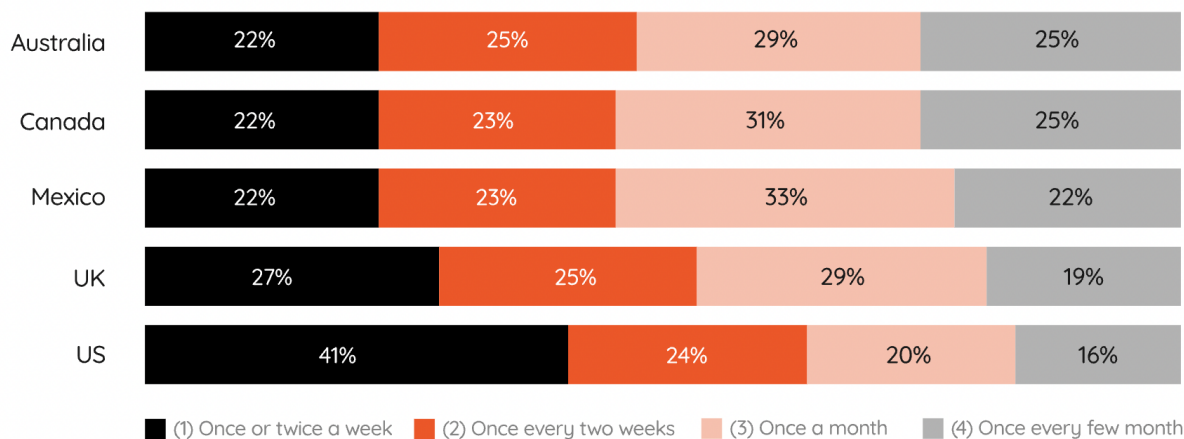


Рис. 3.1. Відповіді респондентів на питання щодо частоти покупок онлайн
Джерело: <https://offer.clear.sale/global-ecommerce-consumer-behavior>

Також цікавий вибір платформи користувачами в залежності від їх країни розташування. Як видно з рисунків 3.2 та 3.3 користувачі Мексики та США або США та Канади попарно демонструють доволі різну поведінку, хоча територіально країни межують, що не лише підтверджує важливість врахування локації для покращення користувацького досвіду споживача, а і специфікацію її не просто до регіону, а до країни місцезнаходження.

Do you usually order from a local website or from overseas?

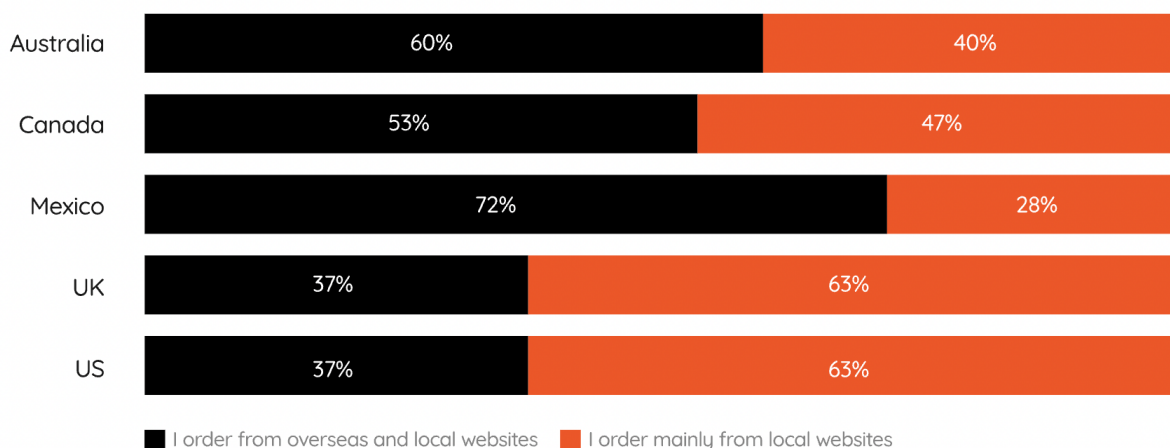


Рис. 3.2. Відповіді респондентів на питання щодо країни-походження товарів, які вони замовляють

Джерело: <https://offer.clear.sale/global-ecommerce-consumer-behavior>

What device are you most likely to use when shopping online?

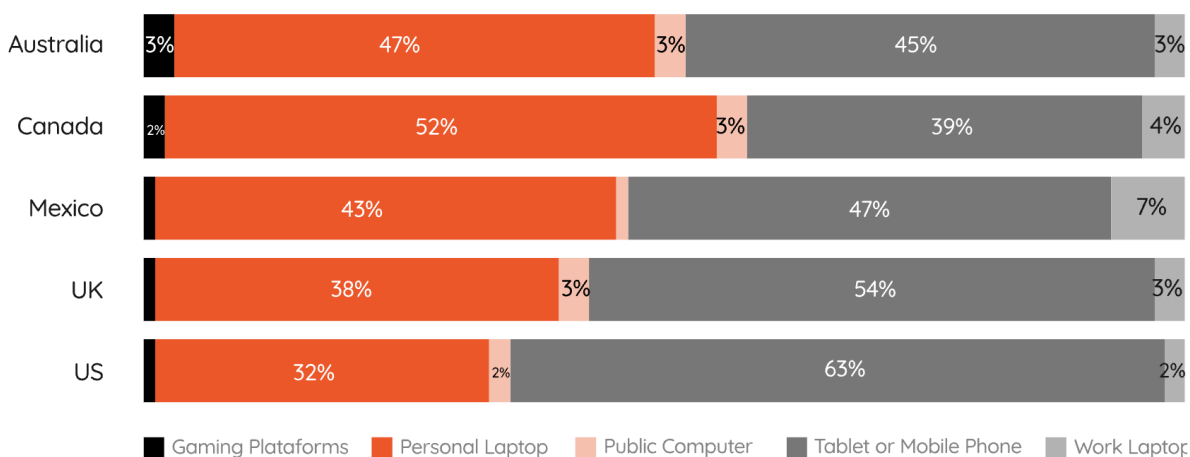


Рис. 3.2. Відповіді респондентів на питання щодо девайсу, через який вони здійснюють покупки онлайн

Джерело: <https://offer.clear.sale/global-ecommerce-consumer-behavior>

Найчастіше демографічна фільтрація застосовується саме в маркетингу для залучення специфічної цільової категорії користувачів, зокрема при налаштуванні PPC реклами, або з метою здешевлення закупки трафіку, адже ціна на різних сегментах за цільову дію може відрізнятись [49].

Коли мова йде про електронну комерцію, де об'єктом покупки користувача є контент (текстового або медіа- формату) то тут поведінкові особливості, а саме

культурні вподобання можуть відігравати одну з ключових ролей. Наприклад, країни Далекого Сходу (Японія, Південна Корея) вирізняються двома дещо протилежними поведінковими трендами: американізація та етноцентризм.

Останній означає процес, коли людина надає перевагу виключно об'єктам власної культури. Етноцентризм є поняттям з соціальної антропології і відображає тенденцію оцінювати і розуміти інші культури з перспективи своєї власної культури. Цей процес може мати вплив на сприйняття та оцінку інших культур, що веде до виникнення певних перекосяв або обмежень у розумінні та сприйнятті різних культурних проявів. Етноцентризм може впливати на вибір об'єктів культури, які споживаються, таких як музика, кіно, література, мистецтво, мода та інші. [50] З цього можна зробити висновок, що їх цікавитиме переважно контент англійською мовою та рідними мовами, що означає значно меншу частку переглядів контенту з Європи.

Іншим важливим прикладом поведінкових особливостей, що залежать від регіону є поточне неприйняття більшістю українців російськомовного контенту у зв'язку з війною Росії проти України. Окрім значного зниження кількості переглядів певного сегменту контенту, а отже і виручки платформи, бізнес ризикує отримати багато негативних відгуків та зіпсувати свій імідж рекомендуючи контент країни-ворога. Як показує практика, сучасні стрімінгові платформи дещо нехтували цим важливим поведінковим аспектом у 2022-2023 роках продовжуючи не лише отримувати доходи з контенту створеного росіянами, а й рекомендувати подібний контент українцям, що потерпали від збройного вторгнення Росії.

Третім фактором, що свідчить на користь врахування локації користувача при наданні рекомендацій є локальні обмеження на певний вид контенту. Наприклад, цензура контенту у Китаї сприятиме тому, що контент «забороненого» типу не буде користуватися популярністю серед жителів цієї країни.

Зокрема, можемо розглянути концепцію побудови гібридної рекомендаційної системи на базі компанії, що є тристороннім маркетплейсом.

Тристоронній маркетплейс - це онлайн-платформа, яка об'єднує трьох різних учасників: покупців, продавців і адміністратора (платформу). У такому маркетплейсі покупці можуть знаходити продавців, здійснювати покупки, а продавці можуть розміщувати свої товари або послуги на платформі. Адміністратор платформи забезпечує правильну роботу маркетплейсу, забезпечує безпеку транзакцій і надає підтримку користувачам. До особливостей тристороннього маркетплейсу належать:

- Багатостороння взаємодія: Тристоронній маркетплейс залучає трьох різних учасників - покупців, продавців і адміністратора. Кожен з цих учасників має свої потреби та мети, і маркетплейс забезпечує їх взаємодію для здійснення торгівлі товарами або послугами.
- Мережевий ефект: Тристоронній маркетплейс має потенційний мережевий ефект, де зростання числа покупців приводить до збільшення числа продавців, а збільшення числа продавців залучає більше покупців. Це створює позитивний цикл, який забезпечує зростання маркетплейсу.
- Управління транзакціями: Адміністратор маркетплейсу відповідає за управління транзакціями між покупцями та продавцями. Це включає обробку платежів, забезпечення безпеки та вирішення спорів.
- Рейтинги та відгуки: Тристоронні маркетплейси часто надають можливість користувачам залишати відгуки та оцінювати продавців. Це допомагає забезпечити якість товарів і послуг, а також покращує довіру між учасниками.

Саме наявність опції оцінювання, а також можливість впливати на поєднання споживача та постачальника робить тристоронні маркетплейси з одного боку зручними для дослідження роботи рекомендаційних систем, а з іншого боку створює доволі важливу потребу бізнесу мати такий інструмент, що може напряму впливати на обсяги доходів та інші показники діяльності платформи. Отже проблемою, яку покликана вирішити гібридна рекомендаційна система є надання рекомендацій користувачам на основі їх поведінки, інтересів та локації.

Дані про користувачів можна поділити на три типи:

- Атрибути: вік, локація, джерело трафіку з якого юзер потрапив на платформу, пошта тощо. Ці дані збираються за допомогою сторонніх додатків, наприклад GoogleAds, Growmatik або визначаються платформою самостійно за наявності відповідних інструментів та знань (IP-користувача)
- Поведінка: об'єкти (та данні про них), які юзер шукає, обирає, оцінює, переглядає на платформі. Ці данні збираються платформою під час сесій користувача за наявності відповідного трекінгу
- Побажання: опитування при реєстрації, протягом використання продукту, у тому числі через контакт з супортом. Ці данні отримуються безпосередньо від користувача шляхом отримання прямих чи опосередкованих відповідей на поставлені питання

Найпростіші характеристики, як от купівлі/оцінки користувача та його локація за регіоном доступні як малим так і великим бізнесам. Це означає, що побудова рекомендаційної моделі лише на цих простих параметрах буде універсальною в застосуванні. Однак збільшення характеристик, що включені до процесу моделювання потенційно може допомогти ідентифікувати нові когорти користувачів, взаємозв'язки між ними тощо.

Нормування оцінок користувачів є важливим процесом в рекомендаційних системах з метою забезпечення об'єктивності та порівнянності рейтингів. Оцінки користувачів можуть варіювати в широкому діапазоні, і нормування дозволяє привести їх до загальної шкали або стандартизувати їх, щоб здійснювати порівняння між різними оцінками.

Основні причини нормування оцінок користувачів:

- виправлення дисбалансу: Деякі користувачі можуть бути більш схильними до надання високих або низьких оцінок в порівнянні з іншими. Нормування допомагає вирівняти цей небаланс, щоб оцінки були більш об'єктивними.
- Порівняння між користувачами: Нормування дозволяє порівнювати оцінки різних користувачів, навіть якщо вони використовують різні шкали оцінювання.

Це дозволяє встановити спільну основу для порівняння і враховувати вплив кожного користувача на рекомендації.

- Зменшення шуму: В оцінках користувачів можуть бути присутні шумові чинники, такі як помилки або випадкові відхилення. Нормування допомагає зменшити вплив цього шуму на кінцеві рекомендації.
- Врахування ваги оцінок: Нормування може включати також врахування ваги оцінок користувачів, щоб враховувати рівень довіри до кожного користувача або врахувати інші фактори, які впливають на значимість їх оцінок.

Існують різні методи нормування оцінок, такі як центрування (центрування оцінок навколо середнього значення), масштабування (перетворення оцінок до певного діапазону), зважування (врахування ваги оцінок) та інші. Вибір конкретного методу залежить від контексту та особливостей конкретної рекомендаційної системи. У розглянутому випадку система оцінювання змінювалася в часі, тому доцільним буде використання масштабування через Min-Max нормалізацію.

Враховуючи специфіку поняття тристоронній маркетплейс доцільним є використовувати як user-based CF та item-based CF, оскільки дослідження показують, що комбінація кількох моделей збільшує точність гібридної моделі, у поєднанні з демографічною фільтрацією, а саме фільтрацією за країною знаходження користувача, адже цей атрибут є досить важливим чинником, який допомагає ідентифікувати та передбачити майбутню поведінку та інтереси користувача.

Важливою складовою є те, що гібридизація буде відбуватися за рахунок зваженого середнього, а коефіцієнти можна задавати та змінювати, аби гнучко та швидко реагувати на поточні проблеми та виклики як бізнесу так і світу, зокрема збільшуючи вагу однієї з моделей. Наприклад, це дало б змогу відомим компаніям відреагувати на війну в Україні та підтримати українських користувачів та творців контенту. Також важливо відмітити, що самостійне управління ваговими коефіцієнтами моделей залишає бізнесу великий простір для експериментів та A/B тестів в пошуку ідеальної довгострокової комбінації.

Отримана модель може бути оцінена за типовими показниками, такими як:

- Точність (*precision*) - вимірює, наскільки точно модель визначає позитивні приклади серед усіх прикладів, які вона визначає як позитивні. Це кількість правильно класифікованих позитивних прикладів (істинно позитивних), поділена на суму правильно класифікованих позитивних і неправильно класифікованих позитивних прикладів (істинно позитивних та хибно позитивних).

$$prec_p = \frac{TP}{TP + FP} \quad (3.1)$$

- Повнота (*recall, sensitivity*) – вимірює, наскільки ефективно модель визначає всі позитивні приклади. Це кількість істинно позитивних прикладів, поділена на суму істинно позитивних та хибно негативних прикладів.

$$TPR = sensitivity = recall_p = \frac{TP}{TP + FN} \quad (3.2)$$

- F-міра (*F-measure*) або F-скор. - є гармонічним середнім між точністю і повнотою і дозволяє об'єднати їх у один числовий показник. Вона враховує як точність, так і повноту моделі. F-міра обчислюється за формулою (3.3):

$$F_i = \frac{2}{\frac{1}{prec_i} + \frac{1}{recall_i}} = \frac{2 * prec_i * recall_i}{prec_i + recall_i} \quad (3.3)$$

F-міра дає більшу вагу більшій з точності або повноті, що дозволяє збалансувати обидві метрики. Значення F-мери знаходиться в діапазоні від 0 до 1, де 1 відповідає ідеальній точності і повноті, а 0 - найгіршому результату.

Описана концепція може мати ще багато доповнень, зокрема у вигляді використання вагових коефіцієнтів в залежності від LT (*life-time*) користувача, наприклад новим користувачам надавати більше рекомендацій на основі демографічної фільтрації, а згодом на основі його взаємодій з контентом та платформою.

3.2 Розробка гібридної рекомендаційної системи на основі демографічної фільтрації

Для реалізації гібридної системи рекомендацій було обрано мову програмування Python. Використання Python має багато переваг, особливо у контексті розробки програмного забезпечення та аналізу даних. Основні переваги використання Python включають:

- Простота вивчення і використання: Python має чистий і зрозумілий синтаксис, що робить його легким для вивчення, навіть для початківців. Це дозволяє швидко розпочати розробку і аналіз даних без значного зусилля.
- Велика кількість бібліотек: Python має широкий спектр сторонніх бібліотек, таких як NumPy, Pandas, SciPy, scikit-learn, TensorFlow та багато інших. Ці бібліотеки надають потужні інструменти для обробки даних, машинного навчання, наукових обчислень, візуалізації та інших завдань.
- Велика спільнота користувачів: Python має активну та велику спільноту користувачів, що означає, що ви можете легко знайти підтримку, документацію та рішення для своїх проблем. Ви можете приєднатися до форумів, груп в соціальних мережах або спільнот розробників, щоб обмінюватися досвідом і отримувати допомогу.
- Платформонезалежність: Python підтримується на різних операційних системах, таких як Windows, macOS, Linux і інші. Це означає, що ви можете розробляти і запускати свої програми на будь-якій платформі, що спрощує розповсюдження та використання вашого коду.
- Широке застосування: Python використовується в багатьох галузях, включаючи веб-розробку, наукові дослідження, аналітику даних, штучний інтелект, інтернет речей та багато інших. Це робить Python універсальним інструментом для вирішення різноманітних завдань.

Загалом, використання Python дозволяє швидко і ефективно розробляти програмне забезпечення, виконувати складні аналітичні завдання та реалізовувати різноманітні проекти.

Для побудови рекомендаційної системи було використано наступні пакети для Python:

- Pandas - це пакет, який надає широкі можливості для роботи з даними, зокрема з маніпуляціями табличних даних. Він забезпечує потужні структури даних, такі як DataFrame, які дозволяють легко зчитувати, записувати та обробляти дані з різних джерел. Pandas надає зручні функції для фільтрації, сортування, групування та агрегування даних. [51]
- NumPy є основним пакетом для наукових обчислень в Python. Він надає підтримку для масивів та матриць, а також функції для виконання математичних операцій на них. NumPy робить обчислення більш ефективними і зручними, надаючи широкі можливості для роботи з числовими даними. Завдяки своїй швидкості та простоті використання, він є необхідним інструментом для багатьох дослідників та аналітиків даних. [52]
- Scikit-learn є одним з найпопулярніших пакетів машинного навчання. Він надає широкий набір алгоритмів та інструментів для задач класифікації, регресії, кластеризації, підбору параметрів та іншого. Scikit-learn спрощує процес розробки та застосування моделей машинного навчання, надаючи зручний та послідовний інтерфейс для тренування та оцінки моделей. Крім того, він містить інструменти для попередньої обробки даних, валідації моделей та визначення метрик оцінки. [53]
- Recsys - це пакет для рекомендаційних систем, який надає інструменти та алгоритми для створення та розробки рекомендаційних моделей. Він спрощує процес побудови персоналізованих рекомендацій, використовуючи різноманітні підходи, такі як колаборативний фільтр, заснований на змісті, факторизація матриці тощо. Recsys надає реалізації різних алгоритмів рекомендаційних систем, які можна використовувати для аналізу даних та побудови моделей прогнозування. Цей пакет містить функції для завантаження, підготовки та обробки даних, а також для оцінки та тестування рекомендаційних моделей. [54].

Для роботи було використано дані платформи в галузі електронної комерції а саме тристороннього маркетплейсу, специфікою якого є продаж контенту, у проміжок з 01.01.2023 по 01.05.2023. Взаємозв'язок таблиць наведено на рис. 3.4.

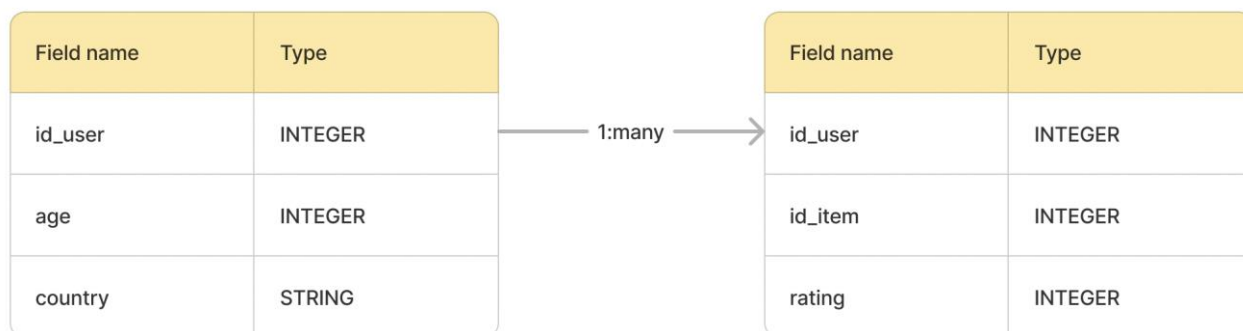


Рис. 3.4. Взаємозв'язок таблиць з вхідними даними

Джерело: складено автором.

Таблиця rating складається з трьох стовпців:

- id_user – ідентифікатор користувача
- id_item – ідентифікатор об'єкту
- rating – оцінка об'єкту за шкалою від 1 до 5.

Вона містить 312915 записів-оцінок від користувачів, що були виставлені впродовж зазначеного періоду часу. (рис 3.5).

	id_user	id_item	rating
0	142317521	81032488	3
1	147567876	69681314	4
2	133003258	56962332	5
3	143601662	74935374	2
4	147690980	73260409	5
...
312910	88790917	32206173	1
312911	132112704	80563952	1
312912	127731894	30089391	1
312913	148908722	79764449	1
312914	83147287	69443135	1

312915 rows x 3 columns

Рис. 3.5. Вигляд таблиці rating

Джерело: складено автором на основі даних компанії.

Після побудови зведеної таблиці стало зрозуміло, що таблиця містить записи про оцінки 10657 користувачів для 13558 унікальних товарів. (рис 3.6)

id_item	19577	209909	328439	343733	561879	779571	785908	832997	838271	846528	...
id_user											
67865342	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
67873453	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
67952159	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
67972227	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
67997066	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
...
150965770	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
150979096	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
150983624	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
150985831	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
150990506	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...

10657 rows × 13558 columns

Рис. 3.6. Матриця оцінок користувачів

Джерело: складено автором

Таблиця users складається з трьох стовпців:

- id_user – ідентифікатор користувача
- age – вік користувача
- country – країна з якої зареєструвався користувач.

	id_user	country	age
0	146534835	PR	54
1	146512054	US	46
2	146521809	US	66
3	146524214	US	56
4	146506092	US	63
...
10648	120544677	US	37
10649	120526747	US	53
10650	129591753	US	49
10651	129598660	AU	36
10652	129617161	US	44

10653 rows × 3 columns

Рис. 3.7. Вигляд таблиці users

Джерело: складено автором на основі даних компанії.

У моделі було вирішено поєднати підходи колаборативної фільтрації на основі користувачів та на основі об'єктів. Перед початком роботи данні були перевірені на наявність несумісних типів, невалідних записів у вигляді пустих рядків, тощо. Також вибірку було розділено на тренувальну та тестову множини (див. Додаток А).

Для реалізації моделей колаборативної фільтрації на основі об'єктів та на основі користувачів була використана модель "Item K Nearest Neighbours". Основна ідея моделі Item K Nearest Neighbours полягає в тому, що для кожного об'єкта (item) визначається його набір "найближчих сусідів" (nearest neighbours) на основі подібності між ними. Подібність може бути визначена за допомогою різних метрик, однак в ході роботи було обрано косинусну подібність (cosine similarity). Ця метрика вимірює кут між векторами, представленими ознаками об'єктів, і використовується для порівняння їх схожості. Після визначення набору найближчих сусідів для кожного об'єкта, модель використовує ці подібні об'єкти для здійснення рекомендацій. Наприклад, якщо користувач виразив зацікавленість певним об'єктом, модель може рекомендувати інші об'єкти, які є "найближчими сусідами" цього об'єкта.

В результаті реалізації моделей з $K=50$ найближчих сусідів було отримано матриці з ймовірностями того, що користувач U придбає товар I за умови перегляду рекомендації. (рис 3.8 та 3.9). Працездатність ItemKNN значно залежить від наявності достатньої кількості даних про рейтинги об'єктів. Якщо матриця оцінок дуже розріджена або неповна, може бути складно знайти достатньо схожих об'єктів для рекомендацій. Тому, відсутність ймовірностей що перевищують 0,5 в обох матрицях можна пояснити розрідженістю даних. Однак це можна потенційно нівелювати використовуючи вищий ваговий коефіцієнт для результатів моделі демографічної фільтрації (додаток Б).

id_item	19577	209909	328439	343733	561879	779571	785908	832997	838271	846528	...	82684411
id_user												
104440609	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0
146912643	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0
134463367	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0
87497855	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0
146903924	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0
...
148793309	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0
143950840	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0
128481565	0.0	0.0	0.0	0.0	0.017839	0.0	0.0	0.0	0.0	0.0	...	0.0
101992233	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0
132331615	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0

3198 rows x 13558 columns

Рис. 3.8. Матриця рекомендацій на основі user-based колаборативної фільтрації

Джерело: складено автором

id_item	19577	209909	328439	343733	561879	779571	785908	832997	838271	846528	...	82684411
id_user												
104440609	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
146912643	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
134463367	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
87497855	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
146903924	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
...
148793309	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
143950840	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
128481565	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
101992233	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
132331615	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0

3198 rows x 13558 columns

Рис. 3.9. Матриця рекомендацій на основі item-based колаборативної фільтрації

Джерело: складено автором

id_item	19577	209909	328439	343733	561879	779571	785908
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...

Рис. 3.10. Матриця рекомендацій на основі демографічної фільтрації

Джерело: складено автором

Для гібридизації рекомендаційних систем було використано вагові коефіцієнти. Через те, що вище досліджувалося питання важливості локальних (місцевих) подій, то ефективним рішенням буде використовувати певні коефіцієнти за замовчуванням, однак залишити можливість компанії швидко змінювати вагові коефіцієнти для тієї чи іншої моделі, аби гнучко реагувати на виклики. Зокрема способом це реалізувати є створення змінних поза частиною моделювання, аби бізнес міг легко змінити їх значення, при цьому не «зачепивши» модель. За замовчуванням були обрані коефіцієнти 0,4, 0,3 та 0,3 відповідно.

На рисунку 3.11 наведено остаточні рішення щодо рекомендацій прийняті гібридною моделлю.

id_item	82753293	82755167	82775758	82822530	82881820	82887526	\
id_user							
104440609	0.0	0.0	0.000000	0.000000	0.000000	0.000000	
146912643	0.0	0.0	0.000000	0.000000	0.000000	0.000000	
134463367	0.0	0.0	0.000000	0.000000	0.000000	0.000000	
87497855	0.0	0.0	0.000000	0.000000	0.000000	0.000000	
146903924	0.0	0.0	0.000000	0.000000	0.000000	0.000000	
...	
148793309	0.0	0.0	0.081884	0.000000	0.000000	0.000000	
143950840	0.0	0.0	0.000000	0.000000	0.000000	0.000000	
128481565	0.0	0.0	0.000000	0.000000	0.000000	0.000000	
101992233	0.0	0.0	0.000000	0.000000	0.071703	0.544035	
132331615	0.0	0.0	0.000000	0.070659	0.000000	0.617723	

Рис. 3.11. Матриця рекомендацій гібридної моделі

Джерело: складено автором

Оскільки одне з основних завдань системи – зменшити навантаження нерелевантною інформацією на користувача, то було вирішено рекомендувати

користувачам протягом кожних півгодини проведених на платформі нову порцію з N товарів з найбільшими ймовірностями придбання. Враховуючи обсяги даних (невелика вибірка) – не є доцільним ставити високий поріг ймовірності, адже це значитиме хоч і високу точність системи, однак дуже малий відсоток покриття (малу кількість рекомендацій).

Застосувавши побудовану модель до тестової вибірки та порівнявши прогнозовані системою рекомендації для вже оцінених об'єктів. Отримані результати наведено на рисунку 3.12.

Точність: 0.7163896000000001
Повнота: 0.4860420772988259
F-міра: 0.5791522228048258

Рис. 3.12. Точність, повнота та F-міра побудованої моделі

Джерело: складено автором

Точність вимірює, наскільки точно модель ідентифікує позитивні екземпляри серед усіх екземплярів, які вона визначає як позитивні. Значення 0,72 означає, що система правильно ідентифікує 72% позитивних рекомендацій серед усіх рекомендацій, які вона зробила. Повнота вимірює, наскільки ефективно модель виявляє всі позитивні екземпляри, що існують у вибірці. У контексті рекомендаційних систем, це означає, наскільки система знаходить і рекомендує всі відповідні об'єкти для користувача. Значення 0,49 означає, що система виявляє лише 49% з усіх позитивних об'єктів, які могли бути рекомендовані користувачеві. F-міра є гармонічним середнім між точністю і повнотою і використовується для оцінки компромісу між цими двома метриками. Вона дозволяє зрозуміти загальну ефективність системи, враховуючи як точність, так і повноту. Значення 0,58 вказує на баланс між точністю і повнотою системи (див. Додаток В)

Отримані результати вказують на можливі точки удосконалення системи, а саме збільшення повноти системи.

ВИСНОВКИ

Результатом дослідження є розробка алгоритму реалізації гібридної рекомендаційної системи на основі демографічної фільтрації з метою створення індивідуального списку товарів для користувача. Проведене дослідження дозволяє сформулювати такі висновки та пропозиції:

1. Електронна комерція в Україні та світі активно розвивалась протягом останніх 10 років. При цьому пік швидкості зростання обсягів продажів припадає на кінець 2020 року, що спричинено епідемією Covid-19 та переходом бізнесів в онлайн через карантинні обмеження
2. Найбільші маркетплейси та стрімінгові платформи України та світу активно використовують рекомендаційні системи, однак постійно стикаються з різноманітними проблемами, вирішення яких призводить до удосконалення рекомендаційних систем. Наприклад, проблема штучного завищення/заниження оцінок.
3. Рекомендаційні системи є досить затребуваним інструментом, тому постійне удосконалення алгоритмів та методів їх роботи призвело до досить розгалуженої схеми видів рекомендаційних систем.
4. Потреба в гібридизації рекомендаційних систем виникає, якщо точність чи ефективність певної окремої моделі не є достатніми для потреб бізнесу, або проблему чи недолік, який має одна модель, - може вирішити інша модель. Основними підходами до гібридизації є комбінування моделей, ансамбль моделей, контент-колаборативний гібрид. Основні методи, що можуть використовуватись в гібридних рекомендаційних системах: контентний аналіз, кластеризація, колаборативний фільтр. Рекомендаційні системи часто поєднують з машинним навчанням, а найновіші дослідження вже включають інтеграцію штучного інтелекту в роботу рекомендаційних систем.
5. Регіон розташування користувача може значним чином впливати на його поведінку, зокрема на контент, який він бажає споживати. Тому врахування цього фактору при розробці рекомендаційної системи є дуже важливим. Яскравим прикладом цього може бути свідома відмова більшості українців споживати

російський контент, при цьому реакція рекомендаційних систем на таку різку зміну поведінки була або повільною або відсутньою, адже алгоритми враховували лише попередню поведінку користувачів.

6. Описано концепцію побудови рекомендаційної системи на основі демографічної фільтрації. При цьому рекомендаційна система може мати ще багато доповнень, зокрема у вигляді використання вагових коефіцієнтів в залежності від LT (life-time) користувача, наприклад новим користувачам надавати більше рекомендацій на основі демографічної фільтрації, а згодом на основі його взаємодій з контентом та платформою.

7. Запропоновано гібридну рекомендаційну систему на основі демографічної фільтрації, яка була реалізована мовою програмування Python, що дає змогу швидко інтегрувати систему в додатки або платформи бізнесів. В системі поєднано контентну фільтрацію на основі користувачів та на основі об'єктів, а також демографічну фільтрацію. Моделі поєднуються між собою вагових коефіцієнтів, що дає змогу бізнесу контролювати «силу» впливу того чи іншого фактору на рекомендації для певної когорти користувачів. Запропонована система має відносно високу точність при гіршому показникові повноти. Це вказує на можливості майбутнього удосконалення системи.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. E-commerce share of retail sales (2021–2026), Oberlo URL: <https://www.oberlo.com/statistics/ecommerce-share-of-retail-sales>
2. Alex Wang, What is the share of E-Commerce in overall retail sales? 2022 URL: <https://www.cbre.com/insights/articles/omnichannel-what-is-the-share-of-e-commerce-in-overall-retail-sales>
3. Good, Nathaniel & Schafer, Ben & Konstan, Joseph & Borchers, Al & Sarwar, Badrul & Herlocker, Jon & Riedl, John. (1999). Combining collaborative filtering with personal agents for better recommendations, Department of Computer Science and Engineering University of Minnesota 439-446. URL: https://www.researchgate.net/publication/221604180_Combining_Collaborative_Filtering_with_Personal_Agents_for_Better_Recommendations
4. Çano, Erion. (2017). Hybrid Recommender Systems: A Systematic Literature Review. Intelligent Data Analysis. 21. 1487-1524. DOI: 10.3233/IDA-163209 URL: <https://content.iospress.com/articles/intelligent-data-analysis/ida163209>
5. Chornous, G., Nikolskyi, I., Wyszynski, M., Kharlamova, G., & Stolarczyk, P. (2021). A hybrid user-item-based collaborative filtering model for e-commerce recommendations. Journal of International Studies, 14(4), 157-173. doi:10.14254/2071-8330.2021/14-4/11
6. Covid-19 and e-commerce: global-review, UNCTAD, URL: <https://unctad.org/publication/covid-19-and-e-commerce-global-review>
7. OECD Policy Responses to Coronavirus (COVID-19): E-commerce in the time of COVID-19, 2020 URL: <https://www.oecd.org/coronavirus/policy-responses/e-commerce-in-the-time-of-covid-19-3a2b78e8/>
8. Impact of COVID Pandemic on eCommerce, International Trade Administration (ITA) URL: <https://www.trade.gov/impact-covid-pandemic-ecommerce>
9. E-commerce and trade, The Digital Watch, URL: <https://dig.watch/topics/e-commerce-and-trade>

10. О. Чайка, Статистика розвитку е-commerce у найбільших регіонах світу, 2022 URL: <https://magazine.ukr-china.com/statystyka-rozvytku-e-commerce-u-najbilshyh-regionah-svitu/>
11. Petros T., Amount of data created, consumed, and stored 2010-2020, with forecasts to 2025, 2022 URL: <https://www.statista.com/statistics/871513/worldwide-data-created/>
12. E-commerce – Ukraine, Statista URL: <https://www.statista.com/outlook/dmo/ecommerce/ukraine>
13. Mazorenko O. How covid-19 pandemic boosts the European and Ukrainian electronic commerce, Simon Kuznets Kharkiv National University of Economics UDC 339.3:004 DOI: <https://doi.org/10.32782/2524-0072/2021-25-59>
14. Симоненко К. Вплив війни на інтернет-торгівлю: як змінювалися онлайн-продажі ритейлерів протягом I півріччя 2022 року, 2022, URL: <https://rau.ua/novyni/vpliv-vijnina-internet/>
15. Kirk W. McLaren, The Future Of E-Commerce: Trends To Watch In 2023 URL: <https://www.forbes.com/sites/forbesmarketplace/2023/03/21/the-future-of-e-commerce-trends-to-watch-in-2023/?sh=6ea6cbb1631e>
16. B. Davis, Ecommerce in 2022: What do the experts predict? URL: <https://econsultancy.com/ecommerce-trends-2022/>
17. Cordoba I., Donovan C., Moulton J, Phillips S., Magni M., Resetting the e-commerce model to achieve profitable growth in Europe, 2022, URL: <https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/resetting-the-e-commerce-model-to-achieve-profitable-growth-in-europe>
18. .O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, Recommendation systems: Principles, methods and evaluation, Egyptian Informatics Journal, 2015, Pages 261-273, ISSN 1110-8665, URL: <https://doi.org/10.1016/j.eij.2015.06.005>
19. Melville, P., Sindhvani, V. (2011). Recommender Systems. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA URL: https://doi.org/10.1007/978-0-387-30164-8_705

20. Голенко М., Вакалюк Т., Область застосування та види рекомендаційних систем, 2021 URL: <https://conf.ztu.edu.ua/wp-content/uploads/2021/01/41-2.pdf>
21. Roy, D., Dutta, M. A systematic review and research perspective on recommender systems. J Big Data 9, 59 (2022). URL: <https://doi.org/10.1186/s40537-022-00592-5>
22. Vatsal P., Recommendation Systems Explained, 2021 URL: <https://towardsdatascience.com/recommendation-systems-explained-a42fc60591ed>
23. Nunung Nurul Qomariyah, Definition and History of Recommender Systems, 2020, URL: <https://international.binus.ac.id/computer-science/2020/11/03/definition-and-history-of-recommender-systems/>
24. Poriya, A., Bhagat, T., Patel, N., & Sharma, R. (2014) Non-Personalized Recommender Systems and User-based Collaborative Recommender Systems. International Journal of Applied Information Systems, 6, 22-27. URL: <https://doi.org/10.5120/ijais14-451122>
25. Hooda, R., Singh, K., & Dhawan, S. (2014) A Study of Recommender Systems on Social Networks and Content-based Web Systems. International Journal of Computer Applications, 97, 23-28. URL: <https://doi.org/10.5120/16996-7128>
26. Woldan, P., Duda, P., & Hayashi, Y. (2020) Visual Hybrid Recommendation Systems Based on the Content-Based Filtering. In Artificial Intelligence and Soft Computing. ICAISC 2020. Lecture Notes in Computer Science, Rutkowski L., Scherer R., Korytkowski M., Pedrycz W., Tadeusiewicz R., Zurada J.M. (eds); Springer: Cham, Switzerland, 12416, 455-465. URL: https://doi.org/10.1007/978-3-030-61534-5_41
27. Resnik, P., Iacovou, N., Suchak, M., & Bergstrom, P. (1994) GroupLens: an open architecture for collaborative filtering of netnews. Proceedings of the 1994 ACM conference on Computer supported cooperative work, North Carolina, USA, October 22-26, 175-181. URL: <https://doi.org/10.1145/192844.192905>
28. Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001) Item-Based Collaborative Filtering Recommendation Algorithms. Proceedings of the 10th International Conference on World Wide Web, Hong Kong, May 1-5. 285- 289. URL: <http://doi.acm.org/10.1145/371920.372071>

29. Shambour, Q. (2020) A deep learning-based algorithm for multi-criteria recommender systems. Knowledge-Based Systems, 211, 106545. URL: <https://doi.org/10.1016/j.knosys.2020.106545>
30. Recommender systems in ecommerce: the quickest way to boost sales, 2022 URL: <https://www.perzonalization.com/blog/recommender-systems-in-ecommerce/>
31. Dakduk S., Horst E., Santalla Z., Molina G., Malavé J., Customer Behavior in Electronic Commerce: A Bayesian Approach, Journal of theoretical and applied electronic commerce research versión On-line ISSN 0718-1876 J. theor. appl. electron. commer. res. vol.12 no.2, 2017 URL: <http://dx.doi.org/10.4067/S0718-18762017000200002>
32. Schafer, Ben & Konstan, Joseph & Riedl, John. (1999). Recommender Systems in E-Commerce. 1st ACM Conference on Electronic Commerce, Denver, Colorado, United States. 10.1145/336992.337035. URL: https://www.researchgate.net/publication/2507550_Recommender_Systems_in_E-Commerce
33. Amazon Books URL: <https://www.amazon.com/books-used-books-textbooks/b?ie=UTF8&node=283155>
34. Reyes J. CDNow: 3 lessons from Fort Washington-based pioneering 1990s-era online music retailer URL: <https://technical.ly/uncategorized/cdnw-3-lessons-from-one-of-pioneering-dot-com-era-online-music-retailer/>
35. eBay URL: <https://www.ebay.com/>
36. Rozetka URL: <https://rozetka.com.ua> .
37. Епіцентр URL: <https://epicentrk.ua>
38. Allo URL: <https://allo.ua/>
39. Khusro, S., Ali, Z., Ullah, I. (2016). Recommender Systems: Issues, Challenges, and Research Opportunities. In: Kim, K., Joukov, N. (eds) Information Science and Applications (ICISA) 2016. Lecture Notes in Electrical Engineering, vol 376. Springer, Singapore. URL: https://doi.org/10.1007/978-981-10-0557-2_112
40. Ricci, Francesco & Rokach, Lior & Shapira, Bracha. (2010). Recommender Systems Handbook. 10.1007/978-0-387-85820-3_1. URL:

https://www.researchgate.net/publication/227268858_Recommender_Systems_Handbook

41. Spertus, Ellen & Sahami, Mehran & Buyukkokten, Orkut. (2005). Evaluating similarity measures: A large-scale study in the Orkut social network. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 678-684. 10.1145/1081870.1081956. URL: https://www.researchgate.net/publication/221653831_Evaluating_similarity_measures_A_large-scale_study_in_the_Orkut_social_network
42. Lathia, Neal & Hailes, Stephen & Capra, Licia. (2008). The effect of correlation coefficients on communities of recommenders. 2000-2005. 10.1145/1363686.1364172. URL: https://www.researchgate.net/publication/221001560_The_effect_of_correlation_coefficients_on_communities_of_recommenders
43. Tran, Thomas T. and Robin Cohen. "Hybrid Recommender Systems for Electronic Commerce." (2000). URL: <https://www.semanticscholar.org/paper/Hybrid-Recommender-Systems-for-Electronic-Commerce-Tran-Cohen/458b8214e835d7de9c1afe39ac17e3b425815ef4>
44. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8), 30-37. Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender Systems Handbook* (pp. 1-35).
45. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331-370. Shao, W., Wang, D., Li, H., Wang, J., & Liu, Y. (2019). An Improved Hybrid Recommendation Algorithm Based on User Preferences. In *Proceedings of the International Conference on Advanced Communication Technology* (pp. 308-313)
46. Melville, P., Mooney, R. J., & Nagarajan, R. (2002). Content-boosted collaborative filtering for improved recommendations. In *Proceedings of the National Conference on Artificial Intelligence* (Vol. 17, No. 1, pp. 187-192) Cantador, I., Bellogín, A., Castells, P., & Mencía, E. L. (2011). Workshops overview. In *Proceedings*

of the International Conference on User Modeling, Adaptation, and Personalization (pp. 391-393).

47. Dietterich, T. G. (2000). Ensemble methods in machine learning. Multiple Classifier Systems, 1857, 1-15. Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. Journal of Artificial Intelligence Research, 11, 169-198
48. 2020-2021 Global Ecommerce Consumer Behavior Report, ClearSale URL: <https://offer.clear.sale/global-ecommerce-consumer-behavior>
49. Sasson D. For PPC ads, filter by demographic data, 2016 URL: <https://www.practicalcommerce.com/for-ppc-ads-filter-by-demographic-data>
50. Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). The Authoritarian Personality. Hofstede, G. (1980). Culture's consequences: International differences in work-related values. Sage Publications. Kim, Y. Y. (2001). Becoming intercultural: An integrative theory of communication and cross-cultural adaptation
51. Офіційна документація бібліотеки pandas (Python), URL: <https://pandas.pydata.org/>
52. Офіційна документація бібліотеки numpy (Python), URL: <https://numpy.org/doc/>
53. Офіційна документація бібліотеки scikit-learn (Python), URL: <https://scikit-learn.org/stable/>
54. Офіційна документація бібліотеки resock (Python), URL: <https://resock.froomle.ai/guides.html>

ДОДАТКИ

Додаток А

Попередня підготовка даних, розподіл даних на тренувальну і тестову вибірки

```
!pip install reckpt
from reckpt.algorithms import ItemKNN
from sklearn.model_selection import train_test_split
from sklearn.metrics import precision_score, recall_score, f1_score
import pandas as pd
import numpy as np
from google.colab import files
from scipy.sparse import csr_matrix
uploaded = files.upload()
rating_data = pd.read_excel('data.xlsx')
users_data = pd.read_excel('data.xlsx','data_users')
rating=rating_data.pivot(index='id_user', columns='id_item', values='rating').fillna(0)
train_data, test_data = train_test_split(rating, test_size=0.3)
print("Розмір тренувальної частини:", len(train_data))
print("Розмір тестової частини:", len(test_data))
```

Застосування item-based рекомендаційної моделі

```
items_knn = ItemKNN(K=50, similarity='cosine', pop_discount=None,  
normalize_X=False, normalize_sim=True)  
items_knn.fit(csr_matrix(train_data.to_numpy()))  
items_knn.get_params()  
item_based = items_knn.predict(csr_matrix(test_data.to_numpy()))  
item_based = pd.DataFrame(item_based.toarray())  
item_based.index = test_data.index  
item_based.columns = test_data.columns  
item_based
```

Гібридизація моделі на основі вагових коефіцієнтів

```
coef1 = 0,3
coef2 = 0,3
coef3 = 0,4
result = item_based.multiply(coef1) + user_based.multiply(coef2) + demog(coef3)
print(result)
binary_table_result = np.where(result > 0.07, 1, 0)
binary_table_test= np.where(test_data > 0, 1, 0)
result_2 = np.where(binary_table_test != 0, binary_table_result, 0)
precision = precision_score(binary_table_test, result_2, average='micro')
recall = recall_score(binary_table_test, result_2, average='micro')
f_measure = f1_score(binary_table_test, result_2, average='micro')
print("Точність:", precision)
print("Повнота:", recall)
print("F-міра:", f_measure)
```