

Міністерство освіти та науки України
Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
(повна назва)

Кафедра _____ Системотехніки _____
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

Рівень вищої освіти _____ другий (магістерський) _____
(рівень вищої освіти)

Розробка та дослідження методів формування рекомендацій для веб-порталу з
інтернет комерції
(тема)

Виконав:

студент групи ІТІМ-21-1

Мілашенко М.В.

(прізвище, ініціали)

спеціальності 122 – Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Інформаційні технології
проекування

(повна назва освітньої програми)

Керівник _____ доц. Хряпкін О.В. _____

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри системотехніки _____

(підпис)

Гребеннік І.В.

(прізвище, ініціали)

2022 р.

Я як студент ХНУРЕ розумію і підтримую політику закладу із академічної доброчесності. Я не надавала і не одержувала недозволену допомогу під час підготовки кваліфікаційної роботи. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

18.12.2022



Мілашенко М.В.

Кваліфікаційна робота не містить відомостей заборонених до відкритого опублікування.

Кваліфікаційна робота виконана у відповідності до стандартів, що діють в Україні.

Попередній захист проведено 18 грудня 2022 р.

Керівник кваліфікаційної роботи



доц. Хряпкін О.В.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Системотехніки
(повна назва)

Рівень вищої освіти другий (магістерський)

Спеціальність 122 – Комп'ютерні науки
(код і повна назва)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма Інформаційні технології проектування
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)


« _____ » _____
2022 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ(ПРОЕКТ)

студентові Милашенко Марії Володимирівни
(прізвище, ім'я, по батькові)

1. Тема роботи Розробка та дослідження методів формування рекомендацій для веб-порталу з інтернет комерції затверджена наказом по університету від « _____ » _____ 2022 р.№ _____
2. Термін подання студентом роботи до екзаменаційної комісії 16 грудня 2022 р.
3. Вихідні дані до роботи: Перелік використовуваних програмних засобів: ОС Windows 10, Paperspace GPU machine using PyTorch 0.4.1.
4. Зміст пояснювальної записки (перелік питань, що потрібно розробити) 4.1 Вступ. 4. 2 Аналіз предметної області. 4. 3 Реалізація та порівняння методів рекомендацій. 4.4 Результати методів рекомендацій 4. 5 Висновок.
5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) 5.1 Приклад рекомендацій Амазону 5.2 Приклад рекомендацій Лінкедін. 5.3 Приклад алгоритму на основі контенту 5.4 Графік порівняння методів на основі оцінки фільмів 5.5 Графік порівняння методів на основі коефіцієнту подібності

6. Консультанти розділів роботи

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
	доц.Хряпкин О.В.		

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	<i>Отримання завдання на кваліфікаційне проектування</i>	08.11.2022	<i>Виконано</i>
2	<i>Аналіз вихідних даних та літератури за темою кваліфікаційної роботи</i>	13.11.2022	<i>Виконано</i>
3	<i>Постановка задачі та вибір методу її вирішення</i>	16.11.2022	<i>Виконано</i>
4	<i>Проведення експериментальних досліджень</i>	18.11.2022	<i>Виконано</i>
5	<i>Оформлення пояснювальної записки</i>	21.11.2022	<i>Виконано</i>
6	<i>Підготовка презентації</i>	21.11.2022	<i>Виконано</i>
7	<i>Подання закінченої роботи науковому керівникові</i>	10.12.2022	<i>Виконано</i>
8	<i>Попередній захист</i>	13.12.2022	<i>Виконано</i>
9	<i>Подання роботи на рецензування</i>	20.12.2022	<i>Виконано</i>
10	<i>Подання роботи до екзаменаційної комісії</i>	23.12.2022	<i>Виконано</i>

Студент



Мілашенко М.В.

(підпис)



Керівник роботи

(підпис)

доц. Хряпкін О.В..

(посада, прізвище, ініціали)

ЗАЯВА
щодо самостійності виконання письмової роботи

Я, Мілащенко Марія Володимирівна
(прізвище, ім'я, по батькові)

посада студентка групи ІТІМ-21-1

-
кафедра Системотехніки

заявляю: моя письмова робота на тему «Розробка та дослідження методів формування рекомендацій для веб-порталу з інтернет комерції»
(назва роботи)

Представлена у екзаменаційну комісію _____
(спеціалізовану вчену раду, екзаменаційну комісію тощо)

Для публічного захисту, виконана самостійно і в ній не міститься елементів плагіату. Всі запозичені з друкованих та електронних джерел, а також із раніше виконаних дослідницьких робіт та захищених кандидатських і докторських дисертацій мають відповідні посилання.

Я Ознайомлений (а) з діючим положенням «Про протидію плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску письмової роботи до захисту та застосування дисциплінарних заходів.

Дата 22.12.2022 р.

Підпис



РЕФЕРАТ

Пояснювальна записка до звіту з кваліфікаційної роботи містить 74 сторінок, 4 рисунки, 26 літературних джерел.

РЕКОМЕНДАЦІЯ ПРОДУКЦІЇ, КОНТЕНТ, РЕКОМЕНДАЦІЙНІ СИСТЕМИ, ІНТЕРНЕТ КОМЕРЦІЯ, ГІБРИДНА МОДЕЛЬ, СИРІ ДАНІ.

Метою роботи є розробка та дослідження методів рекомендацій для веб-порталу з інтернет комерції.

Об'єктом дослідження є аналіз існуючих методів рекомендацій на ринку. Також визначення підходу до вирішення проблеми з вибором найкращого алгоритму рекомендацій. Зроблені порівняння існуючих методів створення рекомендацій на основі вибірки даних, та виокремлений найбільш ефективний алгоритм рекомендацій, який задовольняє поставлену задачу.

Результатом кваліфікаційної роботи є виявлений метод рекомендацій, який найбільше підходить для застосування веб-порталах з інтернет комерції.

ABSTRACT

The explanatory note to the report on attestation practice contains 74 pages, 4 figures, 26 literary sources.

MACHINE LEARNING, CONTENT, RECOMMENDATION SYSTEMS, HYBRID APPROACH, E-COMMERCE, RAW DATA.

The purpose of the work is the development and research of recommendation methods for a web portal for Internet commerce.

The object of the research is the analysis of the existing methods of recommendations on the market. Also defining an approach to solving the problem of choosing the best recommendation algorithm. Comparisons of existing methods of creating recommendations based on data sampling are made, and the most effective algorithm of recommendations that satisfies the given task is singled out.

The result of the certification work is the identified method of recommendations, which is most suitable for use in web portals for Internet commerce.

ЗМІСТ

ВСТУП.....	2
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	3
1.1 Актуальність рекомендаційних систем.....	3
1.2 Огляд аналогів на ринку.....	17
1.3 Методи та алгоритми формування рекомендацій	20
1.4 Постановка завдання	22
2. Огляд систем і інструментів.....	23
2.1 Огляд системи на основі контенту	23
2.2 Огляд системи спільної фільтрації.....	25
2.3 Огляд гібридної системи.....	7
2.4 Огляд інструментів реалізації.....	29
3. РЕАЛІЗАЦІЯ ТА ПОРІВНЯННЯ МЕТОДІВ РЕКОМЕНДАЦІЙ.....	42
3.1 Дослідження методів.....	42
4. РЕЗУЛЬТАТИ МЕТОДІВ РЕКОМЕНДАЦІЙ.....	55
4.1 Порівняння кінцевих результатів методів.....	55
ВИСНОВКИ.....	58
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ.....	59
ДОДАТОК А. Графічний матеріал кваліфікаційної роботи.....	61
ДОДАТОК Б. Текст програми.....	65
ДОДАТОК В. Відомість кваліфікаційної роботи.....	68

**ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,
СКОРОЧЕНЬ І ТЕРМІНІВ**

БД – база даних;

МН – машинне навчання;

СУБД – система управління базою даних.

ВСТУП

Завоювання величезної уваги аудиторії - одна з основних речей для підвищення продажів. І це не станеться, якщо немає змогу розповісти глядачам про послуги, які компанії пропонують. Для цього потрібно дізнатися, що шукає цільова аудиторія продукту чи бізнесу.

Згідно даних дослідження ринок електронної комерції є одним із найдинамічніших в Україні, обсяг ринку у 2020 році виріс на 41% та досягнув \$4 млрд, що становить 8.8% від загального обсягу роздрібної торгівлі. Ринок електронної комерції виріс майже в 3 рази за останні 5 років, прогнозується ріст в 2 рази протягом наступних 5-ти років.

Обсяг е-експорту в 2020 році становив близько \$450 млн. При цьому США є найбільшим імпортером українських товарів – 33% від загального українського е-експорту.

Електроніка та одяг є найбільшими за обсягом та найрозвинутішими секторами електронної торгівлі в Україні. Обсяг електронної комерції одягом в Україні зростав у середньому на 26% з 2016 року і досягнув \$291млн у 2020. Частка електронної комерції у роздрібній торгівлі одягом у 2020 році становила 6.8%.А середній чек складав \$24-31.

Україна має найнижчий рівень витрат однієї людини на е-комерцію в рік – \$104 у порівнянні з сусідніми країнами. З іншого боку - український ринок е-комерції мав найвищий темп росту в Центрально- Східній Європі – 41%. В ТОП-3 за темпами росту е-комерції також увійшли Польща та Чехія, 37% та 29%, відповідно.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Актуальність рекомендаційних систем

У сучасному світі, де Інтернет став важливою частиною життя людини, користувачі стикаються проблеми вибору через велику різноманітність колекцій. Люди шукають велику кількість різноманітних речей, від готелю до варіантів інвестування, а в Інтернеті доступно забагато інформації. Щоб допомогти користувачам впоратися з величезним інформаційним потоком, компанії розгорнули системи рекомендацій для кращого керування пошуком користувачів в їхніх системах. Дослідження в галузі систем рекомендацій тривають досить давно, але інтерес все ще залишається високим через велику кількість магазинів інтернет комерції і та інтересу самих користувачів.

Системи рекомендацій використовуються для надання персоналізованих рекомендацій на основі профілю користувача та попередньої поведінки. Системи рекомендацій є в таких загальновідомих компаніях як Amazon, Netflix і YouTube і загалом широко використовуються в Інтернет-індустрії. Системи рекомендацій допомагають користувачам знаходити та вибирати елементи (наприклад, книги, фільми, ресторани) із великої колекції, доступної в Інтернеті або в інших електронних джерелах інформації. Серед великого набору предметів і опису потреби користувача, вони представляють користувачеві невеликий набір елементів, які найкраще підходять під опис. Так само система рекомендацій фільмів забезпечує рівень комфорту та персоналізація, яка допомагає користувачеві краще взаємодіяти з системою та переглядати фільми, які найкраще відповідає його потребам. Основна мета такої системи — рекомендувати фільми своїм користувачам на основі історії переглядів і оцінок, які вони надають. Система також порекомендує їхні продукти конкретним клієнтам на основі жанру фільмів, який вони віддають перевагу[1]. Спільна фільтрація та фільтрація на основі вмісту є основними підходами до надання рекомендацій користувачів.

Обидва вони найкраще застосовуються в конкретних сценаріях через їхні відповідні властивості.

Інтернет-комерція - це галузь, де вперше широко почали використовуватися системи рекомендацій. Маючи мільйони клієнтів і дані про їх поведінку в Інтернеті, компанії електронної комерції найкраще підходять для створення точних рекомендацій. Є досить багато доменів, де системи рекомендацій використовуються:

1. Роздрібна торгівля
Target лякав покупців ще в 2000-х роках, коли системи Target могли передбачити вагітність ще до того, як матері усвідомлювали свою власну вагітність. Дані про покупки є найціннішими даними, оскільки це найточніша інформація про наміри клієнта. Роздрібні торговці, які мають велику кількість даних про покупки, є лідерами серед компаній, які дають точні рекомендації.
2. ЗМІ
Подібно до електронної комерції, медіа-бізнеси одні з перших починають використовувати рекомендації. Важко побачити новинний сайт без системи рекомендацій.
3. Банківська справа
Продукт для масового ринку, який у цифровому вигляді використовують мільйони. Банківська справа для мас і малих і середніх підприємств є основними для рекомендацій. Знання детальної фінансової ситуації клієнта разом із його минулими вподобаннями в поєднанні з даними тисяч схожих користувачів є досить потужним.
4. Телеком
Він має схожу динаміку з банківською системою. Телекомунікаційні компанії мають доступ до мільйонів клієнтів, кожна взаємодія яких записується. Асортимент їхньої продукції також досить обмежений порівняно з іншими галузями, що робить рекомендації в телекомунікаціях легшою проблемою.

5. Комунальні послуги
Подібна динаміка з телекомунікаційними послугами, але комунальні послуги мають ще вужчий спектр продуктів, що робить рекомендації досить простими.

Системи рекомендацій збирають дані про клієнтів і автоматично аналізують ці дані, щоб створити персоналізовані рекомендації для ваших клієнтів. Ці системи покладаються як на неявні дані, такі як історія веб-перегляду та покупки, так і наявні дані, такі як оцінки, надані користувачем[21].

У після-ковідний час переваги систем рекомендацій стали ще більш очевидними і широко використовуваними. Отже, переваги рекомендаційних систем є наступними:

1. Збільшення продажів/конверсії
Існує дуже мало способів досягти збільшення продажів без збільшення маркетингових зусиль. Налаштувавши автоматизовану систему рекомендацій, ви без будь-яких зусиль отримуєте регулярні додаткові продажі.
2. Підвищення рівня задоволеності користувачів
Найкоротший шлях до продажу є чудовим, оскільки він зменшує зусилля як для вас, так і для вашого клієнта. Системи рекомендацій дозволяють вам скоротити шлях ваших клієнтів до продажу, рекомендуючи їм відповідний варіант іноді ще до того, як вони його шукатимуть.
3. Підвищена лояльність
Змусивши клієнтів витратити більше на вашому веб-сайті, ви можете покращити їхню обізнаність із вашим брендом та користувальницьким інтерфейсом, підвищивши ймовірність того, що вони роблять у вас покупки в майбутньому.
4. Зменшення відтоку клієнтів
Системні електронні листи з рекомендаціями є одним із найкращих способів

повторно залучити клієнтів. Знижки або купони – це інші ефективні, але дорогі способи повторного залучення клієнтів, які можна поєднати з рекомендаціями, щоб підвищити ймовірність переходу клієнтів.

Зростання глобальних роздрібних продажів продовжуватиме зростати та займатиме більшу частку роздрібного ринку. Згідно з дослідженням, проведеним eMarketer і Statista, роздрібні онлайн-продажі досягнуть 6,51 трильйона доларів до 2023 року, а веб-сайти електронної комерції займатимуть 22,3% від загального обсягу роздрібних продажів.

Незважаючи на те, що з 2020 року роздрібній торгівлі було важко, на кожному національному ринку, охопленому eMarketer, спостерігалось двозначне зростання електронної комерції. Тенденція продовжується:

1. У Латинській Америці (включаючи Перу, Бразилію, Аргентину, Чилі, Колумбію та Мексику) у 2022 році продажі електронної комерції склали 104 мільярди доларів, що на 22,4% більше порівняно з 85 мільярдами доларів у 2021 році.
2. Індійський ринок електронної комерції входить до п'ятірки найбільш швидкозростаючих країн у світі, у 2022 році зростання продажів склало 25,5%.
3. Прогнозується, що протягом наступних років Великобританія постійно зростатиме на 85,7 мільярда доларів (+42,88%).
4. Китай продовжує лідирувати на світовому ринку електронної комерції, на нього припадає 46,3% усіх роздрібних продажів електронної комерції в усьому світі, а загальний обсяг онлайн-продажів трохи перевищив позначку в 2,8 трильйона доларів у 2022 році. Він також має найбільше цифрових покупців у світі, 842,1 мільйона, що становить 39,4% від загального глобального заг.
5. Прогнозується, що ринок електронної комерції США досягне понад 904,9 мільярда доларів у 2022 році, що становить трохи більше третини китайського. Після Китаю та США третім за величиною ринком електронної комерції є

Велика Британія, на яку припадає 4,8% частки роздрібних продажів електронної комерції. За Великобританією йдуть Японія (3%) і Південна Корея (2,5%).

П'ять найкращих ринків електронної комерції не змінилися з 2018 року. Тенденції eMarketer свідчать про те, що ці ринки залишаться в першій п'ятірці до 2025 року.

Кейсі Армстронг, директор з маркетингу бренду ShipBob, що займається електронною комерцією, додає: «Хоча велика увага приділяється центрам електронної комерції в Сполучених Штатах і Канаді, є чому повчитися в інших великих міжнародних гравців, які спостерігають ще більш прискорене зростання електронної комерції. Він додає: «Продавці можуть змінити місце продажу на основі цих даних і попиту на електронну комерцію в цих країнах. У ShipBob саме тому ми відкрили центри виконання замовлення в Канаді та Великобританії та збираємося відкрити ще один в Австралії».

Завдяки глобальній пандемії електронна комерція зараз є найпопулярнішою формою торгівлі у світі. Оскільки багато людей вперше належним чином досліджують електронну комерцію під час глобальної пандемії, вона досягла свого найвищого піку в 2022 році та продовжує зберігати свою популярність, навіть коли магазини починають знову відкриватися.

Загалом за останнє десятиліття відбувся поступовий і стрімкий перехід від фізичних магазинів до онлайн-світу. Було сказано, що зростання електронної комерції на 23% у порівнянні з минулим роком.

Тепер люди можуть купувати від брендів і компаній, від одягу до засобів для чищення, по всьому світу одним натисканням кнопки. У 2022 році це актуально, ніж будь-коли, і електронна комерція стала найпопулярнішою за всю історію: компанії, починаючи від оптових продавців клінінгу та закінчуючи компаніями, що займаються виробництвом харчових добавок і навіть обсмажувачами кави, активно орієнтуються на електронну комерцію.

Ось деякі з причин, чому електронна комерція настільки популярна і залишиться тут надовго.

1. Мобільна адаптивність. Як ніколи, веб-трафік надходить через наші смартфони та мобільні пристрої. Вони є постійною частиною життя майже кожного, включно з тим, коли вони хочуть щось купити. У потязі, в машині чи просто лежачи в ліжку, легкість доступу до сайтів електронної комерції з вашого телефону робить його надзвичайно популярним і простим варіантом. Сайти електронної комерції переконуються, що їхні веб-сайти адаптовані для мобільних пристроїв, і це покращує взаємодію з користувачами та сприяє збільшенню продажів. Лише за останні кілька років у США мобільні покупки зросли з 3 до 31 мільярда доларів. Смартфони також мають пристрої визначення місцезнаходження, які дозволяють краще розуміти клієнтів і більш особистий досвід, про що цей посібник буде розглянуто пізніше.
6. Зручність. Зручність електронної комерції в порівнянні з фізичними магазинами незрівнянна. Сайти електронної комерції надають вам цілодобовий доступ до свого сайту та можливість купувати товари в будь-якому місці в будь-який час з будь-якого місця. Немає простоїв у святкові дні, затримок через погану погоду чи години закриття та довгих черг, з якими доводиться мати справу клієнтам. Електронна комерція є швидкою та безпроблемною, а завдяки її зручності у 2022 році вона стане набагато кращим і масштабнішим варіантом, ніж звичайні магазини. Клієнти можуть читати огляди в Інтернеті та порівнювати пропозиції та розпродажі одним натисканням кількох кнопок, не ходячи від магазину до магазину. Вони можуть робити покупки у вільний час легко та зручно, тому електронна комерція залишається такою популярною.
7. Індивідуальні продукти. В останні роки були створені складні алгоритми, які відстежують і розуміють шаблони клієнтів, коли мова заходить про електронну комерцію. Це дозволяє компаніям і рекламодавцям орієнтуватися на вибрану аудиторію, людей, які, як вони знають, можуть купити їхній продукт. Клієнти можуть отримувати рекомендації на основі своїх індивідуальних уподобань, і це

заохочує клієнтів купувати товари, які здаються абсолютно необхідними, оскільки вони відповідають їхнім особистим уподобанням. Це також вірно у випадку індивідуальних продуктів для працівників. Наприклад, компанії все частіше проводять медичні огляди персоналу, щоб переконатися, що він достатньо здоровий і готовий виконувати свою роботу. Якщо є щось важливе, співробітникам можна порадити придбати певний продукт, щоб зменшити будь-які ризики; це також перейшло до моделі електронної комерції. Замість того, щоб скеровувати співробітників до магазину або торгової точки, їм можна надати код або веб-сайт і зробити покупку там, часто в межах своєї компанії за допомогою внутрішнього процесу електронної комерції.

8. Кілька каналів для роздрібно́ї торгівлі. За останні кілька років електронна комерція створила світ, який означає, що бренди та компанії можуть мати кілька каналів для роздрібних торговців, через які вони можуть купувати їхні товари. Підприємства взаємодіють зі своїми клієнтами через веб-сайт, електронну пошту, соціальні мережі тощо, і це лише онлайн. Численні канали купівлі, отримання та обміну означають, що клієнти, швидше за все, будуть використовувати електронну комерцію, а не звичайні магазини, тому що їм це зручно.

Інтегрований досвід роздрібно́ї торгівлі через різні канали дає клієнтам ще одну причину використовувати електронну комерцію та пояснює її популярність у 2022 році.

Глобальний сплеск електронної комерції через COVID-19 спочатку виник через необхідність. Інтернет-магазини стали практичною альтернативою, оскільки роздрібні магазини закрилися, а люди залишилися, щоб уникнути вірусу. Насправді глобальна електронна комерція зросла з 15% від загального обсягу роздрібних продажів у 2019 році до 21% у 2021 році. Зараз вона становить приблизно 22% від продажів.

Але коли споживачі знову почали робити покупки особисто, інвестори почали запитувати: чи був стрибок Covid одноразовою угодою, чи зростання електронної комерції може тривати?

У довгостроковій перспективі ринок електронної комерції має багато можливостей для зростання, і він може зрости з 3,3 трильйона доларів сьогодні до 5,4 трильйона доларів у 2026 році.

«Ми вважаємо, що стрибок, викликаний Covid, не сповільнить майбутнє зростання електронної комерції», — каже Брайан Новак, аналітик фондового ринку, який займається інтернет-індустрією США. За його прогнозами, частка електронної комерції досягне 27% роздрібних продажів до 2026 року. «В усьому світі ми ще не побачили межі проникнення електронної комерції».

Багато факторів стимулюють зростання, зокрема логістика, володіння мобільними пристроями та розширення ринку. Для інвесторів це означає, що бум електронної комерції, ймовірно, триватиме, пропонуючи можливості для отримання прибутків у багатьох бізнесах, регіонах і вертикалях, і в той час, коли останні оцінки акцій не обов'язково відображають це зростання. Запровадження електронної комерції триває. Зростання цифрової комерції означає постійну зміну того, як люди здійснюють покупки. Насправді галузева модель Morgan Stanley разом з іншими даними свідчить про те, що електронна комерція продовжуватиме набирати обертів навіть у країнах, де покупки в Інтернеті вже популярні.

У Південній Кореї, завдяки добре розвиненій платіжній та логістичній інфраструктурі, онлайн-продажі вже складають 37% усієї роздрібної діяльності. Але зростання там не закінчено. Електронна комерція в Південній Кореї може зрости до 45% протягом наступних п'яти років завдяки доставці їжі та варіантам доставки в той же день.

Подібним чином у США електронна комерція може досягти 31% продажів до 2026 року проти 23% зараз, оскільки звичайні магазини закриваються, а споживачі віддають перевагу зручності.

Ринки електронної комерції на ранніх стадіях і нові сегменти також готові до значного зростання. Наприклад, у деяких частинах Південно-Східної Азії та Латинської Америки електронна комерція може зрости на 17% і 20% відповідно протягом наступних п'яти років і щорічно зростати.

Якщо говорити про сегменти, то електроніка, яка лідирує в усіх категоріях електронної комерції, планується збільшити з 38% роздрібних продажів до 45% світових роздрібних продажів. Цифрові продажі також зростають у нових галузях, зокрема в сфері краси, одягу та продуктів. «Хоча в певних країнах і галузях існують перешкоди, ми вважаємо, що ці бар'єри й надалі зникатимуть», — говорить Ендрю Рубен, який займається роздрібною торгівлею та електронною комерцією в Латинській Америці. За кадром удосконалення в усьому – від цифрових платежів до ланцюга поставок і можливостей виконання – покращують взаємодію з клієнтами, що сприяє подальшим змінам у поведінці споживачів.

«Одним із найголовніших факторів диференціації платформ електронної комерції може бути ланцюжок поставок і можливості виконання замовлення, що може сприяти кращому обслуговуванню клієнтів», — каже Гарі Ю, який займається телекомунікаційними та інтернет-акціями в Китаї, де доставка продуктів зростає швидше, ніж у Китаї. - комерція в цілому.

Використання Інтернету та розширене підключення також є значними рушійними силами, особливо на ринках, що розвиваються, де населення стає молодшим і проводить більше часу в Інтернеті, ніж їхні колеги на розвинених ринках. Наприклад, споживачі в Колумбії та Бразилії проводять в Інтернеті в середньому більше п'яти годин щодня, створюючи значну можливість для роздрібних торговців вийти на новий ринок клієнтів. Лояльність до бренду знизилася після пандемії, частково через дефіцит продуктів, і тепер споживачі охочіше пробують нові продукти та послуги, ніж раніше. Це в першу чергу на ринку споживчих товарів, де клієнти більш схильні ризикувати свіжими брендами продуктів харчування та напоїв. Однак це додатково підтверджується загальним зростанням кількості людей, які використовують електронну комерцію для купівлі продуктів,

причому деякі сектори спостерігають еквівалент п'ятирічного зростання протягом п'ятимісячного періоду.

Використання штучного інтелекту (AI) для покращення досвіду покупок для користувачів також стало невід'ємною частиною розумних стратегій електронної комерції. Купуючи та збираючи дані, компанії можуть пристосувати досвід покупок до окремих користувачів, надаючи їм пропозиції щодо продуктів, які відповідають їхнім уподобанням та історії покупок. Постійне вдосконалення аналітичних можливостей штучного інтелекту, а також його зростаюча повсюдність зробили це завдання ще більш керованим і доступнішим для компаній, які бажають інвестувати в технології.

Важливість електронної комерції для світової економіки незаперечна, як і необхідність для компаній адаптуватися до її тенденцій, щоб підтримувати підтримку клієнтів і досягати нових висот успіху. У той час як великі компанії отримують вигоду від нових технологій, однією з переваг сучасної електронної комерції є її сила як вирівнювача. Стартапи, малі підприємства, онлайн-магазини для мам і попів — усі вони отримують конкурентні переваги, які просто не були можливими десять років тому.

Низькі бар'єри входу на ринок і доступні операційні витрати роблять онлайн-магазини та інструменти електронної комерції доступними майже для будь-кого. Якщо ви розумієте ринок і його тенденції, у вас може бути формула успіху. Сучасний ринок електронної комерції дуже відрізняється від того, як він виглядав лише кілька років тому. Продажі електронної комерції становили 19% станом на лютий 2020 року. Лише рік потому цифри різко зросли до 35%, і це не сповільниться найближчим часом. Насправді очікується, що у 2022 році обсяг електронної комерції в США досягне 1,065 трильйона доларів.

У той час як глобальні блокування змусили закрити звичайні магазини, онлайн-магазини були зайняті як ніколи, використовуючи різні моделі електронної

комерції, щоб задовольнити запити споживачів. Більше брендів навіть запускають цифрову комерцію для кількох бізнес-моделей.

Деякі B2B-підприємства вперше прийняли D2C, тоді як інші бренди D2C «випробовують воду» з B2B2C, щоб допомогти їм забезпечити всі шляхи виходу на ринок. У 2017 році Nike покладалася на 30 000 роздрібних продавців, щоб донести свій продукт до кінцевих споживачів, але до 2019 року їй знадобилося лише 40 стратегічних роздрібних партнерів по всьому світу, запровадивши роздрібну торгівлю D2C. Проте його цифрові продажі зросли на 82%.

Незалежно від того, чи є ви брендом, який швидко розвивається, чи традиційним роздрібним продавцем, знання типів електронної комерції, які підходять для вашого бізнесу, є одним із ключових кроків, які допоможуть розвивати ваш бізнес. У цій статті ми розглянемо типи бізнес-моделей електронної комерції, включаючи відповідні приклади, щоб підживити ваше натхнення.

Електронну комерцію B2B також називають цифровою комерцією B2B. Ця форма торгівлі, коли ділові операції відбуваються всередині платформи онлайн-закупівель або будь-якого іншого онлайн-простору. 71% покупців B2B надають переваги покупкам на платформах електронної комерції B2B, що працюють з напруженим традиційним процесом.

Ця бізнес-модель електронної комерції спрощує закупівлі для оптових і роздрібних торговців, забезпечуючи безпечний профіль для поповнення замовлень і запиту пропозицій. Продавці можуть переключити свою увагу з продажу на освітній продукт та створення нових рішень і побудувати конкретні стосунки з покупцями. За даними Market Watch, світовий ринок електронної комерції B2B у 2021 році мав ринкову вартість у 8 523,3 мільярда доларів США, а до 2027 року, за прогнозами, досягає 18 771,4 мільярда доларів США, що робить його прибутковим ринком для брендів. Крім того, компанії B2B інвестують у створення магазинів електронної комерції для свого бізнесу, далі це дозволяє їм розширити потік доходу.

Завдяки електронній комерції B2B постачальники можуть отримати доступ до цілого класу покупців, готових платити за свої рішення. Вони також можуть запропонувати індивідуальні та індивідуальні пропозиції продуктів на основі групи клієнтів, поведінки покупців або профілю клієнта, що призведе до збільшення продажів і кращої рентабельності інвестицій. Однак наступна електронна комерція B2B має справу з діловими транзакціями, вона має обмежену ринкову базу, ніж модель B2C, що ускладнює початок для малого та середнього бізнесу електронної комерції. Крім того, транзакції між компаніями передбачають тривалий цикл продажів, після чого в процесі роботи беруть участь численні зацікавлені сторони. Отже, ця бізнес-модель електронної комерції не підходить для брендів, які потрібен швидкий час виконання.

Електронна комерція B2C – це транзакція між продавцем або роздрібним продавцем і кінцевим клієнтом на платформі електронної комерції. Сьогодні це найпоширеніший вид електронної комерції. Kaleido прогнозує, що транскордонна електронна комерція B2C досягне 1,2 трильйона доларів у 2022 році, що означає, що цей тип електронної комерції відкриває багато унікальних можливостей.

Клієнти купують товари в Інтернеті в автономному магазині електронної комерції або на ринку, як-от Amazon. Це дозволяє клієнтам купувати товар де завгодно, а не відвідувати традиційний звичайний магазин.

B2C – це сегмент електронної комерції, орієнтований на споживачів. У той час як B2B електронна комерція зосереджена на побудові відносин з покупцями, B2C торговці залучають (і утримують) клієнтів за допомогою конкурентоспроможних цін, яскравих зображень продукту та знижок.

Таким чином, бренди та роздрібні продавці у сфері B2C серйозно ставляться до інвестування в зручний, добре пов'язаний і яскравий веб-сайт. Оскільки клієнти B2C роблять менші замовлення та швидше приймають рішення про купівлю, продавці витрачають невелику суму на маркетинг і просування своїх продуктів.

Хоча постачальникам легше виконувати завдання під час трансакцій B2B, для електронної комерції B2C зазвичай потрібен спеціальний відділ, відповідальний за до- та післяпродажне обслуговування, щоб відповідати на запити та вирішувати будь-які проблеми. Крім того, цей тип електронної комерції стосується продажу товарів невеликими обсягами, що вимагає індивідуальної доставки та виробництва для кожної покупки. Електронна комерція D2C або DTC — це модель електронної комерції, яка передбачає продаж безпосередньо кінцевому споживачу без продажу через роздрібного продавця, дистриб'ютора, оптового продавця чи інші торгові точки. У цій моделі компанії обходять діяльність роздрібних торговців і посередників і розвивають справжні стосунки зі своїми кінцевими клієнтами.

І покупці, і продавці можуть отримати вигоду від моделі D2C. Покупці отримують персоналізовані та індивідуальні пропозиції продуктів без особливих проблем. З іншого боку, бренди вчаться на даних клієнтів, створюють сегменти та продають більше продуктів.

У минулому бренди продавали товари роздрібним торговцям або посередникам, які підтримували відносини з клієнтами. Це означає, що бренди не мають доступу до цінних даних про тенденції купівлі, демографічні показники та місцезнаходження клієнтів, щоб забезпечити персоналізований досвід для покупців.

У ринковій моделі електронної комерції і покупці, і продавці здійснюють трансакції на закритій централізованій платформі. Після кожної завершеної трансакції ринок бере комісію як з постачальників, так і з споживачів, але він не зберігає запаси; він співпрацює з виробниками та брендами, щоб відправляти від їх імені.

Незважаючи на те, що ця модель існує вже багато років, у 2021 році вона відчула величезний сплеск, коли відомі імена, як-от Майклз, оголосили про плани захопити її. Існуючі ринки, як-от ринок Walmart, також значно зросли, збільшивши кількість продавців на його платформі вдвічі в середині 2020 року. Крім того, експерти прогнозують, що до 2024 року ринки B2C у всьому світі можуть досягти приблизно

3,5 трильйонів доларів. В електронній комерції на ринковому майданчику ринок надає постачальникам доступ із можливістю натискання для отримання замовлень від споживачів, відображаючи їхні продукти, включаючи їхні характеристики та ціни. Як посередник, ринок обробляє всі «необхідні транзакції» для постачальника, включаючи онлайн-оплату, імпорт та експорт, логістику, страхування, просування та захист споживачів. Крім того, ринок надає споживачам широкий вибір продуктів від різних постачальників і різної якості та ціни. Для клієнтів у всьому світі інфляція стала основною проблемою для 40% респондентів — вище бідності та соціальної нерівності (31%), безробіття та робочих місць (26%) і коронавірусу (12%).

Хоча люди все ще витрачають гроші, глобальні бренди не звільнені від тиску інфляції. Таким чином, якщо ви плануєте використовувати транскордонну електронну комерцію, вам потрібно врахувати вплив інфляції в різних країнах. У недавньому інтерв'ю виданню Glossy генеральний директор Olarplex Джу Е Вонг пояснює, що навіть із зростанням чистих продажів на 38,6% у другому кварталі 2022 року Olarplex не застрахований від глобальної інфляції, макроекономічних і геополітичних ситуацій. У результаті Джу Е Вонг каже, що хоча бізнес багатий на готівку, вони інвестуватимуть у технології (наприклад, рішення для електронної комерції) і технологічні можливості, щоб сприяти зростанню. Очікується, що до 2023 року роздрібні продажі електронної комерції в Азіатсько-Тихоокеанському регіоні будуть більшими, ніж у решті світу разом узятих. Це пов'язано з: швидкою урбанізацією та технологічним прогресом; більше 85% нового середнього класу, що приростає, проживає в Азіатсько-Тихоокеанському регіоні; і безліч урядових і приватних ініціатив у Китаї. Обсяг продажів електронної комерції в Китаї у 2022 році склав приблизно 2,8 трильйона доларів, що більш ніж удвічі перевищує ринок США. Що стосується B2B, виробництво в Азіатсько-Тихоокеанському регіоні та Китаї пережило ренесанс. Як наслідок, диспропорція B2B стає ще чіткішою.

1.2 Огляд аналогів на ринку

Доречним буде привести приклади компаній, які використовують системи рекомендацій на регулярній основі і отримують від цього користі та кращі результати, конкуруючи з іншими компаніями на ринку.

Адже це підкреслює вищенаведену користь різних рекомендаційних систем, їх місце на ринку та широка використовуваність таких технологій і алгоритмів реальними бізнесами та компаніями. Є досить великий список загальновідомих компаній, які регулярно користуються різноманітними рекомендаційними системами задля кращої взаємодії з користувачами та підвищення прибутку. Такими компаніями є:

1. Amazon.com

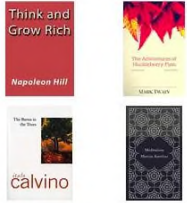



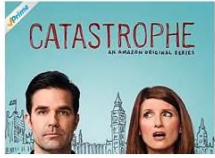



Amazon.com використовує рекомендації спільної фільтрації по елементах на більшості сторінок свого веб-сайту та в кампаніях електронної пошти. За даними McKinsey, 35% покупок Amazon здійснюються завдяки системам рекомендацій. Кілька прикладів, коли Amazon використовує системи рекомендацій:

2.

Netflix

Це ще одна компанія, що керується даними, яка використовує системи рекомендацій для підвищення задоволеності клієнтів. Те саме дослідження McKinsey, яке ми згадували вище, підкреслює, що 75% переглядів Netflix обумовлені рекомендаціями. Насправді Netflix настільки одержимий наданням найкращих результатів для користувачів, що вони провели змагання з вивчення

Recommended for you, Thomas

 <p>Literature & Fiction 62 ITEMS</p>	 <p>Exercise & Fitness Equipment 8 ITEMS</p>	 <p>Health, Fitness & Dieting Books 37 ITEMS</p>	 <p>Tableware 12 ITEMS</p>
 <p>Prime Video – Unlimited Streaming for Prime Members 12 ITEMS</p>	 <p>Coffee, Tea & Espresso 98 ITEMS</p>	 <p>Biographies & Memoirs 17 ITEMS</p>	 <p>Engineering Books 7 ITEMS</p>

даних під назвою Netflix Prize, де той, хто має найточніший алгоритм рекомендації фільмів, виграє приз у розмірі 1 000 000 доларів США.

3.

Spotify

Ця компанія щотижня створює новий персоналізований список відтворення для кожного передплатника під назвою «Відкривайте щотижня», який є персоналізованим списком із 30 пісень на основі унікальних музичних смаків користувачів.

Придбання Echo Nest, музичного інтелектуального стартапу та аналітики даних, дає їм змогу створити систему музичних рекомендацій, яка використовує три різні типи моделі рекомендацій: Спільна фільтрація: фільтрація пісень шляхом порівняння даних про прослуховування користувачами з історією прослуховування інших користувачів.

Обробка природної мови: пошук в Інтернеті інформації про певних виконавців і пісні. Потім кожному виконавцю чи пісні призначається динамічний список найпопулярніших термінів, який змінюється щодня та зважується за релевантністю. Потім механізм визначає, чи схожі два музичні твори або виконавці. Аналіз аудіофайлу: алгоритм аналізує характеристики кожного

окремого аудіофайлу, включаючи темп, гучність, тональність і тактовий розмір, і дає відповідні рекомендації.

4.

LinkedIn

Як і будь-який інший канал соціальних мереж, LinkedIn також використовує рекомендації типу «Ви також можете знати» або «Вам також може сподобатися».

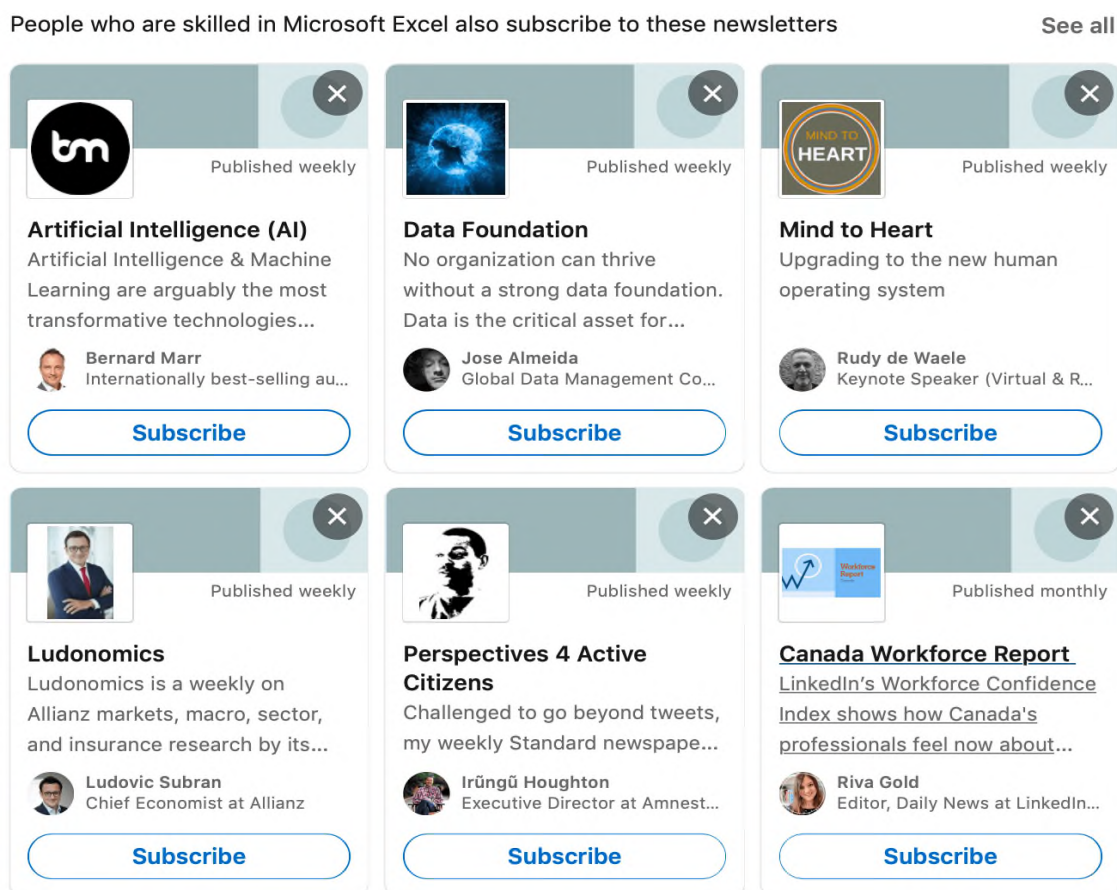


Рисунок 1.2 - Приклад рекомендацій Лінкедін

1.3 Методи та алгоритми формування рекомендацій

Внутрішній штучний інтелект створює модель онлайн-особистості кожного клієнта. Алгоритм персоналізації не обмежується історією покупок. Він також може обробляти цифри з публікацій у соціальних мережах, рішень про купівлю на інших сайтах і запитів Google. Ці складні математичні моделі спираються на показники відкритості, сумлінності, екстраверсії, приємності та невротизму,

пов'язані з даними, отриманими в результаті мільйонів транзакцій клієнтів, із привабливими характеристиками продуктів[18]. Але власник сайту електронної комерції ніколи не бачить ці розрахунки, виконані за мілісекунди за допомогою програмного забезпечення для персоналізації, доданого до задньої частини вашого сайту.

Завдання полягає в тому, щоб продовжувати знаходити варіанти, які дозволять вам рекомендувати продукти, які забезпечують хорошу прибуток і які сподобаються клієнтам. Програмне забезпечення для персоналізації продукту обчислює цифри, тож на сайті представлені продукти, які нмають вищі шанси сподобатися клієнтам. Тобто сайт показуватиме клієнтам лише ті продукти, які вони хочуть купити. Правильно виконана персоналізація продукту зменшує ваші витрати на маркетинг і збільшує утримання клієнтів[17].

Персоналізовані рекомендації щодо продуктів створюються на основі даних користувачів, взятих із веб-сайту. В ідеалі показується клієнтам лише ті продукти, які вони дійсно хотіли б купити. Чим більш персоналізований досвід, тим краще.

Коли все зроблено правильно, персоналізовані рекомендації щодо продуктів можуть заощадити гроші та збільшити утримання клієнтів[11]. Звіти за 2018 рік показує, що покупці, які повернулися до запропонованого продукту (наприклад, натиснувши його та прочитавши опис), мали на 55% більше шансів зробити покупку під час цього сеансу покупок. Для нових клієнтів ця цифра становила 70%.

Механізми рекомендацій, тобто програми, які аналізують дані як про продукти, так і про користувачів, забезпечують такі функції. Не тільки роздрібні торговці використовують ці механізми, веб-сайти потокового відео використовують їх для створення списків рекомендованих кліпів для перегляду своїми користувачами[2].

Існує 3 основних типи механізмів рекомендацій:

1. Спільні системи фільтрації
2. Системи фільтрації на основі контенту

3. Гібридні системи рекомендацій

Системи рекомендацій зазвичай розглядають такі дані, як:

1. Пошукові запити клієнта
2. Історія їх покупок
3. Що зараз у їх кошику для покупок
4. Їхня соціальна поведінка (лайки, поширення тощо)
5. Їх географічне положення
6. Сегмент аудиторії клієнта (демографічні показники)

Спільна фільтрація є першим методом, який розглядається. Спільна фільтрація аналізує дані від кількох клієнтів (статистика працює найкраще, якщо кількість клієнтів становить принаймні 1500), щоб створити панель продуктів, які, ймовірно, сподобаються клієнтам найбільше. Він використовує досвід усіх клієнтів, щоб рекомендувати продукти для кожного з клієнтів[3]. Прикладом є відвідувач сайту, який переглянув камеру DSLR, система може запропонувати людині об'єктиви, які інші клієнти придбали з цією камерою.

Спільна фільтрація покладається не лише на об'єднання даних між клієнтами. Вона фільтрує рекомендації продукту за історією пошуку, за географічним розташуванням і за історією клієнта з моделлю та сайтом. Його головна перевага полягає в тому, що він додає інформацію від кількох клієнтів до персоналізованих рекомендацій для одного клієнта.

Фільтрація на основі вмісту генерує рекомендації «Якщо вам сподобалося це, то вам також може сподобатися те». Він розглядає модель рішення про покупку окремого клієнта, а не дані групи клієнтів. Фільтрація на основі вмісту розглядає дані лише від одного клієнта за раз. Приклад такої фільтрації: припустимо, що клієнт купив мигдаль у березні, волоські горіхи у квітні, кешью у травні та

повернувся на ваш сайт, щоб знову замовити горіхи в червні. Фільтрація на основі вмісту може відображати макадамії та фісташки як додаткові продукти, які їм можуть сподобатися.

Гібридні системи рекомендацій поєднують спільну фільтрацію та фільтрацію на основі вмісту. Вони використовують дані схожих клієнтів, а також дані з попередньої історії пошуку та покупок клієнта. Сервіси потокового передавання відео зазвичай проводять спільну перевірку (щоб визначити, які нові відео, ймовірно, вам сподобаються) у поєднанні з перевіркою на основі вмісту (що зіставляє доступні на даний момент відео з відео, яким ви давали високі оцінки в минулому). Цей вид скринінгу також може враховувати географічне розташування та заявлені симпатії та антипатії.

1.4 Постановка завдання

Завданням кваліфікаційної роботи є розробка та дослідження формування рекомендацій для веб-порталу з інтернет комерції. Системі, маючи набір даних про вибір на поведінку користувачів на сайті інтернет-комерції, потрібно використати цей набір даних для подальшого аналізу та рекомендацій продукції.

Основні характеристики для створення рекомендацій на основі даних про користувачів це подібність продуктів та рейтинг продуктів. Спираючись на ці характеристики, створюється можливість для рекомендації найбільш подібних продуктів для клієнта.

Аналіз на порівняння методів рекомендацій містить в собі аналіз алгоритму рекомендацій, їх недоліки, переваги та складності реалізації. Результатом експерименту є вибір найефективнішого методу рекомендацій, спираючись на такі характеристики як коефіцієнт подібності та рейтинг рекомендованої продукції.

Таким чином, для розробки та дослідження формування рекомендацій необхідно виконати такі завдання:

1. Визначення існуючих методів рекомендацій
2. Вибір декількох методів рекомендацій для експерименту
3. Визначення набору даних про користувачів, який буде

використовуватися для подальшого експерименту

4. Реалізація кожного методу рекомендацій окремо на базі набору даних
5. Порівняння результатів різних методів на основі обраних характеристик
6. Визначення найбільш ефективного методу рекомендацій

2 ОГЛЯД СИСТЕМ І ІНСТРУМЕНТІВ

2.1 Огляд системи на основі контенту

Система рекомендацій на основі контенту — це певний тип систем рекомендацій, які базуються на явних знаннях про асортимент товарів, уподобаннях користувачів і критеріях рекомендацій (тобто, який товар слід рекомендувати в якому контексті). Ці системи застосовуються в сценаріях, коли неможливо застосувати альтернативні підходи, такі як спільна фільтрація та фільтрація на основі вмісту.

Основною перевагою систем рекомендацій, заснованих на знаннях, є відсутність проблем із холодним пуском (розгоном). Відповідним недоліком є потенційне вузьке місце отримання знань, викликане необхідністю визначення рекомендаційних знань у явний спосіб. Цей метод повністю обертається навколо порівняння інтересів користувачів із функціями продукту. Рекомендовано продукти, функції яких найбільше збігаються з інтересами користувачів.

Враховуючи важливість функцій продукту в цій системі, важливо обговорити, як визначаються улюблені функції користувача. Тут можна використовувати два методи (можливо в комбінації). По-перше, користувачам можна надати список функцій, з яких вони можуть вибрати те, що їм найбільше подобається. По-друге, алгоритм може відстежувати продукти, які користувач вибрав раніше, і додавати ці функції до даних користувачів[4].

Подібним чином особливості продукту можуть визначити самі розробники продукту. Крім того, користувачів можна запитати, які функції, на їхню думку, найбільше ідентифікують продукти. Цей метод повністю обертається навколо

порівняння інтересів користувачів із функціями продукту. Рекомендовано продукти, функції яких найбільше збігаються з інтересами користувачів.

Враховуючи важливість функцій продукту в цій системі, важливо обговорити, як визначаються улюблені функції користувача.

Тут можна використовувати два методи (можливо в комбінації). По-перше, користувачам можна надати список функцій, з яких вони можуть вибрати те, що їм найбільше подобається. По-друге, алгоритм може відстежувати продукти, які користувач вибрав раніше, і додавати ці функції до даних користувачів.

Подібним чином особливості продукту можуть визначити самі розробники продукту. Крім того, користувачів можна запитати, які функції, на їхню думку, найбільше ідентифікують продукти.

	Feature 1	Feature 2	Feature 3	Feature 4
Product 1	1		1	2
Product 2		1	4	
Product 3	3			1
User Data	2		1	1

Рисунок 2.1 - Приклад алгоритму на основі контенту

Після того, як функціям продукту та інтересам користувачів присвоєно числове значення, будь то двійкове значення 1 або 0 або довільне число, потрібно визначити метод визначення подібності між продуктами та інтересами користувачів. Дуже простою формулою буде скалярний добуток. Для розрахунку скалярного добутку слід використовувати наступну формулу:

$$\sum_{i=1}^d p_i u_i, (2.1)$$

де p_i — значення функції продукту, а u_i і i — значення інтересу користувача в стовпці i .

2.2 Огляд системи спільної фільтрації

Спільна фільтрація (CF) — це техніка, яка використовується системами рекомендацій. Спільна фільтрація має два значення: вузьке та більш загальне. У новішому, вужчому розумінні, спільна фільтрація — це метод автоматичного прогнозування (фільтрування) щодо інтересів користувача шляхом збору інформації про вподобання або смак від багатьох користувачів (співпраця). Припущення, що лежить в основі підходу спільної фільтрації, полягає в тому, що якщо особа А має таку саму думку, що й особа Б щодо певного питання, А, швидше за все, буде мати думку В з іншого питання, ніж думка випадково обраної особи. Наприклад, спільна система рекомендацій фільтрації для уподобань у телевізійних програмах може робити прогнози щодо того, яке телешоу має сподобатися користувачеві, враховуючи частковий список смаків цього користувача (подобається чи не подобається). Зауважте, що ці прогнози є специфічними для користувача, але використовують інформацію, отриману від багатьох користувачів. Це відрізняється від простішого підходу надання середньої (неконкретної) оцінки для кожного цікавого елемента, наприклад, на основі кількості голосів[5].

У більш загальному сенсі спільна фільтрація — це процес фільтрації інформації або шаблонів за допомогою методів, що включають співпрацю між кількома агентами, точками зору, джерелами даних тощо. Застосування спільної фільтрації зазвичай включають дуже великі набори даних. Методи спільної фільтрації застосовувалися до багатьох різних типів даних, включаючи: дані зондування та моніторингу, наприклад, під час розвідки корисних копалин, зондування навколишнього середовища на великих територіях або використання декількох датчиків; фінансові дані, такі як фінансові установи, які об'єднують багато фінансових джерел; або в електронній комерції та веб-додатках, де основна

увага приділяється даним користувача тощо. Залишок цього обговорення зосереджено на спільній фільтрації даних користувача, хоча деякі методи та підходи також можуть застосовуватися до інших основних програм.

Підхід використовує дані рейтингу користувачів для обчислення подібності між користувачами або елементами. Типовими прикладами цього підходу є CF на основі сусідів і рекомендації Top N на основі елементів/користувачів. Наприклад, у підходах, орієнтованих на користувача, значення оцінок, які користувач u дає елементу i , обчислюється як сукупність оцінок деяких схожих користувачів щодо елемента:

$$r_{u,i} = \text{aggr}_{u' \in U} r_{u',i} \quad (2.2.1)$$

де U позначає набір перших N користувачів, які найбільше схожі на користувача u , який оцінив елемент i . Ось деякі приклади функції агрегації:

$$r_{u,i} = \frac{1}{N} \sum_{u' \in U} r_{u',i} \quad (2.2.2)$$

$$r_{u,i} = k \sum_{u' \in U} \text{simil}(u, u') r_{u',i} \quad (2.2.3)$$

, де r - середня оцінка користувача u для всіх елементів, оцінених u .

Алгоритм на основі сусідства обчислює подібність між двома користувачами або елементами та створює прогноз для користувача, беручи середньозважене значення всіх оцінок. Важливою частиною цього підходу є обчислення подібності між елементами або користувачами. Для цього використовуються численні показники, такі як кореляція Пірсона та векторна косинус-подібність. Подібність кореляції Пірсона двох користувачів x, y визначається як:

$$\text{simil}(x, y) = \frac{\sum_{i \in I_{xy}} (r_{x,i} - \bar{r}_x)(r_{y,i} - \bar{r}_y)}{\sqrt{\sum_{i \in I_{xy}} (r_{x,i} - \bar{r}_x)^2} \sqrt{\sum_{i \in I_{xy}} (r_{y,i} - \bar{r}_y)^2}} \quad (2.2.4)$$

де I_{xy} — набір елементів, оцінених користувачами x і y .

Косинусний підхід визначає косинусну подібність між двома користувачами x і y як:

$$\text{simil}(x, y) = \cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \times \|\vec{y}\|} = \frac{\sum_{i \in I_{xy}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i \in I_x} r_{x,i}^2} \sqrt{\sum_{i \in I_y} r_{y,i}^2}} \quad (2.2.5)$$

Алгоритм перших N рекомендацій на основі користувача використовує векторну модель на основі подібності для визначення k найбільш схожих користувачів на активного користувача. Після того, як знайдено k найбільш схожих користувачів, їхні відповідні матриці елементів користувача агрегуються для визначення набору елементів, які будуть рекомендовані. Популярним методом пошуку схожих користувачів є хешування, чутливе до локальності, яке реалізує механізм найближчого сусіда в лінійному часі[6].

2.3 Огляд гібридної системи

Гібридна система рекомендацій - це рекомендації, якими користуються більшість систем, поєднуючи спільну фільтрацію, фільтрацію на основі вмісту та інші підходи. Немає жодних причин, чому кілька різних технік одного типу не можуть бути змішані. Гібридні підходи можна реалізувати декількома способами: створюючи прогнози на основі контенту та на основі співпраці окремо, а потім об'єднуючи їх; шляхом додавання можливостей на основі контенту до підходу на основі співпраці (і навпаки); або шляхом уніфікації підходів в одну модель. Декілька досліджень, які емпірично порівнюють ефективність гібридних методів із суто спільними методами та методами, заснованими на вмісті, продемонстрували, що гібридні методи можуть надати більш точні рекомендації, ніж чисті підходи. Ці методи також можна використовувати для подолання деяких загальних проблем у системах рекомендацій, таких як холодний запуск і проблема розрідженості, а також вузьке місце інженерії знань у підходах, заснованих на знаннях[7].

Netflix є хорошим прикладом використання гібридних систем рекомендацій. Веб-сайт надає рекомендації, порівнюючи звички перегляду та пошуку подібних користувачів (тобто спільна фільтрація), а також пропонуючи фільми, які мають спільні характеристики з фільмами, які користувачі високо оцінили (фільтрація на основі вмісту).

2.4 Огляд інструментів реалізації

Paperspace — це рішення для віртуалізації, розроблене, щоб допомогти компаніям будь-якого розміру створювати, навчати та розгортати моделі машинного навчання. Використовуючи модуль Core, менеджери можуть створювати та підключати користувачів, мережі та машини для керування серверами за допомогою централізованої інформаційної панелі. Адміністратори можуть запускати Photoshop, 3DStudio Max, Catia, SolidWorks, After Effects та інші високопродуктивні робочі станції в єдиному інтерфейсі.

PyTorch — це оптимізована тензорна бібліотека для глибокого навчання за допомогою графічних і центральних процесорів. Деяка частина роботи стосується створення рекомендацій користувача щодо фільмів за допомогою кількох різних методів матриці факторизації.

Python — це мова програмування високого рівня загального призначення. Його філософія дизайну наголошує на читабельності коду з використанням значних відступів. Python динамічно типізується та збирає сміття. Він підтримує кілька парадигм програмування, включаючи структуроване (зокрема процедурне), об'єктно-орієнтоване та функціональне програмування. Її часто описують як мову «батареї включені» через її повну стандартну бібліотеку[10].

Гвідо ван Россум почав працювати над Python наприкінці 1980-х як наступником мови програмування ABC і вперше випустив її в 1991 році як Python

0.9.0. Python 2.0 був випущений у 2000 році та представив нові функції, такі як розуміння списків, збирання сміття з визначенням циклу, підрахунок посилань і підтримка Unicode. Python 3.0, випущений у 2008 році, був основною версією, яка не повністю сумісна з попередніми версіями.

Python був задуманий наприкінці 1980-х років Гвідо ван Россумом із Centrum Wiskunde & Informatica (CWI) у Нідерландах як наступник мови програмування ABC, яка була натхненна SETL, здатною обробляти винятки (від початок плюс нові можливості в Python 3.11) і взаємодія з операційною системою Amoeba. Його реалізація почалася в грудні 1989 року. Ван Россум ніс виключну відповідальність за проект як провідний розробник до 12 липня 2018 року, коли він оголосив про свою «постійну відпустку» від своїх обов'язків «доброзичливого диктатора Python на все життя» — титул, який спільнота Python надала йому, щоб відобразити його довгострокові зобов'язання як головного органу, який приймає рішення в проекті. У січні 2019 року активні розробники ядра Python обрали керівну раду з п'яти членів для керівництва проектом.

Python 2.0 був випущений 16 жовтня 2000 року з багатьма основними новими функціями. Python 3.0, випущений 3 грудня 2008 року, з багатьма основними функціями, портованими до Python 2.6.x і 2.7.x. Випуски Python 3 включають утиліту 2to3, яка автоматизує переклад коду Python 2 на Python 3.

Кінець терміну служби Python 2.7 спочатку був призначений на 2015 рік, а потім перенесений на 2020 рік через занепокоєння тим, що великий обсяг існуючого коду не можна буде легко перенести на Python 3. Для нього не буде випущено жодних додаткових патчів безпеки чи інших покращень. Наразі підтримуються лише версії 3.7 і новіші. У 2021 році Python 3.9.2 і 3.8.8 було прискорено, оскільки всі версії Python (включно з 2.7) мали проблеми з безпекою, що призводило до можливого віддаленого виконання коду та отруєння веб-кешу.

У 2022 році Python 3.10.4 і 3.9.12 було прискорено, а також 3.8.13 і 3.7.13 через багато проблем з безпекою. Коли Python 3.9.13 було випущено в травні 2022

року, було оголошено, що серія 3.9 (приєднується до старих серій 3.8 і 3.7) отримуватиме лише виправлення безпеки в майбутньому. 7 вересня 2022 року було випущено чотири нові випуски через потенційну атаку на відмову в обслуговуванні: 3.10.7, 3.9.14, 3.8.14 і 3.7.14.

Станом на листопад 2022 року Python 3.11.0 є поточною стабільною версією, і серед помітних змін порівняно з 3.10 є те, що він на 10–60% швидший і значно покращено звітування про помилки. У Python 3.12 (alpha 2) покращено повідомлення про помилки. Python є багатопарадигмальною мовою програмування. Об'єктно-орієнтоване програмування та структуроване програмування повністю підтримуються, і багато їхніх функцій підтримують функціональне та аспектно-орієнтоване програмування (включаючи метапрограмування та метаоб'єкти). Багато інших парадигм підтримуються через розширення, включаючи проектування за контрактом і логічне програмування.

Python використовує динамічну типізацію та комбінацію підрахунку посилань і збирача сміття, що визначає цикл, для керування пам'яттю. Він використовує динамічне розпізнавання імен (пізні зв'язування), яке зв'язує імена методів і змінних під час виконання програми.

Його дизайн пропонує певну підтримку функціонального програмування в традиції Lisp. Він має функції фільтра, відображення та зменшення; списки розуміння, словники, набори та генераторні вирази. Стандартна бібліотека має два модулі (itertools і functools), які реалізують функціональні інструменти, запозичені з Haskell і Standard ML.

Його основна філософія коротко викладена в документі The Zen of Python (PEP 20), який містить такі афоризми, як:

Красиве краще, ніж потворне.

Явне краще, ніж неявне.

Просте краще, ніж складне.

Складне краще, ніж складне.

Читабельність має значення.

Замість того, щоб побудувати всю свою функціональність у своєму ядрі, Python був розроблений таким чином, щоб бути розширюваним за допомогою модулів. Ця компактна модульність зробила його особливо популярним як засіб додавання програмованих інтерфейсів до існуючих програм. Бачення Ван Россума невеликої базової мови з великою стандартною бібліотекою та інтерпретатором, що легко розширюється, впливає з його розчарування в АВС, який підтримував протилежний підхід.

Python прагне до простішого, менш захищеного синтаксису та граматики, надаючи розробникам можливість вибору методології кодування. На відміну від девізу Perl «є більше, ніж один спосіб зробити це», Python приймає філософію «повинен бути один — і бажано тільки один — очевидний спосіб зробити це». Алекс Мартеллі, співробітник Python Software Foundation і автор книжки про Python, написав: «Описати щось як «розумне» не вважається компліментом у культурі Python».

Розробники Python прагнуть уникати передчасної оптимізації та відкидають латки для некритичних частин еталонної реалізації CPython, які забезпечуватимуть незначне збільшення швидкості за рахунок ясності. Коли швидкість важлива, програміст Python може перенести критичні за часом функції до модулів розширення, написаних такими мовами, як C; або використовуйте PyPy, оперативний компілятор. Також доступний Cython, який перекладає сценарій Python на C і здійснює прямі виклики API на рівні C до інтерпретатора Python.

Розробники Python прагнуть, щоб ним було цікаво користуватися. Це відображено в його назві — данина британській комедійній групі Monty Python — і в інколи грайливих підходах до посібників і довідкових матеріалів, таких як приклади, які посилаються на спам і яйця (посилання на скетч Monty Python). стандартних foo і bar.

Поширеним неологізмом у спільноті Python є pythonic, який має широкий діапазон значень, пов'язаних зі стилем програми. «Pythonic» код може добре використовувати ідіоми Python, бути природним або демонструвати вільне володіння мовою, або відповідати мінімалістичній філософії Python і наголосу на

читабельності. Код, який важко зрозуміти або читається як груба транскрипція з іншої мови програмування, називається непітонічним. Python використовує качиний тип і має типізовані об'єкти, але нетипові імена змінних. Обмеження типу не перевіряються під час компіляції; скоріше, операції з об'єктом можуть завершитися помилкою, що означає, що він невідповідного типу. Незважаючи на динамічну типізацію, Python строго типізований, забороняючи операції, які не є чітко визначеними (наприклад, додавання числа до рядка), а не мовчки намагаючись зрозуміти їх сенс.

Python дозволяє програмістам визначати власні типи за допомогою класів, які найчастіше використовуються для об'єктно-орієнтованого програмування. Нові екземпляри класів створюються шляхом виклику класу (наприклад, `SpamClass()` або `EggsClass()`), і класи є екземплярами типу метакласу (сам по собі є екземпляром), що дозволяє метапрограмування та рефлексію.

До версії 3.0 у Python було два типи класів (обидва використовували однаковий синтаксис): старий та новий стиль, поточні версії Python підтримують лише новий стиль семантики.

Довгостроковий план полягає в підтримці поступового друку. Синтаксис Python дозволяє вказувати статичні типи, але вони не перевіряються в реалізації за замовчуванням, CPython. Експериментальна додаткова статична перевірка типів, `myru`, підтримує перевірку типів під час компіляції.

Велика стандартна бібліотека Python надає інструменти, які підходять для вирішення багатьох завдань, і її зазвичай називають однією з її найбільших переваг. Для додатків, що працюють в Інтернеті, підтримується багато стандартних форматів і протоколів, таких як MIME і HTTP. Він включає модулі для створення графічних інтерфейсів користувача, підключення до реляційних баз даних, генерації псевдовипадкових чисел, арифметики з десятковими дробами довільної точності, маніпулювання регулярними виразами та модульного тестування.

Деякі частини стандартної бібліотеки охоплюються специфікаціями, наприклад, реалізація інтерфейсу шлюзу веб-сервера (WSGI) `wsgiref` відповідає

PEP 333, але більшість визначено їх кодом, внутрішньою документацією та наборами тестів. Однак, оскільки більша частина стандартної бібліотеки є крос-платформним кодом Python, лише деякі модулі потребують змін або переписування для варіантних реалізацій.

Більшість реалізацій Python (включно з CPython) включають цикл читання–оцінка–друк (REPL), що дозволяє їм функціонувати як інтерпретатор командного рядка, для якого користувачі вводять оператори послідовно та отримують результати негайно.

Python також поставляється з інтегрованим середовищем розробки (IDE) під назвою IDLE, яке більше орієнтоване на початківців.

Інші оболонки, включаючи IDLE та IPython, додають додаткові можливості, такі як покращене автозавершення, збереження стану сеансу та підсвічування синтаксису.

Окрім стандартних настільних інтегрованих середовищ розробки, існують IDE на основі веб-браузера, включаючи SageMath, для розробки наукових і математичних програм; PythonAnywhere, IDE на основі браузера та середовище хостингу; і Canopy IDE, комерційна IDE, яка наголошує на наукових обчисленнях.

Розробка Python здійснюється в основному за допомогою процесу Python Enhancement Proposal (PEP), основного механізму для пропонування основних нових функцій, збору внесків спільноти щодо проблем і документування проектних рішень Python. Стиль кодування Python розглядається в PEP 8. Спільнота Python і керівна рада переглядають і коментують видатні PEP.

Удосконалення мови відповідає розробці еталонної реалізації CPython. Список розсилки python-dev є основним форумом для розробки мови. Конкретні питання спочатку обговорювалися в інструменті відстеження помилок Roundup, розміщеному фондом. У 2022 році всі проблеми та обговорення перенесено на GitHub. Спочатку розробка велася на власному сховищі вихідного коду під керуванням Mercurial, доки Python не перемістився на GitHub у січні 2017 року.

Публічні випуски CPython бувають трьох типів, які розрізняються за тим, яка частина номера версії збільшується:

Зворотно несумісні версії, де очікується, що код зламаний і його потрібно перенести вручну. Перша частина номера версії збільшується. Такі випуски трапляються нечасто — версія 3.0 була випущена через 8 років після 2.0. За словами Гвідо ван Россума, версія 4.0 навряд чи коли-небудь з'явиться.

Основні або «функціональні» випуски в основному сумісні з попередньою версією, але містять нові функції. Друга частина номера версії збільшується. Починаючи з Python 3.9, очікується, що ці випуски випускатимуться щорічно. Кожна основна версія підтримується виправленнями помилок протягом кількох років після випуску.

Випуски з виправленнями, які не вводять нових функцій, з'являються приблизно кожні 3 місяці та створюються, коли достатню кількість помилок було виправлено на початковому етапі з моменту останнього випуску. У цих випусках також виправлено вразливості системи безпеки. Третя й остання частина номера версії збільшується.

Багато альфа-, бета-версій і релізів-кандидатів також випускаються як попередній перегляд і для тестування перед остаточними випусками. Хоча для кожного випуску існує приблизний графік, вони часто затримуються, якщо код не готовий. Команда розробників Python відстежує стан коду, запускаючи великий набір модульних тестів під час розробки.

Основна наукова конференція з Python — PyCon. Існують також спеціальні наставницькі програми Python, наприклад PyLadies. Python може служити мовою сценаріїв для веб-додатків, наприклад, через `mod_wsgi` для веб-сервера Apache. Завдяки інтерфейсу шлюзу веб-сервера стандартний API розвинувся для полегшення роботи цих програм. Такі веб-фреймворки, як Django, Pylons, Pyramid, TurboGears, web2py, Tornado, Flask, Bottle і Zope, підтримують розробників у розробці та обслуговуванні складних програм. Pyjs і IronPython можна використовувати для розробки клієнтських програм на базі Ajax. SQLAlchemy можна використовувати як засіб відображення даних у реляційну базу даних.

Twisted — це платформа для програмування зв'язку між комп'ютерами, яка використовується (наприклад) Dropbox.

Такі бібліотеки, як NumPy, SciPy і Matplotlib, дозволяють ефективно використовувати Python у наукових обчисленнях, а спеціалізовані бібліотеки, такі як Biopython і Astropy, забезпечують функціональні можливості для певної області. SageMath — це система комп'ютерної алгебри з інтерфейсом ноутбука, програмованим на Python: її бібліотека охоплює багато аспектів математики, включаючи алгебру, комбінаторику, числову математику, теорію чисел і числення. OpenCV має прив'язки Python із багатим набором функцій для комп'ютерного зору та обробки зображень.

Python зазвичай використовується в проектах штучного інтелекту та машинного навчання за допомогою таких бібліотек, як TensorFlow, Keras, Pytorch і scikit-learn. Як мова сценаріїв із модульною архітектурою, простим синтаксисом та інструментами обробки форматowanego тексту, Python часто використовується для обробки природної мови.

Python також можна використовувати для створення ігор з такими бібліотеками, як Pygame, які можуть створювати 2D-ігри.

Python був успішно вбудований у багато програмних продуктів як мова сценаріїв, у тому числі в програмне забезпечення методу кінцевих елементів, таке як Abaqus, тривимірні параметричні моделі, такі як FreeCAD, пакети тривимірної анімації, такі як 3ds Max, Blender, Cinema 4D, Lightwave, Houdini, Maya, modo, MotionBuilder, Softimage, компонувальник візуальних ефектів Nuke, програми 2D-зображень, такі як GIMP, Inkscape, Scribus і Paint Shop Pro, і програми для нотного запису, як-от scorewriter і capella. GNU Debugger використовує Python як гарний принтер для відображення складних структур, таких як контейнери C++. Esri рекламує Python як найкращий вибір для написання сценаріїв у ArcGIS. Він також використовувався в кількох відеоіграх і був прийнятий як перша з трьох доступних мов програмування в Google App Engine, дві інші — Java і Go.

Багато операційних систем містять Python як стандартний компонент. Він постачається з більшістю дистрибутивів Linux, AmigaOS 4 (використовує Python

2.7), FreeBSD (як пакет), NetBSD і OpenBSD (як пакет) і може використовуватися з командного рядка (термінал). Багато дистрибутивів Linux використовують інсталюатори, написані на Python: Ubuntu використовує інсталюатор Ubiquity, тоді як Red Hat Linux і Fedora Linux використовують інсталюатор Anaconda. Gentoo Linux використовує Python у своїй системі керування пакетами Portage.

Python широко використовується в індустрії інформаційної безпеки, включаючи розробку експлоїтів.

Більшість програмного забезпечення Sugar для One Laptop per Child XO, розробленого в Sugar Labs з 2008 року, написано на Python. У проекті одноплатного комп'ютера Raspberry Pi використано Python як основну мову програмування.

LibreOffice містить Python і має намір замінити Java на Python. Його Python Scripting Provider є основною функцією, починаючи з версії 4.0 від 7 лютого 2013 року.

У різних галузях існує велика різноманітність варіантів використання Python. Звичайно, перше, що спадає на думку, коли ми думаємо про найпоширеніші способи використання Python, це для створення веб-, мобільних і настільних програм, а також для їх тестування. Але Python — це мова, яка служить багатьом цілям. Загалом, Python ідеально підходить для таких сфер використання:

1. Розробка веб-додатків
2. Наука про дані
3. Сценарії
4. Програмування бази даних
5. Швидке створення прототипів

Python підходить для всіх форм програмування, що сприяє швидкому зростанню бази користувачів. Скрипти міжплатформної оболонки, швидка автоматизація, проста веб-розробка, аналіз і візуалізація даних, штучний інтелект і машинне навчання – це деякі приклади.

Часто фахівці використовують Python для кращого виконання різноманітних завдань у різних дисциплінах. Кращої продуктивності, серед іншого, можна досягти за допомогою автоматизації. Фінанси, страхування та маркетинг є

основними сферами, у яких люди стикаються з необхідністю виконувати повторювані та нудні завдання: переглядати, копіювати, перейменовувати та завантажувати файли на сервер, завантажувати веб-сайти чи аналізувати дані. Натомість програміст може написати сценарій на Python і автоматизувати все це.

Крім того, не обов'язково бути розробником програмного забезпечення, щоб використовувати Python. Мова дозволяє полегшити аналіз і візуалізацію даних. Він має багату екосистему, що включає ефективні бібліотеки для обробки даних і, отже, допомагає спеціалістам із обробки даних у виконанні складних числових обчислювальних операцій.

Не дарма найбільші компанії світу використовують Python. Pixar використовує його для створення фільмів, Google для сканування сторінок, Netflix для доставки вмісту та Spotify для рекомендації пісень. Ця мова має багато переваг:

1. Простота. Простий і зрозумілий синтаксис Python спонукає початківців вивчати цю мову сценаріїв. З певної точки зору може здатися природним і визначеним, що Python може перетворитися на лінгва франка програмування, показуючи, що всі інші його противники застаріли. Його код легко зрозуміти, поширювати та підтримувати. Немає багатослівності, мова легко вивчається.

6. Потужний інструментарій. За своєю суттю програми на Python є текстовими файлами, що містять інструкції для інтерпретатора та написані в текстовому редакторі або IDE. IDE є повнофункціональними та пропонують такі вбудовані інструменти, як перевірка синтаксису, налагоджувачі та браузері коду, текстові редактори зазвичай не включають функції IDE, але їх можна налаштувати. Python також має величезний набір сторонніх пакетів, бібліотек і фреймворків, які полегшують процес розробки. Таким чином, ці можливості оптимізації роблять Python чудовим для великомасштабних проєктів.

7. Швидкість розвитку. Тут ми маємо на увазі швидкість бізнесу та показник часу виходу на ринок. Python — це динамічна мова сценаріїв, тому вона не призначена для написання програм з нуля, а в першу чергу призначена для підключення компонентів. Компоненти призначені для багаторазового використання, а інтерфейси між компонентами та сценаріями чітко визначені.

Усе це прискорює розробку програмного забезпечення завдяки Python, що робить мову надзвичайно лаконічною та продуктивною.

8. Гнучкість. Хоча Python робить наголос на простоті та читабельності коду, а не на гнучкості, у цій мові це все одно є. Python можна використовувати в різних проектах. Це дозволяє розробникам вибирати між об'єктно-орієнтованим і процедурним режимами програмування. Python також гнучкий у типі даних. Їх 5: число, рядок, список, кортеж і словник, і кожен тип підданих відповідає одному з цих кореневих типів. У результаті пошуковий аналіз даних стає легше проводити завдяки гнучкості Python.

9. Портативність. Python створено для переносимості. Його програми підтримуються на будь-якій сучасній комп'ютерній ОС. Завдяки високорівневому характеру мови сценарій Python інтерпретується, тому його можна написати для подальшої інтерпретації однаково добре в Linux, Windows, Mac OS і UNIX, не вимагаючи коригувань. Програми на Python також дозволяють реалізувати портативні GUI.

10. Сильна громада. Python має базу користувачів, що швидко зростає, і фактично вказує на те, що таке сильна спільнота. Є тисячі учасників потужного інструментарію Python — Pythonists. В онлайн-сховище вже завантажено майже 200 000 програмних пакетів, створених на замовлення. Усе це означає, що велика підтримуюча спільнота є як причиною, так і наслідком попиту на мову.

Усі мови програмування мають свої недоліки. Незважаючи на всі переваги, які пропонує Python як мова програмування, є недоліки, якими слід скористатися:

1. Швидкість як інтерпретована мова. Хороша новина полягає в тому, що цей недолік можна виправити з появою PyPy, яка обіцяє приріст продуктивності.

7. Динамізм Python запобігає виявленню семантичних помилок заздалегідь. Але такі інструменти, як PyChecker, можуть перевіряти наявність помилок, що робив би компілятор таких мов, як C або Java.

8. Потоковість є менш продуктивною в Python, ніж в інших мовах. Багатопотоковість може стати можливою з Jython, але незмінність не надто важлива в Python, тому однопотоковий паралелізм працює добре.

9. Залежність від сторонніх бібліотек і фреймворків. Існує чимало широко використовуваних ресурсів сторонніх розробників, які по суті не є Pythonic. Це фактично суперечить девізу Python.

Python був створений Гвідо ван Россумом і вперше випущений 20 лютого 1991 року. Хоча ви знаєте пітона як велику змію, назва мови програмування Python походить від старого комедійного серіалу BBC під назвою «Летючий цирк Монті Пайтона».

Однією з дивовижних особливостей Python є той факт, що це фактично робота однієї людини. Зазвичай нові мови програмування розробляють і публікують великі компанії, в яких працює багато професіоналів, і через правила авторського права дуже важко назвати когось із людей, які беруть участь у проекті. Python є винятком. Звичайно, Гвідо ван Россум не розробляв і розвивав усі компоненти Python самостійно. Швидкість, з якою Python поширився по всьому світу, є результатом безперервної роботи тисяч (дуже часто анонімних) програмістів, тестувальників, користувачів (багато з них не ІТ-спеціалісти) та ентузіастів, але слід сказати, що саме Перша ідея (зерно, з якого пророс Python) прийшла в голову одній – Гвідо.

Python підтримується Python Software Foundation, некомерційною членською організацією та спільнотою, яка займається розробкою, вдосконаленням, розширенням і популяризацією мови Python та її середовища. У 1999 році Гвідо ван Россум визначив свої цілі щодо Python:

1. легка та інтуїтивно зрозуміла мова, така ж потужна, як і в основних конкурентів;

10. відкритий код, тому будь-хто може зробити свій внесок у його розвиток;

код, зрозумілий як проста англійська;

підходить для повсякденних завдань, що забезпечує короткий час розробки.

Через 20+ років стає зрозуміло, що всі ці наміри здійснилися. Деякі джерела кажуть, що Python є першою за популярністю мовою програмування у світі, а інші стверджують, що це третя мова. У будь-якому випадку, він займає перше місце в першій десятці індексів популярності мов програмування TIOBE Programming Community і PYPL. (Лютий 2022) Python не є молодою мовою.

Python є чудовим вибором для:

1. Веб- та Інтернет-розробка (наприклад, фреймворки Django та Pyramid, мікрофреймворки Flask та Bottle)
11. Наукові та числові обчислення (наприклад, SciPy – колекція пакетів для цілей математики, науки та інженерії; Ipython – інтерактивна оболонка, яка підтримує редагування та запис робочих сеансів)
12. Освіта (це чудова мова для навчання програмуванню)
13. Графічні інтерфейси робочого столу (наприклад, wxWidgets, Kivy, Qt)
14. Розробка програмного забезпечення (контроль збирання, управління та тестування – Scons, Buildbot, Apache Gump, Roundup, Trac)
15. Бізнес-додатки (ERP та системи електронної комерції – Odoo, Tryton)
16. Ігри (наприклад, серія Battlefield, Sid Meier's Civilization IV...), веб-сайти та служби (наприклад, Dropbox, UBER, Pinterest, BuzzFeed...)

Дві з найбільш використовуваних версій мають Python 2.x і 3.x. Між ними є велика конкуренція, і обидва, здається, мають досить багато різних баз шанувальників. Для різних цілей, таких як розробка, створення сценаріїв, генерація та тестування програмного забезпечення, ця мова використовується. Завдяки його елегантності та простоті такі провідні технологічні організації, як Dropbox, Google, Quora, Mozilla, Hewlett-Packard, Qualcomm, IBM і Cisco, впровадили Python.

Python пройшов довгий шлях, щоб стати найпопулярнішою мовою програмування у світі.

3 РЕАЛІЗАЦІЯ ТА ПОРІВНЯННЯ МЕТОДІВ РЕКОМЕНДАЦІЙ

3.1 Дослідження методів

Першим кроком потрібно вибрати набір даних, на базі якого буде проводитися експеримент.

Для порівняння методів буде використовуватися сирий набір даних. Використовуючи набір даних MovieLens 20M, був розроблений підхід для рекомендацій від елемента до елемента (від фільму до фільму), яка рекомендує фільми, схожі на певний вхідний фільм. Отже, опис даних:

1. 20 мільйонів оцінок і 465 000 тегів застосували до 27 000 фільмів 138 000 користувачів.
2. Включає дані геному тегів із 12 мільйонами оцінок релевантності в 1100 тегах.
3. Випущено 4/2015; оновлено 10/2016 для оновлення links.csv і додавання даних геному тегів.

Якщо розглядати дані детальніше, то цей набір даних (ml-20m) описує активність користувачів у MovieLens (служби рекомендацій фільмів). Він містить 20000263 оцінки та 465564 додатки тегів для 27278 фільмів. Ці дані були створені 138493 користувачами в період з 9 січня 1995 року по 31 березня 2015 року. Цей набір даних було створено 17 жовтня 2016 року.

Користувачі були обрані випадковим чином для включення. Усі вибрані користувачі оцінили принаймні 20 фільмів. Демографічна інформація не включена. Кожен користувач представлений ідентифікатором, а інша інформація не надається.

Дані містяться в шести файлах:

1. genome-scores.csv,

4. genome-tags.csv,
5. links.csv,
6. movies.csv,
7. ratings.csv
8. tags.csv

Другим кроком, визначаємо який алгоритм (в випадку цього експеримента використовується косинусна подібність) буде використовуватися.

Щоб створити гібридну модель, ми об'єднали результати автокодувальника, який вивчає вбудовування фільмів на основі вмісту з даних тегів, і нейронної мережі глибокого вбудовування об'єктів, яка вивчає вбудовування фільмів на основі спільного використання з даних оцінок.

Спільні рекомендації покладаються на дані, створені користувачами під час взаємодії з елементами. Приклади цього:

1. Користувачі оцінюють фільми за шкалою від 1 до 5
2. Користувачі, які купують або навіть переглядають товари на сайті роздрібної торгівлі в Інтернеті
3. Користувачі, які реагують «великий палець вгору» або «великий палець вниз» на пісні в онлайн-сервісі потокового передавання музики
4. Гортання ліворуч або праворуч на сайті знайомств

У контексті рекомендації фільмів спільні фільтри виявляють тенденції того, як схожі користувачі оцінюють фільми на основі профілів оцінювання. Дані про рейтинги можна розкласти або обробити іншим чином за допомогою різноманітних методів, щоб остаточно знайти вбудовані користувачі та фільми в спільному прихованому просторі. Вбудовані фільми, які описують їхнє розташування в

прихованому просторі, потім можна використовувати для створення рекомендацій від фільму до фільму.

Однією з переваг спільних даних є те, що вони завжди «самогенеруються» — користувачі створюють дані для вас природним чином під час взаємодії з елементами. Це може бути цінним джерелом даних, особливо у випадках, коли високоякісні функції предмета недоступні або їх важко отримати. Ще одна перевага спільних фільтрів полягає в тому, що вони допомагають користувачам знаходити нові елементи, які знаходяться за межами підпростору, визначеного їхнім історичним профілем.

Однак у спільних фільтрів є деякі недоліки, наприклад добре відома проблема холодного запуску. Спільним фільтрам також важко точно рекомендувати нові або нішеві елементи, оскільки ці елементи зазвичай не мають достатньо даних про взаємодію між користувачем.

Рекомендатори вмісту покладаються на характеристики елементів, щоб давати рекомендації. Приклади цього:

1. Створені користувачем теги на фільмах
2. Колір предмета
3. Текстовий опис або відгук користувача про товар

Фільтри вмісту, як правило, більш надійні проти упередженості популярності та проблеми холодного запуску. Вони можуть легко порекомендувати нові чи нові речі на основі нішевих смаків. Однак у рекомендації елементів фільтри вмісту можуть рекомендувати лише елементи з функціями, подібними до оригінального елемента. Це обмежує обсяг рекомендацій, а також може призвести до появи елементів із низькими оцінками.

У контексті рекомендувача фільму до фільму спільний фільтр відповідає на запитання: «Які фільми мають подібний профіль оцінки користувача?», а фільтр вмісту відповідає на запитання: «Які фільми мають подібні характеристики?». Створивши гібридний рекомендувач, ми спробували створити систему, яка рекомендує фільми, які інші користувачі оцінювали подібним чином, водночас надаючи рекомендації по темі на основі особливостей цього фільму.

Третім кроком розглянемо детальніше косинусну подібність.

Використовується косинуса подібність, щоб кількісно визначити подібність між фільмами. Подібність косинусів коливається від -1 до 1 і обчислюється як скалярний добуток двох векторів, поділений на їхні величини.

$$\text{sim}(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad (3.1)$$

Вектори вбудовування фільму, що вказують в одному напрямку, отримують високі бали косинусної подібності. Ідея полягає в тому, що напрямки через прихований концептуальний простір відображають суть фільмів. Простий приклад, який допоможе уявити це: якщо «наукова фантастика» та «романтика» є вимірами в латентному просторі, то фільми з подібним співвідношенням наукової фантастики до романтичності показуватимуть у тому самому напрямку й, отже, отримують високу оцінку. оцінки косинусної подібності.

Четвертим кроком, приведена частину коду.

Цей код створює матрицю косинусної подібності з вбудованих фільмів і виводить перші n найбільш схожих фільмів для певного вхідного фільму.

```
from sklearn.metrics.pairwise import cosine_similarity
import pandas as pd

class SimilarityPredictions(object):
    '''This class calculates a similarity matrix from latent
embeddings.
```

Input: embeddings - a pandas dataframe of items and latent dimensions.

```
'''  
  
def __init__(self, embeddings):  
    self.embeddings = embeddings  
    self.ids = embeddings.index.tolist()  
  
    def calculate_cosine_similarity_matrix(self):  
        '''Calculates a cosine similarity matrix from the  
embeddings'''  
        similarity_matrix = pd.DataFrame(cosine_similarity(  
            X=self.embeddings,  
            index=self.ids)  
            similarity_matrix.columns = self.ids  
            return similarity_matrix  
  
    def predict_similar_items(self, seed_item, n):  
        '''Use the similarity_matrix to return n most similar  
items.'''  
        similar_items =  
pd.DataFrame(self.similarity_matrix.loc[seed_item])  
        similar_items.columns = ["similarity_score"]  
        similar_items =  
similar_items.sort_values('similarity_score', ascending=False)  
        similar_items = similar_items.head(n)  
        similar_items.reset_index(inplace=True)  
        similar_items = similar_items.rename(index=str,  
columns={"index": "item_id"})  
        return similar_items.to_dict()
```

Перший експеримент проводиться на основі методу з рекомендаціями за основі контенту.

Дані MovieLens містять близько 500 тисяч створених користувачами тегів фільмів. Відповідно до README MovieLens: «Кожен тег зазвичай є одним словом або короткою фразою. Значення, значення та призначення конкретного тегу визначає кожен користувач».

Ці дані групуються за фільмами, а теги об'єднуються, щоб створити корпус документів. У цьому корпусі документи є комбінацією всіх тегів для певного фільму. Нижче наведено фрагмент документа тегів для «Володаря кілець: Братство

персня». Як бачите, такі слова/фрази, як «пригоди», «фентезі» та «за мотивами книги», з'являються часто. Ці дані також містять імена акторів та автора книги.

“adventure characters epic fantasy world fighting photography Action adventure atmospheric based on a book based on book beautifully filmed ensemble cast fantasy fantasy world high fantasy imdb top 250 magic music nature nothing at all Oscar (Best Cinematography) Oscar (Best Effects - Visual Effects) scenic stylized Tolkien wizards adventure atmospheric ensemble cast fantasy fantasy world magic stylized wizards Watched adapted from:book author:J. R. R. Tolkein based on book epic fantasy middle earth faithful to book fantasy good versus evil high fantasy joseph campbell's study of mythology influenced magic atmospheric boring high fantasy Action adventure atmospheric based on a book beautifully filmed fantasy high fantasy magic music mythology romance stylized time travel Viggo Mortensen wizards Peter Jackson Peter Jackson music must see Tolkien high fantasy Myth Tolkien wizards Ian McKellen bast background universe”

Потім ми перетворили документи тегів у представлення частоти термінів — зворотної частоти документа (TF-IDF). TF-IDF перетворює неструктуровані текстові дані на числові функції, які легше обробляти алгоритмами машинного навчання. Грубо кажучи, частота кожного терміна в документі масштабується кількістю документів, що містять цей термін. У результаті слова, які відрізняють фільми (наприклад, «інопланетянин», «фентезі»), матимуть більшу вагу, ніж слова, які використовуються для опису всіх фільмів (наприклад, «фільм», «акторський склад»).

Векторизатор TF-IDF у Sci-kit learn дозволяє вибрати діапазон nграмів для розгляду. Для цього проекту ми виявили, що уніграми достатньо, але при цьому кількість наших функцій залишається керованою[12].

У просторі TF-IDF кожен вимір відображає, наскільки важливе певне слово для фільму. Це представлення не є ідеальним, оскільки кодування однієї концепції (наприклад, інопланетянин/інопланетянин) є фрагментованим і розкиданим у кількох вимірах. Ми хотіли б стиснути дані TF-IDF у нижчий вимірний простір, де концепції об'єднані в спільні виміри. У стисненому просторі сподіваються, що кожен вимір представлятиме складні та надійні концепції.

Для виконання стиснення ми вирішили використовувати автокодер. Автокодери — це нейронні мережі, де виходи ідентичні входам. У нашій архітектурі (показаний на малюнку нижче) високовимірний вхідний сигнал TF-IDF поступово стискається в 100-вимірний центральний прихований шар. Ця перша половина мережі є «кодером». Друга половина мережі, «декодер», намагається реконструювати вихідний вхід. Встановлюючи функцію втрати середнього квадрата помилки та виконуючи зворотне поширення, мережа (зокрема кодер) навчається функції, яка відображає дані в нижньому вимірному просторі таким чином, що більшість інформації зберігається.

Використовуючи систему рекомендацій фільмів на основі контенту, отримане автокодером, нижче наведено 20 найбільш схожих фільмів на «Володар пернів: Братство персня» за косинусною подібністю.

Таблиця 3.1 – Рекомендації на основі контенту

	Рекомендація	Жанр	ratings_count	avg_ratings	similarity_score
1	Lord of the Rings: The Two Towers, The (2002)	Adventure Fantasy	33947	4.1075	0.98616
2	Chronicles of Narnia: The Lion, the Witch and the Wardrobe, The (2005)	Adventure Children Fantasy	7112	3.47883	0.982

3	Dark Crystal, The (1982)	Adventure Fantasy	7980	3.5909	0.9756
4	Lord of the Rings: The Return of the King, The (2003)	Action Adventure Drama Fantasy	31577	4.14238	0.9683
5	Harry Potter and the Sorcerer's Stone	Adventure Children Fantasy	17239	3.6143	0.96784
6	Tales from Earthsea	Adventure Animation Fantasy	131	3.16793	0.9613
7	Almighty Thor	Adventure Fantasy	19	3.2368	0.959
8	Chronicles of Narnia: Prince Caspian	Adventure Children Fantasy	1551	3.3239	0.957
9	Magic Voyage of Sindbad	Adventure Fantasy	3	3.3333	0.9569
10	Harry Potter and the Deathly Hallows: Part 2	Action Adventure Drama	3983	3.9506	0.9565
11	Conan the Destroyer	Action Adventure Fantasy	1791	2.8283	0.954
12	Spiderwick Chronicles	Adventure Children Drama	683	3.330	0.8740
13	Dragonslayer	Action Adventure	921	3.3121	0.8532
14	Harry Potter and the Chamber of Secrets	Fantasy	14469	3.579	0.852

1 5	Kiki's Delivery Service	Fantasy	2	2.5	0.852
1 6	Day The Earth Froze	Adventure Fantasy	2	3.4516	0.851
1 7	Willow	Adventure Fantasy	9217	3.75	0.8515
1 8	Thief of Bagdad	Adventure Fantasy	167	3.7455	0.85110
1 9	Conan the Barbarian	Adventure Fantasy	6713	3.216	0.850

В таблиці можна побачити рекомендований список фільмів з характеристиками, які приймалися до уваги під час аналізу та створення рекомендації. Цей список складається з 19 фільмів.

Другий експеримент проводиться с спільним типом рекомендацій.

Повністю підключена нейронна мережа використовується для пошуку фільмів і вставок користувача. У цій архітектурі матриця розміру вбудовування користувача ($n_users, n_factors$) і матриця розміру вбудовування фільму ($n_movies, n_factors$) ініціалізуються випадковим чином і згодом вивчаються за допомогою градієнтного спуску. Кожна точка даних навчання є індексом користувача, індексом фільму та рейтингом (за шкалою від 1 до 5). Індеси користувача та фільму використовуються для пошуку векторів вбудовування (які є рядками матриць вбудовування). Потім ці вектори об'єднуються та передаються як вхідні дані до нейронної мережі.

Оскільки вихідний сигнал мережі обмежений значенням 1–5, на вихідному рівні використовується масштабована сигмоїдна функція активації. Передача лінійної активації через сигмовид надає моделі трохи більше гнучкості, а отже, полегшує її навчання. Наприклад, якщо модель намагається вивести 5, замість того, щоб примушувати її виводити значення, дуже близьке до 5 із лінійного обчислення, використання сигмоїда дозволяє їй виводити будь-яке високе значення (оскільки всі вхідні дані ~ 6 або більше відображаються на $\sim 1,0$ за сигмоїдною функцією). Альтернативою було б зупинитися на сигмоїді та дозволити моделі навчитися виводити правильний рейтинг безпосередньо з лінійного виводу.

Альтернативою було б зупинитися на сигмоїді та дозволити моделі навчитися виводити правильний рейтинг безпосередньо з лінійного виводу.

Альтернативою було б зупинитися на сигмоїді та дозволити моделі навчитися виводити правильний рейтинг безпосередньо з лінійного виводу.

Після того, як мережа буде задовільно навчена, ви матимете надійний фільм і користувачі вбудовування, готові до використання для різноманітних практичних завдань. Використовуючи спільне вбудовування фільмів, навчене цією нейронною мережею, нижче наведено 20 найбільш схожих фільмів на «Володар пернів: Братство персня» за косинусною подібністю.

Таблиця 3.2 – Рекомендації на основі спільних даних

	Рекомендація	Жанр	ratings_count	avg_ratings	similarity_score
1	Star Wars: Episode IV	Action Adventure	54502	4.190	0.9784
2	Raiders of the Lost Ark	Action Adventure	43295	4.219	0.9765

3	Lord of the Rings: The Return of the King	Action Adventure	31577	4.142	0.9747
4	Star Wars: Episode V	Drama War	45313	4.188	0.9745
5	Schindler's List	Action Crime	50054	4.310	0.9720
6	Lord of the Rings: The Two Towers	Action Adventure	33947	4.1075	0.969
7	Shawshank Redemption	Action Crime	63366	4.446	0.9681
8	Saving Private Ryan	Drama War	37110	4.064	0.9655
9	Good Will Hunting	Action Adventure	28324	4.032	0.9639
10	Matrix	Action Adventure	51334	4.187	0.9635
11	Inception	Action Crime	14023	4.156	0.9625
12	Dark Knight	Drama War	20438	4.220	0.961
13	Princess Bride	Comedy Drama	32586	4.1767	0.96141
14	Silence of the Lambs	Action Adventure	63299	4.177	0.9611
15	Batman Begins	Comedy Drama	18686	3.9707	0.859

16	Slumdog Millionaire	Drama War	9208	3.950	0.8581
17	Toy Story	Comedy Drama	49695	3.92123	0.857
18	Star Wars: Episode VI	Action Crime	46839	4.004	0.856
19	Life Is Beautiful	Action Adventure	18156	4.175	0.855

Третій експеримент є пошуком фільмів на основі гібридної моделі.

Як видно вище, спільні рекомендації щодо «Володаря пернів» здаються насамперед популярними та високо оціненими фільмами-блокбастерами з яскравими бойовиками та пригодницькими темами. Схоже, на рекомендації вмісту менше впливає популярність і вони можуть містити деякі приховані перлини жанру фентезі.

Щоб поєднати результати вмісту та спільних моделей, треба усереднити косинусну подібність. Поєднуючи спільні результати та результати, засновані на вмісті - можемо давати рекомендації, які, спираються на сильні сторони обох методів. Нижче наведено зведені рекомендації для «Володаря кілець».

Таблиця 3.3 – Рекомендації на основі гібридної системи

	Рекомендація	Жанр	ratings _count	avg_rat ings	similarity_ score
1	Lord of the Rings: The Two Towers	Adventure Fantasy	33947	4.1075	0.97
2	Lord of the Rings: The Return of the King	Action Adventure	31577	4.142	0.9715

3	Star Wars: Episode V	Action Adventure	45313	4.188	0.947
4	Star Wars: Episode IV - A New Hope	Action Adventure	54502	4.1906	0.944
5	Princess Bride	Action Adventure	32586	4.176	0.93
6	Raiders of the Lost Ark	Action Adventure	43295	4.2190	0.934
7	Star Wars: Episode VI	Action Adventure	46839	4.0046	0.933
8	WALL·E	Adventure Animation	12176	4.038	0.9306
9	Batman Begins	Action Crime	18686	3.970	0.925
10	Harry Potter and the Deathly Hallows	Action Adventure	3983	3.950	0.924
11	Untouchables	Action Crime	15442	3.884	0.922
12	How to Train Your Dragon	Adventure Animation	4761	4.00	0.921
13	Guardians of the Galaxy	Action Adventure	2049	3.854	0.920
14	Batman: The Dark Knight Returns	Action Animation	373	3.882	0.92047
15	Up	Adventure	9264	4.53826	0.920
16	Dark Knight	Action Crime	20438	4.220	0.918
17	Matrix	Action Sci-Fi	51334	4.187	0.917
18	Natural, The	Drama	4964	3.80288	0.9146
19	Incredibles	Action Adventure	19890	3.908	0.9134

За спостереженнями, гібридний фільтр є ефективним вибором, щоб охоплювати прогалини між жанрами та постійно рекомендувати фільми з високими рейтингами. Також можна побачити, що спільний фільтр об'єднує результати двох попередніх методів (контент та спільний). Тобто кожен результат кожного окремого методу дає унікальний список фільмів. Більшість фільмів схожа, але кожен метод має деякі фільми, специфічні саме до алгоритму. Наприклад, метод контенту рекомендує декілька фільмів, які є схожими за контентом до первинного фільму, але ці фільми не потрапляють до результатів інших двох методів.

4 РЕЗУЛЬТАТИ МЕТОДІВ РЕКОМЕНДАЦІЙ

4.1 Порівняння кінцевих результатів методів

Для підведення результатів будуть використовуватися наступні параметри:

1. `similarity_score` - є коефіцієнтом подібності, який може приймати вигляд будь-якого дробного числа від 0 до 1
9. `avg_ratings` - є середньою оцінкою фільма, яку дали користувачі фільма після перегляду.

Перше порівняння буде оснований на коефіцієнт подібності.

Таблиця 4.1 – Порівняння методів на основі коефіцієнта подібності

Контент-метод	Спільний метод	Гібридний метод
0.98616	0.9784	0.97
0.982	0.9765	0.9715
0.9756	0.9747	0.947
0.9683	0.9745	0.944
0.96784	0.9720	0.93
0.9613	0.969	0.934
0.959	0.9681	0.933
0.957	0.9555	0.9306
0.9569	0.9439	0.925
0.9565	0.9435	0.924
0.954	0.9425	0.922

0.8740	0.941	0.921
0.8532	0.93141	0.920
0.852	0.9111	0.92047
0.852	0.859	0.920
0.851	0.8581	0.918
0.8515	0.857	0.917
0.85110	0.856	0.9146
0.850	0.855	0.9134

Середнє значення отримане для кожного методу:

Таблиця 4.2 – Порівняння методів на основі середніх значень коефіцієнту подібності

Контент-метод	Спільний метод	Гібридний метод
0.9189	0.9238	0.9302

Друге порівняння має на меті привести середню оцінку фільмів, і побачити який метод має найкращу оцінку рекомендованих фільмів:

Таблиця 4.3 – Порівняння методів на основі середньої оцінки фільмів

Контент-метод	Спільний метод	Гібридний метод
4.182	4.095	4.214

Для візуалізації порівняних методів та визначення переможця приведені графіки:

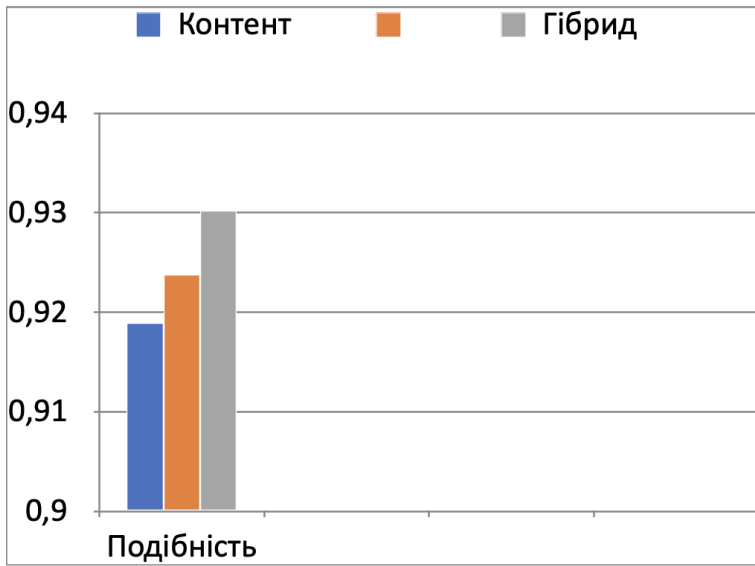


Рисунок 4.1 - Порівняння методів на основі коефіцієнта подібності

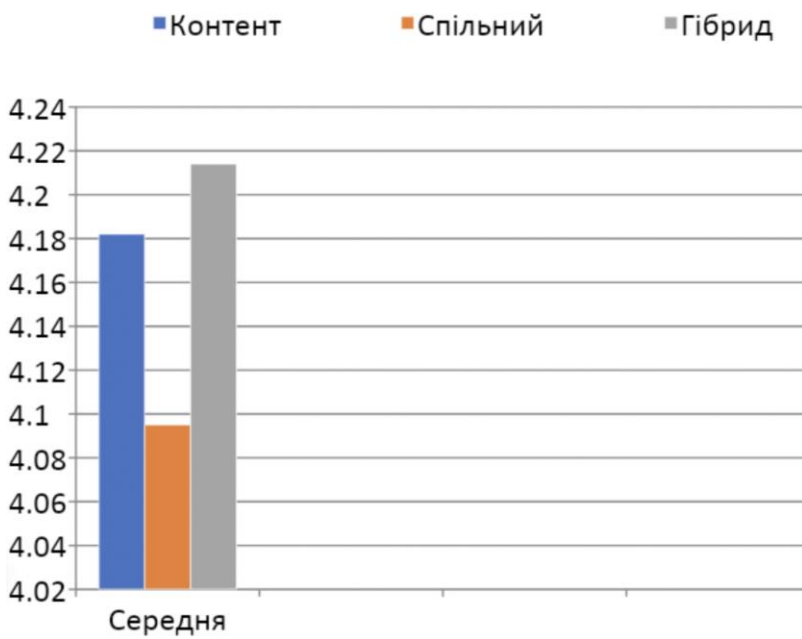


Рисунок 4.2 - Порівняння методів на основі середнього рейтингу

ВИСНОВКИ

Метою магістерської роботи стало дослідження та розробка методів формування рекомендацій для веб-порталу з інтернет комерції.

Як результат, було розроблене та дослідження формування рекомендацій для веб-порталу з інтернет комерції.

Були визначені та використані характеристики, на базі яких було проведений експеримент та які визначили найбільш ефективний метод рекомендацій.

Був проведений аналіз та порівняння методів рекомендацій, які містили в собі:

1. аналіз алгоритму рекомендацій;
2. їх недоліки;
3. переваги;
4. складності реалізації.

Результатом експерименту став вибір найефективнішого методу рекомендацій. Таким методом став гібридний метод формування рекомендацій, спираючись на такі характеристики як коефіцієнт подібності та рейтинг рекомендованої продукції.

Отже, для розробки та дослідження формування рекомендацій були виконані такі завдання:

1. Визначені існуючі методи рекомендацій
5. Обрані декілька методів рекомендацій для експерименту
6. Визначений набір даних про користувачів, який був використаний для експерименту
7. Реалізований кожен метод рекомендацій окремо на базі набору даних
8. Порівняні результати різних методів на основі обраних характеристик
9. Визначений найбільш ефективний метод рекомендацій, який став гібридний методре

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. A. M. Turing (1950) Computing Machinery and Intelligence. Mind 49: 433-460.
2. Artificial intelligence : a modern approach/ Stuart Russell, Peter Norvig. Includes bibliographical references and index. ISBN 0-13-103805-2 1. Artificial intelligence I. Norvig, Peter. II. Title.Q335.R86 1995
3. Lecture 1: Introduction and Scope [Electronic resource] - Mode of access: <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/lecture-videos/lecture-1-introduction-and-scope/>
4. Deep Blue [Electronic resource] - Mode of access: <https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/>
5. AlphaGo [Electronic resource] - Mode of access: <https://deepmind.com/research/case-studies/alphago-the-story-so-far>
6. 4 Types of Artificial Intelligence [Electronic resource] - Mode of access: <https://www.bmc.com/blogs/artificial-intelligence-types/>
7. AI, Deep Learning, and Machine Learning: A Primer [Electronic resource] - Mode of access: <https://a16z.com/2016/06/10/ai-deep-learning-machines/>
8. Executive Office of the President National Science and Technology Council Committee on Technology [Electronic resource] - Mode of access: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
9. Яблоков К. В. Исторические компьютерные игры как способ моделирования исторической информации // История и математика: Анализ и моделирование социально-исторических процессов / Ред. Малков С. Ю., Гринин Л. Е., Коротаяев А. В. М.: КомКнига/УРСС, 2007. С. 263—303
10. Ender's Game (1/10) Movie CLIP - The Mind Game (2013) HD [Electronic resource] - Mode of access: <https://www.youtube.com/watch?v=oDRFKZVZwcA>
11. Red Dead Redemption 2 [Electronic resource] - Mode of access: <https://www.rockstargames.com/ru/games/reddeadredemption2>

12. Dr Mike Cook [Electronic resource] - Mode of access: <http://eecs.qmul.ac.uk/profiles/cookmichael.html>
13. James Scott Nick Polson Artificial intelligence 2019p. c 273-280.
14. Woodcock, S. 2015. Game AI: The state of the industry, available online at <http://www.gamasutra.com/features/20001101/>.
15. G. N. Yannakakis. Preference Learning for Affective Modeling. In Proceedings of the Int. Conf. on Affective Computing and Intelligent Interaction, pages 126–131, Amsterdam, The Netherlands, September 2009. IEEE
16. A. J. Champanand. AI Game Development. New Riders Publishing, 2014.
17. G. N. Yannakakis and J. Togelius. Experience-Driven Procedural Content Generation. IEEE Transactions on Affective Computing, 2:147–161, 2011.
18. Z. Zeng, M. Pantic, G. Roisman, and T. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Trans. Pattern Analysis and Machine Intelligence, 31(1):39–58, 2019.
19. DeepMind [Electronic resource] - Mode of access: <https://ru.wikipedia.org/wiki/DeepMind>
20. Quake III Arena [Electronic resource] - Mode of access: https://ru.wikipedia.org/wiki/Quake_III_Arena
21. Markov decision process [Electronic resource] - Mode of access: https://en.wikipedia.org/wiki/Markov_decision_process
22. Multi-armed bandit [Electronic resource] - Mode of access: https://en.wikipedia.org/wiki/Multi-armed_bandit
23. The Upper Confidence Bound Algorithm [Electronic resource] - Mode of access: <https://banditalgs.com/2016/09/18/the-upper-confidence-bound-algorithm/>
24. Обучение с подкреплением [Электронный ресурс] – Режим доступа: http://neerc.ifmo.ru/wiki/index.php?title=mobileaction=toggle_view_desktop
25. Unity (игровой движок) [Электронный ресурс] – Режим доступа: <https://ru.wikipedia.org/wiki/Unity/>
26. About ML-Agents package [Electronic resource] - Mode of access: <https://docs.unity3d.com/Packages/com.unity.ml-agents@2.0/manual/index.html>