

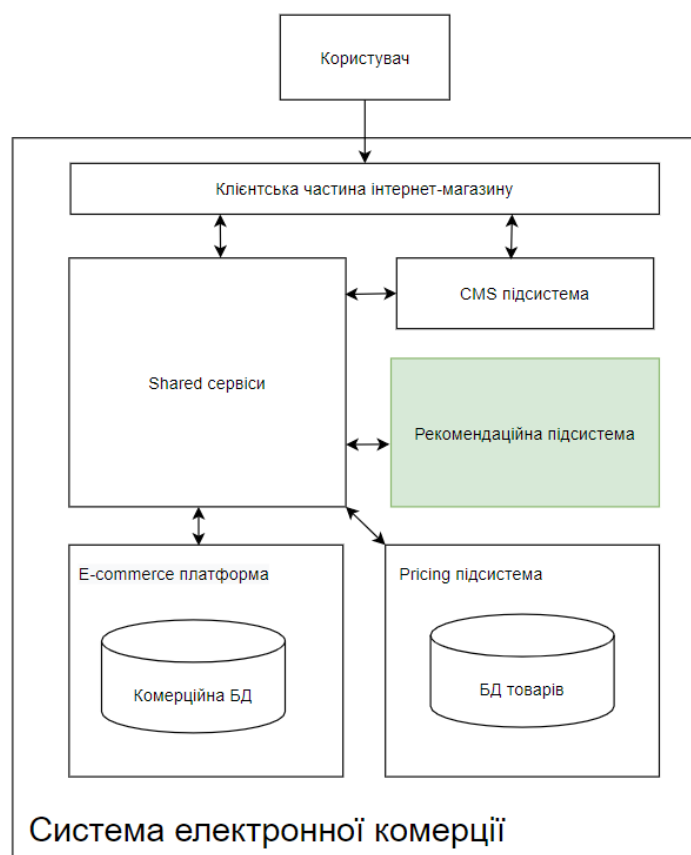
ДОДАТОК А
ГРАФІЧНИЙ МАТЕРІАЛ

Цілі та задачі досліджень

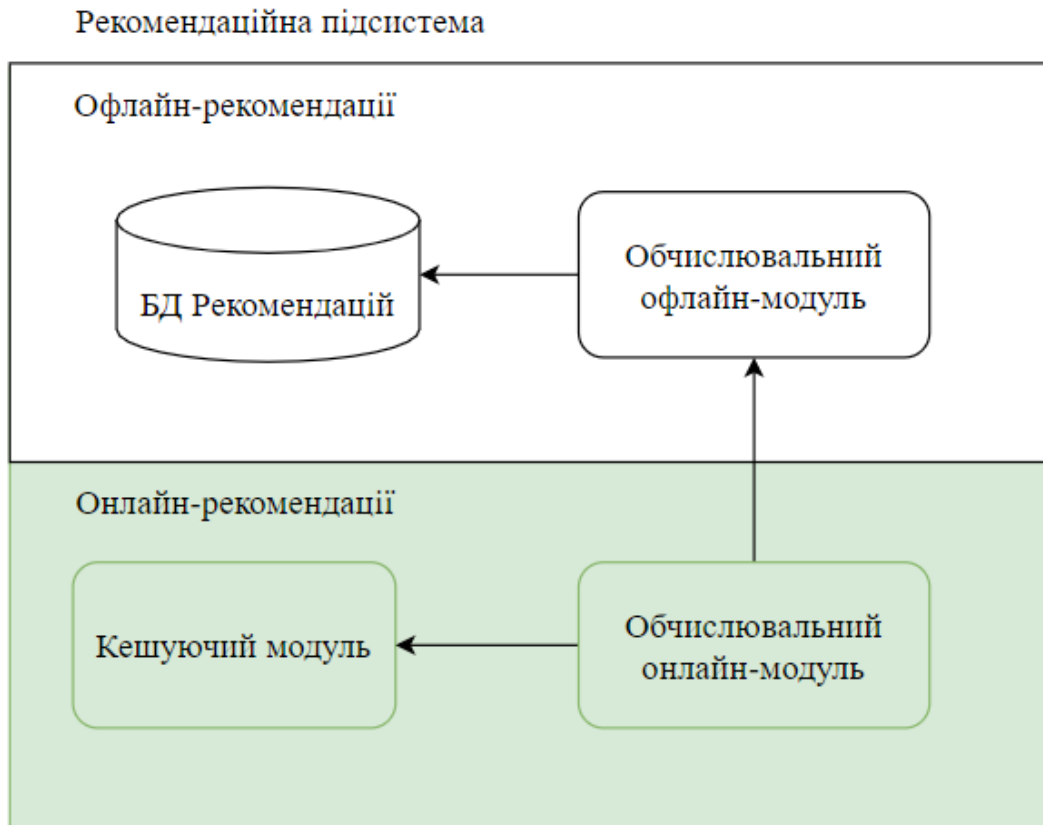
Об'єкт дослідження	Процес формування рекомендації для користувачів систем електронної комерції.
Предмет дослідження	Методи формування онлайн-рекомендацій в системах електронної комерції.
Мета роботи	Удосконалення методів формування онлайн-рекомендацій на основі контенту(контентної фільтрації) та колаборативної фільтрації використовуваних у процесі формування онлайн-рекомендацій для підвищення точності та загального враження користувачів від системи електронної комерції
Научна новизна	Удосконалений метод формування рекомендацій використовує негативний неявний зв'язок для колаборативної фільтрації, а також комбінує її з неявним зворотнім зв'язком та контентну фільтрацію, що дозволяє частково вирішити проблему «холодного старту» для нових користувачів та підвищити точність надання рекомендацій.
Практичні результати	Підвищення точності надання рекомендацій на 4 відсотки. Часткове вирішення проблеми «холодного старту» для нових користувачів.
Задачі дослідження	<ol style="list-style-type: none"> 1. Дослідження особливостей рекомендаційної підсистеми в системі електронної комерції; 2. Дослідження методів формування онлайн-рекомендацій в системах електронної комерції; 3. Дослідження методів колаборативної фільтрації у рекомендаційних системах; 4. Дослідження методів контентної фільтрації у рекомендаційних системах; 5. Удосконалення методу колаборативної фільтрації з моделями прихованих факторів; 6. Розробка технології використання удосконаленого методу; 7. Експериментальна перевірка удосконалення та комбінування методів формування онлайн-рекомендацій.

Структура інформаційної системи електронної комерції

Система електронної комерції розподіляється на наступні підсистеми:
Pricing підсистема, E-commerce платформа, Shared сервіси, CMS підсистема,
Клієнтська частина інтернет-магазину та Рекомендаційна підсистема.



Особливості онлайн-рекомендацій



Офлайн- рекомендації мають менше обмежень на обсяг даних та обчислювальну складність алгоритмів, оскільки воно виконується пакетним способом із розслабленими вимогами до часу. Однак між оновленнями це може легко застаріти, оскільки найновіші дані не включені. Одне з ключових питань архітектури персоналізації полягає в тому, як поєднувати та керувати обчисленнями в Інтернеті та в режимі офлайн безперешкодно.

Онлайн-рекомендації можуть краще реагувати на останні події та взаємодію користувачів. Це може обмежити обчислювальну складність використовуваних алгоритмів, а також обсяг даних, які можна обробити.

Методи формування онлайн-рекомендацій

1. Коллаборативна фільтрація (англ. Collaborative Filtering);
2. Метод фільтрації на основі змісту (англ. Content-Based Filtering);

Коллаборативна фільтрація - це один з методів побудови прогнозів (рекомендацій), який рекомендує активному користувачеві елементи, які сподобались іншим користувачам зі схожими смаками в минулому. Схожість смаку двох користувачів обчислюється на основі подібності в історії рейтингів користувачів.

Переваги методу: Не має потреби аналізувати дані користувачів. Не має потреби аналізувати дані предметів. Якість надання рекомендацій збільшується разом із збільшенням кількості взаємодій користувачів з предметами.

На основі змісту: система вчиться рекомендувати предмети, схожі на ті, що сподобались користувачеві в минулому або, з якими він взаємодіє зараз.

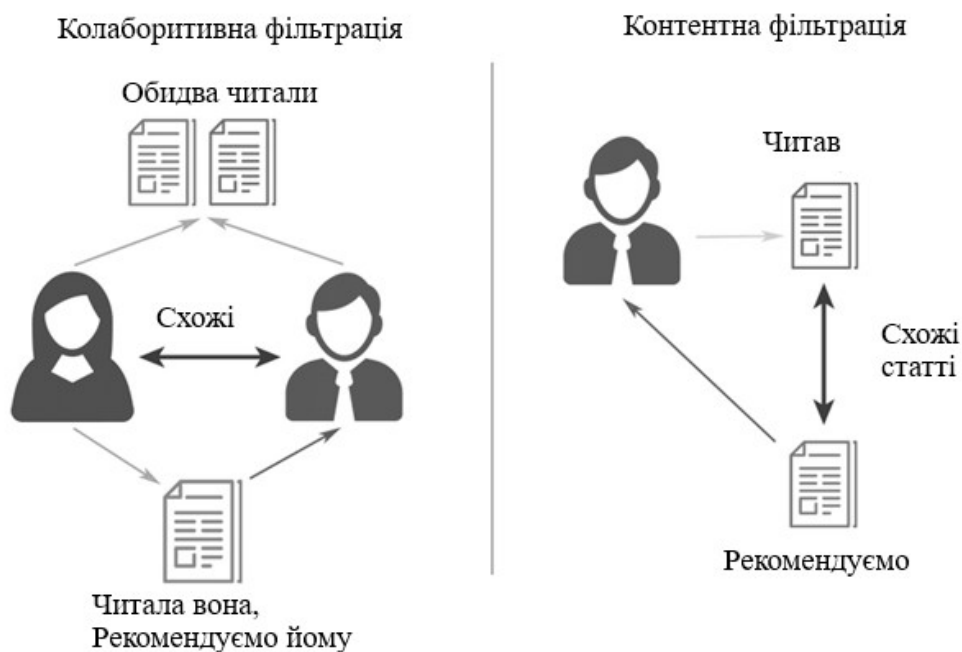
Переваги: Не потребує заповненості профілю користувача для формування рекомендацій. Системі не потрібно обробляти будь-яку інформацію про користувача. Не має проблеми «холодного старту» для користувачів.

Постановка задачі

Існуючі методи формування рекомендацій зазвичай враховують окремо дані споживача або характеристики товарів, історію активності тощо. Наприклад, колаборативна фільтрація страждає від проблеми «холодного старту» для користувачів, та у випадку з неявним зворотнім зв'язком, не використовує негативний відгук. Тоді як контентна фільтрація не бере до уваги вподобання користувачів.

Проблема: удосконалення методу колаборативної фільтрації бере до уваги негативний зворотній зв'язок, наприклад повернення товарів назад до магазину. А також комбінує її з контентною фільтрацією, щоб мати можливість рекомендувати користувачу товари у випадках, коли його вподобання невідомі.

Коллаборативна та контентна фільтрація



Різниця полягає в методі того, як ми визначаємо „подібність” між об’єктами (як правило, товарами).

У фільтрації на основі вмісту ми використовуємо властивості об’єктів та пов’язуємо подібні та показуємо їх, тоді як у коллаборативній фільтрації ми зазвичай використовуємо дані про те, що якимось чином пов’язано зовнішнім об’єктом сортування (наприклад, придбані разом онлайн-покупцем), і показуємо їх в упорядкованому списку.

Для прикладу ми скористаємося пропозицією книг.

Фільтрування вмісту: ви показуєте всі книги, що мають одного автора, одного видавця, той самий жанр та найбільш подібну кількість сторінок.

Коллаборативна фільтрація: ви аналізуєте, які книги читали люди, які читали ті самі книги, що й поточний користувач, а ті, що мають найбільшу кількість тих самих книг, знаходяться вгорі списку.

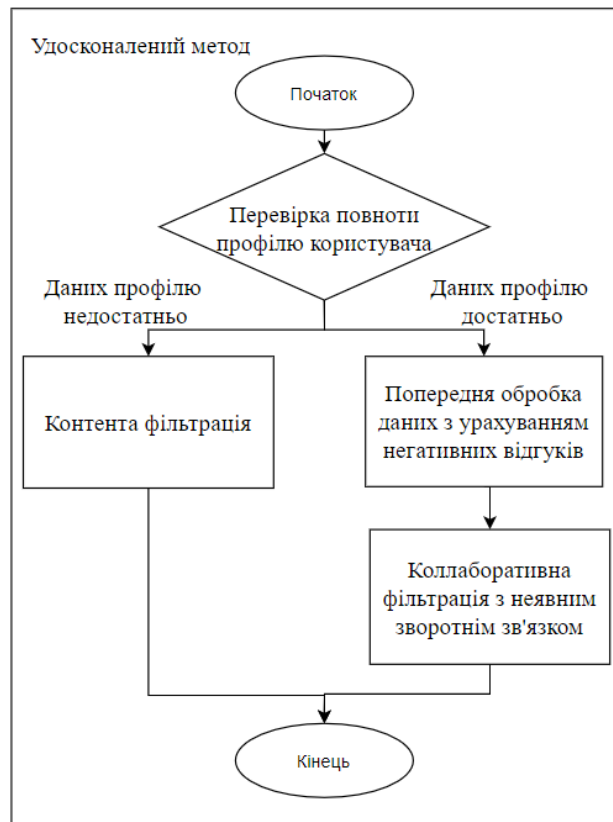
Порівняння методів формування онлайн-рекомендацій

Метод формування рекомендацій	Коллаборативна фільтрація	Контентна фільтрація	Модифікований метод
Характеристики			
Відсутність потреби обробляти дані товарів	+	-	+/-
Відсутність потреби обробляти дані користувачів	+	+	+
Випадковість(англ. Serendipity)	+	-	+
Відсутність «холодного старту» для нових користувачів	-	+	+
Відсутність «холодного старту» для нових товарів	+	-	+/-
Відсутність великої розрідженості даних	-	+	+/-

Наведений порівняльний аналіз методів формування рекомендацій дозволяє зробити висновок про доцільність поєднання контентної та коллаборативної фільтрації.

Підходи, які використовуються при розробці удосконаленого методу

Підхід, що використовується	Призначення
Коллаборативна фільтрація з неявним зворотнім зв'язком	Підвищена точність Використання негативних відгуків
Контентна фільтрація	Підвищена швидкість Можливість надавати рекомендації за незаповненого профілю користувача



Вхідні дані методу

Удосконалений метод використовує наступні вхідні дані.

Для колаборативної фільтрації використовуються дані про активність користувачів у інтернет магазині

$$E = \{u_j, i_k, e_l\}, \quad (1)$$

де u_i - дані споживача; i_k - ідентифікатор товару, e_l - тип події(переглянув, додав у корзину і т.д.).

Для контентної фільтрації використовуються дані про товари магазину у наступному вигляді

$$G = \{t, i_k, p_{n_n}, p_{v_n}\}, \quad (2)$$

де t – позначка часу, i_k - ідентифікатор товару, p_{n_n} - назва властивості товару, p_{v_n} - значення властивості товару.

Удосконалений метод

Метод, що досліджується, містить наступні фази:

- Фаза 1. Вибір використовуваного методу;
- фаза 2. Формування рекомендацій за допомогою колаборативної фільтрації;
- фаза 3. Формування рекомендацій за допомогою контентної фільтрації.

Розглянемо ці фази одна за одною.

Під час фази 1. Ми робимо вибір алгоритму формування рекомендацій в залежності від повноти профілю користувача. Для цього отримаємо кількість взаємодій з системою для поточного користувача.

$$r_u = \sum_{i \in I_u} r_{ui} \quad (1)$$

Тут r_u – сума усіх взаємодій користувача u ,

Тепер отримаємо множину усіх сум взаємодій користувачів системи.

$$(2)$$

Після цього порівняємо суму взаємодій поточного користувача з медіаною взаємодій з усіх користувачів.

$$r_u > \text{median}(r) \quad (3)$$

де median – функція медіани.

У такому випадку, коли сума взаємодій поточного користувача буде більшою за медіану взаємодій усіх користувачів, тоді ми використовуємо алгоритм колаборативної фільтрації з неявним відгуком. Якщо меншою – тоді використовуємо контентну фільтрацію.

Далі розпишемо етапи фази 2, а саме формування рекомендацій для колаборативної фільтрації.

Етап 1.

Нам потрібно формалізувати поняття впевненості, яке вимірюють змінні r_{ui} . Складемо з цих оцінок матрицю оцінок R розміром M на N , де M - кількість користувачів, а N - кількість елементів. Ця матриця досить розріджена, оскільки більшість користувачів взаємодіють лише з декількома елементами. Немодифікована версія алгоритму, зазвичай, розцінює негативні значення відгуків, як помилкові. Для модифікованого методу, вони такими вважатися не будуть, а будуть використовуватися так само, як і у випадках роботи з явними відгуками.

Результатом цього етапу стане матриця розміром M на N , де кожен елемент матриці r_{ui} – це неявна оцінка користувача u для предмету i .

Етап 2.

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u), \quad (4)$$

де x_u – вектор факторів користувача u , Y - матриця прихованих факторів предметів, λ – константа, що обмежує впевненість у наданні рекомендацій, $p(u)$ – вектор, що зберігає усі переваги користувача u , C^u – діагональна матриця, що містить у собі усі змінні впевненості для переваг $p(u)$.

$$y_i = (X^T C^i X + \lambda I)^{-1} X^T C^i p(i) \quad (5)$$

де x_u – вектор факторів предметів i , X - матриця прихованих факторів користувачів λ – константа, що обмежує впевненість у наданні рекомендацій, $p(i)$ – вектор, що зберігає усі переваги для предмету i , C^i – діагональна матриця, що містить у собі усі змінні впевненості для переваг $p(i)$.

Результатом цього етапу стануть матриці прихованих факторів користувачів та предметів X та Y .

Удосконалений метод

Етап 3.

$$R' = XY \quad (1)$$

Результатом цього етапу стане матриця R' , для якої кожен елемент r_{ui} – це оцінка предмету i користувачем u .

Тепер розберемо фазу 3, а саме - контенту фільтрацію.

Етап 1.

$$S_i = (f_{i1}, f_{i2}, \dots, f_{in}), \quad (2)$$

де S_i – рядок, що репрезентує предмет i , а f_{in} – значення особливості n предмету i .

Результатом цього етапу стане набір рядків, що позначають особливості предметів.

Етап 2.

$$, \quad (3)$$

де a – це число входжень слова t в документ, A – загальне число слів в даному документі.

$$, \quad (4)$$

де n – число документів в колекції, D – число документів з колекції D , в яких зустрічається t (коли n не дорівнює 0)

$$x_i = tf(i, d) * idf(i, D), \quad (5)$$

де x_i – оцінка TF-IDF для предмету i .

Результатом цього етапу стане набір оцінок TF-IDF для кожного з предметів для усього набору предметів.

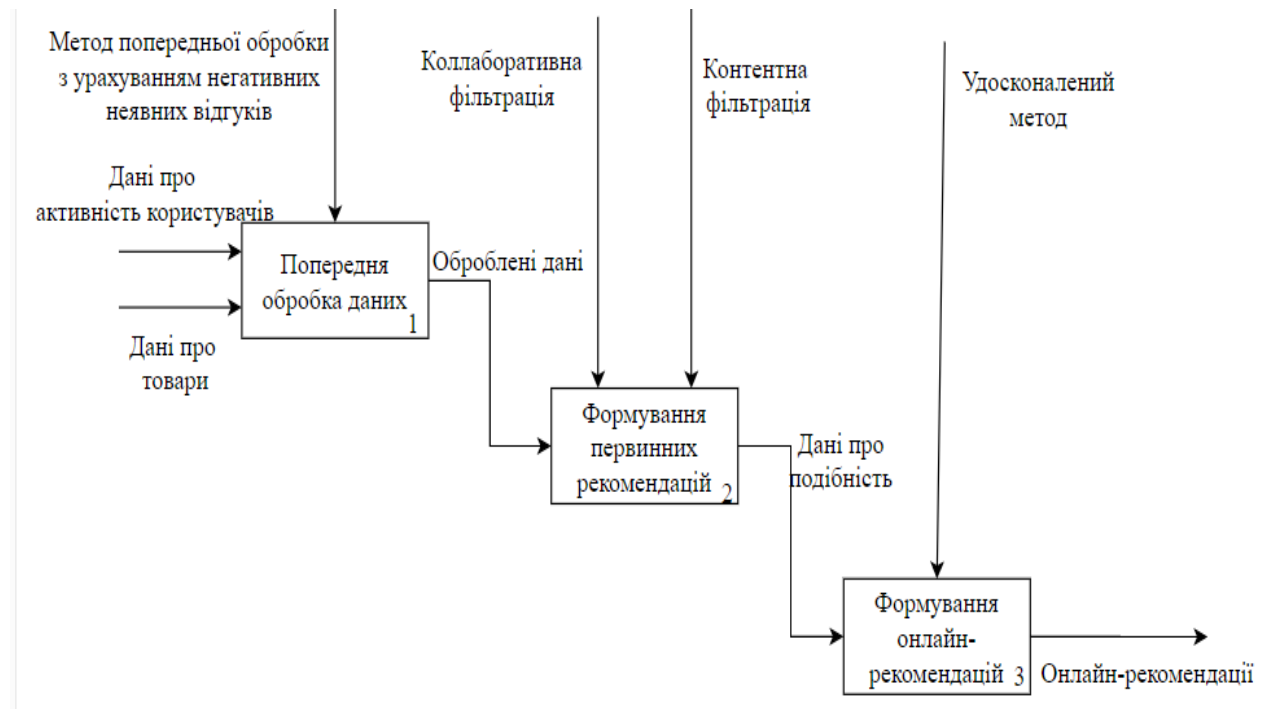
Етап 3.

$$k(x, x) = x_i^T x_i \quad (6)$$

Результатом, цього етапу стануть значення схожості предметів між собою, на основі яких і будуть сформовані рекомендації.

Не залежно від обраного методу, наприкінці виконання алгоритму отримуємо набір рекомендованих товарів для користувача. Різниця, таким чином, буде лише у даних, завдяки яким цей набір було складено: історія активності користувача, або набір атрибутів товарів.

Інформаційна технологія використання удосконаленого методу



Експериментальна перевірка методу

Метод	Значення AUC-показника	Час формування рекомендацій у секундах
Коллаборативна фільтрація з неявним зворотнім зв'язком	0,814	7,03
Модифікована коллаборативна фільтрація	0,857	7,65
Контентна фільтрація	Не обчислюється	0,6

Показник AUC (англ. Area under ROC curve, площа під ROC-кривої) - площа, обмежена ROC-кривої і віссю частки помилкових позитивних класифікацій. Чим вище показник AUC, тим якісніше класифікатор.

AUC -це відношення вірно визначених рекомендацій до всього набору даних.

Опис вибірки для експерименту

Для експерименту була використана вибірка, складається з трьох файлів: файлу з даними поведінки, файлу із властивостями елементів та файлу, що описує дерево категорій.

Наступний рядок означає що користувач з ідентифікатором рівним 1, переглянув елемент з $id = 355908$ на 1433221332117 (мітка часу Unix).

1433221332117	1	view	355908
---------------	---	------	--------

Наступний рисунок описує чорний письмовий стіл з дуба з ідентифікатором 460429 та номером категорії 1338.

1435460400000	460429	categoryid	1338
1441508400000	460429	color	black
1439089200000	460429	name	Bangoo black writing table
1431226800000	460429	material	Oak

Рядок «100,200» означає, що ідентифікатор категорії = 1 має батьківський з ідентифікатором категорії = 200

Рядок «300» означає, що ідентифікатор категорії не має батьківського елемента у дереві

100	200
300	

Приклад без впровадження удосконаленого методу

```
AUC score. Classic collaborative filtering: 0.814  
Process finished with exit code 0
```

Показник AUC для немодифікованого методу колаборативної фільтрації

```
User purchase history  
StockCode Action Description  
22906 transaction 12 Indoor/Outdoor Rug Assorted 6-ft x 8-ft  
13258 transaction CANVAS Darby Outdoor Rug, 8 x 10-ft  
30466 view PDG Storage Deck Box, 99-gal  
  
Recommendations  
StockCode Description  
23133 CANVAS Quince Outdoor Rug 5 x 7-ft  
22862 CANVAS Outdoor Rug 5 x 7-ft  
22597 CHRISTMAS MUSICAL ZINC HEART  
90114 SUMMER DAISIES BAG CHARM  
21693 SMALL HAMMERED SILVER CANDLEPOT  
--- Execution time 7.65806865692139 seconds ---
```

Рекомендації для користувача сформовані за методом колаборативної
фільтрації

Приклад застосування удосконаленого методу

```
AUC score. Modified collaborative filtering: 0.857

Process finished with exit code 0
```

Показник AUC для модифікованого методу колаборативної фільтрації

```
User purchase history
  StockCode Action          Description
  22906      transaction      12 Indoor/Outdoor Rug Assorted 6-ft x 8-ft
  13258      transaction      CANVAS Darby Outdoor Rug, 8 x 10-ft
  30466      view              PDG Storage Deck Box, 99-gal

Recommendations
  StockCode          Description
  23133              CANVAS Quince Outdoor Rug 5 x 7-ft
  22862              CANVAS Outdoor Rug 5 x 7-ft
  90204              CANVAS Tribune Outdoor Rug 5-ft x 7-ft
  90114              SUMMER DAISIES BAG
  22855              FINE WICKER HEART
--- Execution time 7.65806865692139 seconds ---
```

Рекомендації для користувача сформовані за методом колаборативної
фільтрації

```
Recommending products similar to Indoor/Outdoor Rug, Assorted, 6-ft x 8-ft Durable, ribbed and needle punch indoor/outdoor
-----
Recommended: CANVAS Outdoor Rug, 5 x 7-ft CANVAS Outdoor Rug features a PVC construction with 100% polyester border Suitable
Recommended: CANVAS Quince Outdoor Rug, 5 x 7-ft CANVAS Quince Outdoor Rug is 100% woven for quality and durability in the
Recommended: Concord Precut Rugs, 3-ft x 4-ft Concord Precut Rugs are ideal for indoor or outdoor use Tough, durable texture
Recommended: Wyoming climbing poster Platinum Charcoal Mat is designed for indoor and outdoor use Protects floors and sun
Recommended: Command Clear Décor Clips and Strips Clear Décor Clips and Strips are great for seasonal decoration and work
--- Execution time 0.6360299587249756 seconds ---
```

Рекомендації складені шляхом контентної фільтрації

Висновки

Магістерська робота присвячена вивченню проблем підвищення точності рекомендацій з урахуванням відмов від товарів а також вибору схожих користувачів та схожих товарів

В результаті виконання магістерської роботи були досліджені методи формування онлайн-рекомендацій у системах електронної комерції.

Були проаналізовані існуючі методи формування онлайн-рекомендацій, а саме: метод колаборативної фільтрації, заснований на неявному відгуці, а також метод контентної фільтрації, заснований на алгоритмі TF-IDF та косинусах подібності.

Була розглянута проблема колаборативної фільтрації з використанням негативного неявного відгуку, а також заміщення її контентною фільтрацією у разі, коли поведінковий профіль користувача заповнений недостатньо для точного формування рекомендацій. Це дозволило трохи збільшити точність надання рекомендацій для користувачів, у яких такий профіль заповнений достатньо повно, а також знизити час формування рекомендацій, коли даних профілю недостатньо.

В практичному аспекті метод дозволяє покращити показник AUC від 0,814 до 0,857, тобто трохи більше ніж на 4 %, а також у випадку коли профіль користувача заповнений недостатньо повно надавати рекомендації використовуючи контентну фільтрацію.

Результати дослідження по магістерській роботі представлено на міжнародному молодіжному форуму «Радіоелектроніка та молодь у XXI столітті» 2020р. опубліковані тези доповіді «Дослідження методів формування онлайн-рекомендацій у системах електронної комерції».