

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

бакалавр

(назва освітнього ступеня)

на тему: *Комп'ютеризована система формування рекомендацій
на основі вподобань споживача*

Виконав: студент IV курсу, групи СІс-44
спеціальності 123 «Комп'ютерна інженерія»

(шифр і назва спеціальності)

(підпис)

Денисов Д.В.

(прізвище та ініціали)

Керівник

(підпис)

Тим С.В.

(прізвище та ініціали)

Нормоконтроль

(підпис)

Луцик Н.С.

(прізвище та ініціали)

Завідувач кафедри

(підпис)

Осухівська Г.М.

(прізвище та ініціали)

Рецензент

(підпис)

Млинко Б.Б.

(прізвище та ініціали)

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних систем та мереж
(повна назва кафедри)

ЗАТВЕРДЖУЮ
Завідувач кафедри
Осхівська Г.М.
(підпис) (прізвище та ініціали)
« » 2021 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня бакалавр
(назва освітнього ступеня)

за спеціальністю 123 «Комп'ютерна інженерія»
(шифр і назва спеціальності)

студенту Денисову Денису Васильовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Комп'ютеризована система формування рекомендацій на основі вподобань споживача

Керівник роботи Тиш Євгенія Володимирівна, к.т.н., доцент
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «10» лютого 2021 року № 4.7-97

2. Термін подання студентом завершеної роботи 23.06.2021 р.

3. Вихідні дані до роботи Підходи до формування рекомендацій, метрики оцінювання рекомендацій, бібліотеки для формування рекомендацій, особливості колаборативної фільтрації

4. Зміст роботи (перелік питань, які потрібно розробити)

Вступ. 1. Аналіз вимог технічного завдання та принципів організації комп'ютеризованих рекомендаційних систем. 2. Моделі та архітектура комп'ютеризованих систем формування рекомендацій на основі вподобань споживача. 3. Програмне забезпечення комп'ютеризованої системи формування рекомендацій. 4. Безпека життєдіяльності, основи охорони праці. Висновки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1. Приклади застосування рекомендаційних систем та матриця крос-табуляції

2. Класифікація методів колаборативної фільтрації при формуванні рекомендацій

3. Метрики подібності на основі коефіцієнта Жаккара та косинуса кута

4. Метрика подібності на основі коефіцієнта кореляції Пірсона

5. Архітектура комп'ютеризованої системи формування рекомендацій

6. Результати оцінювання моделей формування рекомендацій та приклад пропозицій

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
<i>Безпека життєдіяльності, основи охорони праці</i>	<i>Пилипець М.І., д.т.н, проф. каф. МТ</i>		

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	<i>Розробка та аналіз технічного завдання</i>	<i>10.02-19.02.2021</i>	
2	<i>Аналіз принципів організації рекомендаційних систем</i>	<i>23.02-05.03.2021</i>	
3	<i>Обґрунтування методів колаборативної фільтрації при формуванні рекомендацій</i>	<i>06.03-22.03.2021</i>	
4	<i>Метрики визначення подібності товарів і користувачів</i>	<i>23.03-28.03.2021</i>	
5	<i>Проектування архітектури комп'ютеризованої системи формування рекомендацій</i>	<i>29.03-12.04.2021</i>	
6	<i>Реалізація програмного забезпечення комп'ютеризованої системи</i>	<i>13.04-15.05.2021</i>	
7	<i>Розробка інструкцій з налаштування параметрів комп'ютеризованої системи</i>	<i>15.05-23.05.2021</i>	
8	<i>Безпека життєдіяльності, основи охорони праці</i>	<i>24.05-01.06.2021</i>	
9	<i>Оформлення кваліфікаційної роботи</i>	<i>02.06-12.06.2021</i>	
10	<i>Попередній захист кваліфікаційної роботи</i>	<i>13.06-18.06.2021</i>	
11	<i>Захист кваліфікаційної роботи</i>	<i>21.06-27.06.2021</i>	

Студент

_____ (підпис)

Денисов Денис Васильович

_____ (прізвище та ініціали)

Керівник роботи

_____ (підпис)

Тиш Євгенія Володимирівна

_____ (прізвище та ініціали)

АНОТАЦІЯ

Комп'ютеризована система формування рекомендацій на основі вподобань споживача// Кваліфікаційна робота на здобуття освітнього ступеня бакалавр // Денисов Денис Васильович// ТНТУ, спеціальність 123 «Комп'ютерна інженерія»// Тернопіль, 2021 // с.– 73, рис. – 36 , табл. – 5, аркушів А1 – 6, бібліогр. – 20.

Ключові слова: рекомендація, вподобання, споживач, система.

При розробці комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів використано методи машинного навчання та загальнодоступний набір даних для проведення експериментальних досліджень.

Підхід на основі колаборативної фільтрації передбачає формування матриці крос-табуляції, що містить інформацію про покупця (ідентифікатор), товар і відповідні транзакції щодо здійснення покупки (кількість покупок), або кількісної оцінки, яка інтерпретує вподобання користувача.

У роботі спроектовано архітектуру комп'ютеризованої системи з можливістю інтеграції із системами електронної комерції. Основними програмними компонентами рекомендаційної системи є модуль препроцесингу даних та безпосередньо модуль формування рекомендацій.

Програмне забезпечення комп'ютеризованої системи реалізовано засобами мови програмування Python та за допомогою бібліотек: pandas та numpy для опрацювання даних; turicreate для проведення вибору та оцінки моделі; sklearn для розділення даних на навчальну і тестову вибірки.

Окрім цього, досліджено потенційні метрики для визначення подібності товарів і користувачів, і як наслідок обґрунтовано використання косинусної міри кута та коефіцієнта кореляції Пірсона.

ABSTRACT

Computer-aided system of recommendations formation based on customer preferences // Bachelor's thesis // Denysov Denys Vasylyovych// TNTU, speciality 123 «Computer engineering»// Ternopil, 2021 // p.– 73 , fig. – 36 , tab. – 5, posters A1 – 6, ref. – 20.

Keywords: recommendation, preference, consumer, system.

The development of a computerized system for making recommendations based on consumer preferences used machine learning methods and a publicly available set of data for experimental research.

The collaborative filtering approach involves the formation of a cross-tabulation matrix that contains information about the buyer (identifier), the product and the relevant transactions for the purchase (number of purchases), or a quantitative assessment that interprets the user's preferences.

The paper designs the architecture of a computerized system with the possibility of integration with e-commerce systems. The main software components of the recommendation system are the data preprocessing module and the recommendation generation module itself.

The software of the computerized system is implemented by means of Python programming language and by means of libraries: pandas and numpy for data processing; turicreate for model selection and evaluation; sklearn to divide the data into training and test samples.

In addition, potential metrics for determining the similarity of goods and users are investigated, and as a result, the use of the cosine measure of the angle and Pearson's correlation coefficient is substantiated.

ЗМІСТ

ПЕРЕЛІК ОСНОВНИХ УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ І СКОРОЧЕНЬ	8
ВСТУП	9
РОЗДІЛ 1 АНАЛІЗ ВИМОГ ТЕХНІЧНОГО ЗАВДАННЯ ТА ПРИНЦИПІВ ОРГАНІЗАЦІЇ КОМП'ЮТЕРИЗОВАНИХ РЕКОМЕНДАЦІЙНИХ СИСТЕМ ..	10
1.1 Аналіз вимог технічного завдання до проектування комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів.....	10
1.2 Принципи та особливості побудови рекомендаційних систем	17
РОЗДІЛ 2 МОДЕЛІ ТА АРХІТЕКТУРА КОМП'ЮТЕРИЗОВАНИХ СИСТЕМ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ НА ОСНОВІ ВПОДОБАНЬ СПОЖИВАЧА.....	22
2.1 Методи колаборативної фільтрації і сфера їх використання.....	22
2.2 Метрики подібності при визначенні схожих товарів та покупців у рекомендаційній системі.....	26
2.3 Процедури формування рекомендацій на основі вподобань споживачів ..	31
2.4 Проектування архітектури комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів.....	35
2.5 Препроцесинг даних при формуванні рекомендацій споживачу.....	37
РОЗДІЛ 3 ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ КОМП'ЮТЕРИЗОВАНОЇ СИСТЕМИ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ	46
3.1 Реалізація моделей для формування рекомендацій на основі популярності товарів	46

					КС КРБ 123.169.00.00 ПЗ			
Змн.	Арк.	№ докум.	Підпис	Дата	Комп'ютеризована система формування рекомендацій на основі вподобань споживача	Лім.	Арк.	Аркуші
Розроб.		Денисов Д.В.					6	
Перевір.		Тиш Є.В.				ТНТУ, каф. КС, гр. СІс-44		
Реценз.								
Н. Контр.		Луцик Н.С.						
Затверд.		Осухівська Г.М.						

3.2	Реалізація моделей колаборативної фільтрації при формуванні персональних рекомендацій	52
3.3	Тестування та оцінювання якості моделей формування рекомендацій.....	59
РОЗДІЛ 4 БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ОХОРОНИ ПРАЦІ ..		66
4.1	Вимоги до режимів праці і відпочинку при роботі з ВДТ	66
4.2	Вплив електромагнітних полів (ЕМП) на людину та заходи щодо зменшення їх впливу на обслуговуючий персонал	68
ВИСНОВКИ		73
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....		74
Додаток А. Технічне завдання		

					КС КРБ 123.169.00.00 ПЗ	Арк.
						7
Змн.	Арк.	№ докум.	Підпис	Дата		

ПЕРЕЛІК ОСНОВНИХ УМОВНИХ ПОЗНАЧЕНЬ,
СИМВОЛІВ І СКОРОЧЕНЬ

БД	База даних
КС	Комп'ютерна система
ПЗ	Програмне забезпечення
СКБД	Система керування базами даних
ML	Machine Learning
CF	Collaborative Filtering
IEEE	Institution of Electrical and Electronics Engineers
ISO	International Organization for Standartization

					<i>КС КРБ 123.169.00.00 ПЗ</i>	Арк.
						8
Змн.	Арк.	№ докум.	Підпис	Дата		

ВСТУП

Сучасна епоха цифрової економіки вимагає впровадження новітніх інструментів управління процесами менеджменту, залучення та створення комплексу програмно-апаратних рішень для їх підтримки, а також розширення кількості споживачів різних послуг чи товарів.

Враховуючи глобальне зростання і доступність об'єму інформаційних ресурсів, а також розвиток методів і засобів машинного навчання, зараз спостерігається інтенсивність їхнього застосування для вирішення найрізноманітніших задач. Так, наприклад, для залучення більшої кількості користувачів та формування цільових пропозицій для конкретного користувача, платформа Youtube формує персоналізовані ранжовані пропозиції відео контенту з врахування їхніх вподобань. За подібним принцип функціонують соціальні мережі Facebook, Instagram, Tik Tok, LinkedIn які, наприклад, автоматично генерують профілі користувачів для додавання у друзі. Іншим прикладом є формування рекомендацій щодо вступу у професійні об'єднання на основі вподобань користувачів. Аналізуючи запити користувачів пошукова система Google видає релевантні результати з врахуванням інтересу користувачів та особливостей їхнього портрету у мережі Internet.

Враховуючи тенденції, які сьогодні склались у різних секторах інформаційних технологій та економіки, актуальною задачею є розробка і впровадження комп'ютеризованих рекомендаційних систем. Вони дозволяють підвищити якість обслуговування користувачів, збільшити і стимулювати кількість продаж товарів і як наслідок забезпечити зростання дохідності бізнесу.

У кваліфікаційній роботі розглянуто принципи проектування та реалізації комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів для стимулювання продажу товарів і послуг.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						9
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 1 АНАЛІЗ ВИМОГ ТЕХНІЧНОГО ЗАВДАННЯ ТА ПРИНЦИПІВ ОРГАНІЗАЦІЇ КОМП'ЮТЕРИЗОВАНИХ РЕКОМЕНДАЦІЙНИХ СИСТЕМ

1.1 Аналіз вимог технічного завдання до проектування комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів

Комп'ютеризована система формування рекомендацій на основі вподобань споживача призначена для підвищення продажів у системах онлайн комерції та оффлайн магазинів шляхом реалізації моделі цільової реклами.

Такі системи також є ефективним інструментом для стимулювання продажів в залежності сезону та кон'юнктури користувачів. Окрім цього, комп'ютеризовані системи такого класу є базисом для планування проведення акційних заходів та управління активами підприємства торгівлі.

Перевагами комп'ютеризованих систем формування рекомендацій на основі вподобань споживача є власне побудова конкретних цільових пропозицій споживачам на основі їхнього «кошика» товарів. Це дає змогу сформувати споживчий портрет на рівні окремого користувача та виділити кластери подібних товарів і покупців.

Для забезпечення ефективності рекомендаційних систем необхідний доволі великий набір даних і чим він більший, тим точніше формування рекомендацій для покупців та клієнтів закладів торгівлі. Комп'ютеризовані рекомендаційні системи дозволяють визначати набори товарів, які доповнюють один одного і у результаті цього можна одержати групи крос-продуктів.

Стимулювання продажу одного виду товару пов'язано із зміною таких факторів як ціна, сезонність, збільшення доходу за рахунок обороту товарів.

					КС КРБ 123.169.00.00 ПЗ			
<i>Змн.</i>	<i>Арк.</i>	<i>№ докум.</i>	<i>Підпис</i>	<i>Дата</i>				
<i>Розроб.</i>		<i>Денисов Д.В.</i>			<i>Аналіз вимог технічного завдання та принципів організації комп'ютеризованих рекомендаційних систем</i>	<i>Літ.</i>	<i>Арк.</i>	<i>Аркуші</i>
<i>Перевір.</i>		<i>Тиш Є.В.</i>					10	
<i>Реценз.</i>						<i>ТНТУ, каф. КС, гр. СІс-44</i>		
<i>Н. Контр.</i>		<i>Луцик Н.С.</i>						
<i>Затверд.</i>		<i>Осухівська Г.М.</i>						

Зважаючи на наведені сфери застосування комп'ютеризованих рекомендаційних систем можна виділити основні напрямки їхнього застосування:

- стимулювання продажів товарів;
- виявлення подібних товарів або їхніх груп;
- виявлення подібних споживачів та формування відповідних кластерів;
- підвищення ефективності крос-продаж.

Метою створення комп'ютеризованої системи формування рекомендацій на основі вподобань споживача є підвищення ефективності процесів продажу товарів за рахунок впровадження моделі цільової реклами для кінцевих покупців.

Для того, щоб досягти поставленої мети потрібно розв'язати такі задачі:

- аналіз підходів до побудови рекомендаційних систем;
- обґрунтування вибору мови програмування та відповідних середовищ реалізації рекомендаційної системи;
- підготовка та препроцесинг даних для моделювання ;
- визначення характеристик та екосистеми функціонування рекомендаційної системи;
- програмна реалізація моделі машинного навчання для формування пропозицій споживачу;
- обґрунтування та аналіз метрик для оцінки якості формування рекомендацій;
- визначення потенційних шляхів інтеграції рекомендаційної системи у наявну інформаційну інфраструктуру;
- проведення експериментальних досліджень щодо точності та стійкості формування рекомендацій.

До основних задач і функцій комп'ютеризованої системи, яка розробляється у кваліфікаційній роботі належить формування рекомендацій товарів для споживачів на основі їхніх вподобань. Вхідними даними при цьому є набір продукції, яка була придбана покупцем протягом певного періоду часу.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						11
Змн.	Арк.	№ докум.	Підпис	Дата		

Основна задача, яку має розв'язати система, полягає у реалізації моделі на основі алгоритмів машинного навчання для формування пропозицій з наперед визначеною точністю та адекватністю. Окрім цього, сформовані рекомендації повинні бути ранжованими і кінцевому споживачу має відобразитися 5 пропозицій з найвищою точністю.

До непрямих функцій комп'ютеризованої системи належать:

- виявлення подібних товарів;
- групування товарів в однотипні кластери;
- визначення споживачів з подібними смаками;
- формування кластерів схожих покупців;
- створення базису для прийняття рішень щодо крос-продажів (cross-selling) та керування процесом стимулювання продажів (up-selling).

При розробці комп'ютеризованої системи формування рекомендацій на основі вподобань споживача необхідно передбачити можливість інтеграції із зовнішніми системами, наприклад з платформами електронної комерції.

До складу системи повинно входити апаратне і програмне забезпечення. Апаратні пристрої системи включають сервер, на якому знаходяться дані про покупки споживачів, та клієнтських машин – з можливістю доступу до мережі інтернет. Характеристики сервера та комп'ютера для моделювання повинні володіти високими технічними характеристиками, що обумовлено необхідністю опрацювання великої кількості інформації та можливих алгоритмів перебору даних.

До вимог, які в цілому висуваються до комп'ютеризованої системи формування рекомендацій, належать вимоги щодо забезпечення функціональності зчитування даних про покупки споживачів, виявлення суміжних товарів та подібних покупців і їхніх кошиків. Як наслідок, на основі такої інформації потрібно сформувати пропозиції продукції, яка або вже існує в кошику і користувач давно не здійснював її придбання, або ще не має запису про наявний товар з групи подібних покупців.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						12
Змн.	Арк.	№ докум.	Підпис	Дата		

Комп'ютеризована система формування рекомендацій на основі споживчого кошику повинна використовувати на апаратному і програмному рівні архітектуру і технологію клієнт-сервер. Для реалізації системи доцільним є використання середовища програмування Jupyter Notebook, а також відповідних розширень, наприклад, платформи GraphLab. В якості мови програмування для реалізації моделі рекомендаційної системи може бути використана мова Python або R.

Перелік основних вимог, які висуваються до комп'ютеризованої системи виглядає таким чином:

- можливість одержання інформації про покупки споживачів з відповідної бази даних;
- можливість представлення даних у форматі csv;
- здатність фільтрації кошиків споживачів;
- можливість виявлення подібних товарів та споживачів;
- здатність ранжування результуючої вибірки подібних товарів;
- можливість формування 5-ти найкращих товарів для кожного користувача у наборі даних;
- здатність до співіснування та інтеграції із зовнішніми системами;
- можливість оцінювання якості сформованих рекомендацій на основі обґрунтованих метрик точності.

До структури комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів висуваються вимоги щодо наявності вхідного набору даних, який містить ідентифікатори покупців та кількості придбаних позицій товарів. Реалізація логіки формування пропозицій покладається на модель машинного навчання на основі алгоритму популярних товарів (popularity model) при появі нових покупців та колаборативної фільтрації (collaborative filtering) при наявності споживчого кошику покупця. Функціонування системи забезпечуються наявністю таких основних блоків:

- джерело даних – фрагмент бази даних, що містить інформацію про покупців і товари;

					КС КРБ 123.169.00.00 ПЗ	Арк.
						13
Змн.	Арк.	№ докум.	Підпис	Дата		

– інтелектуальний модуль формування персональних рекомендацій для кожного покупця.

Інтелектуальний модуль повинен використовувати такі бібліотеки як:

- pandas;
- numpy;
- turicreate;
- sklearn.

Комп'ютеризована система формування рекомендацій повинна забезпечувати сталість функціонування та результатів опрацювання даних і видавати ранжований список товарів для кожного конкретного споживача.

Доступ до даних інтелектуального модуля формування пропозицій споживача повинен бути забезпечений або локально – при зберіганні даних та самого модуля на одному хості, або за допомогою доступу до мережі Інтернет із вказанням конкретної адреси сервера, на якому розміщено файл з даними у форматі csv.

При інтеграції інтелектуального модуля з платформою електронної комерції необхідно забезпечити авторизований доступ до відповідних ресурсів.

Вимогами щодо діагностики комп'ютеризованої системи є:

- вимоги перевірки наявності зв'язку інтелектуального модуля з джерелом даних;
- вимоги коректності формування рекомендацій споживача у відповідності до його вподобань;
- здатність до виявлення та ремонтпридатності шляхом застосування системного логування подій.

При появі нового покупця, персональні рекомендації повинні бути сформовані на основі вибірки товарів, які найчастіше купують користувачі, інформація про кошики яких вже є відомою. В інших випадках – повинен виконуватись алгоритм колаборативної фільтрації. За наявності помилок при функціонуванні комп'ютеризованої системи повинні застосовуватись механізми

					КС КРБ 123.169.00.00 ПЗ	Арк.
						14
Змн.	Арк.	№ докум.	Підпис	Дата		

та інструменти відновлення працездатності системи, і відповідно, її складові частини повинні відповідати характеристиці ремонтпридатності.

Модернізація проектованої системи можлива лише за потреби інтеграції зовнішніх систем з особливими вимогами, або у випадку зміни структури джерела даних. Зміна джерела даних може бути пов'язана з додавання додаткових властивостей про товари або про споживачів, які дозволяють збільшити інформативність і точність формування персональних пропозицій.

До перспектив розвитку комп'ютеризованої системи входить побудова та інтеграція модулів з визначення додаткових прихованих властивостей товарів, виявлення сезонної складової і трендів, що в комплексі могли б забезпечити вищу точність і робастність алгоритмів машинного навчання.

До вимог надійності комп'ютеризованої системи, яка розробляється у кваліфікаційній роботі належать:

- час стабільного та безперервного функціонування на рівні 23 год./добу;
- здатність до відновлення працездатності після виникнення помилок та збоїв у роботі апаратного забезпечення;
- можливість забезпечення авторизованого доступу до інтелектуального модуля та джерела даних;
- можливість тимчасового відключення роботи інтелектуального сервісу;
- можливість реінтеграції системи із зовнішніми чи сторонніми сервісами;
- адекватність надання результатів на запит користувачів;
- можливість надійного розпаралелювання та здатності використовувати графічні ядра при формуванні рекомендацій.

Вимогами до функцій і задач, які покликана виконувати комп'ютеризована система формування рекомендацій на основі вподобань споживачів є:

- здатність зчитування та фільтрації інформації з джерела даних;
- можливість реалізації «холодного» старту при формуванні рекомендацій;

					КС КРБ 123.169.00.00 ПЗ	Арк.
						15
Змн.	Арк.	№ докум.	Підпис	Дата		

- здатність формувати рекомендаційні на основі алгоритму колаборативної фільтрації;
- можливість визначення подібних товарів;
- можливість підвищення крос-продажів;
- можливість стимулювання росту продажів;
- здатність формувати групи подібних покупців;
- здатність забезпечувати і підтримувати цілісність бази даних;
- забезпечення точності формування рекомендацій на рівні не нижче, ніж 80%;
- можливість ранжування рекомендованих товарів.

Рекомендованими вимогами до апаратного забезпечення при проектуванні комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів є наступні:

- процесор, що відповідає наступним технічним характеристикам: Intel Core i5 4300M з частотою 2,60 ГГц або 2,59 ГГц (1 сокет, 2 ядра, 2 потоки на ядро)

- оперативна пам'ять – 8 ГБ;

Альтернативними конфігураціями при реалізації рекомендаційної системи є:

- Xeon E5-2698 v3 з частотою 2,3 ГГц (2 сокети, 16 ядер кожен, 1 потік на Ядро, або Xeon Phi 7210 з частотою 1,3 ГГц (1 сокет, 64 ядра, 4 потоки на ядро);

- 32 ГБ або 64 ГБ об'єм оперативної пам'яті;

- накопичувач (жорсткий диск) з наявністю не менше, ніж 2-3 ГБ простору.

Операційні системи для функціонування комп'ютеризованої системи можуть бути будь-якого типу (Windows, Linux, MacOS) однак повинні підтримувати мову програмування Python відповідної версії, а також перелік бібліотек, які описані у пункті структури та функціонування комп'ютеризованої системи.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						16
Змн.	Арк.	№ докум.	Підпис	Дата		

1.2 Принципи та особливості побудови рекомендаційних систем

У зв'язку з наявністю та накопиченням великої кількості інформації у сфері електронної комерції, провадженні бізнес-процесів у торгівлі, формуванні відгуків користувачів про товари і послуги виникає новий напрям в інформаційних технологіях, який покликаний стимулювати продажі товари та формувати персональні пропозиції споживачу в залежності від його вподобань.

Для опрацювання великих об'ємів даних необхідне застосування засобів автоматизації та впровадження рекомендаційних систем. Областю застосування таких комплексів можуть бути не тільки електронна чи оффлайн комерція, але й діяльність у галузі інформаційної безпеки, автомобілебудування, ранжування товарів і послуг. Рекомендаційні системи, в даному випадку, виконують функцію підтримки прийняття рішень. Одним з підходів до побудови рекомендаційних систем є застосування методів та інструментів колаборативної фільтрації. Даний підхід базується на аналізі вподобань споживачів, а сам алгоритм використовує історичні дані щодо, наприклад, попередніх покупок користувачів, виставлених ними оцінок та відгуків, а також передбачає можливість формування рекомендацій на основі визначених груп покупців з подібним споживацьким кошиком.

Зараз методи колаборативної фільтрації використовують багато алгоритмів, що умовно можуть належати до одного з двох класів, а в загальному випадку, класифікація методів колаборативної фільтрації представлена на рис. 1.1.

Перша група методів передбачає статистичний аналіз оцінок або кількості придбаних товарів і носить назву Memory-based алгоритмів. Алгоритми даного класу основну увагу акцентують на виявленні сукупності покупців, які подібні до того, якому необхідно сформувавши рекомендацію.

Одним з доволі простих та ефективних екземплярів цього класу методів є підхід, який передбачає виявлення K найближчих сусідів. Його суть полягає в

					КС КРБ 123.169.00.00 ПЗ	Арк.
						17
Змн.	Арк.	№ докум.	Підпис	Дата		

тому, що алгоритм орієнтується на визначення заданої кількості подібних покупців та враховує попередні вподобання цільового користувача.

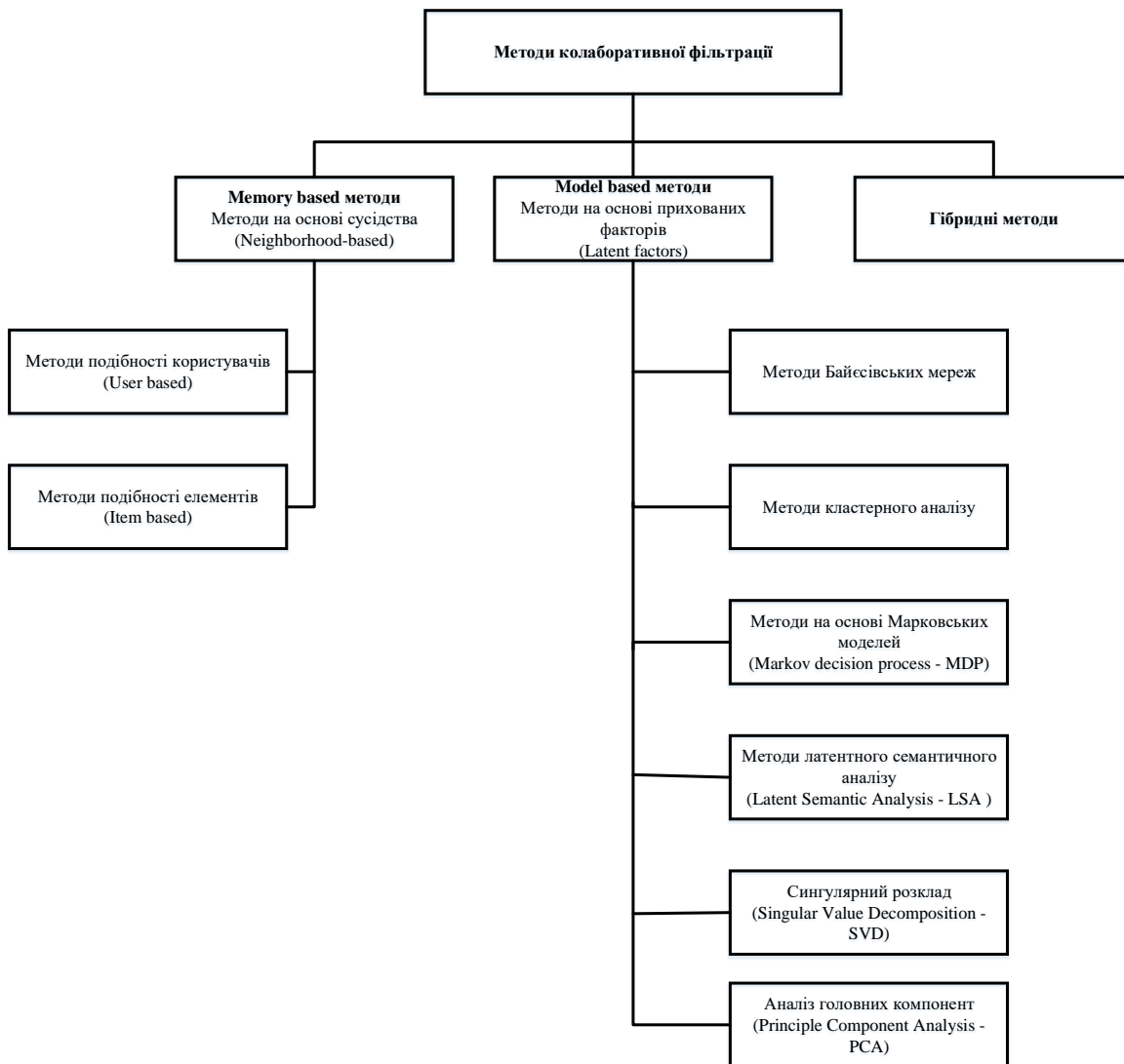


Рисунок 1.1 – Методи колаборативної фільтрації

Кількісне значення міри подібності споживчих кошиків покупців формується на основі метрик подібності, які характерні для цього класу задач.

Інший клас методів побудови рекомендаційних систем передбачає аналіз моделей даних і носить назву Model-based. Характерною і відмінною рисою цих алгоритмів є те, що на початковому етапі будується модель опису вподобань споживачів, визначаються зв'язки споживач-товар, а після цього створюються рекомендації для цільового покупця.

При формуванні рекомендацій на основі підходу Model-based спочатку відбувається навчання моделі, а після того модель починає працювати у режимі реального часу. Для реалізації таких алгоритмів можуть ефективно використовуватись підходи, засновані на методах теорії імовірності, аналізу латентних властивостей, а також методи кластерного аналізу.

Застосування адекватних метрик при побудові рекомендаційних систем є важливим аспектом забезпечення точності формування персональних пропозицій товарів. Як показує практика, найбільш ефективними серед них є наступні:

- метрики, що базуються на розрахунку Евклідової відстані або відстані Хеммінга;
- метрики на основі косинуса кута, центрованого косинуса (кореляції Пірсона) або кореляції Спірмена;
- метрика Жаккара (Jaccard coefficient);

При застосуванні колаборативної фільтрації для визначення міри подібності користувачів досягається висока точність формування рекомендацій. Проте, така фільтрація вимагає залучення значних ресурсів апаратного забезпечення, зокрема оперативної пам'яті. Це зумовлено складністю обчислень для формування ранжованого списку персональних пропозицій цільовому покупцю. Характерною особливістю при обчисленні міри подібності є те, що вона калькулюється у поточний момент часу, що пов'язано з появою транзакції лише в момент формування рекомендації. User-based метод доцільно використовувати у тому випадку, коли кількість даних для опрацювання є не надто великою.

Різновидом колаборативної фільтрації є підхід Item-based, який не вимагає формування рекомендації в реальному часі. При цьому метрика подібності цільового товару з наявними у базі даних, може обчислюватись заздалегідь за наперед визначеним графіком. Це обумовленим тим, що вектори популярності товарів є доступними до безпосередньої побудови пропозиції.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						19
Змн.	Арк.	№ докум.	Підпис	Дата		

Отже, алгоритм Item-based більш доцільно використовувати при формуванні рекомендацій товарів, оскільки завжди існує можливість попереднього гнучкого препроцесингу інформації.

Як для User-based, так і для Item-based алгоритмів потрібно завжди зберігати матриці оцінок щодо вподобань споживачів. Це породжує ще одну важливу задачу, яка полягає у коректному формуванні рекомендацій для нових покупців або нових товарів, які ще не мають своїх рейтингових оцінок. Обмеженість застосування методів колаборативної фільтрації пов'язана також з опрацюванням великих масивів інформації.

У випадку великої розрідженості матриці та надлишковості даних варто виконувати кластеризацію, тобто поділ на групи покупців і товарів з однаковими профілями оцінок.

Для прикладу, один і той же стиль музики чи жанр фільму буде подобатися визначеному набору користувачів, інші набори стилів чи жанрів відповідно подобатимуться іншому кластеру користувачів. У зв'язку з цим можна спостерігати деяку концентрацію та утворення класів користувач-фільм, користувач-стиль музики чи у нашому випадку – користувач-товар. При цьому для ефективного використання апаратних ресурсів потрібно забезпечити зниження розмірності матриці оцінок. Для вирішення цього завдання використовують методи, які належать до групи Model-based методів (рис. 1.1). Одним з шляхів формування підвбірок, або підматриць загального набору покупців і товарів – об'єднання користувачів або товарів у класи за допомогою накладання штучних індексів. І в такому випадку формування рекомендацій відбуваються у межах визначеного класу чи кластеру генеральної сукупності.

Перевагою застосування кластерного аналізу для виявлення груп товарів чи покупців є те, що ці моделі є більш гнучкими і масштабованими. Вони виконують перевірку споживчого кошика покупця у відносно невеликій кількості відфільтрованих даних і не передбачають завантаження в оперативну пам'ять усіх записів з бази даних. Реалізація алгоритму кластеризації покупців чи товарів відбувається в оффлайн режимі.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						20
Змн.	Арк.	№ докум.	Підпис	Дата		

При застосуванні колаборативної фільтрації, предметна область повинна описувати вподобання користувачів щодо товарів чи послуг. Вподобання споживача можна представити у вигляді кількісної оцінки товару i , в загальному випадку, формування трьохкомпонентної структури: покупець-товар-значення оцінки. Кількісне вираження вподобання споживача може використовувати різні шкали, що обумовлюється типом рекомендаційного сервісу.

Часто в мережі Інтернет для висловлення вподобання користувач може використовувати цілочисельну шкалу від 0 до 5, або застосовувати бінарне представлення – «подобається» чи «не подобається».

Сукупність усіх трикомпонентних структур утворює розріджену матрицю оцінок. У випадку, коли покупець не висловив свою думку про оцінку товару, тобто її значення є невідомим, то такі пари утворюють невідомі оцінки матриці. Типову структуру набору даних при застосуванні колаборативної фільтрації показано у вигляді табл. 1.1.

Таблиця 1.1 – Структура даних колаборативної фільтрації

	Товар 1	Товар 2	Товар 3
Покупець 1	5	3	?
Покупець 2	4	?	2
Покупець 3	?	?	5

Застосування методів колаборативної фільтрації дає змогу розв’язати дві основні задачі:

- визначення невідомих оцінок шляхом прогнозування значення на основі вподобань покупця;
- формування пропозиції цільовому споживачу – визначення відранжованого набору з K товарів для цільового споживача.

РОЗДІЛ 2 МОДЕЛІ ТА АРХІТЕКТУРА КОМП'ЮТЕРИЗОВАНИХ СИСТЕМ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ НА ОСНОВІ ВПОДОБАНЬ СПОЖИВАЧА

2.1 Методи колаборативної фільтрації і сфера їх використання

При побудові комп'ютеризованої системи формування рекомендацій, яка б враховувала персоналізацію та вподобання користувачів доцільним є застосування методів колаборативної фільтрації. Реалізація таких систем можлива лише за наявності даних, приклад структури і пояснення яких показано на рис. 2.1.

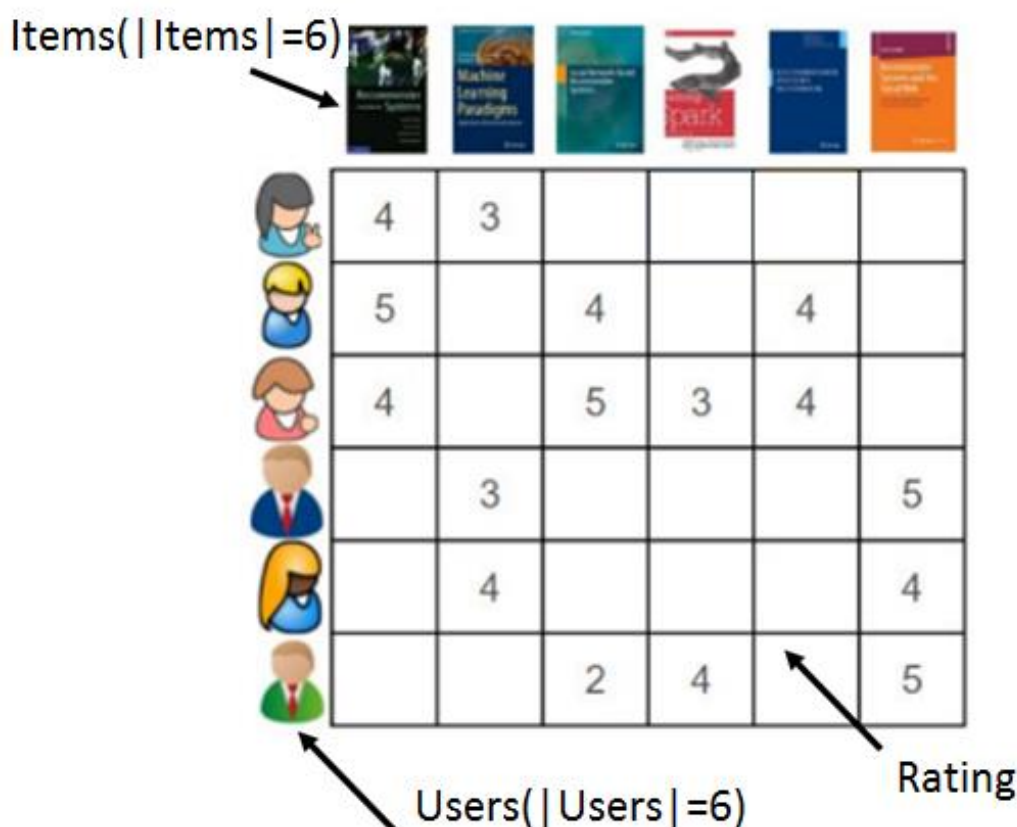


Рисунок 2.1 – Формат представлення даних при застосуванні методів колаборативної фільтрації

					КС КРБ 123.169.00.00 ПЗ		
Змн.	Арк.	№ докум.	Підпис	Дата			
Розроб.		Денисов Д.В.			Лім.	Арк.	Аркуші
Перевір.		Тиш Є.В.				22	
Реценз.					ТНТУ, каф. КС, гр. СІс-44		
Н. Контр.		Луцик Н.С.					
Затверд.		Осухівська Г.М.					

На рис. 2.1 інтерпретація позначень відповідає наступним сутностям:
 Users – сукупність покупців, інформація про яких наявна у системі;
 Items – набір товарів або послуг, які придбав або оцінив користувач;
 Rating – множина оцінок або кількісних значень, що відображає вподобання користувача чи кількість придбаного товару;

$Matrix = ||m_{ui}||$ – матриця відповідності покупців і товарів (крос-табуляція) розміром $|Users| \times |Items|$

$m_{ui} = Aggr \{ (user_t, items_t, rating_t) \in K | user_t = user, items_t = item \}$ – представляє собою деякий агрегатор, який визначає споживчий кошик покупця (запис «user-item»).

Прикладними областями використання рекомендаційних систем на основі колаборативної фільтрації можуть бути мережі супермаркетів, бібліотеки, онлайн інструменти з продажу найрізноманітніших товарів і послуг, онлайн платформи навчання та ряд інших.

Приклад та візуалізація предметних областей можливого використання комп'ютеризованих систем формування рекомендацій на основі вподобань користувачів показано на рис. 2.2 і 2.3.

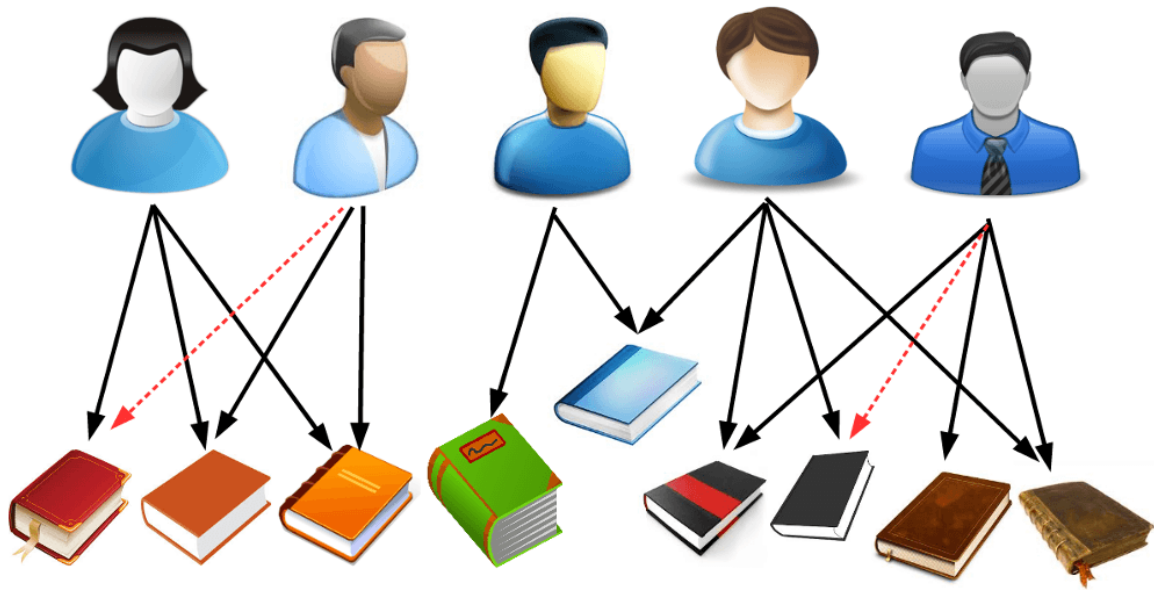


Рисунок 2.2 – Рекомендації при купівлі книг

Змн.	Арк.	№ докум.	Підпис	Дата

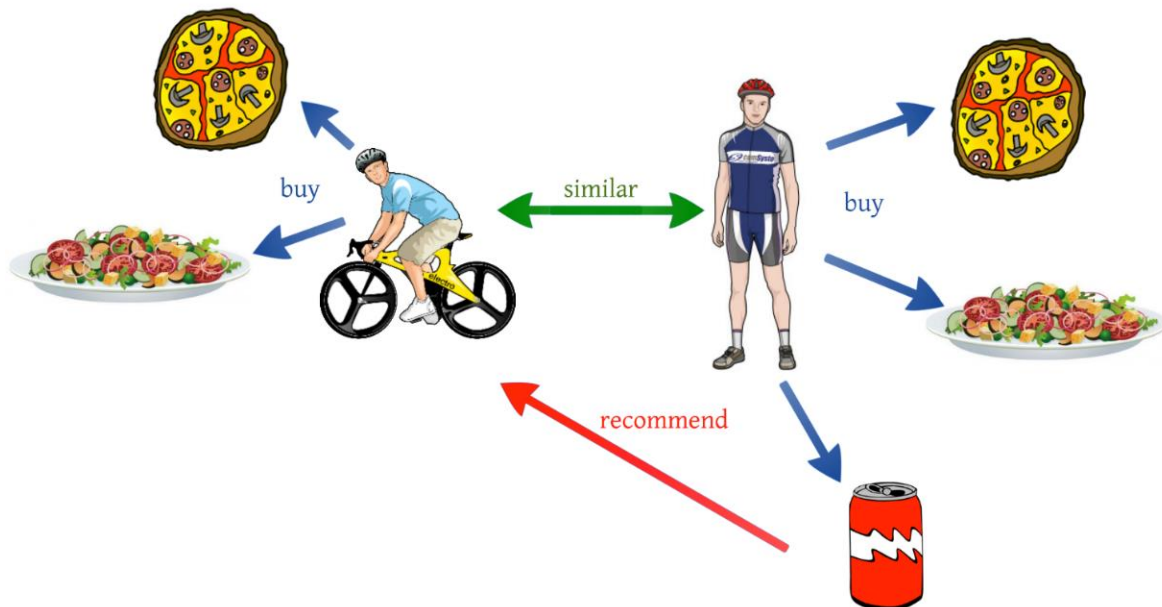


Рисунок 2.3 – Рекомендації крос-товарів у сфері здорового харчування

Для продуктових магазинів та супермаркетів, систем онлайн торгівлі вхідний набір даних мав би вигляд, як показано на рис. 2.4. Зазвичай, дані про товари, транзакції і користувачів у таких системах зберігаються у вигляді реляційних баз даних, які можна легко експортувати у файли із зручними наборами для побудови рекомендацій.

У даному випадку інтерпретація сутностей предметної області при побудові рекомендаційної системи формується наступним чином:

Users – покупці магазину;

Items – номенклатура товарів магазину;

Rating_{ui} = [покупець user придбав товар item].

За допомогою колаборативної фільтрації, при наявності такої структури даних, можна забезпечити стимулювання продажу одного і того ж виду товару – так званий up-sell. Окрім цього, такі методи дозволяють формувати набори супутніх товарів до тих, які вже придбав покупець, що носить назву cross-selling.

За допомогою колаборативної фільтрації в області торгівлі товарами і послугами можна досягти таких переваг над конкурентами:

- кількісна оцінка рекомендації товару для конкретного покупця;
- формування набору товарів, які враховують вподобання покупців;

Змн.	Арк.	№ докум.	Підпис	Дата

- можливість формувати пропозиції щодо супутніх товарів;
- забезпечення нотифікації покупців про появу нових товарів або знижок на товари;
- формування груп подібних покупців;
- формування груп подібних товарів;
- врахування необхідності проведення заходів щодо стимулювання продажів.

	+	+		+		
	+		+		+	
	+		+	+	+	
		+				+
		+				+
			+	+		+

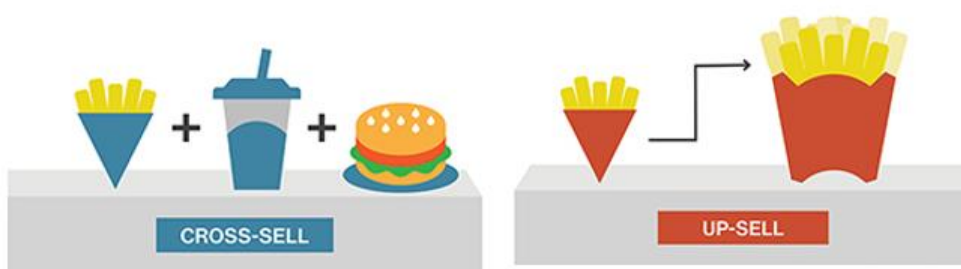


Рисунок 2.4 – Приклад формування даних для продуктових магазинів

Основною ціллю застосування колаборативної фільтрації у закладах торгівлі є збільшення продажів товарів і як наслідок зростання дохідної частини. Окрім цього, формування персональних пропозицій для конкретного покупця збільшує шанси на придбання товарів і послуг.

2.2 Метрики подібності при визначенні схожих товарів та покупців у рекомендаційній системі

При застосуванні колаборативної фільтрації можливе використання наступних метрик якості формування рекомендацій:

- коефіцієнт подібності Жаккара;
- косинусна міра кута;
- центрований косинус або коефіцієнт кореляції Пірсона.

Коефіцієнт подібності Жаккара визначається відношенням кількості елементів двох множин на їх перетині до загальної кількості елементів при об'єднанні множин. Виходячи з цього, міру подібності $sim(A, B)$ двох множини A і B можна записати у вигляді наступної формули::

$$sim(A, B) = |r_A \cap r_B| / |r_A \cup r_B| \quad (2.1)$$

Ефективність застосування коефіцієнта Жаккара практично обгрунтовано для задач пошуку подібних текстів, які містяться у великих корпусах даних та при реалізації систем антиплагіату. Окрім цього, коефіцієнт Жаккара для розв'язання таких задач може бути застосований разом з алгоритмом шинглів. Оптимально розрахунок коефіцієнта (2.1) виконувати на основі алгоритму MinHash. Приклад формування рекомендації фільмів для користувачів із використання коефіцієнта Жаккара показано на рис. 2.5.












							
	Film 1	Film 2	Film 3	Film 4	Film 5	Film 6	Film 7
A 	4			5	1		
B 	5	5	4				
C 				2	4	5	
D 		3					3

Рисунок 2.5 – Приклад застосування коефіцієнта Жаккара

Аналізуючи рис. 2.5 можна побачити, що кількісним показником схожості користувача А з іншими користувачами є наступні міри:

- подібність користувача А з користувачем В – 20%;
- схожість користувача А з користувачем С – 50%.

Звідси випливає, що вподобання користувача А є більш подібними до вподобань користувача С.

Основним недоліком використання коефіцієнта Жаккара є те, що він не враховує кількісного значення оцінки при купівлі товарів чи їх оцінюванні. Проте, він є ефективним інструментом тоді, коли значення покупок інтерпретуються у бінарній системі «придбав товар або не придбав товар». Також його можна використовувати у випадку так званого «холодного старту» для формування найбільш популярних товарів у наборі товарів чи послуг при появі нового клієнта.

Значення у матриці крос-табуляції, що інтерпретують кількість придбаного товару кожним користувачем або рівень їхнього вподобання можна представити у вигляді вектора у деякому просторі. Тоді міра кута (рис. 2.6), зокрема його косинуса, дозволяє виражати подібність між товарами, фільмами, або іншими сутностями предметної області, а також схожість між покупцями.

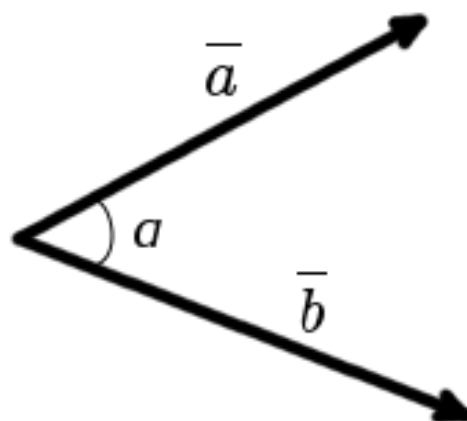


Рисунок 2.6 – Міра подібності на основі кута між векторами

Формула за знаходження косинуса кута між двома векторними представленнями при знаходженні міри подібності представлена нижче.

$$\cos(\alpha) \equiv \frac{\bar{a} \cdot \bar{b}}{|\bar{a}| \cdot |\bar{b}|} \quad (2.2)$$

При застосуванні метрики косинуса кута для набору даних з попереднього прикладу (рис. 2.3) потрібно заповнити нулями невідомі (порожні значення) у комірках матриці крос-табуляції, як показано на рис. 2.7.

							
	Film 1	Film 2	Film 3	Film 4	Film 5	Film 6	Film 7
A 	4	0	0	5	1	0	0
B 	5	5	4	0	0	0	0
C 	0	0	0	2	4	5	0
D 	0	3	0	0	0	0	3

Рисунок 2.7 – Заповнення матриці крос-табуляції при використанні метрики подібності косинус кута

Як і в попередньому випадку потрібно знайти міру подібності користувача А, який тепер представляється вектором \bar{a} з користувачами В (вектор \bar{b}) і С (вектор \bar{c}).

$$\bar{a} = (4, 0, 0, 5, 1, 0, 0);$$

$$\bar{b} = (5, 5, 4, 0, 0, 0, 0);$$

$$\bar{c} = (0, 0, 0, 2, 4, 5, 0);$$

$$|\bar{a}| = \sqrt{4^2 + 5^2 + 1^2} = \sqrt{20};$$

$$|\bar{b}| = \sqrt{5^2 + 5^2 + 4^2} = \sqrt{66};$$

$$|\bar{c}| = \sqrt{2^2 + 4^2 + 5^2} = \sqrt{45};$$

$$\bar{a} \cdot \bar{b} = (4 \cdot 5 + 0 \cdot 5 + 0 \cdot 4 + 5 \cdot 0 + 0 \cdot 1) = 20;$$

$$\bar{a} \cdot \bar{c} = (4 \cdot 0 + 5 \cdot 2 + 1 \cdot 4 + 0 \cdot 5) = 14;$$

$$\text{sim}(A, B) = \frac{20}{\sqrt{42 \cdot 66}} = 0,38;$$

$$\text{sim}(A,C) = \frac{14}{\sqrt{42 \cdot 45}} = 0,32.$$

У результаті проведених розрахунків отримуємо, що міра подібності користувача А з користувачем В становить 38%, а користувача А з користувачем С – 32%.

Недолік використання метрики косинус кута полягає в тому, що при заповненні невідомих даних нулями можлива втрата інформативності і точності формування рекомендацій.

З метою усунення недоліку, який притаманний метриці косинуса кута, доцільним є використання метрику на основі коефіцієнта кореляції Пірсона, яка також відома як метрика центрованого косинуса. Розрахунок метрики подібності на основі коефіцієнта кореляції Пірсона виконується за формулою:

$$\text{sim}(u, u') = \frac{\sum_{i \in I(u, u')} (r_{ui} - \bar{r}_u) \cdot (r_{u'i} - \bar{r}_{u'})}{\sqrt{\sum_{i \in I(u, u')} (r_{ui} - \bar{r}_u)^2 \cdot \sum_{i \in I(u, u')} (r_{u'i} - \bar{r}_{u'})^2}} \quad (2.3)$$

Для застосування коефіцієнта кореляції Пірсона вхідний набір даних повинен відповідати представленню, як показано на рис. 2.8

							
	Film 1	Film 2	Film 3	Film 4	Film 5	Film 6	Film 7
A 	4	0	0	5	1	0	0
B 	5	5	4	0	0	0	0
C 	0	0	0	2	4	5	0
D 	0	3	0	0	0	0	3

Рисунок 2.8 – Вхідний набір даних для визначення міри подібності на основі коефіцієнта кореляції Пірсона

Після цього, для кожного користувача обчислюється середнє значення його оцінок чи кількості придбаних товарів.

$$\text{Avg}(A) = (4 + 5 + 1) / 3 = 10 / 3;$$

$$\text{Avg}(B) = (5 + 5 + 4) / 3 = 14 / 3;$$

$$\text{Avg}(C) = (2 + 4 + 5) / 3 = 11 / 3;$$

$$\text{Avg}(D) = (3 + 3) / 2 = 3.$$

Розрахувавши середнє значення, потрібно трансформувати матрицю крос-табуляції з вигляду, приведеного на рис. 2.8 у вигляд, як продемонстровано на рис. 2.9. Трансформація виконується шляхом виконання операції різниці між значенням оцінки конкретного товару і середнім значенням оцінок конкретного користувача чи покупця.












	 Film 1	 Film 2	 Film 3	 Film 4	 Film 5	 Film 6	 Film 7
A 	2/3			5/3	-7/3		
B 	-2/3	1/3	-2/3				
C 				-5/3	1/3	4/3	
D 		0					0

Рисунок 2.9 – Трансформована вхідна матриця

Далі, над одержаними у матриці (рис 2.9) значеннями, виконується процедура обчислення косинуса кута у відповідності до формули 2.3. Внаслідок виконання такої операції одержують наступні міри подібності:

- схожість вподобань користувача А з вподобаннями користувача В становить 0,09;
- схожість вподобань користувача А з вподобаннями користувача С дорівнює -0,56.

У результаті одержуємо, що користувач А більш подібний до користувача В.

На основі розглянутих прикладів використання метрик подібності, можна зробити висновок про те, що найбільш ефективним є використання метрики на основі коефіцієнта кореляції Пірсона. Дана міра подібності найбільш адекватно відображає вподобання користувачів. Проте, практичне використання метрик подібності визначається типом вхідного набору даних, особливостями предметної області та розрідженістю даних у матриці крос-табуляції.

Отже, провівши аналіз та обґрунтувавши метрики подібності об'єктів, можна перейти безпосередньо до формалізації задачі прогнозування значення оцінок та формування рекомендацій користувачам на основі методів колаборативної фільтрації.

2.3 Процедури формування рекомендацій на основі вподобань споживачів

При формуванні рекомендацій на основі вподобань споживачів з використанням колаборативної фільтрації вводять ряд означень. Перше поняття стосується "холодного старту", тобто коли у користувача порожня споживацька корзина.

Припускають, що вектор $rating_x$ містить значення оцінки або кількість покупок деякого споживача x . Сукупність всіх користувачів формує множину N , що містить k споживачів, вподобання яких найбільш схожі до вподобань споживача x і їхній кошик містить оцінки товару $item$. Задача формування рекомендацій полягає у прогнозуванні вподобань на наборі I для споживача x .

Найпростішим способом формування прогнозу рекомендації товарів для нового споживача полягає в обчисленні середньої оцінки для кожного товару:

$$rating_{x_{item}} = 1/k \sum_{y \in N} rating_{y_{item}} \quad (2.4)$$

де k – загальна кількість оцінок для товару y_{item} ;

$rating_{y_{item}}$ – оцінки y_{item} матриці «користувач-товар».

					КС КРБ 123.169.00.00 ПЗ	Арк.
						31
Змн.	Арк.	№ докум.	Підпис	Дата		

Обчислення та результат застосування формули (2.4) для прогнозування оцінок щодо вподобань нового користувача E при перегляді фільмів приведено нижче та показано на рис. 2.10.

$$r_{E1} = 1/2(4 + 5) = 4,5;$$

$$r_{E2} = 1/2(5 + 3) = 4;$$

$$r_{E3} = 4;$$

$$r_{E4} = 1/2(5 + 2) = 3,5;$$

$$r_{E5} = 1/2(1 + 4) = 2,5;$$

$$r_{E6} = 5;$$

$$r_{E7} = 3.$$

							
	Film 1	Film 2	Film 3	Film 4	Film 5	Film 6	Film 7
A 	4			5	1		
B 	5	5	4				
C 				2	4	5	
D 		3					3
E 	4,5	2,67	4	3,5	2,5	5	3

Рисунок 2.10 – Результат прогнозування оцінок за формулою (2.4)

Основною перевагою підходу щодо прогнозування вподобань покупців на основі так званої моделі популярності товарів є те, що з появою нового споживача існує можливість сформулювати йому пропозицію, тобто частково розв'язується проблема «холодного старту». Наведений приклад демонструє застосування колаборативної фільтрації для матриці «користувач-товар».

Проте, застосування моделі популярності товарів не забезпечує персоналізації рекомендацій за наявності споживчих кошиків покупців. Для

вирішення цієї задачі потрібно використовувати інший підхід до прогнозування оцінок і відповідно вподобань користувачів.

Для обчислення кількісного значення оцінок використовується наступна формула:

$$r_{xi} = \frac{\sum_{j \in N(i;x)} sim(i, j) \cdot r_{xj}}{\sum_{j \in N(i;x)} sim(i, j)} \quad (2.5)$$

На прикладі використаємо формулу 2.5 для вхідного набору даних, яка представляє собою матрицю крос-табуляції «товар-користувач», яка показана на рис.2.11.

		<i>users</i>											
		1	2	3	4	5	6	7	8	9	10	11	12
<i>Films</i>	1	1		3			5			5		4	
	2			5	4			4			2	1	3
	3	2	4		1	2		3		4	3	5	
	4		2	4		5			4			2	
	5			4	3	4	2					2	5
	6	1		3		3			2			4	

Рисунок 2.11 – Вхідний набір даних матриці крос-табуляції «товар-користувач»

Порожні комірки матриці крос-табуляції інтерпретують невідомі значення оцінок товарів, тобто такі, які не були оцінені користувачами. Відповідно жовті – представляють собою значення, які містять оцінки фільмів. Задачу формування рекомендації можна сформулювати наступним чином: надати користувачу під номером 5 рекомендації щодо перегляду фільмів 1, 3, 6 або по-іншому спрогнозувати оцінку цих фільмів відповідним користувачем.

При такій постановці задачі, спочатку необхідно знайти фільми, які є схожими до фільму під номером 1. Для цього варто скористатись метрикою центрованого косинуса кута.

Алгоритм обчислення міри подібності між фільмами наступний:

– усереднюються оцінки для усіх кортежів таблиці (рис. 2.11). Для першого фільму середнє значення буде $m_1 = (1 + 3 + 5 + 5 + 4)/5 = 3,6$. У результаті отримують вектор з компонентами – $(-2,6;0;-0,6;0;0;1,4;0;0;1,4;0;0,4;0)$;

– розраховується косинус кута між усіма, одержаними векторами.

Кількісні значення обчислених коефіцієнтів схожості показано на рис. 2.12 червоним кольором.

		<i>users</i>												
		1	2	3	4	5	6	7	8	9	10	11	12	<i>sim (1,m)</i>
<i>Films</i>	1	1		3			5			5		4		1
	2			5	4			4			2	1	3	-0,18
	3	2	4		1	2		3		4	3	5		<u>0,41</u>
	4		2	4		5			4			2		-0,1
	5			4	3	4	2					2	5	-0,31
	6	1		3		3			2			4		<u>0,59</u>

Рисунок 2.14 – Обчисленні значення міри подібності фільмів

На основі представлених на рис. 2.14 коефіцієнтів подібності фільмів видно, що найбільш схожими до фільму під номером 1 є фільми під номером 3 і 6. Далі застосувавши формулу (2.5) можна спрогнозувати значення оцінок, які мав би поставити користувач під номером 5.

$$r_{15} = (0,41 * 2 + 0,59 * 3) / (0,41 + 0,59) = 2,6$$

Обчислені прогнозовані значення оцінок користувача 5 для фільмів 1, 3 та 6 записано у комірці червоного кольору матриці крос-табуляції, як показано на рис. 2.15.

		<i>users</i>												
		1	2	3	4	5	6	7	8	9	10	11	12	<i>sim (1,m)</i>
<i>Films</i>	1	1		3		2,6	5			5		4		1,00
	2			5	4			4			2	1	3	-0,18
	3	2	4		1	2		3		4	3	5		<u>0,41</u>
	4		2	4		5			4			2		-0,10
	5			4	3	4	2					2	5	-0,31
	6	1		3		3			2			4		<u>0,59</u>

Рисунок 2.15 –Прогнозовані значення оцінок фільмів

Розглянутий підхід визначення подібності між фільмами та прогнозування значень оцінок може ефективно застосовуватись у комп'ютеризованих системах формування рекомендацій на основі вподобань споживачів. Однак для реалізації такої системи потрібно спроектувати її архітектуру та інтегрувати обгрунтовані методи колаборативної фільтрації.

2.4 Проектування архітектури комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів

Проектування архітектури комп'ютеризованої системи формування рекомендацій є важливою стадією реалізацією системи, оскільки визначає структуру модулів та їх взаємодію. Окрім цього, варто розуміти якими вхідними та вихідними даними повинна оперувати система побудови рекомендацій для одержання ефективного результату персоналізації пропозицій конкретному споживачу.

Враховуючи розглянуті аспекти та особливості формування рекомендацій на основі колаборативної фільтрації та відповідних метрик визначення подібності товарів чи користувачів, архітектуру комп'ютеризованої системи формування рекомендацій запропоновано та показано на рис. 2.16.

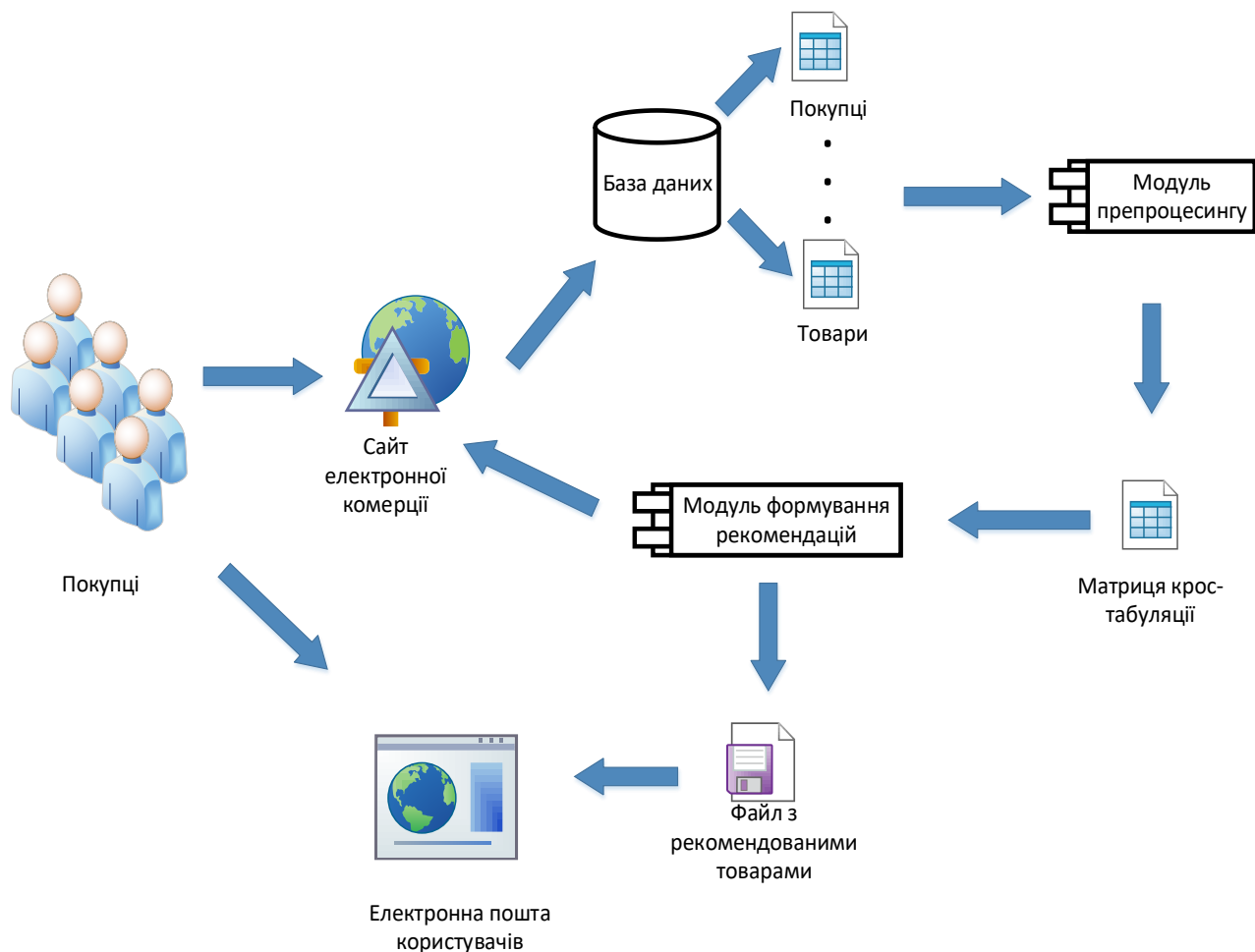


Рисунок 2.16 – Архітектура комп’ютеризованої системи формування рекомендацій на основі вподобань споживачів

Алгоритм функціонування комп’ютеризованої системи, згідно архітектури наведеної на рис. 2.16, передбачає можливість формування покупок за допомогою сайту електронної комерції.

Здійснивши покупку, інформація про користувача, зокрема його ідентифікатор, а також відповідну транзакцію записується у базу даних. У базі даних повинні бути наявними або таблиці, які зберігаються інформацію про користувачів і таблиці, або відповідні поля більш складної структури, наприклад механізм фіксації транзакцій.

Модуль препроцесингу призначений для забезпечення «чистоти даних» і фільтрації даних таким чином, щоб одержати матрицю крос-табуляції.

Після цього, матриця крос-табуляції подається на вхід модуля формування рекомендацій, у якому реалізовано алгоритми колаборативної фільтрації та

механізм обчислення метрик подібності товарів та користувачів. Сформовані рекомендації одночасно пересилаються в особистий кабінет споживача та на їхню електронну пошту. Тобто основними архітектурними компонентами комп'ютеризованої системи формування рекомендацій є:

- модуль препроцесингу даних;
- модуль формування рекомендацій;
- база даних транзакцій користувачів.

Перше, що потрібно реалізувати – це модуль препроцесингу даних для одержання матриці крос-табуляції. Вимоги до апаратного і програмного забезпечення для проектування та функціонування комп'ютеризованої системи формування рекомендацій на основі вподобань споживача наведено у технічному завданні та у підрозділі 1.1.

2.5 Препроцесинг даних при формуванні рекомендацій споживачу

При реалізації комп'ютеризованої системи формування рекомендацій на основі вподобань споживача необхідно провести експериментальні дослідження. Для прикладу, продуктова торгівельна мережа впроваджує мобільний додаток, який дозволяє своїм клієнтам робити замовлення ще до того, як їм навіть доведеться заходити в магазин. Додаток має можливість формувати рекомендації: коли клієнт вперше натискає на сторінці “замовлення”, можна сформулювати пропозицію певної кількості найкращих товарів, які потрібно додати до своєї кошика: одноразовий посуд, свіже м'ясо, чіпси тощо.

Інструмент також зможе здійснювати пошук список рекомендацій на основі вподобань цільового користувача. Процедура формування такого набору наступна:

- зчитування або введення ідентифікатора клієнта;
- повернення результату у вигляді ранжованого списку товарів (ідентифікаторів товарів), які користувач, швидше за все, захоче покласти у свій (порожній) “кошик”.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						37
Змн.	Арк.	№ докум.	Підпис	Дата		

Реалізацію рекомендаційного сервісу необхідно почати з імпорту таких модулів:

- pandas та numpy для обробки даних;
- turicreate для проведення вибору та оцінки моделі;
- sklearn для розділення даних на навчальну і тестову вибірки.

Лістинг 2.1 – Імпорт необхідних бібліотек

```
import pandas as pd
import numpy as np
import time
import turicreate as tc
from sklearn.cross_validation import train_test_split
import sys
sys.path.append("../")
```

Наступний крок полягає у зчитуванні наборів даних, на основі якого будуть формуватись рекомендації. У даному випадку генеральну сукупність даних формують два файли у форматі csv:

- users.csv – містить список 1000 ідентифікаторів покупців, які є вхідними даними для побудови рекомендаційної системи;
- transaction.csv – містить дані про транзакції (покупки) покупців з набору users.csv.

Для того, щоб зчитати дані із зазначених вище файлів використовується скрипт, програмний код якого наведений у лістингу 2.2.

Лістинг 2.2 – Зчитування даних

```
customers = pd.read_csv('../data/users.csv')
transactions = pd.read_csv('../data/transaction.csv')
```

Результат зчитування набору даних показано на рис. 2.17.

					КС КРБ 123.169.00.00 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		38

```
print(customers.shape)
customers.head()
```

(1000, 1)

	customerId
0	1553
1	20400
2	19750
3	6334
4	27773

```
print(transactions.shape)
transactions.head()
```

(62483, 2)

	customerId	products
0	0	20
1	1	2 2 23 68 68 111 29 86 107 152
2	2	111 107 29 11 11 11 33 23
3	3	164 227
4	5	2 2

Рисунок 2.17 – Результат зчитування даних

Наступний етап полягає у реалізації препроцесингу даних. У даному випадку необхідно зробити декомпозицію значень розділених прямою лінією у стовпці товарів на окремі рядки та підрахувати кількість товарів, придбаних покупцем

Створення таблиці з інформацією про користувачів, товари та цільові полі дані відіграє важливу роль, оскільки вона є вхідним набором даних для подальшого формування рекомендацій. Пропонується створити таблицю, що міститиме наступні стовпці:

- customerId;
- productId;

– purchase_count.

Фрагмент коду для трансформації вхідного набору даних наведено у лістингу 2.3.

Лістинг 2.3 – Трансформація вхідного набору даних

```
data = pd.melt(transactions.set_index('customerId')
['products'].apply(pd.Series).reset_index(),
              id_vars=['customerId'],
              value_name='products') \
    .dropna().drop(['variable'], axis=1) \
    .groupby(['customerId', 'products']) \
    .agg({'products': 'count'}) \
    .rename(columns={'products': 'purchase_count'}) \
    .reset_index() \
    .rename(columns={'products': 'productId'})
data['productId'] = data['productId'].astype(np.int64)
```

Результат трансформації вхідного набору приведено на рис. 2.18.

```
print(data.shape)
data.head()
```

(133585, 3)

	customerId	productId	purchase_count
0	0	1	2
1	0	13	1
2	0	19	3
3	0	20	1
4	0	31	2

Рисунок 2.18 – Результат трансформації вхідних даних

Далі необхідно сформувати маркери для позначення того чи купував конкретний товар клієнт чи ні. У випадку, коли покупець здійснив покупку, тоді значення `purchase_dummy` маркується 1.

Для маркування куплених товарів написано програмний код, який наведено у лістингу 2.4.

Лістинг 2.4 – Функція маркування купівлі товарів

```
def create_data_dummy(data):
    data_dummy = data.copy()
    data_dummy['purchase_dummy'] = 1
    return data_dummy
data_dummy = create_data_dummy(data)
```

Нормалізація кількості покупок кожного користувача не спрацює, оскільки клієнти можуть мати різну частоту покупок і не мають однакового смаку. Однак можна нормалізувати товари за частотою покупок для всіх покупців. При цьому необхідно створити матрицю відповідності товарів та користувачів, як показано у лістингу 2.5.

Лістинг 3.5 – Матриця відповідності товарів та покупців

```
df_matrix = pd.pivot_table(data, values='purchase_count',
    index='customerId', columns='productId')
```

Результат одержаної матриці показано на рис. 2.19.

productId	0	1	2	3	4	5	6	7	8	9	...	290	291	292	293	294	295	296	297	298	2	
customerId																						
0	NaN	2.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	6.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	1.0	NaN	NaN	1.0	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Рисунок 2.19 – Матриця «Покупець-товар»

Безпосередньо програмний код нормалізації матриці, приведеної на рис. 2.19, показано нижче (лістинг 2.6), а одержаний результат продемонстровано на рис. 2.20.

Лістинг 2.6 – Код нормалізації значень матриці

```
df_matrix_norm = (df_matrix-df_matrix.min())/(df_matrix.max()-df_matrix.min())
```

productId	0	1	2	3	4	5	6	7	8	9	...	290	291	292	293	294	295	296	297	298	
customerId																					
0	NaN	0.1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	0.166667	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Рисунок 2.20 – Нормалізована матриця «Покупець-товар

Для трансформації нормалізованої матриці як вхідних даних для побудови моделі формування рекомендації використано скрипт, який наведено у лістингу 2.7.

Лістинг 2.7 – Формування вхідних даних моделі побудови рекомендацій

```
# create a table for input to the modeling
d = df_matrix_norm.reset_index()
d.index.names = ['scaled_purchase_freq']
data_norm = pd.melt(d, id_vars=['customerId'],
value_name='scaled_purchase_freq').dropna()
print(data_norm.shape)
data_norm.head()
```

Результат сформованого набору для навчання моделі показано на рис. 2.21.

(133585, 3)

	customerId	productId	scaled_purchase_freq
9	9	0	0.133333
25	25	0	0.133333
32	33	0	0.133333
35	36	0	0.133333
43	44	0	0.133333

Рисунок 2.21 – Результат формування вхідних даних для навчання моделі

Попередні кроки, пов'язані з нормалізацією і трансформацією матриці можна об'єднати в окремий логічний програмний модуль у вигляді функції (лістинг 2.8), яка показана нижче.

Лістинг 2.8 – Функція нормалізації матриці

```
def normalize_data(data):  
    df_matrix = pd.pivot_table(data, values='purchase_count',  
                               index='customerId', columns='productId')  
    df_matrix_norm = (df_matrix -  
df_matrix.min()) / (df_matrix.max() - df_matrix.min())  
    d = df_matrix_norm.reset_index()  
    d.index.names = ['scaled_purchase_freq']  
    return pd.melt(d, id_vars=['customerId'],  
value_name='scaled_purchase_freq').dropna()
```

У результаті проведених маніпуляцій одержано нормалізований споживчий кошик в інтервалі від 0 до 1. Відповідно найбільшій кількості покупок товару відповідає 1, а у випадку, коли товар не купувався конкретним покупцем – 0.

Декомпозиція одержаних даних на навчальні і тестові вибірки є важливою частиною оцінювання процесу прогнозування значень при формуванні рекомендацій на основі моделі колаборативної фільтрації. Загально прийнятим підходом до розбиття є такий, що більша частина даних використовується для навчання моделі, а менша – для тестування. Найбільш оптимальним є декомпозиція генеральної сукупності у співвідношенні 80:20

Навчальна частина вибірки буде використана для розробки моделі прогнозування, а інша – а для оцінки ефективності моделі. Для формування навчальної тестової вибірки використовується код, наведений у лістингу 2.9.

Лістинг 2.9 – Декомпозиція даних на навчальну і тестову вибірки.

```
def split_data(data):  
    '''  
    Splits dataset into training and test set.  
  
    Args:  
        data (pandas.DataFrame)  
  
    Returns  
        train_data (tc.SFrame)  
        test_data (tc.SFrame)  
    '''  
    train, test = train_test_split(data, test_size = .2)  
    train_data = tc.SFrame(train)  
    test_data = tc.SFrame(test)  
    return train_data, test_data
```

Тепер можна маніпулювати трьома наборами даних:

- набір, що містить дані з обчисленою кількістю покупок – purchase counts;
- набір даних «користувач-товар» з бінарно промаркованими товарами – purchase dummy;
- набір нормалізованих покупок товарів – scaled purchase counts.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						44
Змн.	Арк.	№ докум.	Підпис	Дата		

Лістинг формування навчальних тестових наборів для трьох, описаних вище, наборів даних показано у лістингу 2.10.

Лістинг 2.10 – Формування навчальних і тестових вибірок для 3 наборів даних

```
train_data, test_data = split_data(data)
train_data_dummy, test_data_dummy = split_data(data_dummy)
train_data_norm, test_data_norm = split_data(data_norm)
```

Далі потрібно перейти до безпосередньої реалізації моделей формування рекомендацій на основі вподобань користувачів.

					КС КРБ 123.169.00.00 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		45

РОЗДІЛ 3 ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ КОМП'ЮТЕРИЗОВАНОЇ СИСТЕМИ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ

3.1 Реалізація моделей для формування рекомендацій на основі популярності товарів

Перш ніж застосовувати більш складний підхід на основі колаборативної фільтрації, потрібно побудувати базову модель для порівняння та оцінки рекомендацій. Оскільки базовий рівень, як правило, використовує дуже простий підхід, тому варто обрати методи, які використовуються поза ним, якщо вони дійсно демонструють відносно кращу точність та меншу складність. В якості базової моделі для порівняння з моделлю колаборативної фільтрації використаємо підхід, який базується на формуванні рекомендацій найбільш популярних товарів.

Більш складним, але загальним підходом для формування пропозицій товарів є колаборативна фільтрація. Однак на даному етапі потрібно оголосити змінні, які будуть представляти відповідні моделі (лістинг 3.1).

Лістинг 3.1 – Оголошення моделей для формування рекомендацій

```
# константи для визначення імен
user_id = 'customerId'
item_id = 'productId'
users_to_recommend = list(customers[user_id])
n_rec = 10 # кількість товарів, що входять у набір рекомендаційних
n_display = 30 # відобразити перших 30 рядків результуючого набору
```

Змн.	Арк.	№ докум.	Підпис	Дата	КС КРБ 123.169.00.00 ПЗ			
Розроб.		Денисов Д.В.			Програмне забезпечення комп'ютеризованої системи формування рекомендацій	Літ.	Арк.	Аркуші
Перевір.		Тиш Є.В.					46	
Реценз.						ТНТУ, каф. КС, гр. СІс-44		
Н. Контр.		Луцик Н.С.						
Затверд.		Осухівська Г.М.						

За допомогою бібліотеки Turicreate доволі просто побудувати моделі і для цього достатньо написати функцію, що буде використовуватися для усіх моделей (лістинг 3.2).

Лістинг 3.2 – Функція оголошення моделей для формування рекомендацій

```
def model(train_data, name, user_id, item_id, target,
users_to_recommend, n_rec, n_display):
    if name == 'popularity':
        model = tc.popularity_recommender.create(train_data,
                                                    user_id=user_id,
                                                    item_id=item_id,
                                                    target=target)

    elif name == 'cosine':
        model = tc.item_similarity_recommender.create(train_data,
                                                       user_id=user_id,
                                                       item_id=item_id,
                                                       target=target,
                                                       similarity_type='cosine')

    elif name == 'pearson':
        model = tc.item_similarity_recommender.create(train_data,
                                                       user_id=user_id,
                                                       item_id=item_id,
                                                       target=target,
                                                       similarity_type='pearson')

    recom = model.recommend(users=users_to_recommend, k=n_rec)
    recom.print_rows(n_display)
    return model
```

Модель на основі популярності товарів формує рекомендації таким чином, що кожному покупцю пропонується придбати товари, які найбільш часто купувались.

При використанні набору даних, що містить дані з обчисленою кількістю покупок (purchase count), програмний код виглядає, як показано у лістингу 3.3, а одержаний результат продемонстрований на рис. 3.1.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						47
Змн.	Арк.	№ докум.	Підпис	Дата		

Лістинг 3.3 – Побудова моделі популярності товарів на основі purchase count

```
name = 'popularity'
target = 'purchase_dummy'
pop_dummy = model(train_data_dummy, name, user_id, item_id,
target, users_to_recommend, n_rec, n_display)
```

customerId	productId	score	rank
1553	248	3.1111111111111111	1
1553	37	3.0988142292490117	2
1553	34	3.030888030888031	3
1553	132	3.015873015873016	4
1553	0	2.959847036328872	5
1553	3	2.853249475890985	6
1553	27	2.769230769230769	7
1553	110	2.7349397590361444	8
1553	32	2.650717703349282	9
1553	82	2.609865470852018	10
20400	248	3.1111111111111111	1
20400	37	3.0988142292490117	2
20400	34	3.030888030888031	3
20400	132	3.015873015873016	4
20400	0	2.959847036328872	5
20400	3	2.853249475890985	6
20400	27	2.769230769230769	7
20400	110	2.7349397590361444	8
20400	32	2.650717703349282	9
20400	82	2.609865470852018	10
19750	248	3.1111111111111111	1
19750	37	3.0988142292490117	2
19750	34	3.030888030888031	3
19750	132	3.015873015873016	4
19750	0	2.959847036328872	5
19750	3	2.853249475890985	6
19750	27	2.769230769230769	7
19750	110	2.7349397590361444	8
19750	32	2.650717703349282	9
19750	82	2.609865470852018	10

Рисунок 3.1 – Результат формування 10 рекомендацій для покупця з ідентифікатором 1553

Результат застосування моделі популярності товарів на наборі даних «користувач-товар» з бінарно промаркованими товарами показано на рис. 3.2.

customerId	productId	score	rank
1553	17	1.0	1
1553	21	1.0	2
1553	13	1.0	3
1553	1	1.0	4
1553	76	1.0	5
1553	47	1.0	6
1553	101	1.0	7
1553	25	1.0	8
1553	174	1.0	9
1553	186	1.0	10
20400	17	1.0	1
20400	21	1.0	2
20400	13	1.0	3
20400	1	1.0	4
20400	76	1.0	5
20400	47	1.0	6
20400	101	1.0	7
20400	25	1.0	8
20400	174	1.0	9
20400	186	1.0	10
19750	17	1.0	1
19750	21	1.0	2
19750	13	1.0	3
19750	1	1.0	4
19750	76	1.0	5
19750	47	1.0	6
19750	101	1.0	7
19750	25	1.0	8
19750	174	1.0	9
19750	186	1.0	10

Рисунок 3.2 – Результат формування рекомендацій на наборі purchase dummy

На рис. 3.3 наведено результати формування рекомендацій для набору нормалізованих покупок.

customerId	productId	score	rank
1553	226	0.7931034482758621	1
1553	247	0.3358208955223881	2
1553	230	0.31532846715328416	3
1553	125	0.26029411764705845	4
1553	294	0.2573643410852709	5
1553	248	0.25555555555555554	6
1553	276	0.24807692307692308	7
1553	165	0.23036649214659685	8
1553	155	0.22807017543859645	9
1553	72	0.2268993839835729	10
20400	226	0.7931034482758621	1
20400	247	0.3358208955223881	2
20400	230	0.31532846715328416	3
20400	125	0.26029411764705845	4
20400	294	0.2573643410852709	5
20400	248	0.25555555555555554	6
20400	276	0.24807692307692308	7
20400	165	0.23036649214659685	8
20400	155	0.22807017543859645	9
20400	72	0.2268993839835729	10
19750	226	0.7931034482758621	1
19750	247	0.3358208955223881	2
19750	230	0.31532846715328416	3
19750	125	0.26029411764705845	4
19750	294	0.2573643410852709	5
19750	248	0.25555555555555554	6
19750	276	0.24807692307692308	7
19750	165	0.23036649214659685	8
19750	155	0.22807017543859645	9
19750	72	0.2268993839835729	10

Рисунок 3.3 – Результат формування рекомендацій на наборі нормалізованих даних покупок користувачів

При побудові наведених вище моделей виконувалось прогнозування рекомендацій на основі популярності товарів. У результаті цього для різних наборів даних одержано записи, які відображають перші 30 стрічок для 1000 користувачів із 10 рекомендаціями. Ці 30 записів включають 3 користувачів та рекомендовані для них товари, а також оцінки та ранги відсортовані у порядку спадання.

В загальному випадку, хоча різні моделі мають різний список рекомендацій, кожному користувачеві рекомендується однаковий список із 10 товарів. Це пов'язано з тим, що популярність товару обчислюється шляхом визначення товарів, які найбільш часто купували усі користувачів наявні у наборі даних.

Якщо проаналізувати наведений нижче приклад групування (рис. 3.4), то товари з ідентифікаторами 132, 248, 37 та 34 є найбільш популярними (найбільш продаваними) серед споживачів. При цьому обчислювалась кількість транзакцій по товару, що поділена на кількість клієнтів. Звідси видно, що ці товари купуються в середньому у навчальному наборі транзакцій принаймні 3 рази (так само, як перший показник популярності за `purchase_count`).

```
train.groupby(by=item_id)['purchase_count'].mean().sort_values(ascending=False).head(20)
```

```
productId
248    3.111111
37     3.098814
34     3.030888
132    3.015873
0      2.959847
3      2.853249
27     2.769231
110    2.734940
32     2.650718
82     2.609865
10     2.608955
230    2.584615
129    2.584337
226    2.557823
245    2.554455
58     2.489362
68     2.432203
54     2.405109
18     2.400000
91     2.393162
Name: purchase_count, dtype: float64
```

Рисунок 3.4 – Згруповані за популярністю товари

Побудувавши просту модель формування рекомендацій на основі обчислення значень найбільш популярних товарів, далі необхідно удосконалити процедуру формування рекомендацій шляхом застосування методів колаборативної фільтрації.

3.2 Реалізація моделей колаборативної фільтрації при формуванні персональних рекомендацій

При використанні методів колаборативної фільтрації для формування рекомендацій необхідно встановити, яким чином подібні покупці купують товари. Для прикладу, якщо покупець 1 і покупець 2 придбали подібні товари, припустимо, що перший придбав товари А, В, С, а другий – А, В, то логічним є формування пропозицію товару С для другого покупця.

Для того, щоб виявити схожість між користувачами (міру їхньої подібності) потрібно виконати наступні кроки:

1. Побудувати матрицю споживчого кошика користувача таким чином, щоб значення індексів представляли унікальні ідентифікатори клієнтів, а значення стовпців - унікальні ідентифікатори товарів.

2. Сформувати матрицю подібності між товарами. Ідея полягає в тому, щоб підрахувати, наскільки товар схожий на інший товар. Для цього існує ряд способів і методів, які використовують метрики подібності наприклад косинусна міра кута між векторами, або коефіцієнт кореляції Пірсона. Щоб розрахувати схожість продуктів А та В, варто звернути увагу на всіх клієнтів, які оцінили обидва ці елементи. Наприклад, як А, так і В були оцінені клієнтами 1 і 2. Після цього формуються два вектори елементів, v_1 для елемента А та v_2 для елемента В, у просторі користувача (1, 2). Далі знаходиться косинус або кут/відстань за Пірсоном між цими векторами. Нульовий кут або коли вектори накладаються буде відповідати значенню косинуса кута 1. Це говорить про загальну схожість (або для кожного користувача в усіх товарах однаковий рейтинг), а кут 90 градусів означає, що його косинус дорівнює 0 тобто подібність між товарами відсутня.

3. Спрогнозувати ймовірність придбання товару, що відмінний від тих, які є в кошику покупця. У наведеному вище прикладі потрібно розрахувати оцінку товару С для другого покупця. Товар С буде цільовим елементом. Щоб обчислити значення оцінки, необхідно провести зважування розрахованого показника подібності між цільовим товаром та іншими товарами, які вже

					КС КРБ 123.169.00.00 ПЗ	Арк.
						52
Змн.	Арк.	№ докум.	Підпис	Дата		

придбав клієнт. Коефіцієнт зважування – це кількість покупок, яку містить запис користувача з вже придбаним товаром. Далі ця зважена сума масштабується із сумою мір подібності, щоб розрахований рейтинг, що залишався в межах заздалегідь визначених інтервалів. Таким чином, прогнозований рейтинг для товару С і покупця 2 буде розрахований із застосуванням метрик подібності.

Для формування рекомендацій із застосування метрики подібності «косинус кута» на наборі з врахуванням кількості придбаного товару (purchase count) реалізовано фрагмент коду, який показаний у лістингу 3.4.

Лістинг 3.4 – Формування рекомендацій на наборі purchase count з використанням метрики косинус кута

```
name = 'cosine'  
target = 'purchase_count'  
cos = model(train_data, name, user_id, item_id, target,  
users_to_recommend, n_rec, n_display)
```

Змінна cos зберігає модель рекомендацій, входними параметрами якої є:

- train_data – навчальний набір даних на якому формується рекомендація;
- name – метрика, на основі якої визначається подібність товарів;
- user_id – ідентифікатор покупця (користувача);
- item_id – ідентифікатор товару;
- target – цільова змінна, у даному випадку стовбець purchase_count, що містить значення кількості придбаного товару;
- user_to_recommend – список покупців, яким необхідно сформувати пропозиції;
- n_rec – кількість товарів, які входять у рекомендацію;
- n_display – кількість рекомендацій товарів, які потрібно відобразити.

Результат виконання, наведеного у лістингу 3.4 коду, показано на рис. 3.5.

customerId	productId	score	rank
1553	2	0.1110002875328064	1
1553	35	0.08345692157745362	2
1553	1	0.07202702760696411	3
1553	5	0.06954102516174317	4
1553	17	0.062017965316772464	5
1553	61	0.06159121990203857	6
1553	21	0.051773107051849364	7
1553	41	0.048556816577911374	8
1553	47	0.04801986217498779	9
1553	82	0.04561170339584351	10
20400	1	0.049841105937957764	1
20400	182	0.0481337308883667	2
20400	13	0.0420384407043457	3
20400	56	0.04202163219451904	4
20400	215	0.04182213544845581	5
20400	131	0.036169588565826416	6
20400	265	0.03536689281463623	7
20400	54	0.03426623344421387	8
20400	6	0.033963680267333984	9
20400	160	0.033889055252075195	10
19750	38	0.08994142214457194	1
19750	48	0.08908288677533467	2
19750	2	0.08164189259211223	3
19750	1	0.07568801442782085	4
19750	79	0.075407475233078	5
19750	36	0.06567970911661784	6
19750	5	0.05915198723475138	7
19750	21	0.05801957845687866	8
19750	14	0.04799354076385498	9
19750	13	0.04312247037887573	10

Рисунок 3.5 – Результат формування рекомендацій за метрикою косинус кута на наборі даних purchase count

Результат застосування метрики подібності на основі косинуса кута та набору «користувач-товар» з бінарно промаркованими товарами показано на рис. 3.6.

customerId	productId	score	rank
1553	2	0.11069482564926147	1
1553	1	0.07363476355870564	2
1553	5	0.06431001424789429	3
1553	21	0.06199751297632853	4
1553	17	0.0569839080174764	5
1553	61	0.05539397398630778	6
1553	29	0.05539158980051676	7
1553	105	0.053917884826660156	8
1553	15	0.05295040210088094	9
1553	143	0.051859756310780845	10
20400	17	0.0	1
20400	21	0.0	2
20400	13	0.0	3
20400	1	0.0	4
20400	76	0.0	5
20400	47	0.0	6
20400	101	0.0	7
20400	25	0.0	8
20400	174	0.0	9
20400	186	0.0	10
19750	48	0.10387962063153584	1
19750	38	0.09720247983932495	2
19750	36	0.0862055520216624	3
19750	1	0.08478485544522603	4
19750	79	0.07974439859390259	5
19750	2	0.07456593712170918	6
19750	194	0.05650122960408529	7
19750	21	0.05510311325391134	8
19750	14	0.05398238698641459	9
19750	76	0.04635966817537943	10

Рисунок 3.6 – Результат формування рекомендацій на наборі purchase dummy з використанням метрики косинуса кута

Застосування метрики косинуса кута на нормованому наборі покупок товарів показано на рис. 3.7.

customerId	productId	score	rank
1553	18	0.0	1
1553	2	0.0	2
1553	225	0.0	3
1553	174	0.0	4
1553	9	0.0	5
1553	147	0.0	6
1553	78	0.0	7
1553	4	0.0	8
1553	270	0.0	9
1553	267	0.0	10
20400	18	0.0	1
20400	2	0.0	2
20400	225	0.0	3
20400	174	0.0	4
20400	9	0.0	5
20400	147	0.0	6
20400	78	0.0	7
20400	4	0.0	8
20400	270	0.0	9
20400	267	0.0	10
19750	18	0.0	1
19750	2	0.0	2
19750	225	0.0	3
19750	174	0.0	4
19750	9	0.0	5
19750	147	0.0	6
19750	78	0.0	7
19750	4	0.0	8
19750	270	0.0	9
19750	267	0.0	10

Рисунок 3.7 – Результат формування рекомендацій на нормованому наборі покупок із використанням метрики подібності косинус кута

Ще однією метрикою, яка може використовуватись при визначенні міри подібності як користувачів, так і товарів є коефіцієнт кореляції Пірсона. Розрахунок даного коефіцієнта наведено у другому розділі, по-іншому даний коефіцієнт називають центрованим косинусом кута. Програмний код для формування рекомендацій на основі коефіцієнта кореляції Пірсона має вигляд, подібний до того, який наведено у лістингу 3.4. Однак відмінність полягає в тому, що значення параметра `name='pearson'`, а не `'cosine'`. Рекомендації, сформовані із використанням коефіцієнта кореляції Пірсона на наборі з кількістю куплених товарів кожним користувачем показано на рис. 3.8.

Змн.	Арк.	№ докум.	Підпис	Дата

customerId	productId	score	rank
1553	2	0.1110002875328064	1
1553	35	0.08345692157745362	2
1553	1	0.07202702760696411	3
1553	5	0.06954102516174317	4
1553	17	0.062017965316772464	5
1553	61	0.06159121990203857	6
1553	21	0.051773107051849364	7
1553	41	0.048556816577911374	8
1553	47	0.04801986217498779	9
1553	82	0.04561170339584351	10
20400	1	0.049841105937957764	1
20400	182	0.0481337308883667	2
20400	13	0.0420384407043457	3
20400	56	0.04202163219451904	4
20400	215	0.04182213544845581	5
20400	131	0.036169588565826416	6
20400	265	0.03536689281463623	7
20400	54	0.03426623344421387	8
20400	6	0.033963680267333984	9
20400	160	0.033889055252075195	10
19750	48	0.08908288677533467	1
19750	38	0.08239569266637166	2
19750	2	0.08164189259211223	3
19750	1	0.07568801442782085	4
19750	79	0.075407475233078	5
19750	36	0.06567970911661784	6
19750	21	0.05801957845687866	7
19750	5	0.048634866873423256	8
19750	14	0.04799354076385498	9
19750	33	0.04798781871795654	10

Рисунок 3.8 – Результат застосування коефіцієнта кореляції Пірсона на наборі purchase count

Для набору «користувач-товар» з бінарно промаркованими товарами та нормованого даних про кількість придбаних товарів кожним покупцем, результати формування рекомендацій виглядають, як показано на рис. 3.9 та рис. 3.10.

customerId	productId	score	rank
1553	2	0.11069482564926147	1
1553	1	0.07363476355870564	2
1553	5	0.06431001424789429	3
1553	21	0.06199751297632853	4
1553	17	0.0569839080174764	5
1553	61	0.05539397398630778	6
1553	29	0.05539158980051676	7
1553	105	0.053917884826660156	8
1553	15	0.05295040210088094	9
1553	143	0.051859756310780845	10
20400	17	0.0	1
20400	21	0.0	2
20400	13	0.0	3
20400	1	0.0	4
20400	76	0.0	5
20400	47	0.0	6
20400	101	0.0	7
20400	25	0.0	8
20400	174	0.0	9
20400	186	0.0	10

Рисунок 3.9 – Рекомендації для набору «користувач-товар» з бінарно промаркованими товарами

customerId	productId	score	rank
1553	18	0.0	1
1553	2	0.0	2
1553	225	0.0	3
1553	174	0.0	4
1553	9	0.0	5
1553	147	0.0	6
1553	78	0.0	7
1553	4	0.0	8
1553	270	0.0	9
1553	267	0.0	10
20400	18	0.0	1
20400	2	0.0	2
20400	225	0.0	3
20400	174	0.0	4
20400	9	0.0	5
20400	147	0.0	6
20400	78	0.0	7
20400	4	0.0	8
20400	270	0.0	9
20400	267	0.0	10

Рисунок 3.10 – Рекомендації для нормованого дата сету

3.3 Тестування та оцінювання якості моделей формування рекомендацій

Для оцінювання моделей формування рекомендацій можна використати наступні метрики:

- RMSE (середньоквадратичне відхилення);
- точність (precision)
- повнота (recall).

RMSE – метрика, що дозволяє кількісно виразити похибку прогнозованих значень. Чим менше значення цієї метрики, тим краще сформована рекомендація

Метрику повноти формування рекомендації можна розглянути на прикладі: якщо покупець придбав 5 товарів, а рекомендаційна система запропонувала 3 з них, тоді повнота становить 0,6.

Точність формування рекомендацій інтерпретується наступним прикладом: якщо покупцю сформована рекомендація з 5 товарів, з яких він купує 4, то точність дорівнює 0,8.

Розглянемо випадок, коли у рекомендації входять усі товари. Це означає, що пропозиція кожному покупцю безсумнівно буде включати ті товари, які їм сподобались та були частково придбані. У цьому випадку одержуємо 100% повноту, однак це не вказує на те, що модель є коректною. Потрібно враховувати критерій точності формування рекомендації: якщо сформовано пропозицію з 300 товарів, а покупець купує лише 3 з них, то точність становить тільки 0,1%, що є доволі низьким показником і свідчить про те, що модель не є хорошою. Отже, задача побудови хорошої моделі формування рекомендацій полягає у оптимізації як повноти, так і точності. При цьому чим ближче їхні значення до 1, тим кращою є модель.

Для використання метрик оцінювання рекомендаційних моделей спершу потрібно створити початкові змінні для оцінки моделі, що показано у лістингу 3.5.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						59
Змн.	Арк.	№ докум.	Підпис	Дата		

Лістинг 3.5 – Моделі оцінювання якості формування рекомендацій

```
models_w_counts = [popularity_model, cos, pear]
models_w_dummy = [pop_dummy, cos_dummy, pear_dummy]
models_w_norm = [pop_norm, cos_norm, pear_norm]
names_w_counts = ['Popularity Model on Purchase Counts', 'Cosine
Similarity on Purchase Counts', 'Pearson Similarity on Purchase
Counts']
names_w_dummy = ['Popularity Model on Purchase Dummy', 'Cosine
Similarity on Purchase Dummy', 'Pearson Similarity on Purchase
Dummy']
names_w_norm = ['Popularity Model on Scaled Purchase Counts',
'Cosine Similarity on Scaled Purchase Counts', 'Pearson
Similarity on Scaled Purchase Counts']
```

Далі потрібно порівняти значення одержаних метрик для різних моделей формування рекомендацій. Програмний код їхнього порівняння приведений у лістингу 3.6.

Лістинг 3.6 – Порівняння моделей за метриками якості формування рекомендацій

```
eval_counts = tc.recommender.util.compare_models(test_data,
models_w_counts, model_names=names_w_counts)
eval_dummy = tc.recommender.util.compare_models(test_data_dummy,
models_w_dummy, model_names=names_w_dummy)
eval_norm = tc.recommender.util.compare_models(test_data_norm,
models_w_norm, model_names=names_w_norm)
```

Результат порівняння моделей на основі метрики RMSE при застосуванні мір подібності косинуса кута і коефіцієнта кореляції Пірсона, а також на основі моделі популярності товарів при формуванні рекомендацій показано на рис. 3.11.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						60
Змн.	Арк.	№ докум.	Підпис	Дата		

формує лише однаковий список рекомендованих товарів кожному користувачеві. Застосування моделі популярності є доцільним при появі нового користувача (покупця), який ще не здійснював жодних покупок. Значення метрик точності і повноти приведені на рис. 3.12 для різних наборів даних: «Абсолютна кількість придбаних товарів», «Придбані-непридбані товари», «Нормалізований набір покупок». Однак, оскільки оцінка рекомендацій для нормалізованих даних про покупки дорівнює нулю і є константою, то потрібно обрати модель на наборі даних «Придбані-непридбані товари».

Насправді RMSE мало чим відрізняється у моделях на наборі «Придбані-непридбані товари» та «Нормалізований набір покупок». Оскільки, значення середньоквадратичного відхилення є більшим у випадку, якщо використовувати коефіцієнт кореляції Пірсона, ніж метрика косинуса кута, тому доцільно обрати модель з меншим значенням RMSE. В якості моделі формування рекомендацій на основі вподобань споживачів обрано модель, що використовує метрику косинуса кута на наборі «Придбані-непридбані товари».

Для управління процесом виводу рекомендацій з можливістю експорту даних у csv формат, а також реалізації функції, яка повертає список рекомендацій із значенням ідентифікатора клієнта необхідно реалізувати програмний код, наведений у лістингу 3.7. При цьому обґрунтовану модель потрібно застосувати для цілого вхідного набору даних.

Лістинг 3.7 – Застосування моделі для вхідного набору даних

```
final_model =  
tc.item_similarity_recommender.create(tc.SFrame(data_norm),  
                                     user_id=user_id,  
                                     item_id=item_id,  
                                     target='purchase_dummy',  
                                     similarity_type='cosine')  
recom = final_model.recommend(users=users_to_recommend, k=n_rec)  
recom.print_rows(n_display)
```

					КС КРБ 123.169.00.00 ПЗ	Арк.
						62
Змн.	Арк.	№ докум.	Підпис	Дата		

У результаті виконання коду лістингу 3.7, одержимо результат, як показано на рис. 3.13.

(10000, 4)

	customerId	productId	score	rank
0	1553	226	0.772044	1
1	1553	247	0.337553	2
2	1553	230	0.327010	3
3	1553	125	0.256670	4
4	1553	294	0.252761	5

Рисунок 3.18 – Результат застосування обраної моделі для всього набору даних (відображено рекомендації для покупця з ідентифікатором 1553)

Для зберігання рекомендацій на основі вподобань користувачів у бажаному файлі у файлі csv формату потрібно виконати код, наведений у лістингу 3.8.

Лістинг 3.8 – Функція формування вихідного формату файлу csv

```
def create_output(model, users_to_recommend, n_rec, print_csv=True):
    recommendation = model.recommend(users=users_to_recommend, k=n_rec)
    df_rec = recommendation.to_dataframe()
    df_rec['recommendedProducts'] = df_rec.groupby([user_id])[item_id]
    \
        .transform(lambda x: '|'.join(x.astype(str)))
    df_output = df_rec[['customerId', 'recommendedProducts']]
    .drop_duplicates() \
        .sort_values('customerId').set_index('customerId')
    if print_csv:
        df_output.to_csv('../output/option1_recommendation.csv')
        print("An output file can be found in 'output' folder with
name 'option1_recommendation.csv'")
    return df_output
```

Результат форматування вихідного файлу з рекомендаціями показано на рис. 3.14.

```
recommendations finished on 1000/1000 queries. users per second: 38980.3
An output file can be found in 'output' folder with name 'option1_recommendation.csv'
(1000, 1)
```

	recommendedProducts
customerid	
4	2 82 249 14 86 215 39 194 111 8
11	0 51 2 103 31 169 13 226 11 271
12	44 109 170 1 82 2 19 276 118 47
16	14 162 1 47 17 105 21 223 118 0
21	48 38 93 36 79 2 50 144 1 0

Рисунок 3.14 – Вихідний формат рекомендацій

Реалізація функції формування рекомендацій покупцю за вказаним ідентифікатором приведена у лістингу 3.9.

Лістинг 3.9 – Функція формування рекомендацій за ідентифікатором покупця

```
def customer_recomendation(customer_id):
    if customer_id not in df_output.index:
        print('Customer not found.')
        return customer_id
    return df_output.loc[customer_id]
```

Результат виконання функції лістингу 3.9 показаний на рис. 3.15.

```
customer_recomendation(4)
recommendedProducts    2|82|249|14|86|215|39|194|111|8
Name: 4, dtype: object

customer_recomendation(21)
recommendedProducts    48|38|93|36|79|2|50|144|1|0
Name: 21, dtype: object
```

Рисунок 3.15 – Результат виконання функції customer_recomendation ()

При практичній реалізації комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів використано підхід колаборативної фільтрації з метриками косинусна кута і коефіцієнта кореляції Пірсона. Виконано порівняння моделі з базовою моделлю на основі популярності товарів. Експерименти поставлено на трьох наборах даних, які включають «Абсолютну кількість придбаних товарів», «Придбані-непридбані товари», «Нормалізований набір покупок». Останній набір відображає нормалізовану частоту покупок як цільову змінну.

Використовуючи метрики RMSE, точності і повноти у роботі проведено оцінювання якості моделі та виявлено вплив персоналізації рекомендацій.

					<i>КС КРБ 123.169.00.00 ПЗ</i>	Арк.
						65
<i>Змн.</i>	<i>Арк.</i>	<i>№ докум.</i>	<i>Підпис</i>	<i>Дата</i>		

РОЗДІЛ 4 БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ОХОРОНИ ПРАЦІ

4.1 Вимоги до режимів праці і відпочинку при роботі з ВДТ

При організації праці, пов'язаної з використанням ВДТ ЕОМ і ПЕОМ, для збереження здоров'я працюючих, запобігання професійним захворюванням і підтримки працездатності передбачаються внутрішньозмінні регламентовані перерви для відпочинку.

Внутрішньозмінні режими праці і відпочинку містять додаткові нетривалі перерви в періоди, що передують появі об'єктивних і суб'єктивних ознак стомлення і зниження працездатності.

При виконанні робіт, що належать до різних видів трудової діяльності, за основну роботу з ВДТ вважається така, що займає не менше 50% робочого часу. Впродовж робочої зміни мають передбачатися:

- перерви для відпочинку і вживання їжі (обідні перерви);
- перерви для відпочинку і особистих потреб (згідно з трудовими нормами);
- додаткові перерви, що вводяться для окремих професій з урахуванням особливостей трудової діяльності.

За характером трудової діяльності розрізняють три професійні групи, згідно з діючим класифікатором професій (ДК-003-95 і Зміна N1 до ДК-003-95):

- розробники програм (інженери-програмісти) виконують роботу переважно з відеотерміналом та документацією при необхідності інтенсивного обміну інформацією з ЕОМ і високою частотою прийняття рішень. Робота характеризується інтенсивною розумовою творчою працею з підвищеним напруженням зору, концентрацією уваги на фоні нервово-емоційного напруження, вимушеною робочою позою, загальною гіподинамією,

					КС КРБ 123.169.00.00 ПЗ			
<i>Змн.</i>	<i>Арк.</i>	<i>№ докум.</i>	<i>Підпис</i>	<i>Дата</i>				
<i>Розроб.</i>		<i>Денисов Д.В.</i>			<i>Безпека життєдіяльності, основи охорони праці</i>	<i>Літ.</i>	<i>Арк.</i>	<i>Аркуші</i>
<i>Перевірів</i>		<i>Тиш Є.В.</i>					66	
<i>Консульт.</i>		<i>Пилипець М.І.</i>				<i>ТНТУ, каф. КС, гр. СІс-44</i>		
<i>Н. Контр.</i>		<i>Луцик Н.С.</i>						
<i>Затверд.</i>		<i>Осухівська Г.М.</i>						

періодичним навантаженням на кисті верхніх кінцівок. Робота виконується в режимі діалогу з ЕОМ у вільному темпі з періодичним пошуком помилок в умовах дефіциту часу;

– оператори електронно-обчислювальних машин виконують роботу, пов'язану з обліком інформації, одержаної з ВДТ за попереднім запитом, або тієї, що надходить з нього, супроводжується перервами різної тривалості, пов'язана з виконанням іншої роботи і характеризується напруженням зору, невеликими фізичними зусиллями, нервовим напруженням середнього ступеня та виконується у вільному темпі;

– оператор комп'ютерного набору виконує одноманітні за характером роботи з документацією та клавіатурою і нечастими нетривалими переключеннями погляду на екран дисплея, з введенням даних з високою швидкістю. Робота характеризується як фізична праця з підвищеним навантаженням на кисті верхніх кінцівок на фоні загальної гіподинамії з напруженням зору (фіксація зору переважно на документи), нервово-емоційним напруженням.

Правилами встановлюються такі внутрішньозмінні режими праці та відпочинку при роботі з ЕОМ при 8-годинній денній робочій зміні в залежності від характеру праці:

– для розробників програм із застосуванням ЕОМ слід призначати регламентовану перерву для відпочинку тривалістю 15 хвилин через кожен годину роботи за ВДТ;

– для операторів із застосуванням ЕОМ слід призначати регламентовані перерви для відпочинку тривалістю 15 хвилин через кожні дві години;

– для операторів комп'ютерного набору слід призначати регламентовані перерви для відпочинку тривалістю 10 хвилин після кожної години роботи за ВДТ.

У всіх випадках, коли виробничі обставини не дозволяють застосувати регламентовані перерви, тривалість безперервної роботи з ВДТ не повинна перевищувати 4 години.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						67
Змн.	Арк.	№ докум.	Підпис	Дата		

При 12-годинній робочій зміні регламентовані перерви повинні встановлюватися в перші 8 годин роботи аналогічно перервам при 8-годинній робочій зміні, а протягом останніх 4-х годин роботи, незалежно від характеру трудової діяльності, через кожен годину тривалістю 15 хвилин.

Для зниження нервово-емоційного напруження, втомлення зорового аналізатора, поліпшення мозкового кровообігу, подолання несприятливих наслідків гіподинамії, запобігання втомі доцільно деякі перерви використовувати для виконання комплексу вправ, які наведені у Державних санітарних правилах і нормах роботи з візуальними дисплейними терміналами електронно-обчислювальних машин ДСанПІН 3.3.2.007-98.

4.2 Вплив електромагнітних полів (ЕМП) на людину та заходи щодо зменшення їх впливу на обслуговуючий персонал

Сучасний рівень розвитку техніки означає високий рівень використання різноманітної радіоелектронної апаратури. В основному всю радіоапаратуру можна розділити на дві великі групи: засоби зв'язку і засоби обробки та зберігання даних в найрізноманітнішому вигляді.

Основну небезпеку для людини становить високочастотне електромагнітне випромінювання. Вплив випромінювання низької частоти вважається безпечним, хоча існують версії, що вплив випромінювання низької частоти проявляється через значно довший період, співрозмірний із тривалістю життя людини.

Основними джерелами електромагнітного випромінювання в електронно-обчислювальній техніці є генератори тактової частоти, сигнали яких використовуються для синхронізації роботи електронних схем обчислювальної техніки. Випромінювання цих пристроїв є малопотужним, хоча і високочастотним. Низький рівень випромінювання диктується тим, щоби мінімізувати взаємний вплив сигналів одних кіл на інші, а, отже, і мінімізувати спотворення.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						68
Змн.	Арк.	№ докум.	Підпис	Дата		

Більш потужний електромагнітний фон випромінюють блоки живлення обчислювальної техніки, оскільки більшість з них – імпульсні. Але, як правило, блоки живлення не розміщені в безпосередній близькості до користувача і тому не становлять особливої небезпеки для людини.

Системний блок комп'ютера не несе особливої небезпеки для життя та здоров'я користувача. Найбільшу небезпеку несе підвищене значення напруги, що подається з електромережі на блок живлення системного блоку. Несприятливий вплив на користувача може здійснювати шум, що створюється при роботі вентиляторів та накопичувачів системного блоку.

Електромагнітні випромінювання, які виникають при роботі електронних компонентів блоку мають незначні рівні, тому можуть створювати хіба що радіочастотні перешкоди.

Більшість електронних компонентів комп'ютера живляться напругою "+5 В", двигуни накопичувачів "+12 В", деякі пристрої "- 5 В" та "- 12 В". Такі значення напруги не несуть небезпеки для людини. В той же час на вхід блоку живлення подається напруга електромережі (220 В), яка є небезпечною для людини, з точки зору її можливого ураження. Тому до блоку живлення висувається низка вимог електробезпеки. Зокрема, електропроводи та кабелі повинні мати надійну ізоляцію, а на випадок короткого замикання чи інших аварійних режимів в електричній схемі блоку живлення повинні бути передбачені елементи захисту. Сам же він знаходиться в корпусі, який перекриває доступ до струмоведучих елементів блоку.

Дисплей є одним з основних блоків комп'ютера, що призначений для візуалізації інформації. Від його характеристик в значній мірі залежить працездатність та стан здоров'я користувача комп'ютера.

Концепція нормування електромагнітних полів і випромінювань передбачає:

- вироблення єдиної системи нормативних значень гранично допустимих рівнів електромагнітних полів і випромінювань;
- захист природних ресурсів від втрат, обумовлених дією цих полів на різні компоненти природного середовища;

					КС КРБ 123.169.00.00 ПЗ	Арк.
						69
Змн.	Арк.	№ докум.	Підпис	Дата		

– запобігання значним функціональним порушенням екосистем в результаті прямої або непрямой дії полів на ті або інші компоненти цих систем.

Монітор і ПЕОМ повинні забезпечувати потужність експозиційної дози рентгенівського випромінювання в будь-якій точці на відстані 0,05м від екрана і корпусу монітора при будь-яких положеннях регульовальних пристроїв не вище $7,74 \times 10$ А/кг, що відповідає еквівалентній дозі, рівної 0,1 мбер/година (100 мкр/годину).

Дисплеї, сконструйовані на основі ЕПТ, є джерелом кількох видів електромагнітного випромінювання та полів, а саме:

- іонізуючого (рентгенівського) випромінювання;
- оптичного випромінювання;
- електромагнітних випромінювань та полів.

Джерелом рентгенівського випромінювання дисплею є ЕПТ, зокрема люмінофорне покриття екрана. Електрони, які летять з великою швидкістю, різко сповільнюються шаром люмінофора. Це й спричинює рентгенівське випромінювання, енергія якого обмежується потенціалом розгону електронів. При цьому можлива поява лише, так званого "м'якого" рентгенівського випромінювання з довжиною хвилі 2 - 5 нм, яке ефективно поглинається скляним екраном.

Оптичне випромінювання виникає в результаті взаємодії електронів з шаром люмінофору. Дослідження показали, що в процесі роботи дисплея, окрім видимого випромінювання, мають місце і інші види оптичного-випромінювання (ультрафіолетове та інфрачервоне).

Електромагнітні випромінювання та поля різних діапазонів частот (високих, низьких та вкрай низьких) виникають в системах горизонтальної та вертикальної розгортки та в результаті дії електронного променя.

Основним джерелом статичних електричних полів є висока напруга (6 - 15 кВ), яка подається на блок анодів та внутрішню поверхню екрана ЕПТ.

Реальна інтенсивність, напруженість, рівень та інші параметри кожного виду електромагнітного випромінювання залежить від технічної конструкції конкретного дисплея, режимів його роботи, екранування та інших факторів.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						70
Змн.	Арк.	№ докум.	Підпис	Дата		

Центральним елементом лазерного принтера є експонуючий барабан. Його поверхня покрита, як правило, органічним світлочутливим непровідним матеріалом. Спочатку поверхня експонуючого барабана заряджається від'ємним зарядом. Далі в дію вступає лазерний промінь, що відтворює інформацію, отриману принтером з комп'ютера. В місцях, на які потрапив промінь лазера, відбувається зміна знаку електричного заряду. За допомогою валиків подачі і перенесення тонера, останній потрапляє на експонуючий барабан. Тонер прилипає лише в місцях, що були опромінені лазером. З касети до барабану подається папір. Транспортуючий валик притискає його до барабана, при цьому тонер переходить на папір.

Вузол фіксування складається, частіше за все, з двох барабанів, один із яких нагрівається лампою до температури близько 200 °С. Під дією температури тонер розплавляється і назавжди зв'язується з папером. Далі віддрукований аркуш подається у приймальний відділ принтера. Надлишковий тонер, що не перейшов на папір, забирається з барабана у вузлі очистки. В кінці циклу світлочутливе покриття барабана розряджається за допомогою лампи.

Таким чином, розглянувши роботу лазерного принтера, можна виділити такі його шкідливі та небезпечні фактори за порядком їх проявлення в розглянутому вище процесі друкування:

- електростатичні заряди;
- лазерний промінь;
- висока температура.

Для захисту людини від шкідливого впливу електромагнітних полів приймаються нормативи, стандарти та норми, які пов'язані із захистом людини від небезпечного впливу. Вони завжди представляють собою компроміс між перевагами використання нових технологій і техніки та можливим ризиком, спричиненим їхнім використанням.

ГОСТ 12.1.006-84 “Електромагнітні поля радіочастот” охоплює діапазон частот 60 кГц-300МГц. Він встановлює, що оцінка ЕМП в діапазоні 60 кГц-300МГц проводиться окремо з електричних і магнітних складових поля. Допустимі рівні протягом робочого дня по електричній складовій не повинні

					КС КРБ 123.169.00.00 ПЗ	Арк.
						71
Змн.	Арк.	№ докум.	Підпис	Дата		

перевищувати 50 В/м знижуючись ступенями 5 В/м на міру підвищення частоти. По магнітній складовій встановлені рівні тільки для окремих ділянок діапазону: 5 А/м для частот 60 кГц-1.5 МГц та 0,3 А для частот 30-50 МГц. Допускається перевищення цих стандартів, але не більше ніж двократно, при скороченні робочого дня не менш як на 50% [14].

Для частот 300 МГц-30 ГГц гранично допустимі значення щільності визначаються як результат ділення нормованої величини енергетичного навантаження за робочий день на час впливу. Енергетичне навантаження протягом робочого дня не повинно перевищувати 200 мкВт год./ см².

Електромагнітні поля сильно впливають на людський організм. Вони негативно впливають майже на усі системи організму. Тому треба створювати певні методи захисту від їх дії. Найпоширенішими методами є [14]:

- зменшення щільності потоку енергії, якщо дозволяє даний технологічний процес або обладнання;
- захист часом (тобто обмеження часу знаходження у зоні джерела ЕМП);
- захист відстанню;
- екранування робочого місця чи джерела;
- раціональне планування робочого місця;
- застосування засобів попереджувальної сигналізації;
- застосування засобів особистого захисту.

Для зменшення впливу електромагнітних полів на персонал, який знаходиться у зоні дії деяких радіоелектронних засобів необхідним є ряд захисних заходів: організаційні, інженерно-технічні та лікувально-профілактичні.

На етапі проектування системи взаємне розміщення об'єктів має бути забезпечено таким чином, щоб інтенсивність опромінення була мінімальною. Окрім цього, потрібно також зменшити час перебування персоналу у зоні опромінення. Потужність джерел випромінювання має бути найменшою з можливих.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						72
Змн.	Арк.	№ докум.	Підпис	Дата		

ВИСНОВКИ

При розробці комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів використано методи машинного навчання та загальнодоступний набір даних для проведення експериментальних досліджень.

Для досягнення мети роботи спочатку було створено технічне завдання на проектування системи і проведено його аналіз. Виявивши вимоги до комп'ютеризованої системи, далі було проведено аналіз методів щодо можливості автоматичного формування рекомендацій і встановлено, що ефективним інструментом є застосування методів колаборативної фільтрації.

Підхід на основі колаборативної фільтрації передбачає формування матриці крос-табуляції, що містить інформацію про покупця (ідентифікатор), товар і відповідні транзакції щодо здійснення покупки (кількість покупок), або кількісної оцінки, яка інтерпретує вподобання користувача.

У роботі спроектовано архітектуру комп'ютеризованої системи з можливістю інтеграції із системами електронної комерції. Основними програмними компонентами рекомендаційної системи є модуль препроцесингу даних та безпосередньо модуль формування рекомендацій.

Програмне забезпечення комп'ютеризованої системи реалізовано засобами мови програмування Python та за допомогою бібліотек:

- pandas та numpy для опрацювання даних;
- turicreate для проведення вибору та оцінки моделі;
- sklearn для розділення даних на навчальну і тестову вибірки.

Окрім цього, досліджено потенційні метрики для визначення подібності товарів і користувачів, і як наслідок обгрунтовано використання косинусної міри кута та коефіцієнта кореляції Пірсона, які, як показала практика, найбільше враховують вподобання споживачів. Результати експериментальних дослідження також наведені у кваліфікаційній роботі. Для кожного споживача було запропоновано ранжований список пропозицій з 5 товарів.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						73
Змн.	Арк.	№ докум.	Підпис	Дата		

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Фоусет Т., Провост Ф. Data Science для бизнеса: Як збирати, аналізувати і використовувати дані. Наш формат. Київ. 2019. 400 с.
2. К. О Нил, Шатт Р. Data Science. Инсайдерская информация для новичков. Включая язык R. Издательский дом "Питер". 2018. 368 с.
3. Рекомендательные системы. URL: <http://www.numberscompany.ru/products/recommenders> (дата звернення 20.03.2021 р.)
4. Setten M., Pokraev S., Koolwaaij J. Setten M. Context-aware recommendations in the mobile tourist application compass. Heidelberg, vol. 3137. 2004. pp. 515–548
5. Гомзин А., Коршунов А. Системы рекомендаций: обзор современных подходов. Труды ИСП РАН. 2012. URL: <http://cyberleninka.ru/article/n/sistemy-rekomendatsiy-obzorsovremennyh-podhodov> (дата звернення 27.03.2021 р.)
6. Савчук Т.О., Сакалюк А.В. Застосування кластерного аналізу для колаборативної фільтрації. Вісник Хмельницького національного університету. №1. 2011. С. 186-192.
7. Linden G., Smith B., York J. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing. vol. 7. No. 1. 2003. pp. 76–80.
8. Python-recsys on Github. URL: <https://github.com/ocelma/python-recsys> (дата звернення 15.04.2021 р.)
9. Рекомендательные системы: Часть 1. Введение в подходы и алгоритмы URL: <https://www.ibm.com/developerworks/ru/library/os-recommender1/> (дата звернення 22.04.2021 р.)
10. Preprocessing data. URL: <https://scikit-learn.org/stable/modules/preprocessing.html#preprocessing> (дата звернення 28.04.2021 р.)
11. API reference. URL: <https://pandas.pydata.org/docs/reference/index.html> (дата звернення 03.05.2021 р.)
12. NumPy Reference. URL: <https://numpy.org/doc/stable/reference/index.html> (дата звернення 08.05.2021 р.)

					КС КРБ 123.169.00.00 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		74

13. Turi Create API Documentation. . URL: <https://apple.github.io/turicreate/docs/api/> (дата звернення 15.05.2021 р.)

14. Барсегян А. Анализ данных и процессов. 3 изд. БХВ-Петербург. 2009. 512 с.

15. Breese J., Heckerman D., Kadie C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. Proc. 14th Conf. Uncertainty in Artificial Intelligence. 1998. pp. 223-234.

16. Adomavicius G. На пути к новому поколению рекомендационных систем: обзор имеющихся систем и возможные инновации. IEEE Transactions on Knowledge and Data Engineering. Vol. 17. No. 6. 2005. с. 78-86

17. Лексин В.А. Анализ клиентских сред: выявление скрытых профилей и оценивание сходства клиентов и ресурсов. Математические методы распознавания образов-13. М. МАКС Пресс. 2007. С. 488-491

18. Kurucz M., Benczur A., Csalogany K. Methods for large scale SVD with missing values. Proceedings of KDD Cup and Workshop. 2007. pp. 122-129.

19. ДСанПіН 3.3-2.007-98 Державні санітарні правила і норми роботи з візуальними дисплейними терміналами електронно-обчислювальних машин. - Київ, 1999. - 18с.

20. НПАОП 0.00-1.28-10 «Правила охорони праці під час експлуатації електронно-обчислювальних машин». – Київ, 2010. – 8 с.

					КС КРБ 123.169.00.00 ПЗ	Арк.
						75
Змн.	Арк.	№ докум.	Підпис	Дата		

Додаток А.
Технічне завдання

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії

Кафедра комп'ютерних систем та мереж

“Затверджую”

Завідувач кафедри КС

_____ Осухівська Г.М.

“ ___ ” _____ 2021 р

КОМП'ЮТЕРИЗОВАНА СИСТЕМА ФОРМУВАННЯ РЕКОМЕНДАЦІЙ НА
ОСНОВІ ВПОДОБАНЬ СПОЖИВАЧА

ТЕХНІЧНЕ ЗАВДАННЯ

на 11 листках

Вид робіт:

Кваліфікаційна робота

На здобуття освітнього ступеня «Бакалавр»

Спеціальність 123 «Комп'ютерна інженерія»

«УЗГОДЖЕНО»

Керівник кваліфікаційної роботи

_____ к.т.н., доц. Тиш Є.В.

« ___ » _____ 2021 р.

«ВИКОНАВЕЦЬ»

Студент групи СІс-44

_____ Денисов Д.В.

« ___ » _____ 2021 р.

Тернопіль 2021

1 Загальні відомості

1.1 Повна назва та її умовне позначення

Повна назва теми кваліфікаційної роботи: «Комп'ютеризована система формування рекомендацій на основі вподобань споживача».

Умовне позначення кваліфікаційної роботи: КС КРБ 123.169.00.00

1.2 Виконавець

Студент групи СІс-44, факультету комп'ютерно-інформаційних систем і програмної інженерії, кафедри комп'ютерних систем та мереж, Тернопільського національного технічного університету імені Івана Пулюя, Денисов Денис Васильович.

1.3 Підстава для виконання роботи

Підставою для виконання кваліфікаційної роботи є наказ по університету (№ 4.7-97 від 10.02.2021 р.)

1.4 Планові терміни початку та завершення роботи

Плановий термін початку виконання кваліфікаційної роботи – 10.02.2021 р.

Плановий термін завершення виконання кваліфікаційної роботи – 23.06.2021 р.

1.5 Порядок оформлення та пред'явлення результатів роботи

Порядок оформлення пояснювальної записки та графічного матеріалу здійснюється у відповідності до чинних норм та правил ІСО, ГОСТ, ЕСКД, ЕСПД та ДСТУ.

Пред'явлення проміжних результатів роботи з виконання кваліфікаційної роботи здійснюється у відповідності до графіку, затвердженого керівником роботи.

Попередній захист кваліфікаційної роботи відбувається при готовності роботи на 90% , наявності пояснювальної записки та графічного матеріалу.

Пред'явлення результатів кваліфікаційної роботи відбувається шляхом захисту на відповідному засіданні ЕК, ілюстрацією основних досягнень за допомогою графічного матеріалу.

2 Призначення і цілі створення системи

2.1 Призначення системи

Комп'ютеризована система формування рекомендацій на основі вподобань споживача призначена для підвищення продажів у системах онлайн комерції та оффлайн магазинів шляхом реалізації моделі цільової реклами.

Такі системи також є ефективним інструментом для стимулювання продажів в залежності сезону та кон'юнктури користувачів. Окрім цього, комп'ютеризовані системи такого класу є базисом для планування проведення акційних заходів та управління активами підприємства торгівлі.

Перевагами комп'ютеризованих систем формування рекомендацій на основі вподобань споживача є власне побудова конкретних цільових пропозицій споживачам на основі їхнього «кошика» товарів. Це дає змогу сформувати споживчий портрет на рівні окремого користувача та виділити кластери подібних товарів і покупців.

Для забезпечення ефективності рекомендаційних систем необхідний доволі великий набір даних і чим він більший, тим точніше формування рекомендацій для покупців та клієнтів закладів торгівлі. Комп'ютеризовані рекомендаційні системи дозволяють визначати набори товарів, які доповнюють один одного і у результаті цього можна одержати групи крос-продуктів.

Стимулювання продажу одного виду товару пов'язано із зміною таких факторів як ціна, сезонність, збільшення доходу за рахунок обороту товарів. Зважаючи на наведені сфери застосування комп'ютеризованих рекомендаційних систем можна виділити основні напрямки їхнього застосування:

- стимулювання продажів товарів;
- виявлення подібних товарів або їхніх груп;
- виявлення подібних споживачів та формування відповідних кластерів;
- підвищення ефективності крос-продаж.

2.2 Мета створення системи

Метою створення комп'ютеризованої системи формування рекомендацій на основі вподобань споживача є підвищення ефективності процесів продажу товарів за рахунок впровадження моделі цільової реклами для кінцевих покупців.

Для того, щоб досягти поставленої мети потрібно розв'язати такі задачі:

- аналіз підходів до побудови рекомендаційних систем;
- обґрунтування вибору мови програмування та відповідних середовищ реалізації рекомендаційної системи;
- підготовка та препроцесинг даних для моделювання ;
- визначення характеристик та екосистеми функціонування рекомендаційної системи;
- програмна реалізація моделі машинного навчання для формування пропозицій споживачу;
- обґрунтування та аналіз метрик для оцінки якості формування рекомендацій;
- визначення потенційних шляхів інтеграції рекомендаційної системи у наявну інформаційну інфраструктуру;
- проведення експериментальних досліджень щодо точності та стійкості формування рекомендацій.

2.3 Характеристика об'єкту

2.3.1 Основні задачі та функції об'єкту

До основних задач і функцій комп'ютеризованої системи, яка розробляється у кваліфікаційній роботі належить формування рекомендацій товарів для споживачів на основі їхніх вподобань. Вхідними даними при цьому є набір продукції, яка була придбана покупцем протягом певного періоду часу.

Основна задача, яку має розв'язати система, полягає у реалізації моделі на основі алгоритмів машинного навчання для формування пропозицій з наперед визначеною точністю та адекватністю. Окрім цього, сформовані рекомендації повинні бути ранжованими і кінцевому споживачу має відобразитися 5 пропозицій з найвищою точністю.

До непрямих функцій комп'ютеризованої системи належать:

- виявлення подібних товарів;
- групування товарів в однотипні кластери;
- визначення споживачів з подібними смаками;
- формування кластерів схожих покупців;
- створення базису для прийняття рішень щодо крос-продажів (cross-selling)

та керування процесом стимулювання продажів (up-selling).

При розробці комп'ютеризованої системи формування рекомендацій на основі вподобань споживача необхідно передбачити можливість інтеграції із зовнішніми системами, наприклад з платформами електронної комерції.

До складу системи повинно входити апаратне і програмне забезпечення. Апаратні пристрої системи включають сервер, на якому знаходяться дані про покупки споживачів, та клієнтських машин – з можливістю доступу до мережі інтернет. Характеристики сервера та комп'ютера для моделювання повинні володіти високими технічними характеристиками, що обумовлено необхідністю опрацювання великої кількості інформації та можливих алгоритмів перебору даних.

3 Вимоги до системи

3.1 Вимоги до системи в цілому

До вимог, які в цілому висуваються до комп'ютеризованої системи формування рекомендацій, належать вимоги щодо забезпечення функціональності зчитування даних про покупки споживачів, виявлення суміжних товарів та подібних покупців і їхніх кошиків. Як наслідок, на основі такої інформації потрібно сформувавши пропозиції продукції, яка або вже існує в кошику і користувач давно не здійснював її придбання, або ще не має запису про наявний товар з групи подібних покупців.

Комп'ютеризована система формування рекомендацій на основі споживчого кошику повинна використовувати на апаратному і програмному рівні архітектуру і технологію клієнт-сервер. Для реалізації системи доцільним є використання середовища програмування Jupyter Notebook, а також відповідних розширень, наприклад, платформи GraphLab. В якості мови програмування для реалізації моделі рекомендаційної системи може бути використана мова Python або R.

Перелік основних вимог, які висуваються до комп'ютеризованої системи виглядає таким чином:

- можливість одержання інформації про покупки споживачів з відповідної бази даних;
- можливість представлення даних у форматі csv;
- здатність фільтрації кошиків споживачів;
- можливість виявлення подібних товарів та споживачів;
- здатність ранжування результуючої вибірки подібних товарів;
- можливість формування 5-ти найкращих товарів для кожного користувача у наборі даних;
- здатність до співіснування та інтеграції із зовнішніми системами;
- можливість оцінювання якості сформованих рекомендацій на основі обґрунтованих метрик точності.

3.1.1 Вимоги до структури та функціонування системи

До структури комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів висуваються вимоги щодо наявності вхідного набору даних, який містить ідентифікатори покупців та кількості придбаних позицій товарів. Реалізація логіки формування пропозицій покладається на модель машинного навчання на основі алгоритму популярних товарів (popularity model) при появі нових покупців та колаборативної фільтрації (collaborative filtering) при наявності споживчого кошику покупця. Функціонування системи забезпечуються наявністю таких основних блоків:

- джерело даних – фрагмент бази даних, що містить інформацію про покупців і товари;
- інтелектуальний модуль формування персональних рекомендацій для кожного покупця.

Інтелектуальний модуль повинен використовувати такі бібліотеки як:

- pandas;
- numpy;
- turicreate;
- sklearn.

Комп'ютеризована система формування рекомендацій повинна забезпечувати сталість функціонування та результатів опрацювання даних і видавати ранжований список товарів для кожного конкретного споживача.

3.1.2 Вимоги до способів та засобів зв'язку між компонентами системи

Доступ до даних інтелектуального модуля формування пропозицій споживача повинен бути забезпечений або локально – при зберіганні даних та самого модуля на одному хості, або за допомогою доступу до мережі Інтернет із вказанням конкретної адреси сервера, на якому розміщено файл з даними у форматі csv.

При інтеграції інтелектуального модуля з платформою електронної комерції необхідно забезпечити авторизований доступ до відповідних ресурсів.

3.1.3 Вимоги по діагностуванню системи

Вимогами щодо діагностики комп'ютеризованої системи є:

- вимоги перевірки наявності зв'язку інтелектуального модуля з джерелом даних;
- вимоги коректності формування рекомендацій споживача у відповідності до його вподобань;
- здатність до виявлення та ремонтпридатності шляхом застосування системного логування подій.

При появі нового покупця, персональні рекомендації повинні бути сформовані на основі вибірки товарів, які найчастіше купують користувачі, інформація про кошики яких вже є відомою. В інших випадках – повинен виконуватись алгоритм колаборативної фільтрації. За наявності помилок при функціонуванні комп'ютеризованої системи повинні застосовуватись механізми та інструменти відновлення працездатності системи, і відповідно, її складові частини повинні відповідати характеристиці ремонтпридатності.

3.1.4 Перспективи розвитку, модернізація системи

Модернізація проектованої системи можлива лише за потреби інтеграції зовнішніх систем з особливими вимогами, або у випадку зміни структури джерела даних. Зміна джерела даних може бути пов'язана з додавання додаткових властивостей про товари або про споживачів, які дозволяють збільшити інформативність і точність формування персональних пропозицій.

До перспектив розвитку комп'ютеризованої системи входить побудова та інтеграція модулів з визначення додаткових прихованих властивостей товарів, виявлення сезонної складової і трендів, що в комплексі могли б забезпечити вищу точність і робастність алгоритмів машинного навчання.

3.1.5 Вимоги до надійності системи

До вимог надійності комп'ютеризованої системи, яка розробляється у кваліфікаційній роботі належать:

- час стабільного та безперервного функціонування на рівні 23 год./добу;
- здатність до відновлення працездатності після виникнення помилок та збоїв у роботі апаратного забезпечення;
- можливість забезпечення авторизованого доступу до інтелектуального модуля та джерела даних;
- можливість тимчасового відключення роботи інтелектуального сервісу;
- можливість реінтеграції системи із зовнішніми чи сторонніми сервісами;
- адекватність надання результатів на запит користувачів;
- можливість надійного розпаралелювання та здатності використовувати графічні ядра при формуванні рекомендацій.

3.1.6 Вимоги до функцій та задач, які виконує система

Вимогами до функцій і задач, які покликана виконувати комп'ютеризована система формування рекомендацій на основі вподобань споживачів є:

- здатність зчитування та фільтрації інформації з джерела даних;
- можливість реалізації «холодного» старту при формуванні рекомендацій;
- здатність формувати рекомендаційні на основі алгоритму колаборативної фільтрації;
- можливість визначення подібних товарів;
- можливість підвищення крос-продажів;
- можливість стимулювання росту продажів;
- здатність формувати групи подібних покупців;
- здатність забезпечувати і підтримувати цілісність бази даних;
- забезпечення точності формування рекомендацій на рівні не нижче, ніж 80%;
- можливість ранжування рекомендованих товарів.

3.1.7 Вимоги до апаратного забезпечення

Рекомендованими вимогами до апаратного забезпечення при проектуванні комп'ютеризованої системи формування рекомендацій на основі вподобань споживачів є наступні:

- процесор, що відповідає наступним технічним характеристикам: Intel Core i5 4300M з частотою 2,60 ГГц або 2,59 ГГц (1 сокет, 2 ядра, 2 потоки на ядро)

- оперативна пам'ять – 8 ГБ;

Альтернативними конфігураціями при реалізації рекомендаційної системи є:

- Xeon E5-2698 v3 з частотою 2,3 ГГц (2 сокети, 16 ядер кожен, 1 потік на Ядро, або Xeon Phi 7210 з частотою 1,3 ГГц (1 сокет, 64 ядра, 4 потоки на ядро);

- 32 ГБ або 64 ГБ об'єм оперативної пам'яті;

- накопичувач (жорсткий диск) з наявністю не менше, ніж 2-3 ГБ простору.

3.1.8 Вимоги до програмного забезпечення

Операційні системи для функціонування комп'ютеризованої системи можуть бути будь-якого типу (Windows, Linux, MacOS) однак повинні підтримувати мову програмування Python відповідної версії, а також перелік бібліотек, які описані у пункті структури та функціонування комп'ютеризованої системи.

4 Вимоги до документації

Документація повинна відповідати вимогам ЄСКД та ДСТУ

Комплект документації повинен складатись з:

- пояснювальної записки;

- графічного матеріалу:

1. Приклади застосування рекомендаційних систем та матриця крос-табуляції.

2. Класифікація методів колаборативної фільтрації при формуванні рекомендацій.

3. Метрики подібності на основі коефіцієнта Жаккара та косинуса кута

4. Метрика подібності на основі коефіцієнта кореляції Пірсона

5. Архітектура комп'ютеризованої системи формування рекомендацій
6. Результати оцінювання моделей формування рекомендацій та приклад пропозицій

5 Стадії та етапи проектування

Таблиця 1 – Стадії та етапи виконання кваліфікаційної роботи бакалавра

№ етапу	Назва етапу виконання кваліфікаційної роботи	Термін виконання
1.	Розробка та аналіз технічного завдання	10.02-19.02.2021
2.	Аналіз принципів організації рекомендаційних систем	23.02-05.03.2021
3.	Обґрунтування методів колаборативної фільтрації при формуванні рекомендацій	06.03-22.03.2021
4.	Метрики визначення подібності товарів і користувачів	23.03-28.03.2021
5.	Проектування архітектури комп'ютеризованої системи формування рекомендацій	29.03-12.04.2021
6.	Реалізація програмного забезпечення комп'ютеризованої системи	13.04-15.05.2021
7.	Розробка інструкцій з налаштування параметрів комп'ютеризованої системи	15.05-23.05.2021
8.	Безпека життєдіяльності, основи охорони праці	24.05-01.06.2021
9.	Оформлення кваліфікаційної роботи	02.06-12.06.2021
10.	Попередній захист кваліфікаційної роботи	13.06-18.06.2021

6 Додаткові умови виконання кваліфікаційної роботи

Під час виконання кваліфікаційної роботи у дане технічне завдання можуть вноситися зміни та доповнення.