

Міністерство освіти і науки України

Харківський національний університет радіоелектроніки

Факультет \_\_\_\_\_ Комп'ютерних наук \_\_\_\_\_  
(повна назва)

Кафедра \_\_\_\_\_ Інформаційних управляючих систем \_\_\_\_\_  
(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти \_\_\_\_\_ другий (магістерський) \_\_\_\_\_

\_\_\_\_\_ Дослідження гібридних методів формування рекомендацій у ІТ- проектах  
з розробки рекомендаційних систем \_\_\_\_\_  
(тема)

Виконав:

студент 2 курсу, групи УПГІТМ-20-1

\_\_\_\_\_ Потехіна Н. В. \_\_\_\_\_  
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки

(код і повна назва спеціальності)

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

Освітня програма Управління проектами в  
галузі інформаційних технологій  
(повна назва освітньої програми)

Керівник проф. Чала О.В. \_\_\_\_\_  
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри \_\_\_\_\_  
(підпис)

\_\_\_\_\_ Петров К.Е. \_\_\_\_\_  
(прізвище, ініціали)

2021р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук  
(повна назва)

Кафедра Інформаційних управляючих систем  
(повна назва)

Рівень вищої освіти другий (магістерський)

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

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

Освітня програма Управління проектами в галузі інформаційних технологій  
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_  
(підпис)

« \_\_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ р.

**ЗАВДАННЯ**  
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Потехіній Наталії Вікторівні  
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження гібридних методів формування рекомендацій у ІТ- проектах з розробки рекомендаційних систем

затверджена наказом по університету від 05.11 2021 р. № 1646 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 13 12 2021 р.

3. Вихідні дані до роботи Науково-технічні публікації та інтернет джерела з тематики кваліфікаційної роботи

4. Перелік питань, що потрібно опрацювати в роботі вступ, – дослідження характеристик ІТ- проектів рекомендаційних систем; аналіз методів побудови рекомендацій; дослідження проблем побудови рекомендацій при використанні методів фільтрації; дослідження гібридних методів побудови рекомендацій; удосконалення методу гібридного методу побудови рекомендацій для побудови рекомендацій в режимі онлайн; планування проекту розробки гібридного методу побудови рекомендацій; експериментальна перевірка удосконаленого гібридного методу побудови рекомендацій в режимі онлайн у проектах рекомендаційних систем висновки.

## КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз літератури та Інтернет-джерел	08.11.2021	
2	Постановка задачі	09.11.2021-11.11.2021	
3	Обробка матеріалу	12.11.2021-13.11.2021	
4	Дослідження іт- проектів з розробки рекомендаційних систем	14.11.2021-15.11.2021	
5	Дослідження методів гібридних метод формування рекомендацій в онлайн-режимі	15.11.2021-16.11.2021	
6	Проект розробки методу гібридної фільтрації у іт-проектах з розробки рекомендаційних систем	17.11.2021-18.11.2021	
7	Апробація результатів дослідження на прикладі	19.11.2021-20.11.2021	
8	Написання пояснювальної записки	21.11.2021-28.11.2021	
9	Підготовка презентації	29.11.2021-02.12.2021	
10	Перевірка на плагіат	03.12.2021-06.12.2021	
11	Нормоконтроль	07.12.2021-10.12.2021	
12	Захист	14.12.2021	

Дата видачі завдання 08.11 2021 р.

Студент \_\_\_\_\_  
(підпис)

Керівник роботи \_\_\_\_\_ проф. Чала О.В.  
(підпис) (посада, прізвище, ініціали)

## РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи містить: 83 с., 4 розділи, 9 рис., 6 табл., 31 джерело.

### ГІБРИДНІ МЕТОДИ, КОЛАБОРАТИВНА ФІЛЬТРАЦІЯ, РЕКОМЕНДАЦІЙНІ СИСТЕМИ, ФОРМУВАННЯ РЕКОМЕНДАЦІЙ

У роботі виконано огляд методів формування рекомендацій у ІТ проектах рекомендаційних систем. Проаналізовано існуючі гібридні методи побудови рекомендацій. За результатами виконаного аналізу методів та підходів до побудови рекомендацій запропоновано удосконалений гібридний метод формування рекомендацій у ІТ проектах рекомендаційних систем з урахуванням змін у вимогах користувача.

В ході виконаного дослідження отримані такі результати: досліджено характеристик ІТ- проектів рекомендаційних систем; виконано аналіз методів побудови рекомендацій; досліджено проблеми побудови рекомендацій при використанні методів фільтрації; досліджено гібридні методи побудови рекомендацій; удосконалено гібридний метод побудови рекомендацій для роботи в режимі онлайн; виконано планування проекту розробки гібридного методу побудови рекомендацій; виконано експериментальну перевірку удосконаленого гібридного методу побудови рекомендацій в режимі онлайн у проектах рекомендаційних систем.

## **ABSTRACT**

Explanatory note to master qualification work contains 83 pages, 4 sections, 9 pictures, 6 tables, 31 sources.

**COLLABORATIVE FILTRATION, RECOMMENDATION SYSTEMS, FORMATION OF RECOMMENDATIONS, HYBRID METHODS,**

The paper reviews the methods of forming recommendations in IT projects of recommendation systems. The existing hybrid methods of constructing recommendations are analyzed. Based on the results of the analysis of methods and approaches to the construction of recommendations, an improved hybrid method of forming recommendations in IT projects of recommendation systems, considering changes in user requirements, is proposed.

The following results were obtained during the research: the characteristics of IT projects of recommendation systems were studied; the analysis of methods of construction of recommendations is executed; problems of construction of recommendations at use of filtration methods are investigated; hybrid methods of constructing recommendations have been studied; improved hybrid method of building recommendations for online work; the planning of the project of development of a hybrid method of construction of recommendations is executed; an experimental test of an improved hybrid method of constructing recommendations online in the projects of recommendation systems was performed.

## **ПЕРЕЛІК СКОРОЧЕНЬ, УСЛОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ ТА ТЕРМІНІВ**

ДМ – демографічний метод;

КБ – побудова рекомендацій на основі використання баз знань;

КН – фільтрація на основі контенту;

КФ – колаборативна фільтрація;

РС – рекомендаційна система;

ПЗ – програмне забезпечення.

## ЗМІСТ

Вступ.....	8
1 Аналіз ІТ- проектів з розробки рекомендаційних систем .....	10
1.1 Аналіз структури систем побудови рекомендацій .....	10
1.2 Аналіз методів побудови рекомендацій.....	15
1.3 Дослідження проблем побудови рекомендацій при використанні методів фільтрації .....	23
1.4 Дослідження гібридних методів побудови рекомендацій.....	28
1.5 Постановка задачі дослідження .....	35
2 Гібридний метод формування рекомендацій в онлайн-режимі .....	37
2.1 Підхід до побудови рекомендацій з урахуванням змін у вимогах користувачів .....	37
2.2 Удосконалений гібридний метод побудови рекомендацій у режимі онлайн з урахуванням змін вимог користувачів .....	41
3 Проект розробки методу гібридної фільтрації у іт- проектах з розробки рекомендаційних систем .....	47
3.1 Опис проекту розробки гібридного методу.....	47
3.2 Статут проекту.....	48
3.3 Планування проекту.....	50
4 Практичне застосування отриманих результатів.....	57
4.1 Схема експериментальної перевірки удосконаленого методу побудови рекомендацій.....	57
4.2 Результати експериментальної перевірки удосконаленого методу....	58
Висновки.....	62
Перелік джерел посилання.....	63
Додаток А Графічний матеріал.....	67

## ВСТУП

Рекомендаційні системи персоналізують контент, що надається користувачеві. Вони враховують обмеженість знань користувачів та спрощують їх вибір у складному інформаційному середовищі. Найбільшого поширення такі системи набули у сфері електронної комерції. Вказані системи використовуються для підтримки вибору споживача в Інтернет-магазинах, системах бронювання готелів, системах продажу потокового відео.

У цій галузі вони формують рейтинговий список товарів та послуг з урахуванням прогнозованих інтересів користувачів. Цей список містить обмежений набір елементів, упорядкованих для зменшення очікуваної корисності для споживача.

Для побудови рекомендацій використовуються методи колаборативної фільтрації, фільтрації на основі контенту, на основі демографічних даних, а також фільтрації з використанням знань про предметну область.

Для підвищення ефективності побудови рекомендацій використовують гібридні методи. Проведений аналіз показав, що гібридні методи складаються з комбінації традиційних методів побудови рекомендацій. За рахунок такого комбінування вхідні дані фільтруються в два етапи. На першому етапі відбирають множину кандидатів, а на другому формують список об'єктів. Такі рекомендації будуються в онлайн та офлайн-режимах. В онлайн-режимі враховуються актуальні дані про користувачів.

Такі рекомендації мають менші вимоги до точності, але вони мають бути швидко адаптовані до поточних потреб користувача. Тому при удосконаленні гібридного методу побудови рекомендацій для роботи в режимі онлайн доцільно враховувати зміни у виборі користувачів з часом.

Об'єктом дослідження в рамках магістерської кваліфікаційної роботи є процес формування рекомендацій у IT- проектах рекомендаційних систем.

Предметом дослідження є методи формування рекомендацій пояснень у проектах рекомендаційних систем.

Метою даної роботи є дослідження гібридних методів формування рекомендацій для побудови рекомендацій в онлайн-режимі.

# 1 АНАЛІЗ ІТ- ПРОЕКТІВ З РОЗРОБКИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ

## 1.1 Аналіз структури систем побудови рекомендацій

Рекомендаційна система допомагає користувачам вибрати ті продукти, які їм персонально підходять. Система рекомендацій, по суті, узгоджує інформацію користувача з інформацією про предмет, яка може його зацікавити [1]. Рекомендаційна система використовує два базових сценарії побудови рекомендацій:

- рекомендації на основі запитів користувача;
- потокові рекомендації на основі властивостей користувача та відповідного елемента (продукту, інформації).

Сценарій побудови рекомендацій на основі запитів включає пошук відповідності уподобань користувачів щодо покупки та властивостей товару.

Наприклад, якщо користувач шукає маску, відображені в рекомендації елементи мають бути пов'язані з масками. Проте елементів, пов'язаних з масками, може бути дуже багато. Рекомендаційна система необхідна, щоб визначити їх рейтинг, та упорядкувати рекомендацію за рейтингом. Системі потрібно ранжувати товари відповідно до властивостей користувачів, наприклад таких, як їхні улюблені кольори та цінові переваги. Якщо користувач вподобує товари, що відносяться до предметів розкоші, система обов'язково поставить дорогі маски на перше місце для цього користувача. Якщо користувач чутливий до цін, рекомендаційній системі може знадобитися поставити дешевші маски з більш низькою вартістю на перше місце для цього користувача.

Потокові рекомендації все частіше стають основним режимом взаємодії між багатьма програмами та їхніми користувачами. Зокрема, стрічки новин на домашніх сторінках можуть мати рекомендації відповідно до щоденних уподобань користувача. Наприклад, якщо користувач любить

новини, пов'язані з баскетболом, то на сайті новин йому можуть порекомендувати більше спортивного контенту. Рекомендаційна модель для потокових рекомендацій використовує машинне навчання. Очікується, що модель дізнається як про налаштування користувача, так і про властивості елемента [2].

Узагальнений процес побудови рекомендацій зображено на рис.1.1.



Рисунок 1.1 – Процес побудови рекомендацій

Розглянемо приклад використання рекомендаційної системи для платформи новин. Користувач А відвідує платформу. Ця платформа містить тисячі новин, які називаються елементами або об'єктами. Кожен елемент має ідентифікатор. Рекомендаційна система фільтрує елементи для користувача А із загальної кількості в десятки тисяч елементів. Типова система рекомендацій, заснована на збігу та рейтингу, зазвичай має дві підсистеми.

Перша – це підсистема побудови моделі, яка співставляє користувачів та елементи. Друга підсистема ранжує результати співставлення та визначає рейтинг елементів.

Перша підсистема виконує попередню фільтрацію тисяч елементів, щоб вибрати ті елементи, які можуть сподобатися користувачеві А. Наприклад, якщо до короткого списку після фільтрації потрапили 1000 позицій, відомо лише, що користувачу А можуть сподобатися 1000 елементів. Але невідомо, які з них є улюбленими для користувача А. Друга підсистема ранжує цю 1000 позицій на основі уподобань користувача А, щоб створити остаточний список елементів, які будуть доставлені цьому користувачеві. Тому в рекомендаційній системі перша підсистема забезпечує попередню фільтрацію для визначення загального обсягу рекомендованих елементів. Це прискорює ранжування елементів на основі властивостей підсистемою рейтингування та робить зворотний зв'язок відносно рекомендацій більш ефективним для користувачів.

Професійна система рекомендацій повинна бути здатною надавати зворотній зв'язок щодо рекомендацій протягом десятків мілісекунд після отримання запиту користувача. Оновлення потокових рекомендацій також може тривати десятки мілісекунд, а потім система повинна негайно показати нові рекомендовані елементи [3].

При проектуванні рекомендаційних систем враховують такі вимоги.

По-перше, цільовий клієнт повинен мати додаток, який містить інформацію про активності користувачів та їх вподобання протягом місяця і повинен рекомендувати елементи користувачам. У кожній моделі навчання загальна кількість вибірок для навчання може містити мільйони записів. Для рекомендацій потрібно побудувати загальну модель на основі даних усієї платформи за останній місяць або навіть півроку. У машинному навчанні чим більше об'єм даних, тим точніша модель.

Дані для побудови рекомендацій можна розділити на три типи [4]:

– дані про поведінку користувача;

- дані про поведінку елемента;
- дані взаємодії користувача з елементом.

По-друге, архітектура рекомендаційної системи повинна забезпечувати можливість розгорнути алгоритми як плагіни. Сфера машинного навчання, включаючи побудову рекомендацій, швидко розвивається. Щороку з'являються нові рекомендаційні алгоритми. Тому ці алгоритми потрібно підключати або відключати від усієї системи гнучко. Це демонструє надійність системи, включаючи її здатність підтримувати компонентні алгоритми.

По-третє, продуктивність рекомендаційної системи має бути достатньо високою, щоб забезпечити зворотній зв'язок протягом мілісекунд для кожного запиту.

В-четвертих, архітектура повинна підтримувати еластичне масштабування ресурсів. Наприклад, деякі програми можуть частіше використовуватися під час вечірніх годин пік, коли користувачі повертаються додому з роботи, але рідше в ранкові години. Тоді рекомендаційна модель вимагає більше основних ресурсів у години пік і менше ресурсів у ранкові години. Тому архітектура повинна підтримувати гнучке масштабування базових ресурсів, щоб збалансувати витрати. Хмарні сервіси можуть задовольнити цю вимогу, оскільки вони можуть використовувати еластичні ресурси [5].

Загальну архітектуру рекомендаційної системи наведено на рис. 1.2. Архітектура містить базовий рівень даних. Цей шар містить дані профілю користувача, дані про елементи, дані про поведінку та дані коментарів.

Даними профілю користувача можуть бути зріст і вага користувачів, предмети, які вони придбали, їхні переваги щодо покупок або їх освіта.

Дані про товар – це ціни, кольори та походження товарів. Якщо елемент є відео, дані елемента – це інформація про відео, як-от вміст і теги.

Дані про поведінку стосуються взаємодії між користувачами та елементами. Наприклад, коли користувач дивиться відео, він може поставити

лайк до відео, додати відео до вибраного або заплатити за відео. Ці дії є даними про поведінку користувача [6].

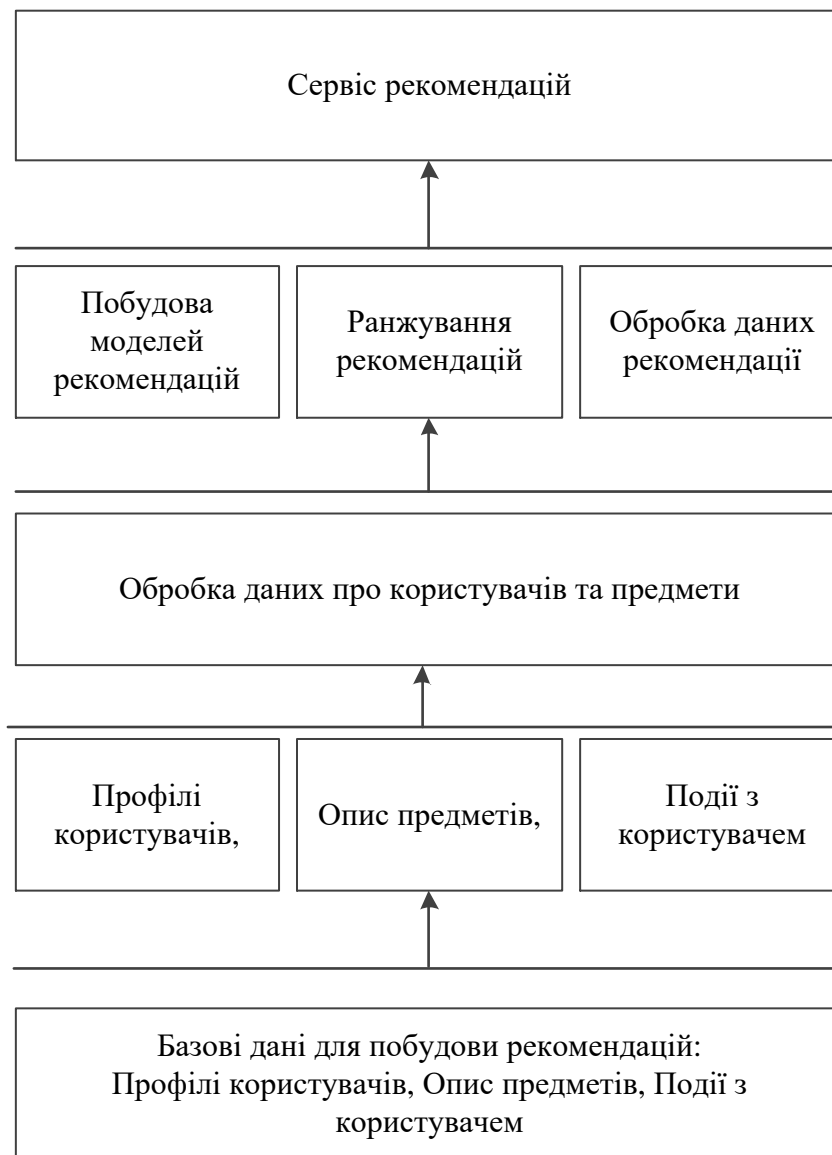


Рисунок 1.2 – Структура комерційної рекомендаційної системи

Дані коментарів можуть включати дані третіх сторін, і може бути доступним не для кожного елемента на кожній платформі.

Ключовими є дані користувача, дані елемента та дані про поведінку.

На рівні обробки та зберігання даних виконується обробка даних, наприклад, ідентифікувати функції користувача, особливості матеріалу та особливості подій. Далі виконується моделювання на основі цих функцій.

Кілька алгоритмів можуть працювати паралельно в модулі відповідності. Наприклад, визначення користувацьких функцій, матеріальних особливостей та особливостей події. Далі буде моделювання на основі цих функцій. Як ми зазначали вище в попередньому розділі, весь процес рекомендацій містить два важливі модулі: відповідність і рейтинг. Кілька алгоритмів можуть працювати паралельно в підсистемі побудови моделі відповідності. Наприклад, визначення користувацьких функцій, матеріальних особливостей та особливостей події. Далі виконується моделювання на основі цих функцій.

Після встановлення відповідності відбувається рейтингування. Виконується видалення дублікатів у рекомендації, A/B-тестування результатів.

На верхньому рівні – служба видачі рекомендацій, яка може рекомендувати рекламу, продукт або користувача. Наприклад, програма соціальної мережі може рекомендувати користувачам дозволити їм стежити один за одним. Реалізація архітектури виконується у хмарі, щоб система відповідала чотирьом основним вимогам до рекомендаційної системи на рівні підприємства.

## 1.2 Аналіз методів побудови рекомендацій

Кінцева мета будь-якої користувача системи, що адаптується до користувача, полягає в тому, щоб надати користувачам інформацію щодо потрібних їм предметів, не запитуючи їх явно. На відміну від традиційного налаштування інформаційних систем, рекомендаційна система реалізує принцип персоналізації. Різниця базується на тому, хто керує створенням профілів користувачів, а також представленням елементів інтерфейсу користувачеві. Під час налаштування користувачі зазвичай керують своїми уподобаннями або вимогами вручну.

При персоналізації профілі користувачів створюються оновлюються рекомендаційною системою автоматично з мінімальним контролем з боку користувача. Такий підхід зменшує кількість часу, який користувач витрачає на пошук необхідних предметів. Процес персоналізації складається з трьох етапів [7]:

- підготовка та перетворення даних;
- виявлення шаблонів у даних;
- побудова рекомендацій.

Рекомендаційні системи виконують фазу виявлення шаблонів у режимі онлайн. На етапі підготовки даних формується інформація про кліки, які можна обробляти за допомогою методів аналізу даних.

Методи, які використовуються для інтелектуального аналізу даних (наприклад, кластеризація, аналіз правил асоціацій та послідовне виявлення шаблонів) застосовуються на етапі виявлення шаблону [8].

Рекомендаційний механізм враховує поточний сеанс користувача та виявлені шаблони для забезпечення персоналізованого контенту. Персоналізований вміст може мати форму рекомендованих посилань, переліку продуктів, цільової реклами. Процес навчання рекомендаційних систем будується на основі використання даних в пам'яті (з відкладеним навчанням) і на основі моделі. У моделях також дані з пам'яті використовуються під час розрахунку рекомендацій. Системи першого типу чутливі до проблем масштабованості. Системи на основі моделі більш масштабовані у великих обсягах даних.

Методи побудови рекомендацій поділяються на групи на основі типу вхідних даних, підходів до створення профілів користувачів і алгоритмічних методів, які використовуються для створення рекомендацій:

- методи побудови рекомендацій на основі знань;
- методи побудови рекомендацій на основі контенту;
- колаборативна фільтрація;
- демографічна фільтрація.

Методи побудови рекомендацій на основі знань використовують правила, які формуються вручну або автоматично з профілів користувачів. Ці методи виявляють фактори, які впливають на вибір товару чи продукту. Багато з існуючих сервісів для електронної комерції використовують системи рекомендацій, засновані на правилах вручну. Ці системи дозволяють адміністраторам сайту встановлювати правила на основі статистичної, психологічної та демографічної інформації про користувачів. У деяких випадках правила дуже залежать від домену та суперечать бізнес-цілям веб-сайту. Ці правила використовуються для покращення вмісту, наданого користувачеві, якщо його профіль відповідає хоча б одній із умов.

Методи побудови рекомендацій на основі правил суттєво залежать від інженерних здібностей розробників систем, що створюють базу правил для специфічних характеристик домену та ринку. Профілі користувачів зазвичай поповнюються шляхом явної взаємодії з користувачами [9].

Ряд робіт досліджував методи навчання для класифікації користувачів на різні групи на основі їх статистичної інформації, а потім визначення необхідних правил для рекомендацій. Ці методи спрямовані на вилучення персоналізованих правил для кожного користувача. Загальний механізм полягає в тому, що користувач оголошує системі свої інтереси, а потім система оцінює кожен із наявних елементів для кожного користувача на основі наявної у нього бази знань. Однією з переваг таких методів є можливість користувачів висловлювати характеристики своїх улюблених предметів. Однією з проблем використання цих методів, на додаток до обмежень інженерії знань, є проблема створення профілів користувачів. Вхідними даними для цих систем є пояснення користувачів щодо їхніх особистих інтересів, і, як наслідок, це упереджені дані. Профілі в цих системах зазвичай статичні, і, отже, продуктивність систем погіршується через деякий час внаслідок старіння профілів користувачів.

Методи рекомендацій на основі контенту (контентна фільтрація) надають користувачам рекомендації на основі порівняння характеристик

товарів або продуктів з предметами, до яких користувач виявив інтерес. Профіль користувача в цих системах представляє пояснення щодо характеристик продукту, які користувач вибрав раніше. Ці пояснення ілюструються набором характеристик або функцій, що описують продукти в профілі користувача. Створення рекомендацій включає порівняння характеристик елементів, які користувач не бачив або не оцінив, з описом вмісту профілю користувача. Користувачеві рекомендуються елементи, які досить схожі на профіль користувача [10].

Системи для формування рекомендацій на основі вмісту зазвичай покладаються на методи аналізу інформації, зокрема класифікацію, кластеризацію та аналіз тексту. У більшості рекомендаційних систем на основі вмісту, особливо у веб-системах та системах електронної комерції, описи вмісту є текстовими функціями, витягнутими з веб-сторінок або описів продуктів. Зазвичай ці системи покладаються на відомі підходи до моделювання документів, які ґрунтуються на пошуку інформації та повторному пошуку інформації. Профілі та елементи користувачів можуть бути показані як зважені вектори слів (наприклад, на основі моделі зважування  $tf.idf$ ). Передбачити інтерес користувача до певного елемента можна на основі обчислення векторної подібності (наприклад, косинусної міри подібності) між вектором профілю користувача та вектором профілю елемента або на основі ймовірнісних методів (наприклад, байєсівських класифікаторів). Крім того, незважаючи на спільні методи фільтрації, профілі користувачів створюються окремо, лише на основі елементів, які бачив або оцінив сам користувач [11]. Однією з проблем методів побудови рекомендацій основі контенту, пов'язаних з попередніми оцінками та інтересами користувачів, є тенденція до уточнення у виборі елементів. Однак дослідження інтересів користувачів показують, що користувачів, як правило, більше цікавлять нові й дивовижні предмети, запропоновані рекомендаційними системами. Крім того, тут не розглядаються практичні зв'язки між предметами, такі як їх спільне використання або доповнення для

виконання конкретного завдання. Інша проблема полягає в тому, що деякі елементи опису контенту не можуть бути представлені за допомогою спеціальних функцій, таких як текстові, тому вони не будуть доступні при використанні таких методів. Колаборативна фільтрація вирішує проблем, що повстають перед методами побудови рекомендацій на основі правил і контенту. Рекомендаційні системи, засновані на фільтрації, досягли прийняттого успіху на сайтах електронної комерції. Використовуються відповідні оцінки поточного користувача (наприклад, рейтинги книг чи фільмів) з подібними користувачами (близькими сусідами), щоб рекомендувати предмети, які ще не бачив/не оцінив цей користувач. У стандартному випадку ці методи використовують данні з пам'яті [12]. Традиційно колаборативна фільтрація використовувала стандартний підхід класифікації на основі використання даних у пам'яті, заснований на методі  $k$ -найближчого сусіда (kNN). У цьому алгоритмі цільовий профіль користувача порівнюється з іншими профілями користувача, щоб визначити перших  $k$  користувачів, які мають схожі інтереси з цим користувачем. Прогнозований рейтинг активного користувача  $u$  для кожного елемента  $j$  розраховується як зважена сума рейтингів подібних користувачів по тому самому елементу:

$$p_{u,i} = \bar{r}_u + \alpha \sum_i w(u,i)(r_{i,j} - \bar{r}_i), \quad (1.1)$$

де  $u$  – активний користувач;

$r_{i,j}$  – оцінки користувача  $i$  для предмету  $j$ ;

$\bar{r}_i$  – середній рейтинг користувача  $i$ ;

$w(u,i)$  – вага користувача  $i$ .

Значення ваги обчислюються шляхом обчислення косинусної подібності, евклідової подібності або кореляції Пірсона для профілів схожих користувачів.

Найважливішою проблемою традиційної колаборативної фільтрації на основі пам'яті є те, що вона не піддається масштабуванню. В алгоритмі kNN формування сусідів має здійснюватися онлайн-методом. Іншими словами, на відміну від методів, заснованих на моделі, в яких фаза навчання моделі виконується на основі даних навчання, фаза моделювання в цих системах виконується як онлайн-завдання. Зі збільшенням кількості користувачів і елементів цей метод може бути неприпустимо повільним для вироблення динамічних рекомендацій під час взаємодії з користувачами [13].

Інша проблема пов'язана з розрідженістю більшості наборів даних. Більше елементів у наборі даних призводить до зменшення щільності профілю користувача. Як наслідок, зменшується ймовірність подібності побачених елементів серед користувачів, що призводить до меншої впевненості в кореляційних розрахунках. Колаборативна фільтрація працює найкраще, коли є явні небінарні оцінки для елементів, тоді як для багатьох веб-сайтів це не так. На деяких веб-сайтах збирати інформацію про користувача для персоналізації легше, використовуючи відвідані сторінки чи продукти, або запитуючи інформацію про продукт чи зміни в кошику для покупок. Ці джерела інформації розглядаються як неявний зворотний зв'язок.

Цей метод також вразливий щодо виникнення проблеми холодного старту, зокрема пов'язаної з використанням нових предметів, коли до набору товарів додається новий предмет або продукт, який не має оцінок. Як результат, він не існує в жодному профілі користувача, і рекомендаційна система не може рекомендувати його жодному користувачеві. Іншою проблемою цих систем є відсутність можливості пояснювати рекомендації користувачам. Оскільки колаборативна фільтрація не використовує інші інформаційні ресурси, такі як вміст або семантичні дані, вона не може пояснити причину рекомендації певного елемента користувачеві [14].

Щоб вирішити проблеми розрідженості та масштабованості, використовуються методи оптимізації, в тому числі методи зменшення розмірності, індексацію схожості та звичайну кластеризацію профілю

користувача в минулому для пошуку у відповідному кластері під час створення рекомендацій.

Іншим методом, який базується на колаборативній фільтрації, є спільна фільтрація на основі елементів. У цьому методі матриця подібності елементів створюється на основі рейтингових даних профілів користувачів офсетним способом. Ця матриця використовується для генерування повторних рекомендацій на етапі онлайн. Іншими словами, замість того, щоб покладатися на схожість між елементами в їх описах вмісту, вона розраховується на основі оцінок користувачів. Кожен елемент відображається у вигляді вектора, а подібність розраховується на основі таких показників, як косинусна подібність, або на основі подібності на основі кореляції, як кореляція Пірсона чи Спірмена. Процес генерування рекомендацій формує оцінку цільового користувача для цільового елемента за зваженою сумою наданих оцінок до подібних елементів. Аналогічно порівнюються профілі елементів.

Більшість методів аналізу даних для персоналізації є розширенням колаборативної фільтрації. У цих методах використовуються попередні профілі або оцінки користувачів як вхідні дані та генерується агрегована модель користувачів. Таку модель можна використовувати з поточним профілем користувача для створення рекомендацій або прогнозування поведінки користувачів у майбутньому [15].

Метод демографічної фільтрації класифікує користувачів на основі особистих атрибутів. Користувачі поділяються на демографічні класи за особистими ознаками. Надалі даються рекомендації на основі отриманих демографічних класів. Ці класи служать вхідними даними для процесу рекомендацій. Мета даного методу полягає в тому, щоб знайти класи людей, яким подобається певний продукт. Якщо людям з класу  $C$  подобається товар  $s$  і є користувач  $c$  із класу  $C$ , який ще не бачив продукт  $s$ , то цей продукт можна рекомендувати людині  $c$ . Клієнти надають персональні дані через опитування, які вони заповнюють під час реєстрації або можуть бути

вилучені з історії покупок користувачів. Перевага демографічного методу в тому, що він не потребує збору складних даних, таких як історія покупок користувачів і оцінки. Слабкі сторони демографічної фільтрації полягають у тому, що отримана класифікація може бути занадто загальною. Це призводить до втрати індивідуальності користувачів. Також цей метод використовує дані, які надають користувачі. Такі дані можуть бути неповними, або ж неправдивими. Класифікація методом демографічної фільтрації створюється відповідно до інтересів клієнта, які змінюються з часом. Однак даний метод не підтримує внесення змін до профілю користувача [16].

Порівняльну характеристику розглянутих методів наведено в табл. 1.1.

Таблиця 1.1 – Методи побудови рекомендацій

Метод	Вхідні дані	Особливості
Колаборативна фільтрація	Рейтинги товарів, надані користувачами	Рекомендації на основі рейтингів схожих користувачів
Контентна фільтрація	Рейтинги товарів, надані користувачами	Рекомендація на основі рейтингів схожих предметів
Демографічна фільтрація	Демографічна інформація щодо користувачів	Відбираються користувачі, які схожі за демографічними ознаками
Побудова рекомендацій на з використанням знань	Знання щодо того, як вибрані предмети відповідають потребам користувача	Визначається зв'язок між предметами та потребами користувача

Розглянуті методи є ефективними в типових ситуаціях. Однак у випадках неповноти, неточності даних при використанні таких методів виникає ряд проблем.

### 1.3 Дослідження проблем побудови рекомендацій при використанні методів фільтрації

Традиційні методи побудови рекомендацій потребують доопрацювання вузьких місць, які можна віднести до неточності або неповноти даних. Узагальнений опис таких проблем наведено в табл. 1.2.

Проблема рейтингів виникає внаслідок того, що більшість користувачів не ставлять жодних оцінок. Тому виникає проблема дослідження, як дізнатися, чи задоволені вони продуктом і наскільки. Є два способи отримання оцінок. Явний спосіб – запитати користувачів після того, як вони купили або переглянули будь-який товар. Інший спосіб – прогнозування рейтингів запропонованих предметів на основі їхніх уподобань щодо інших предметів. елемента. Це метод неявного збору рейтингів.

Проблема холодного старту з'являється, коли в рекомендаційній системі не знайдено інформації про користувача або об'єкт. При використанні колаборативної фільтрації рекомендаційна система потребує обов'язкової інформації про користувача та об'єкт.

Дана проблема складається з трьох підпроблем. Перша з них виникає, коли ми не маємо жодної інформації про нового користувача, який входить в систему. Це відбувається тоді, коли новий користувач реєструється в рекомендаційній системі. Ця проблема відома як проблема холодного запуску нового користувача [17].

Друга підпроблема виникає, коли в рекомендаційну систему вводиться новий об'єкт. Тоді система не може знайти жодних оцінок, пов'язаних із цим елементом.

Таблиця 1.2 – Проблеми побудови рекомендацій, пов'язані із неповнотою і неточністю вхідних даних

Назва	Опис
1. Проблема рейтингів	Більшість користувачів не ставлять жодних оцінок товарам. Тому виникає проблема: як дізнатися, чи задоволені вони продуктом і наскільки.
2. Проблема холодного старту 2.1. Проблема холодного запуску користувача 2.2. Проблема холодного запуску товару 2.3. Проблема холодного запуску рекомендаційної системи	В рекомендаційній системі не знайдено інформації про користувача або об'єкт. Тому неможливо надати рекомендації, порівнюючи вибір користувачів або властивості предметів.
3. Проблема розрідженості вхідних даних	Розрідженість рейтингової матриці, оскільки користувачі не виставляють оцінки товарам. Приводить до неточних рекомендацій при використанні колаборативної фільтрації.
4. Проблема масштабованості	Сповільнення роботи рекомендаційної системи, коли кількість користувачів суттєво збільшується.
5. Проблема надмірної спеціалізації	Виникає, якщо рекомендовані об'єкти занадто схожі один на одного. Користувач може перейти до іншого сайту з більш цікавими пропозиціями.

Кінець таблиці 1.2

6. Проблема зміни даних з часом	Старіння інформації щодо товарів в рекомендаційній системі приводить до неточних рекомендацій
7. Проблема зміни налаштувань користувача	Уподобання користувача змінюються з часом, що приводить до неперсоналізованих рекомендацій
8. Проблема непередбачуваних предметів	Не формуються рекомендації для нових товарів, для яких не вказано категорії з історією перегляду.

Колаборативна фільтрація потребує матриці оцінки об'єктів користувачами, щоб дати рекомендацію, і тому не може надати рекомендацію. Така проблема відома як проблема холодного запуску об'єкту (предмету).

Третя підпроблема виникає, коли ми запустили рекомендаційну систему вперше. У цьому випадку система не має жодної інформації про користувача, а також інформації про товари.

Іншими словами, відсутня матриця оцінок у форматі: «користувачів-оцінка об'єкту», яка потрібна для належної роботи методи колаборативної фільтрації. Ця проблема відома як проблема системи холодного запуску рекомендаційної системи [18].

Для вирішення проблем холодного старту необхідно використовувати гібридні методи. Наприклад, враховувати контент [19].

Проблема розрідженості – це проблема, яка виникає через розрідженість рейтингової матриці. Проблема розрідженості вхідних даних виникає тому, що більшість користувачів користуються рекомендаційною системою, але не дають належним чином оцінку товарам, тобто відсутній зворотній зв'язок. Тому, незважаючи на те, що в системі може бути зареєстровано багато користувачів, може бути мало оцінок від цих

користувачів щодо різних товарів, які їм сподобалися, які вони придбали чи які їм не сподобалися. Користувача часто уникають оцінок, а іноді можуть давати помилкові оцінки, наприклад, надати максимальні 5 зірок або мінімальну 1 зірку, навіть не враховуючи властивості продукту. Такі оцінки використовуються для побудови рекомендацій, які відображають небажані результати для користувача. Користувач може втратити інтерес до цієї рекомендаційної платформи.

Проблема масштабованості виникає при збільшенні масштабів обробки даних. Наприклад, у випадку рекомендаційних систем, масштабованість можна розуміти як ситуацію, коли рекомендаційна система працює дуже добре у випадку кількох користувачів, але коли користувач зростає до декількох десятків тисяч, то рекомендаційна система стає повільною.

Проблеми масштабованості включає апаратну і програмну масштабованість. Вирішення першої проблеми полягає у збільшенні можливостей апаратного забезпечення при збільшенні кількості користувачів. Наприклад, можна змінити на більш потужну модель процесора, збільшити оперативну пам'ять та замінити сервер. Але тільки збільшення можливостей апаратного забезпечення не завжди дає можливість вирішити цю проблему.

Масштабованість програмного забезпечення забезпечується написанням алгоритмів, які ефективно функціонують при розширенні конфігурації апаратного забезпечення буде збільшена за потреби в майбутньому. Також вони мають ефективно виконуватись при збільшенні обсягу даних [19].

Проблема надмірної спеціалізації виникає, якщо рекомендовані об'єкти, занадто схожі один на одного. Наприклад, користувач регулярно купує продукти на веб-сайті покупок. Щоразу, коли він відкриває веб-сайт, система рекомендацій рекомендує один і той же товар, а не нову унікальну пропозицію. Це може змусити користувача змінити веб-сайт і вибрати систему електронної комерції, яка рекомендує пропонує кращі, цікаві та

персоналізовані результати для користувача. Таким чином, якщо з'являється така проблема, то користувач втрачає інтерес до рекомендованих об'єктів. Одним із рішень цієї проблеми є диверсифікація рекомендацій, коли перераховуються несхожі об'єкти, які можуть зацікавити користувача.

Проблема зміни даних з часом виникає внаслідок старіння інформації. Багато рекомендаційних систем збирають дані користувачів або використовують минулу поведінку користувачів. Але аналіз минулої поведінки користувачів не є ефективним інструментом для хорошої рекомендації, оскільки тенденції та інтереси користувачів постійно змінюються, тому системі рекомендацій буде дуже важко реагувати в мінливому середовищі даних. Щоб показати вам, що я маю на увазі, зазвичай молодята перевіряють сукні, косметичні засоби та предмети способу життя, але з плином часу та зміною ситуацій їхній інтерес змінюється до дитячих товарів, таких як підгузки, дитяче харчування тощо, що повністю відрізняється від попередній пошук, який може заплутати систему рекомендацій і привести до відображення результатів, які більше не цікавлять покупця [20].

Проблема зміна налаштувань користувача пов'язана із швидкою зміною його вподобань з часом. Як ми вже спостерігали, наміри користувача щодо перегляду певного елемента можуть відрізнитися через різні проміжки часу. Тому системи рекомендацій, які повністю засновані на вподобаннях користувачів, можуть виконувати неправильні рекомендації. Наприклад, користувач спочатку переглядає книги для себе, а потім переглядає спортивний товар для подарунку. Тоді рекомендаційна система на основі уподобань користувача може рекомендувати неправильні предмети.

Проблема непередбачуваних предметів пов'язана з використанням різних рейтингів. Рекомендація предметів використовується для визначення рейтингового списку передбачуваного списку предметів і непередбачуваного списку предметів. Він може включати незвичайні продукти, які існували раніше або мають відношення до існуючих. У цьому сценарії рекомендаційна

система не зможе рекомендувати продукт нікому з користувачів. Наприклад, якщо в продаж запущено новий продукт, який належить до нової категорії. Оскільки історія перегляду категорії не знайдена, то рекомендації для користувача не формуються. Для вирішення цієї проблеми використовуються методи фільтрації.

Для вирішення цих проблем використовуються гібридні методи формування рекомендацій.

#### 1.4 Дослідження гібридних методів побудови рекомендацій

Гібридні рекомендаційні системи використовують два й більше методів побудови рекомендацій для досягнення більшої ефективності та задоволеності користувачів. На практиці найчастіше колаборативна фільтрація поєднується з додатковими методами для того, щоб уникнути проблеми збільшення розміру матриці вхідних даних. В таблиці 1.3 показані основні гібридні методи [21].

Зважений метод побудови рекомендацій обчислює оцінку рекомендованого пункту за результатами всіх доступних методів, наявних у рекомендаційній системі. Часто використовується лінійна комбінація рекомендаційних балів. Наприклад, гібридна система P-Tango використовує такий гібрид. Спочатку рекомендаціям з колаборативної фільтрації та контенту надається однакова вага. Поступово ваги коригуються, оскільки прогнози щодо оцінок користувачів підтверджуються або спростовуються. Інший гібридний метод не використовує числові оцінки, а використовує в консенсусну схему для результатів рекомендаційних методів колаборативної фільтрації, за змістом та за демографічними даними.

Перевага зваженого гібридного методу полягає в тому, що всі існуючі алгоритми використовуються в процесі побудови рекомендацій у простий спосіб. Кожному методу надається вага і тому просто налаштувати гібридний

метод. Однак зважений метод використовує неявне припущення, що відносне значення різних методів є більш-менш рівномірним у просторі можливих елементів. Однак на практиці спільна рекомендація буде слабшою для предметів з невеликою кількістю оцінок, що надали користувачі.

Таблиця 1.3 – Гібридні методи побудови рекомендацій

Метод	Опис методу
Зважений	Оцінюються та комбінуються результати кількох рекомендаційних методів для кінцевої рекомендації
Перемикання	Рекомендаційна система використовує один з методів побудови рекомендацій в залежності від поточної ситуації
Змішаний	Одночасно представляються рекомендації, які побудовані кількома різними методами
Комбінація ознак	Використовуються ознаки рекомендаційних даних, що отримані із різних джерел
Каскадний	Один метод уточнює рекомендації, які дав інший метод.
Нарощування ознак	Вихід одного методу використовується як вхід для іншого методу.
Використання метарівня	Модель рекомендацій використовується як вхідні дані для методу побудови рекомендацій

Гібридний метод перемикання використовує певний критерій для перемикання між методами рекомендацій.

Наприклад, комбінуються рекомендації за контентом та колаборативну фільтрація. Спочатку використовується метод побудови рекомендацій на основі контенту.

Якщо метод на основі контенту не може побудувати рекомендацію з достатньою впевненістю, тоді використовується колаборативна фільтрація. Даний гібридний метод побудови рекомендацій не вирішує проблему «нового користувача». Проте він не потребує великої кількості прикладів для точної класифікації.

Гібридний метод перемикання може виробляти рекомендації, які семантично не близькі до елементів, які раніше були високо оцінені, але все ще є актуальними.

Наприклад, користувач, який цікавиться розслідуванням щодо монополій, також може бути зацікавлений у інформації щодо злиття фірм. Метод зіставлення контенту не рекомендуватиме історії злиття. Однак інші користувачі, зацікавлені в корпоративній владі в індустрії високих технологій, можуть високо оцінити обидві групи історій. Така комбінація дозволить рекомендаційній системі включити ці результати до рекомендації.

Гібридний метод перемикання спочатку використовує швидку модель, а альтернативний метод використовується лише якщо перший варіант не дає результатів.

Варіант даного методу використовує узгодження між попередніми оцінками користувача та рекомендаціями кожного методу для вибору методу побудови для наступної рекомендації.

Гібридний метод перемикання ускладнює процес рекомендацій, оскільки необхідно визначити критерії перемикання. Це вводить інший рівень параметризації. Однак перевага полягає в тому, що система може бути чутливою до сильних і слабких сторін, що входять до її складу.

Змішаний гібридний метод використовується, якщо доцільно робити велику кількість рекомендацій одночасно. Тоді рекомендації з більш ніж одного методу представляються разом.

Наприклад, для складання рекомендованої програми перегляду телебачення використовуються методи на основі контенту, які використовують текстовий опис телешоу, та інформація про схожі

вподобання інших користувачів. Рекомендації з цих двох методів об'єднуються в рекомендованій програмі.

Змішаний гібридний метод дозволяє уникнути проблеми «нового предмету». Цей метод використовує рекомендації на основі контенту, щоб рекомендувати нові шоу на основі їх описів, навіть якщо ці передачі ще не були оцінені.

Однак метод не вирішує проблему «нового користувача», оскільки і контент, і методи спільної роботи потребують певних даних про переваги користувачів, щоб почати роботу. Але якщо така система інтегрована в цифрове телебачення, вона може відстежувати, що за шоу і як довго дивляться користувачі. Відповідно формуються профілі цих користувачів. Як і попередній метод, змішаний метод може знайти нові елементи, які не можуть бути отримані на основі контенту.

Для використання такого методу потрібно багато рекомендацій. Там, де виникають конфлікти, є певний тип арбітражу між методами. Наприклад, рекомендації на основі вмісту мають перевагу над спільними відповідями. Зазвичай змішаний метод вимагає ранжування елементів або вибору однієї найкращої рекомендації, після чого необхідно застосувати методи їх комбінування.

Гібридний метод комбінації ознак дозволяє досягти об'єднання вмісту та спільної роботи. Для цього спільна інформація розглядається як додаткові ознаки кожного елемента вхідних даних. Потім використовуються методи на основі вмісту обробки такого розширеного набору даних. Однак для досягнення переваг цього методу часто використовується ручна фільтрація змісту.

Метод приводить до значного покращення точності в порівнянні з підходом суто спільної роботи. Даний метод частково використовує дані про спільні інтереси користувачів, що зменшує чутливість системи до кількості користувачів, які оцінили рекомендований елемент.

Каскадний гібридний метод передбачає поетапний процес побудови рекомендацій. Перший метод рекомендацій використовується спочатку для отримання грубого рейтингу кандидатів у рекомендований список, а другий метод уточнює рекомендацію з набору цих кандидатів. Наприклад, в рекомендаціях для ресторанів спочатку використовуються знання про ресторани, щоб давати рекомендації на основі заявлених інтересів користувача.

Отримані рекомендації поміщаються в групи з однаковими перевагами. Далі використовується метод колаборативної фільтрації для додаткової оцінки пропозицій в кожній групі.

Каскадний метод дозволяє рекомендаційній системі уникати використання методу з нижчим пріоритетом для елементів, які розрізняються першим методом, або ж мають досить низькі оцінки і тому ніколи не будуть рекомендовані. Другий крок каскадного методу зосереджується лише на тих елементах, для яких потрібна додаткова перевірка. Тому метод є ефективнішим, ніж зважений гібридний метод, який застосовує свої базові методи до всіх вхідних елементів. Крім того, каскад за своєю природою толерантний до шумів під час роботи другого методу, оскільки оцінки, надані першим методом, можна лише уточнити, а не скасувати.

Гібридний метод нарощування ознак використовує перший базовий метод для визначення рейтингу або класифікації об'єкта. Ця інформація потім включається в обробку наступним методом.

Наприклад, рекомендації книг даються на основі контенту з використанням класифікованих даних, знайдених на Amazon.com. Для класифікації використовується наївний класифікатор тексту Байеса. У текстові дані, що використовуються для рекомендацій, включається інформація про пов'язаних авторів та пов'язані назви. Такі особливості підвищують якість рекомендацій.

Інший варіант методу пов'язаний із використанням фільтрів, заснованих на знаннях. Ці фільтри використовують такі критерії, як кількість

орфографічних помилок та розмір включених повідомлень. Фільтри вносять рейтинги в базу даних рейтингів, яку використовує спільна частина рекомендаційної системи, виконуючи роль штучних користувачів. Такий підхід покращив фільтрацію електронної пошти.

Додаткову функціональність додають модулі-посередники, які можуть використовувати інші методи для збільшення самих даних. Даний гібридний метод відрізняється від комбінації функцій, у якій об'єднуються вихідні дані з різних джерел [22].

Гібридний каскадний метод та метод нарощування ознак принципово відрізняються тим, що у методі нарощування ознак другий базовий метод, використовує вихідні дані першого. У каскадному гібридному методі другий базовий метод не використовує жодних результатів першого методу для формування рейтингу. Замість цього результати двох методів об'єднуються у пріоритетному порядку.

Гібридний метод використання метарівня для опису рекомендацій використовує модель, створену попереднім методом, як вхідні дані для наступного методу.

Цей метод відрізняється від методу нарощування ознак.

Метод використання метарівня подає на вхід другого методу всю модель, а метод нарощування ознак – лише ознаки, отримані в результаті використання першого базового методу.

Метод використання метарівня застосовується, наприклад, для веб-фільтрації. Спочатку виконується фільтрація на основі контенту, тобто на основі їх інтересу для суспільства в цілому. Використовуються моделі всіх користувачів. Потім відбираються специфічні дані для певних користувачів.

Інша реалізація даного підходу використовує двоетапну схему на основі класифікаторів Байєса. Для кожного користувача створюється наївний класифікатор Байєса на основі контенту. Потім параметри класифікаторів для різних користувачів пов'язуються за допомогою регресії.

Перевага даного гібридного методу, особливо для гібриду вміст/спільна робота, полягає в тому, що отримана на першому етапі модель є стислим представленням інтересів користувача [23]. Механізм колаборативної фільтрації, який використовується далі, може працювати з цією цим представленням інформації ефективніше, ніж з необробленим рейтингом. Можливості поєднання базових методів побудови рекомендацій при побудові гібридних методів наведено в табл. 1.4. В таблиці використовуються такі позначення:

- КФ – колаборативна фільтрація;
- КН – фільтрація на основі контенту;
- ДМ – демографічний метод;
- КБ – побудова рекомендацій на основі використання баз знань.

Таблиця 1.4 – Комбінування базових методів побудови рекомендацій в гібридних методах

Гібридні методи побудови рекомендацій	КФ	КФ	КФ	КН	КН	КН	ДМ	ДМ	ДМ	КБ	КБ	КБ
	КН	ДМ	КБ	КФ	ДМ	КБ	КФ	КН	КБ	КФ	КН	ДМ
Зважений	+	+	+		+	+			+			
Перемикання	+	+	+		+	+			+			
Змішаний	+	+	+		+	+			+			
Комбінація ознак	+	+	Ні		+	Ні			Ні			
Каскадний	+	+	+	+	+	+	+	+	+	+	+	+
Нарощування ознак	+	+	+	+	+	+	+	+	+	+	+	+
Використання метарівня	+	Ні	+	+	Ні	+	Ні	+	+	+	+	Ні

Гібридизація підвищує ефективність колаборативної фільтрації та інших базових методів побудови рекомендацій.

Поєднання рекомендацій за змістом та колаборативної фільтрації завжди приводить до проблеми збільшення обчислювальних витрат, оскільки для обох методів потрібна база даних рейтингів. Тим не менш, такі гібриди мають перевагу, оскільки рейтинги в більшості випадків вже існують або можуть бути виведені з даних.

Використання метарівня дозволяє уникнути проблеми розрідженості, стискаючи оцінки для багатьох вхідних даних у модель, яку можна легше порівнювати для різних користувачів.

Методи, засновані на знаннях і корисності, можуть бути використані для гібридизації, оскільки вони стикаються з проблемою збільшення вхідних даних.

### 1.5 Постановка задачі дослідження

Проведений аналіз показав, що існуючі гібридні методи комбінують традиційні методи побудови рекомендацій. Побудова рекомендацій виконується в дві фази. На першій фазі відбувається попередній відбір рекомендованих даних, а на другій – уточнюється список предметів, які будуть представлені користувачеві.

Існуючі гібридні методи побудови рекомендацій, як правило, не враховують зміни вимог та потреб користувачів з часом. Тому такі методи використовуються в офлайн-режимі. Однак це створює проблеми при побудові рекомендацій в режимі онлайн. Такі рекомендації мають менші вимоги до точності, але вони мають бути швидко адаптовані до поточних потреб користувача.

Об'єктом дослідження в рамках магістерської кваліфікаційної роботи є процес формування рекомендацій у ІТ- проектах рекомендаційних систем.

Предметом дослідження є методи формування рекомендацій пояснень у проектах рекомендаційних систем.

Метою даної роботи є дослідження гібридних методів формування рекомендацій для побудови рекомендацій в онлайн-режимі.

Для досягнення мети роботи необхідно вирішити такі задачі:

- дослідження характеристик ІТ- проектів рекомендаційних систем;
- аналіз методів побудови рекомендацій;
- дослідження проблем побудови рекомендацій при використанні методів фільтрації;
- дослідження гібридних методів побудови рекомендацій;
- удосконалення гібридного методу побудови рекомендацій в режимі онлайн;
- планування проекту розробки гібридного методу побудови рекомендацій;
- експериментальна перевірка удосконаленого гібридного методу побудови рекомендацій в режимі онлайн у проектах рекомендаційних систем.

## 2. ГІБРИДНИЙ МЕТОД ПОБУДОВИ РЕКОМЕНДАЦІЙ В ОНЛАЙН-РЕЖИМІ

2.1 Підхід до побудови рекомендацій з урахуванням змін у вимогах користувачів

Рекомендаційні системи здійснюють прогноз та формування рекомендацій на основі наявної інформації про взаємодію двох категорій сутностей: користувачі та об'єкти [24].

Користувачі надають вхідну інформацію для формування рекомендацій. Об'єктами є цікаві для користувача товари, послуги, сайти тощо.

Формування рекомендацій для користувачів виконується як в автономному (офлайн), так і онлайн режимі [25].

У першому випадку використовуються збережені дані про вибір, покупки, оцінки користувачів.

Офлайн-обчислення виконуються в пакетному режимі, що дозволяє використовувати великі обсяги даних та ресурсомісткі моделі та алгоритми. Проте результати цих обчислень можуть застаріти на момент формування рекомендацій. Тому додатково виконуються обчислення онлайн-режимі. Такі обчислення дозволяють враховувати останні зміни інтересів користувачів та, тим самим, краще задовольняти їх запити. Онлайн-обчислення виконуються у реальному часі. Це означає, що в режимі онлайн необхідно враховувати обмеження за обсягом даних і обчислювальної складності застосовуваних методів [26].

Таким чином, проблема побудови онлайн-рекомендацій у рекомендаційних системах з урахуванням змін пріоритетів користувачів потребує удосконалення гібридних методів побудови рекомендацій.

Існуючі підходи до формування рекомендацій використовують інформацію про взаємодію користувачів та об'єктів.

Така взаємодія виникає, наприклад, при покупці товарів онлайн, виставленні оцінок цим товарам, при серфінгу сторінок з описом товарів та послуг. Ці підходи поділяються на дві групи: побудова рекомендацій з урахуванням подібності об'єктів, і навіть з урахуванням подібності споживачів.

У першому випадку вибір об'єктів виконується на основі порівняння їх характеристик із тими об'єктами, які були обрані поточним користувачем рекомендаційної системи [27].

У другому випадку використовується алгоритм колаборативної фільтрації. Основна ідея колаборативної фільтрації полягає в тому, щоб виявити приховані фактори, що відображають зв'язок користувачів та об'єктів. Ці фактори використовуються для пошуку споживачів із подібними запитами. Рекомендації формуються з урахуванням інформації про відомий вибір таких споживачів [28].

Як вхідні дані для колаборативної фільтрації використовується інформація про вибір користувачів зі схожими характеристиками. Для зменшення обсягу оброблюваних даних доцільно враховувати темпоральну динаміку користувачів, зміну їх інтересів та поведінки з часом.

У низці робіт, присвячених колаборативної фільтрації, враховуються темпоральні характеристики користувачів, зокрема вплив сезонних змін на вибір користувачів, врахування довгострокових та короткострокових переваг користувачів. При цьому виконується часткової фільтрації даних для відображення динаміки уподобань користувача. Така фільтрація призначена для побудови короткострокових прогнозів та базується на принципах самонавчання.

Також використовується моделювання поведінки користувачів при придбанні товарів та послуг (купівлі, кліки, перегляд сторінок). Ця поведінка моделюється з використанням Обчислюється близькість об'єктів з урахуванням явного зворотного зв'язку від користувача [29].

Такі гібридні методи, хоча і враховують зміни у поведінці користувача, але не призначені для застосування в онлайн-режимі, оскільки вони вимагають значних обчислень.

Для подолання цього недоліку на першій фазі побудови рекомендацій доцільно зменшити кількість вихідних даних. Для цього потрібно врахувати темпоральний аспект та виділити підмножину даних з урахуванням поточних інтересів групи схожих користувачів.

Надалі рекомендації формуються на основі отриманої підмножини відфільтрованих даних з використанням існуючих підходів, зокрема колаборативної фільтрації.

Такий двох етапний гібридний підхід призначений для побудови онлайн-рекомендацій з урахуванням типової поведінки користувачів для того, щоб використовувати в переліку як популярні об'єкти, так і предмети персонального інтересу користувача.

Удосконалений метод заснований на ідеї, що з послідовності записів про поведінку деякої системи можуть бути виділені підмножини даних, відповідних процесам в її підсистемах, які також відображають типову поведінку користувачів цієї системи. Ці підмножини містять шаблони поведінки користувача у формі темпоральних залежностей.

Відмінність удосконаленого гібридного методу базується на тому, що дані залежності відображають неявні знання користувача про предметну область, а також систему та її елементи.

Це свідчить про доцільність відбору даних, відповідних неявним темпоральним залежностям.

Таким чином, для побудови рекомендацій в онлайн-режимі доцільно виконати попередній відбір підмножини даних, а потім - колаборативну фільтрацію підмножини, що вийшло.

Удосконалений підхід до побудови рекомендацій в режимі онлайн складається із двох фаз. Загальну схему удосконаленого методу показано на рис. 2.1.

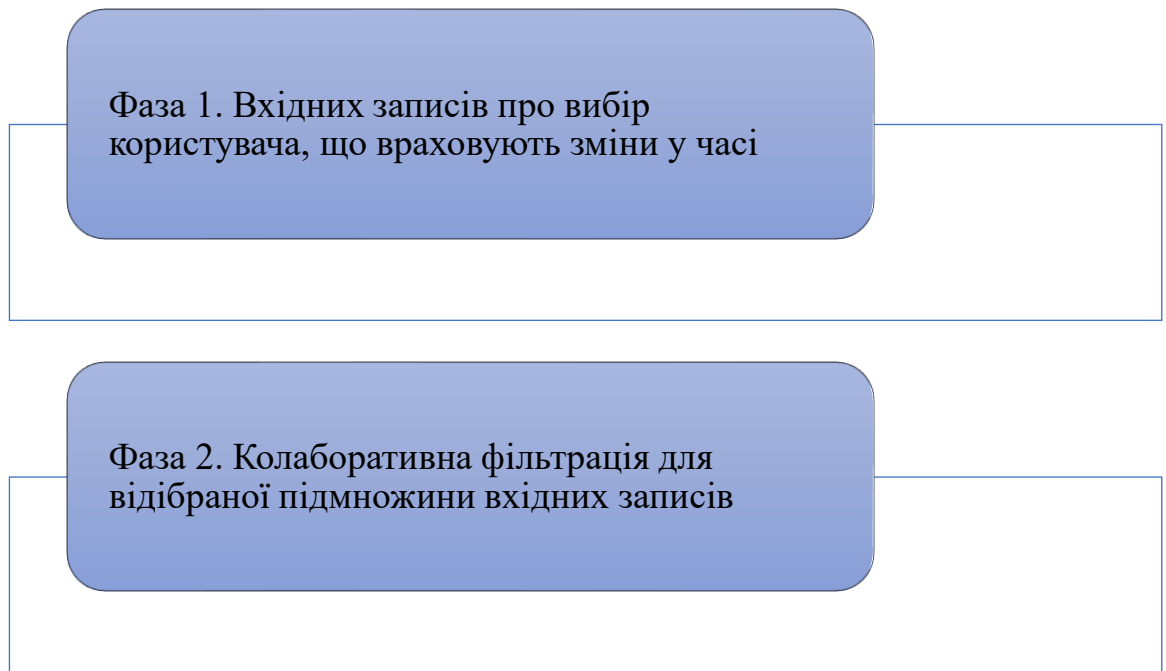


Рисунок 2.1 – Фази гібридного методу побудови рекомендацій

На першій фазі попереднього відбору необхідно знайти таке підмножина вихідних записів про вибір користувача (у вигляді, наприклад, інформації про продаж або оцінок користувачів), яке найкраще відповідає його перевагам з урахуванням зміни цих переваг у часі.

На другій фазі виконується традиційна колаборативна фільтрація для відібраної підмножини релевантних вихідних записів.

На першій фазі виконуються такі дії:

- дані розбиваються на підмножини по інтервалам часу;
- з кожної підмножини відбираються дані про вибір схожих користувачів;
- схожими є користувачі, які:
  - вибрали ті ж товари, що й цільовий користувач;
  - вибрали в тій же або більше кількості, що й цільовий користувач;
  - відбір виконується спочатку на останньому інтервалі, а потім на попередніх інтервалах часу;
  - з множини схожих товарів відбираються найбільш популярні товари;

За результатами першої фази формуються вхідні дані для другої фази.

На другій фазі формуються рекомендації методом колаборативної фільтрації.

## 2.2 Удосконалений гібридний метод побудови рекомендацій у режимі онлайн з урахуванням змін вимог користувачів

В якості вхідних даних удосконаленого методу можуть використовуватись:

- статистика продажів;
- кліки на сторінках сайту;
- інформація про інтереси цільового користувача, яка міститься у записах про покупки.

Набір вхідних даних подається у вигляді багат шарового дводольного графа, що відображає зміну інтересів (покупок, кліків) користувачів на послідовності інтервалів:

$$T = \langle T^1, T^2, \dots, T^k, \dots, T^K \rangle. \quad (2.1)$$

Кожен із цих інтервалів заданий початковою  $t_{st}^k$  і кінцевою  $t_{fn}^k$  мітками часу:

$$T^k = [t_{st}^k, t_{fn}^k], k = \overline{1, K}. \quad (2.2)$$

Кожен  $k$  – шар графа визначає вибір об'єктів користувачами на заданому інтервалі  $T^k$ .

Наприклад, у кожному шарі можуть бути представлені покупки на сайті за годину, день, тиждень.

Кожен  $k$  – шар графу містить підмножину вершин, які відповідають користувачам:

$$Us^k = \{Us_i^k\}, \bigcup_k Us^k = Us. \quad (2.3)$$

Також граф містить підмножину вершин, які відповідають об'єктам:

$$It^k = \{It_j^k\}, \bigcup_k It^k = It. \quad (2.4)$$

Дуги графа  $E^k$ , у кожному шарі задають зв'язки між вершинами користувачів та об'єктів на інтервалі часу  $T^k$ :

$$E^k = \{E_{i,j}^k : \forall k \exists T^k\}, E^k \subseteq Us^k \times It^k, \bigcup_k E^k = E. \quad (2.5)$$

Дуги графа  $E_{i,j}^k$  задаються тільки в тому випадку, якщо користувач  $Us_i^k$  вибирав чи оцінював предмета  $It_j^k$ . Семантика цього відношення відрізняється для позитивного та негативного зворотного зв'язку від користувача.

У разі явного зворотного зв'язку від користувача кожна дуга  $E_{i,j}^k$  відповідає оцінці користувача та мають вагу, рівну рейтингу предмета  $R_{i,j}^k$ .

У разі неявного зворотного зв'язку вага дуги може відповідати кількості куплених предметів  $S_{i,j}^k$ , кількості кліків на сторінці з описом предмета, кількості переходів на сторінку і т.п.

Для неявного зворотного зв'язку вага становить кількість куплених користувачем предметів, оскільки дана інформація завжди доступна в системах електронної комерції.

У тому випадку, якщо є додаткова інформація про кліки та переходи на сторінку, то як вагу можна використовувати зважену суму значень цих показників.

Метод використовує зважений граф для випадку неявного зворотного зв'язку:

$$E^k = \{S_{i,j}^k\}. \quad (2.6)$$

Формування рекомендацій у разі явного зворотного зв'язку виконується за допомогою такої самої послідовності кроків.

Елементи  $P_j^{k,l}$  множини  $P$  відносин між шарами  $k$  та  $l$  задаються через зв'язок між ідентичними предметами  $It_j^k$  і  $It_j^l$ , що належать до цих шарів. Такий вибір взаємозв'язків обумовлений тим, що користувачі можуть вибирати об'єкти лише окремі моменти часу.

У той же час, об'єкти використовуються користувачами на регулярній основі. Наприклад, споживач купує ігрову комп'ютерну мишу раз на кілька місяців чи років. У той самий час ці миші продаються різним користувачам регулярно.

Необхідною умовою для зв'язку  $P_j^{k,l}$  є існування зв'язків  $E_{i,j}^k$  і  $E_{i,m}^l$  на відповідних шарах. Іншими словами, об'єкти  $It_j^k$  і  $It_j^l$  повинні бути обрані  $i$  і  $m$  користувачами протягом інтервалів часу  $T^k$  і  $T^l$ :

$$\begin{aligned} P &= \{P^{k,l}\}, P^{k,l} \subseteq It^k \times It^l, \\ P^{k,l} &= \{P_j^{k,l} : (\forall k, j) \exists E_{i,j}^k, (\forall l, j) \exists E_{i,m}^l\}. \end{aligned} \quad (2.7)$$

Кожен елемент множини  $P$  має вигляд:

$$P^{k,l} = \{P_j^{k,l} : (\forall l, j) \exists E_{i,m}^l, (\forall k, j) \exists E_{i,j}^k\}. \quad (2.8)$$

Багатошаровий граф  $G$ , що описує вхідні дані, включає всі розглянуті вершини і дуги:

$$G = \{U_s, I_t, E, P\}. \quad (2.9)$$

Рекомендації для поточного користувача  $U_s^K$  складаються на останньому із відомих інтервалів  $T^K$  або після його завершення:

$$T^K = [t_{st}^K, t_{fn}^K]. \quad (2.10)$$

Тоді на поточному шарі  $K$  доцільно враховувати лише тих користувачів, які вибирали (купували або, переглядали) такі ж товари, як і цільовий користувач  $U_s^K$ .

Іншими словами, має виконуватися умова  $S_{\alpha,j}^K \geq 1$ . При цьому кількість виборів кожного з об'єктів має бути близькою до покупок користувача  $U_s^K$ , тобто.  $|S_{i,j}^K - S_{\alpha,j}^K| \leq 1$ .

Відповідно, на останньому  $K$  - шарі необхідно сформуванати підграф  $G_\alpha^K$ , який має такий вигляд:

$$G_\alpha^K = \{U_s^K, I_t^K, E^K : (\forall j \forall i) \forall S_{\alpha,j}^K \geq 1 |S_{i,j}^K - S_{\alpha,j}^K| \leq 1\}. \quad (2.11)$$

Фільтрування дозволяє відібрати користувачів зі схожими інтересами.

Зміна інтересів користувачів у часі розраховується на основі динаміки популярності кожного об'єкта, що буде включено до набору відібраних на першій фазі даних.

Удосконалений метод побудови рекомендацій з урахуванням змін потреб користувачів включає наступні етапи.

Етап 1. Розбиття множини вхідних даних  $Dt$  на впорядковану послідовність підмножин  $Dt^k$ . Кожна елемент  $Dt_s^k$  містить мітку часу  $t_s^k$  з інтервалу  $T^k$ :

$$Dt = \left\langle Dt^k = \{Dt_s^k\} : (\forall k) t_s^k \in T^k, k = \overline{1, K} \right\rangle. \quad (2.12)$$

Результатом цього етапу можуть бути, наприклад, підмножини записів про продаж за кожен день (за годину, за тиждень).

Етап 2. Відбір даних із підмножини  $Dt^K$  за умовою зміни інтересів користувачів. Запит на рекомендації від цільового  $\alpha$  – користувача надходить на останньому відомому інтервалі часу  $T^K$ . Підмножина  $Dt^K$  містить найактуальніші дані, що виникли на інтервалі запиту рекомендацій, що є ключовим для онлайн-рекомендацій. Результатом даного етапу є підмножина записів з  $Dt^K$ , яка містить інформацію про вибір користувачів, схожих на вибір користувача  $Us_\alpha^K$ , на останньому інтервалі. Схожість полягає в тому, що користувачі  $Us_i^K$  вибирають стільки ж або на 1 більше екземплярів об'єкта  $It_j^K$ , що і цільовий користувач  $Us_\alpha^K$ .

Етап 3. Відбір підмножин даних  $Dt^{k*}$  з  $Dt^k$ , для схожих користувачів. Для таких користувачів виконується умова  $\exists Us_i^K \in G_\alpha^K$  належності користувача  $Us_i^K$  до графу продажів для цільового користувача :

$$Dt^* = \left\langle Dt^{k*} = \{Dt_s^k\} : (\forall i \forall k) \exists Us_i^K \in G_\alpha^K \right\rangle. \quad (2.13)$$

Результатом даного етапу є підмножина вхідних даних  $Dt^*$ , який містить записи про вибір користувачів, схожих на  $\alpha$  – користувача.

Етап 4. Відбір із підмножини  $Dt^*$  найбільш популярних даних на кількох послідовних інтервалах часу  $T^k$ .

$$Dt^{**} = \left\langle \left\{ Dt^{k*} \right\} : (\forall k) S_{i,j}^{k*} = \max_i (S_{i,j}^k) \right\rangle. \quad (2.14)$$

Результатом даного етапу є підмножина вхідних даних  $Dt^{**}$ , який містить записи про вибір схожими користувачами об'єктів, популярних протягом кількох послідовних інтервалів часу  $T^k$ , включаючи інтервал  $T^K$ .

Етап 5. Формування рекомендацій за допомогою колаборативної фільтрації на основі підмножини вихідних даних  $Dt^{**}$ .

Результатом цього етапу є впорядкований набір рекомендованих об'єктів, який враховує інтереси цільового користувача, і навіть найбільш популярні нині об'єкти схожих користувачів.

### **3. ПРОЕКТ РОЗРОБКИ МЕТОДУ ГІБРИДНОЇ ФІЛЬТРАЦІЇ У ІТ-ПРОЕКТАХ З РОЗРОБКИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ**

#### **3.1 Опис проекту розробки гібридного методу побудови рекомендацій**

Проект розробки програмного додатку для методу гібридної фільтрації розробляється командою із таких спеціалістів:

- менеджер проекту;
- програміст;
- тестувальник (QA).

Оцінка тривалості проекту становить 27 робочих днів. У процесі розробки використовується waterfall - модель життєвого циклу ІТ- проекту з розробки елементів рекомендаційної системи.

Декомпозиція проекту на обов'язкові задачі дала можливість отримати такий перелік задач:

- проектування;
- реалізація удосконаленого гібридного методу побудови рекомендацій у режимі онлайн;
- модульне тестування алгоритму побудови рекомендацій у режимі онлайн;
- інтеграція модуля в рекомендаційну систему;
- тестування інтеграції модуля;
- первинне тестування додатку.

Проектування містить у собі підготовку вхідних даних. Ці дані відбираються із журналу продажів рекомендаційної системи.

Проектування включає:

- визначення вимог;
- формування задачі фільтрації онлайн;
- написання ТЗ.

### 3.2 Статут проекту

1. Причини ініціалізації даного проекту (ідентифікація проблеми, що розв'язується за допомогою розробленого модулю та обґрунтування ініціалізації проекту на підприємстві).

На сьогодні ваговою причиною надання неточних рекомендацій в онлайн режимі є те, що існуючі гібридні методи побудови рекомендацій лише частково враховують швидкі або циклічні зміни потреб користувачів. Це не дає бізнесу у повній мірі надавати якісні рекомендації при зміні потреб користувачів.

Такі зміни відбуваються при переїзді користувачів, завершенні освіти, зміні сімейного стану. Мотивуючою причиною для розробки даного проекту є можливість надати бізнесу удосконалений гібридний метод побудови рекомендацій, який враховує темпоральні зміни у потребах та вимогах користувачів рекомендаційної системи.

2. Сутність даної інноваційної ідеї та розроблений спосіб її застосування для розв'язання проблеми побудови рекомендацій.

В роботі проведено детальний аналіз існуючих базових та гібридних алгоритмів та методів формування пояснень у ІТ проектах побудови рекомендаційних систем. За результатами аналізу виявлено потребу у доопрацюванні методів, оскільки вони не приділяють уваги змінам звичок користувача протягом часу. Але ця властивість є дуже важливою для побудови рекомендацій в режимі онлайн, оскільки ці рекомендації потрібно надавати з урахуванням певного часового інтервалу актуальних вхідних даних.

Побудова рекомендацій ще ускладнюється у випадку незначної кількості даних про користувача рекомендаційної системи. З цих причин було зроблено висновок про важливість удосконалення гібридного методу,

щоб відбирати вхідні дані за результатами аналізу вподобань користувача з часом і потім формувати рекомендації.

3. Мета проекту: створити модуль побудови рекомендацій, який би у складі існуючої рекомендаційної системи дозволив оперативно надавати рекомендації в режимі онлайн з урахуванням останніх змін у поведінці користувачів на основі даних щодо вибору таких покупців за визначений період часу.

4. Очікувані вигоди від даного проекту.

Вигоди від даного проекту базуються на включенні модуля побудови гібридних рекомендацій в онлайн-режимі в роботі рекомендаційної системи. Використання можливостей оперативного уточнення рекомендацій за допомогою даного модулю сприятиме підвищенню довіри користувачів. Це створює можливості розширення компанії, що займається електронною комерцією, за рахунок збільшення замовлень; ріст бази клієнтів такої компанії за рахунок репутації ефективної компанії, що оперативно враховує персональні вимоги клієнтів, здатна швидко запропонувати та актуальні товари на продаж; фінансова мета: збільшення продажів за рахунок підвищення довіри користувачів розробленого модуля.

5. Обмеження проекту.

Початок проекту розробки гібридного методу побудови рекомендацій – не пізніше 8 листопада 2021 року.

Завершення не пізніше 13 Грудня 2021 року.

6. Завдання проекту, критерії їх досягнення.

Перше завдання проекту полягає в розробці удосконаленого гібридного методу побудови рекомендацій в режимі онлайн.

Критерієм досягнення є підвищення точності рекомендацій. Точність визначається за метрикою AUC.

Друге завдання проекту полягає в розробці модулю побудови рекомендацій в режимі онлайн.

Критерієм вирішення завдання є позитивні відгуки від споживачів, що використовують рекомендаційну систему, про задоволення їх очікувань, про більш точні рекомендації товарів, а також про підвищення довіри до результатів роботи рекомендаційної системи.

#### 7. Межі (зміст) проекту.

Межі даного проекту визначаються переліком робіт від початку планування робіт і до передачі модулю після кінцевого тестування. В межі проекту не входять роботи з дистрибуції модуля, інтеграції з рекомендаційною системою, та робота з клієнтами.

### 3.3 Планування проекту

Планування є ключовим процесом при управлінні проектом та визначає усю діяльність по реалізації проекту у часі. Планування представляє собою постійний процес, що призначений для визначення ефективного способу дій для того, щоб досягти поставлених у проекті цілей з урахуванням поточних обмежень.

Процес планування проекту виконується на протязі всього його життєвого циклу. Даний процес починається з розробки попереднього повного плану, що входить до складу концепції проекту розробки гібридного методу побудови рекомендацій. Процес закінчується розробкою детального плану робіт до фази завершення проекту. По мірі виконання проекту відбувається уточнення і деталізація розроблених планів.

На стадії планування проекту задаються методи і засоби, що забезпечують управління виконанням проекту. При цьому розглядається проект як цілісна система, а також по окремих етапах та елементах. Планування зв'язано з такими важливими стадіями процесу управління проектом розробки гібридного методу, як ініціація, а також організація й

контроль виконання проекту, а також аналіз і регулювання й завершення проекту.

Метою планування є побудова моделі щодо реалізації проекту. Стадія планування завершується таким результатом, як статут проекту. Статут охоплює результати планування проекту з усіх функцій управління.

Даний документ є ключовим при реалізації проекту. Цей документ виконує роль як моделі так і плану дій, а також прогнозує етапи виконання проекту та оточення проекту.

При виконанні проекту зазвичай відбуваються зміни в самому проекті, також і за межами проекту. Основним призначенням процесу планування є постійна підтримка реалізації проекту аж до його успішного вчасного завершення робіт.

Об'єктами планування проекту удосконаленого гібридного методу побудови рекомендацій є:

- предметна область;
- організація та комунікації;
- поставки та контракти, зміни;
- ризики;
- якість, час та вартість;
- інтеграційний план;
- інші компоненти проекту.

Project Score – предметною областю проекту є сукупність товарів та послуг, персоналізовані рекомендації яких мають бути забезпечені внаслідок успішного завершення проекту удосконаленого гібридного методу. Предметна область визначається цілями, роботами, а також результатами проекту.

В у життєвому циклі проекту всі наведені складові предметної області, як правило, зазнають змін. Вони можуть коригуватись або уточнюватись. Такі зміни відбуваються не тільки у процесі розробки поточного проекту, а і продовж отримання проміжних результатів проекту.

Організація спільною роботи всіх учасників проекту розробки удосконаленого гібридного методу організовується з використанням календарних планів, а також розкладів робіт. Використовуються такі основні параметри: тривалості робіт проекту, ключові дати та терміни виконання.

Календарним планом є проектний та технологічний документ, який задає:

- перелік та взаємозв'язок робіт проекту;
- послідовність і терміни виконання;
- виконавців і ресурси, що потрібні для успішного виконання робіт проекту.

Виконання робіт із планування відповідно до часових параметрів містить складання календарних планів згідно вимог та обмежень проекту.

Календарні плани розробляються:

- для всього життєвого циклу проекту;
- для етапів проекту;
- для певних рівнів управління;
- для кожного учасника проекту.

Етап прокту	Початок	Тривалість	Загримка	Закінчення
Постановка задачі	08.11.2021	1	0	08.11.2021
Написання ТЗ	09.11.2021	6	0	16.11.2021
Розробка макету	17.11.2021	3	0	19.11.2021
Розробка модулю оцінки	22.11.2021	4	2	25.11.2021
Тестування	25.11.2021	2	-1	26.11.2021
Розробка модулю гібридних рекомендацій	26.11.2021	6	-1	03.12.2021
Тестування	02.12.2021	2	-2	03.12.2021
Інтеграція модулів	07.12.2021	2	3	08.12.2021
Розробка звіту	10.12.2021	1	1	10.12.2021

Рисунок 3.1 – Декомпозиція робіт проекту

Декомпозиція відображає перелік робіт проекту розробки удосконаленого гібридного методу побудови рекомендацій. Декомпозиція також містить тривалість усіх робіт. Декомпозиція може бути виконана на декількох рівнях, тобто роботи можуть бути розбиті на менші роботи.

Форму використання задач наведено на рис. 3.2.

Название задачи	Трудозатраты	Длительность	Начало	Окончание
☐ Постановка задачі	8 ч	1 день	Пн 08.11.21	Пн 08.11.21
<i>Менеджер проекту</i>	8 ч		Пн 08.11.21	Пн 08.11.21
☐ Написання ТЗ	48 ч	6 днів	Вт 09.11.21	Вт 16.11.21
<i>Менеджер проекту</i>	48 ч		Вт 09.11.21	Вт 16.11.21
☐ Розробка макету	24 ч	3 днів	Ср 17.11.21	Пт 19.11.21
<i>Програміст</i>	24 ч		Ср 17.11.21	Пт 19.11.21
☐ Розробка модулю оцінк	32 ч	4 днів	Пн 22.11.21	Чт 25.11.21
<i>Програміст</i>	32 ч		Пн 22.11.21	Чт 25.11.21
☐ Тестування	16 ч	2 днів	Чт 25.11.21	Пт 26.11.21
<i>Тестувальник</i>	16 ч		Чт 25.11.21	Пт 26.11.21
☐ Розробка модулю гібри,	48 ч	6 днів	Пт 26.11.21	Пт 03.12.21
<i>Програміст</i>	48 ч		Пт 26.11.21	Пт 03.12.21
Тестування	0 ч	2 днів	Чт 02.12.21	Пт 03.12.21
☐ Інтеграція модулів	32 ч	2 днів	Вт 07.12.21	Ср 08.12.21
<i>Програміст</i>	16 ч		Вт 07.12.21	Ср 08.12.21
<i>Тестувальник</i>	16 ч		Вт 07.12.21	Ср 08.12.21
☐ Розробка звіту	8 ч	1 день	Пт 10.12.21	Пт 10.12.21
<i>Менеджер проекту</i>	8 ч		Пт 10.12.21	Пт 10.12.21

Рисунок 3.2 – Форма використання задач

Також задається порядок виконання робіт проекту, оцінки тривалості виконання робіт. На схемі зображені дати завершення робіт над проектом. Якщо строки робіт змінюються, то дати наступних робіт здвигаються згідно до таких змін. Також декомпозиція робіт може містити паралельні роботи. У проекті розробки удосконаленого гібридного методу формують критичний шлях, тобто самий довгий послідовний ланцюг робіт проекту. Послідовність

робіт критичного шляху задає дату завершення проекту. Порушення строків робіт критичного шляху змінює цю дату.

Діаграма Ганта, що містить послідовність виконання робіт, представлена на рисунку 3.3.

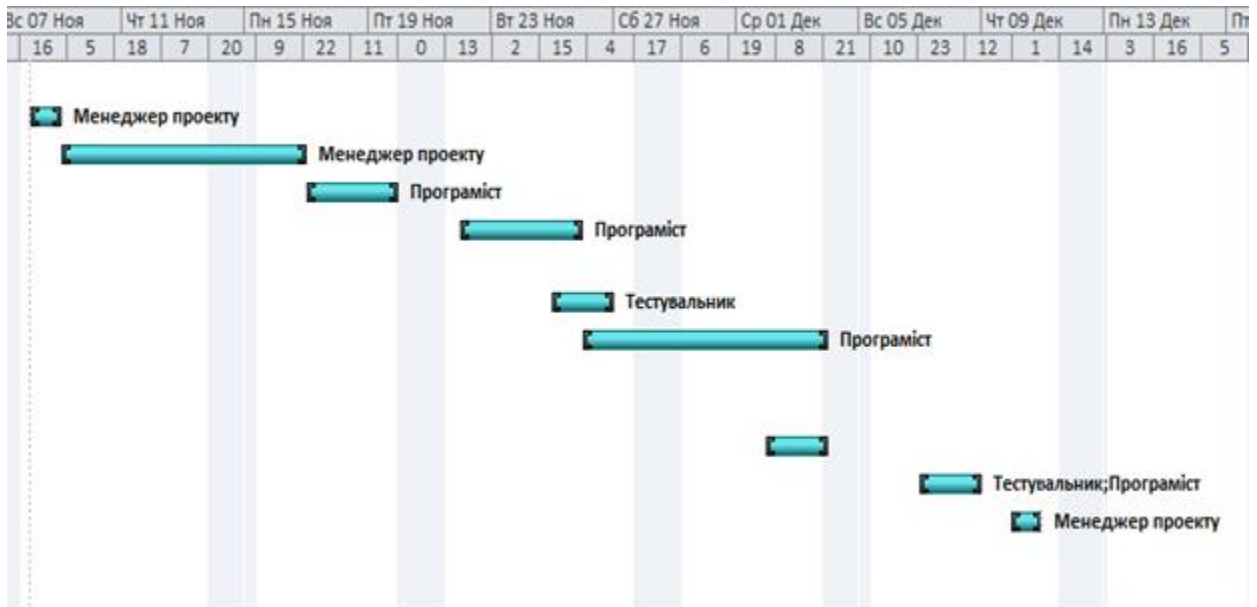


Рисунок 3.3 –Діаграма Ганта

Діаграма Ганта містить розклад календарного робочого плану. Тому дана діаграма дозволяє оцінити взаємозв'язок робіт. Використання даної діаграми зменшує в подальшому число змін у процесі виконання заходів по проекту. Важливо, що дана форма подання робіт робить послідовність завдань за часом більш наочною.

Згідно РМІ, поняття календарного плану та розкладу проекту збігаються. Даний розклад створюється для того, щоб отримати розрахунок і відобразити тривалість проекту через терміни виконання проектних операцій. В розкладі наведено послідовність ланок у процесі планування. Ця послідовність співвідноситься з ієрархією робіт, а також залученими ресурсами: матеріальними, фінансовими і людськими. Ступінь декомпозиції

завдань залежить від можливостей доступу до ресурсів: матеріалів, коштів, а також людських ресурсів. Контроль за послідовністю та термінами виконанням робіт, їх вмістом реалізується з використанням координуючого календарного плану. Проте діаграма Ганта не показує ресурсоемності робіт проекту, лише розклад цих робіт.

Організаційна структура проекту містить схему команди проекту, тобто нової робочої групи. Команда проекту складається з менеджера, у підпорядкування якого є розробник та тестувальник. Тобто команда складається із фахівців з різних підрозділів. Ці фахівці можуть бути повністю зайнятими в проекті, або ж мати часткову зайнятість. Для них визначаються проектні ролі, функції, повноваження і правила взаємодії, а також обов'язки та відповідальність.

Таблиця використання ресурсів (рис. 3.4) містить у собі детальний опис ресурсів проекту.

Назва ресурса	Трудозатрати	Подробности	Декабрь 2021									
			16	19	22	25	28	01	04	07	10	
☐ Не назначен	0 ч	Трудозатр.										
Тестування	0 ч	Трудозатр.										
☐ Менеджер проекту	64 ч	Трудозатр.	8ч									8ч
Постановка задачі	8 ч	Трудозатр.										
Написання ТЗ	48 ч	Трудозатр.	8ч									
Розробка звіту	8 ч	Трудозатр.										8ч
☐ Програміст	120 ч	Трудозатр.	16ч	8ч	24ч	16ч	16ч	24ч			16ч	
Розробка макету	24 ч	Трудозатр.	16ч	8ч								
Розробка модулю о	32 ч	Трудозатр.			24ч	8ч						
Розробка модулю гі	48 ч	Трудозатр.				8ч	16ч	24ч				
Інтеграція модулів	16 ч	Трудозатр.									16ч	
☐ Тестувальник	32 ч	Трудозатр.				16ч					16ч	
Тестування	16 ч	Трудозатр.				16ч						
Інтеграція модулів	16 ч	Трудозатр.									16ч	

Рисунок 3.4 – Таблиця використання ресурсів

Наведена таблиця містить докладну інформацію про використання ресурсів поточного проекту, зокрема яка робота виконується, наскільки завантажені, чи можуть бути використані для інших завдань. Така таблиця використовується для порівняння використання ресурсів у проекті. Для

вирішення цієї задачі ресурси відфільтровуються. Інформацію про ресурси можна переглядати як для відібраної групи, так і для окремих ресурсів. Зайнятість спеціалістів проекту (менеджера, QA та програміста) відображається у годинах.

Таблиця використання ресурсів забезпечує краще розуміння етапів розробки проекту, розподілу робіт серед виконавців, а також витрати часу. У лівій частині таблиці ресурсів представлено коротка інформація щодо назви ресурсу та переліку робіт, що забезпечуються цим ресурсом. Дата та час виконання відображається в правій частині.

## 4. ПРАКТИЧНЕ ЗАСТОСУВАННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

### 4.1 Схема експериментальної перевірки удосконаленого методу побудови рекомендацій

Мета експериментальної перевірки удосконаленого методу – перевірити можливість зменшити обсяг вхідних даних для колаборативної фільтрації з використанням матричної факторизації та зберегти порівняну точність рекомендацій.

Обчислювальна складність колаборативної фільтрації відповідає квадратичній залежності, що свідчить про важливість зниження кількості вхідних даних для побудови онлайн-рекомендацій.

Для оцінки точності рекомендацій використовується метрика AUC (Area Under the Curve), яка визначатиме площу під кривою помилок ROC (Receiver Operating Characteristic) [30].

Чим вище значення AUC, тим вища точність рекомендацій. Фрагмент коду з оцінки точності результатів представлено на рис. 4.1. Реалізацію виконано мовою Python.

```
store_uc = []
popularity_auc = []
pop_items = np.array(test_set.sum(axis = 0)).reshape(-1)
item_vecs = predictions[1]
for user in altered_users:
    training_row = training_set[user,:].toarray().reshape(-1)
    zero_inds = np.where(training_row == 0)
    user_vec = predictions[0][user,:]
    pred = user_vec.dot(item_vecs).toarray()[0,zero_inds].reshape(-1)

    actual = test_set[user,:].toarray()[0,zero_inds].reshape(-1)

    pop = pop_items[zero_inds]
    store_auc.append(auc_score(pred, actual))
    popularity_auc.append(auc_score(pop, actual))

return float('%0.3f'%np.mean(store_auc)),
float('%0.3f'%np.mean(popularity_auc))
```

Рисунок 4.1 – Процедура оцінки точності рекомендацій

В експериментальній перевірці методу використовувався набір даних «Online Retail» із сховища UCI.

Цей набір містить список транзакцій британської мережі магазинів, що продає різноманітні подарунки. Компанія працює із оптовими клієнтами, тому один клієнт може купувати набір різних товарів. Вхідні дані мають таку структуру: номер рахунка-фактури; код товару; Назва товару; кількість одиниць товару одну транзакцію; дата рахунок-фактури; ціна за одиницю; код клієнта; країна. Отже, вхідні дані є даними з неявним зв'язком та містять темпоральні мітки, що дозволяє використовувати удосконалений метод. Повний набір даних містить 541 910 записів. Код цільового користувача, а також стартовий інтервал  $T^k$  вибиралися випадковим чином. Тривалість інтервалів  $T^k$  становила 1 день.

#### 4.2 Результати експериментальної перевірки удосконаленого методу

Було проведено перевірку на кількох вибірках вхідних даних тривалістю від 2 до 6 днів. Ця тривалість залежала від вибраної мінімальної кількості обраного товару при покупці.

Приклади отриманих на наборі «Online Retail» рекомендацій мають вигляд, представлений на рис. 4.2 та рис. 4.3.

Код товару	Опис товару
023167	SMALL CERAMIC TOP STORAGE JAR
123165	LARGE CERAMIC TOP STORAGE JAR
422980	PANTRY SCRUBBING BRUSH
523296	SET OF 6 TEA TIME BAKING CASES

Рисунок 4.2 – Приклад рекомендації новому користувачеві з попереднім відбором даних

Код товару	Опис товару
222963	JAM JAR WITH GREEN LID
023167	SMALL CERAMIC TOP STORAGE JAR
623293	SET OF 12 FAIRY CAKE BAKING CASES
722978	PANTRY ROLLING PIN
823295	SET OF 12 MINI LOAF BAKING CASES
922962	JAM JAR WITH PINK LID
123165	LARGE CERAMIC TOP STORAGE JAR
323294	SET OF 6 SNACK LOAF BAKING CASES

Рисунок 4.3 – Приклад рекомендації новому користувачеві без попереднього відбору даних

Результати оцінки рекомендацій наведено у табл. 4.1.

Отримані після фільтрації вихідної вибірки результуючі набори включають кілька тисяч рядків, що, однак, на два порядки менше розміру вихідної вибірки.

Таблиця 4. 1 – Результати оцінки рекомендацій з урахуванням динаміки інтересів користувачів

Кількість рядків після відбору даних для колаборативної фільтрації, тис. записів	4	8	12
AUC	0,63	0,71	0,73

Для того, щоб порівняти точність рекомендацій у режимі онлайн, з вихідних даних було виділено 3 вибірки аналогічного розміру та проведено оцінку рекомендацій, отриманих в результаті колаборативної фільтрації на основі матричної факторизації.

Результати цієї оцінки наведено у табл. 4.2.

З даних видно, що у разі обліку динаміки інтересів користувачів оцінки отриманих рекомендацій підвищуються приблизно 3%. Однак ці оцінки

залежать від розміру виборки, що була отримана на фазі 1 гібридного методу побудови рекомендацій. Для досягнення мінімально прийнятної точності онлайн-рекомендацій, розмір отриманої вибірки повинен перевищувати 10 тисяч записів.

Таблиця 4.2 – Результати порівняльної оцінки рекомендацій

Кількість рядків, тис. записів	4	8	12
AUC	0,6	0,68	0,71

Таким чином, удосконалення дозволяє знизити обчислювальні витрати, бо вони квадратично залежать від обсягу вхідних даних.

Щоб отримати більш високі значення AUC, необхідно використовувати додаткову інформацію про об'єкти, які вибирає користувач. Така інформація задається властивостями товарів. Тому подальший розвиток методу пов'язаний із поєднанням фільтрації по клієнтам та фільтрації по даним щодо продукції.

Експеримент показав, що удосконалений метод побудови рекомендацій може працювати у режимі он-лайн з урахуванням зміни інтересів користувачів з часом.

При виконанні методу отримуємо послідовність об'єктів, що рекомендуються для користувача, дозволяє спростити прийняття рішень з вибору об'єктів користувачем у випадку, якщо у нього відсутній повний набір знань про характеристики цих об'єктів.

Удосконалений метод виконує попередню фільтрацію вхідних даних з урахуванням вибору схожих користувачів, а також з урахуванням інтервалу часу, на якому інтерес цих користувачів до вибору певних об'єктів істотно не змінюється [31].

Удосконалення дає можливість гнучко зменшити обсяг вхідних даних про відомий вибір користувачів і тим самим знизити обчислювальні витрати на побудову рекомендацій, що є важливим для роботи в режимі онлайн.

Однак розмір попередньо відфільтрованої вибірки  $i$ , як наслідок, оцінка отриманих рішень значною мірою залежить від показників вихідних даних.

Удосконалений метод призначений для застосування в режимі онлайн у рекомендаційній системі та доповнює офлайн-підходи до побудови рекомендацій.

При подальшому удосконаленні методу можна враховувати властивості об'єктів, вибирає користувач рекомендаційної системи. Надалі можна уточнити користувачів після фільтрації властивостей об'єктів.

## ВИСНОВКИ

В магістерській кваліфікаційній роботі досліджено проблему побудови рекомендацій при зміні потреб користувачів.

Виконаний в роботі аналіз показав, що гібридні методи комбінують традиційні методи побудови рекомендацій. Побудова рекомендацій виконується в дві фази. На першій фазі відбувається попередній відбір рекомендованих даних, а на другій – уточняється список предметів, які будуть представлені користувачеві. Аналіз показав, що гібридні методи побудови рекомендацій дають можливість більш ефективно встановлювати зв'язки між потребами користувачів та властивостями товарів.

Дослідження процесу побудови рекомендацій дало можливість зробити висновок, що важливо врахувати те що користувачі змінюють свої потреби з часом.

В роботі удосконалено гібридний метод побудови рекомендацій на основі попередньої фільтрації даних щодо змін вибору користувачів. Метод додатково враховує найбільш популярні товари, що дає можливість формувати рекомендації в режимі онлайн для нових користувачів.

Для імплементації отриманого результату удосконаленого методу було сплановано проект розробки методу гібридної фільтрації.

В роботі виконано експериментальну перевірку гібридного методу на наборі даних із оптових продажів товарів.

Перевірка показала, що точність рекомендацій підвищилась на 2,4%.

Результати кваліфікаційної роботи представлені у матеріалах 25 міжнародного молодіжного форуму «Радіоелектроніка та молодь у XXI столітті», 2021 р.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Методичні вказівки щодо розробки та оформлення кваліфікаційної роботи (для студентів усіх форм навчання другого (магістерського) рівня вищої освіти спеціальності 122 Комп'ютерні науки освітньо-професійної програми «Інформаційні управляючі системи та технології»/ Петров К.Е., Левикін В.М., Чалий С.Ф., Євланов М.В., Саєнко В.І., Міхнов Д.К., Міхнова А.В., Чала О.В. Харків. ХНУРЕ, 2021. – 30 с.
2. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Моделювання контексту в рекомендаційних системах. *Науковий журнал «Проблеми інформаційних технологій»*, 2018, №. 1(023). С. 21-26.
3. Чалий С.Ф., Прібильнова І.Б., Ситуаційна модель користувачького вибору в рекомендаційній системі. *Системи управління, навігації та зв'язку*. – 2019. – Вип. 2(54). – С.159-163.
4. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Моделювання пояснень щодо рекомендованого переліку об'єктів з урахуванням темпорального аспекту вибору користувача. *Системи управління, навігації та зв'язку*, 2019. Том 6 № 58. С. 97-101.
5. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Доповнення вхідних даних рекомендаційної системи в ситуації циклічного холодного старту з використанням темпоральних обмежень типу «NEXT». *Системи управління, навігації та зв'язку*, 2019. Вип. 4(56). С. 105-109.
6. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Модель інтерфейсу пояснень з темпоральними параметрами в рекомендаційній системі. *Системи управління, навігації та зв'язку*. 2020. Вип. 2(60). С. 105-109.
7. Чалий С.Ф., Лещинський В.О., Лещинська І.О., Багаторівнева персоналізація пояснень в рекомендаційних системах. *Сучасні інформаційні системи*. 2020. Том 4, № 2. С. 170-175.

8. Чалий С.Ф., Лещинський В.О., Лещинська І.О., Концепція формування пояснень в рекомендаційних системах за принципом білого ящика. *Збірник наукових праць*. – Полтава: ПНТУ, 2019. – Т. 3 (55). – С. 156-160.

9. Gunes I., Kaleli C., Bilge A., Polat H. Shilling attacks against recommender systems: a comprehensive survey. *Artificial Intelligence Review*, 2014. 42, 767–799.

10. Chala O., Novikova L., Chernyshova L., Method for detecting shilling attacks in e-commerce systems using weighted temporal rules. *EUREKA: Physics and Engineering*. 2019. Vol. 5, 29-36. DOI: 10.21303/2461-4262.2019.00983

11. Elahi M., Ricci F., Rubens N. A survey of active learning in collaborative filtering recommender systems. *Computer Science Review*. 2016, 29-50.

12. Xiang L., Yuan Q. Temporal Recommendation on Graphs via Long-and Short-term Preference Fusion. *KDD'10 of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2010,723-732.

13. Chalvi S., Pribylnova I. The method of constructing recommendations online on the temporal dynamics of user interests using multilayer graph. *EUREKA: Physics and Engineering*. 2019. Vol. 3. P. 13-19.

14. Zhu Y., Lin J., He S., Wang B., Guan Z., Liu H., and Cai D. Addressing the item cold-start problem by attribute-driven active learning,”. *arXiv* 2018, Vol. 3. P. 12-19.

15. Braunhofer M. Hybrid Solution of the Cold-Start Problem in Context-Aware Recommender Systems. *UMAP 2014: User Modeling, Adaptation, and Personalization*. 2014, 484-489.

16. Son Le Hoang. Dealing with the new user cold-start problem in recommender systems: A comparative review. *Information Systems*, 58, 2016, 87-26. Aggarwal C. Recommender Systems. *The Textbook*. New York, 2017, 498 p.

17. Lee, D., Park, S. E., Kahng, M., Lee, S. and Lee, S. Exploiting contextual information from event logs for personalized recommendation. *In Computer and Information Science*. 2010, pp. 121–139.
18. Chalyi S., Leshchynskiy V. Knowledge Representation in the Recommendation System Based on the White Box Principle. *Сучасні інформаційні системи*. 2019. Т. 3, № 3 С. 82-86.
19. Lee, J. S. and Lee, J. C. Music for my mood: A music recommendation system based on context reasoning. *In Smart sensing and context*. 2006, pp. 190–203.
21. Chalyi S., Leshchynskiy V., Leshchynska I. Detailing explanations in the recommender system based on matching temporal knowledge. *Eastern-European Journal of Enterprise Technologies* . 2020, Vol 4, No 2 (106). P. 6-13
22. Elahi M., Ricci F., Rubens N. A survey of active learning in collaborative filtering recommender systems. *Computer Science Review*. 2016, 29-50.
23. Chalyi S., Levykin I., Biziuk A., Vovk A., Bogatov I. Development of the technology for changing the sequence of access to shared resources of business processes for process management support. *Eastern-European Journal of Enterprise Technologies*. 2020. P. 22-29.
24. Claypool, M., P. Le, M. Wased and D. Brown, Implicit interest indicators. *Proceeding of International Conference on Intelligent User Interfaces*, 2001, pp: 33-40.
25. Kalynychenko, O., Chalyi, S., Bodyanskiy, Y., Golian, V., Golian, N. Implementation of search mechanism for implicit dependences in process mining. *IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)*. 2013, pp 25-28.
26. Herlocker, J.L., J.A. Konstan, A. Borchers and J. Riedl, An algorithmic framework for performing collaborative filtering. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. Berkeley, California, USA, 1999, pp: 230-237.

27. Guo, G. Improving the performance of recommender systems by alleviating the data sparsity and cold start problems. *In Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, AAAI Press, 2013*, pp. 3217-3218.
28. Lika, B., Kolomvatsos, K. and Hadjiefthymiades, S. Facing the cold start problem in recommender systems. *Expert Systems with Applications*. Vol. 41, 2014, pp. 2065–2073.
29. Bobadilla, J., Ortega, F., Hernando, A. and Gutiérrez, Recommender systems survey. *Knowledge-Based Systems*, Vol. 46, 2013, pp. 109–132.
30. Об'єктнозорієнтований аналіз і проектування ІКС на основі UML : навч. посібник / В. П. Борисенко, В. М. Левикін, Ю. В. Пономарьов, Т. І. Борисенко; МОН України, Наук.-метод. центр вищої освіти, Харківськ. націон. ун-т радіоелектроніки. Харків: ХНУРЕ, 2004. 80 с.
31. Суботіна Н. В. Дослідження гібридних методів формування рекомендацій в проектах побудови рекомендаційних систем. Матеріали 25 Міжнародного молодіжного форуму «Радіоелектроніка та модуль у XXI столітті». Харків: ХНУРЕ. 2021. С. 88-89.