

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ І. І. МЕЧНИКОВА  
(повне найменування закладу вищої освіти)

Факультет математики, фізики та інформаційних технологій  
(повне найменування факультету)

Кафедра алгебри, геометрії та диференціальних рівнянь  
(повна назва кафедри)

## Кваліфікаційна робота

на здобуття ступеня вищої освіти «магістр»

**«Розробка системи рекомендацій з використанням машинного навчання»**  
(тема кваліфікаційної роботи українською мовою)

**«Development of a recommendation system using machine learning»**  
(тема кваліфікаційної роботи англійською мовою)

Виконала: здобувачка денної форми навчання  
спеціальності 123 – Комп'ютерна інженерія  
(код, назва спеціальності)

Освітня програма Комп'ютерна інженерія  
(назва)

Пікуш Вікторія Андріївна

(прізвище, ім'я, по-батькові здобувача)

Керівник к.ф.-м.н., доц. Савастру О.В. \_\_\_\_\_  
(науковий ступінь, вчене звання, прізвище, ініціали) (підпис)

Рецензент к.т.н., доц. Якімова Н.А.  
(науковий ступінь, вчене звання, прізвище, ініціали)

Рекомендовано до захисту:  
Протокол засідання кафедри

№ \_\_\_\_\_ від \_\_\_\_\_ . \_\_\_\_\_ . 2023 р.

Завідувач кафедри

\_\_\_\_\_ Євтухов В. М.  
(підпис) (прізвище, ім'я)

Захищено на засіданні ЕК № \_\_\_\_\_  
протокол № \_\_\_\_\_ від \_\_\_\_\_ . \_\_\_\_\_ . 2023 р.

Оцінка \_\_\_\_\_ / \_\_\_\_\_ / \_\_\_\_\_  
(за національною шкалою/шкалою ECTS/ бали)

Голова ЕК

\_\_\_\_\_ \_\_\_\_\_  
(підпис) (прізвище, ім'я)

Одеса 2023

## АНОТАЦІЯ

Кваліфікаційна робота присвячена темі «Розробка системи рекомендацій з використанням машинного навчання».

Сучасний світ переповнений інформацією, і однією з ключових задач є її структуризація та представлення користувачам у зручній формі. Системи рекомендацій є невід'ємною частиною багатьох сервісів, починаючи від електронної комерції до стрімінгових платформ.

Метою даної роботи є розробка системи рекомендацій, що використовує методи машинного навчання для підвищення точності та персоналізації пропозицій.

У дослідженні було проведено аналіз існуючих рекомендаційних систем та обрано оптимальні алгоритми машинного навчання, які були адаптовані для розробки запропонованої системи. Підготовка даних для тренування моделі включала вибір та очищення даних, що є критично важливим етапом для забезпечення точності рекомендацій.

В результаті проведених в роботі досліджень розроблено модель з гібридним підходом, що поєднує колаборативну та вмістову фільтрацію. Для покращення ефективності системи запропоновано використання аналізу демографічних даних. Такий підхід дозволив підвищити точність системи порівняно з класичними методами фільтрації.

Навчання та тестування системи відбувалося на даних датасету, отриманих з платформи Kaggle, що дозволило оцінити ефективність моделі. Реалізацію системи рекомендацій книг виконано у вигляді веб-застосунку.

Висновки дослідження підтвердили, що використання запропонованого методу забезпечує значне покращення у точності та персоналізації рекомендацій. Розроблена система відзначається гнучкістю, масштабованістю та може бути інтегрована у широкий спектр платформ електронної комерції, що робить її високо конкурентоспроможною на ринку рекомендаційних систем.

## **ABSTRACT**

The qualification work is devoted to the topic "Development of a recommendation system using machine learning."

The modern world is full of information, and one of the key tasks is its structuring and presentation to users in a convenient form. Recommender systems are an integral part of many services, from e-commerce to streaming platforms.

The purpose of this work is to develop a recommendation system that uses machine learning methods to increase accuracy and personalize offers.

The research analyzed existing recommender systems and selected optimal machine learning algorithms that were adapted to develop the proposed system. Data preparation for model training included data selection and cleaning, which is a critical step to ensure the accuracy of the recommendations.

As a result of the research conducted in the work, a model with a hybrid approach combining collaborative and content filtering was developed. To improve the efficiency of the system, the use of demographic data analysis is proposed. This approach made it possible to increase the accuracy of the system compared to classical filtering methods.

The training and testing of the system took place on the dataset data obtained from the Kaggle platform, which allowed to evaluate the effectiveness of the model. The book recommendation system was implemented in the form of a web application.

The research findings confirmed that the use of the proposed method provides a significant improvement in the accuracy and personalization of recommendations. The developed system is characterized by flexibility, scalability and can be integrated into a wide range of e-commerce platforms, which makes it highly competitive in the recommender system market.

## ЗМІСТ

	Стор.
ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ ТА ТЕРМІНІВ .....	5
ВСТУП.....	6
1 ОГЛЯД ІСНУЮЧИХ СИСТЕМ ТА АЛГОРИТМІВ РЕКОМЕНДАЦІЙ .....	9
1.1 Існуючі сервіси з рекомендаціями.....	9
1.2 Основні підходи до розробки рекомендаційних систем .....	14
1.3 Опис алгоритмів машинного навчання .....	19
1.4 Уточнення постановки задачі .....	23
2 ПРОЄКТУВАННЯ СИСТЕМИ РЕКОМЕНДАЦІЙ .....	25
2.1 Проєктування БД.....	26
2.2 Проєктування системи .....	28
2.3 Модуль рекомендацій .....	30
2.4 Засоби реалізації.....	32
3 РЕАЛІЗАЦІЯ СИСТЕМИ РЕКОМЕНДАЦІЙ.....	37
3.1 Опис класів, системи та моделей.....	37
3.2 Опис роботи системи .....	38
4 ТЕСТУВАННЯ СИСТЕМИ.....	51
ВИСНОВКИ .....	56
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ .....	58

**ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ ТА ТЕРМІНІВ**

МН – машинне навчання

КФ – колаборативна фільтрація

ВФ – вмістова фільтрація

RMSE – середньоквадратична помилка (Root Mean Square Error)

MAE – середня абсолютна помилка (Mean Absolute Error)

API – інтерфейс програмування додатків (Application Programming Interface)

SQL – мова структурованих запитів (Structured Query Language)

БД – база даних

SVD – сингулярне розкладення (Singular Value Decomposition)

ALS – чергування найменших квадратів (Alternating least squares)

## ВСТУП

У епоху діджиталізації, коли користувачі зазнають інформаційного перенасичення, ефективні рекомендаційні системи виступають як важливий інструмент для персоналізації контенту. Здатність точно передбачати інтереси та побажання користувачів стала ключовою перевагою в багатьох сферах, від розваг та медіа сервісів до освіти та бізнесу.

Рекомендаційні системи використовують у багатьох сферах:

- 1) електронна комерція та бізнес: для персоналізації пропозицій товарів та послуг відповідно до інтересів та поведінки користувачів, прогнози та рекомендації інвестицій;
- 2) медіа та розваги: персоналізація стрімінгових сервісів, рекомендації музики, фільмів, ігор, місць відпочинку;
- 3) соціальні мережі: підбір вмісту, який відображається в стрічці новин користувача, рекомендації друзів та груп;
- 4) освіта: персоналізовані навчальні ресурси, матеріали та курси відповідно до інтересів та знань студентів;
- 5) охорона здоров'я: персоналізація медичних рекомендацій, підбір лікування та рекомендації щодо здорового способу життя.

На сьогоднішній день вже багато компаній таких як Google, Amazon, Facebook, Spotify використовують рекомендаційні системи для того, щоб допомагати своїм користувачам з нелегким вибором представлених продуктів. Таким чином відкриваються широкі перспективи для підвищення ефективності бізнесу, наукових досліджень і навчального процесу, підвищується задоволеність користувачів, збільшуються продажі та покращується взаємодія користувачів з продуктами та послугами.

Не зважаючи на значну кількість існуючих систем, актуальність розробки завжди буде високою, оскільки обумовлена зростаючими обсягами даних та потребою у їх ефективній обробці з метою підвищення задоволеності

користувачів і оптимізації бізнес-процесів. Є ряд актуальних проблеми, з якими стикаються рекомендаційні системи [1]:

- 1) холодний старт: визначення персоналізованих рекомендацій для нових користувачів або продуктів, де історія взаємодії відсутня;
- 2) обмеження різноманітності: рекомендація тільки популярного контенту, ігноруючи менш відомі, але потенційно цікаві варіанти;
- 3) фільтрова бульбашка: тенденція систем до обмеження рекомендацій контентом, який відповідає існуючим перевагам користувача, ігноруючи альтернативні або різноманітні точки зору;
- 4) обробка великих даних: забезпечення швидкого та ефективного аналізу великих наборів даних з урахуванням збільшення кількості користувачів та контенту;
- 5) точність та релевантність: підтримка високої точності рекомендацій, що відповідають унікальним інтересам та потребам користувачів;
- 6) упередження та приватність: уникнення упередженості в рекомендаціях та забезпечення приватності даних користувачів;
- 7) взаємодія з користувачами: залучення користувачів для отримання зворотного зв'язку та покращення якості рекомендацій.

Останніми роками було розроблено ряд рішень для поліпшення рекомендацій. Ці рішення включають використання машинного навчання, впровадження алгоритмів глибинного навчання, які дозволяють виявляти складні шаблони у поведінці користувачів, а також використання методів обробки природної мови для кращого розуміння контексту та інтересів користувачів. Також активно розвиваються гібридні системи, які комбінують різні типи фільтрації (колаборативну, контент-орієнтовану та інші), що дозволяє реалізувати більш точні та ефективні системи рекомендацій. Точність цих рішень в значній мірі залежить від якості та обсягу використовуваних даних, а також від вибраних моделей і алгоритмів до конкретних задач. Належна підготовка даних, усунення упередженості в

датасетах і тонке налаштування параметрів моделей машинного навчання є ключовими факторами для підвищення точності [2].

Дана робота присвячена розробці системи рекомендацій для книжкового магазину, яка інтегрує методи машинного навчання з метою підвищення точності, персоналізації та релевантності пропозицій для кінцевих користувачів. Основну увагу буде зосереджено на покращенні методів для вирішення проблеми точності персоналізованих та релевантних рекомендацій. Це включає аналіз різних типів даних користувачів, вивчення методів машинного навчання для прогнозування користувацьких інтересів та вподобань, а також оцінку ефективності різних підходів до рекомендацій. Важливість цієї теми зростає з кожним роком у зв'язку з постійним нарощуванням обсягів даних та розвитком цифрових технологій, що вимагає все більш ефективних методів для аналізу та обробки цієї інформації.

Мета даної роботи полягає у розробці ефективної системи рекомендацій, шляхом реалізації підходящого алгоритму та його покращенні, що допоможе більш точно прогнозувати переваги користувачів і пропонувати їм відповідні продукти або послуги.

Предметом дослідження є методи машинного навчання, алгоритми рекомендацій та шляхи їх покращення.

Об'єктом дослідження є процеси взаємодії користувачів із платформами, де використовуються системи рекомендацій.

Для реалізації мети необхідно виконати наступні задачі:

- 1) глибокий аналіз існуючих систем, методів та алгоритмів машинного навчання для розробки систем рекомендацій;
- 2) вибір та обґрунтування методів машинного навчання;
- 3) збір та підготовка даних для тренування моделі;
- 4) розробка моделі машинного навчання для системи рекомендацій;
- 5) тестування та оцінка ефективності розробленої системи;
- 6) розробка інтерфейсу користувача для інтеграції системи рекомендацій в платформу електронної комерції.

# 1 ОГЛЯД ІСНУЮЧИХ СИСТЕМ ТА АЛГОРИТМІВ РЕКОМЕНДАЦІЙ

## 1.1 Існуючі сервіси з рекомендаціями

Використання рекомендаційної системи може підвищити різні показники ефективності, включаючи обсяг проданих товарів, загальний обсяг продажів, прибутковість, середню тривалість візиту користувача на сайті, вірність клієнтів та інші. Тож рекомендаційні системи вже давно використовуються, особливо в сфері електронної комерції та цифрових сервісів. Вони стали особливо популярними на початку 2000-х, коли компанії, такі як Amazon і Netflix, почали активно впроваджувати їх для персоналізації рекомендацій для своїх користувачів [1]. З тих пір технології рекомендаційних систем значно розвинулися, і вони використовуються в широкому спектрі індустрій для підвищення продажів, покращення користувацького досвіду та забезпечення більш цілеспрямованого маркетингу. На сьогоднішній день одні з найбільших розробок у галузі систем рекомендацій були здійснені наступними компаніями: Amazon, Netflix, Google, Facebook, Spotify, TikTok.

### 1.1.1 Amazon

Один із найвідоміших сервісів, який активно використовує системи рекомендацій – це Amazon. Цей гігант електронної комерції базується на аналізі великого обсягу даних про покупки та перегляди товарів користувачами.

Їх система рекомендацій базується на алгоритмі колаборативної фільтрації товарів. Замість зосередження на схожості між клієнтами, вона використовує кореляції між товарами. Якщо покупці, які купують товар А, також схильні купувати товар Б, то ці два товари вважаються пов'язаними. Алгоритм використовує метрику відносності, засновану на різниці у ймовірностях покупки для формування рекомендацій [3-4].

Переваги:

- Amazon використовує величезні масиви даних про покупки та перегляди для точних рекомендацій;
- здатність адаптувати рекомендації до індивідуальних переваг користувачів.

Недоліки:

- рекомендації можуть бути упереджені через надмірне фокусування на популярних продуктах;
- занепокоєння щодо збору та використання персональних даних користувачів.

### 1.1.2 Netflix

Netflix є одним із лідерів у використанні систем рекомендацій для пропонування фільмів та телесеріалів. Система рекомендацій Netflix базується на аналізі історії переглядів користувачів, інших виборах користувачів із схожими смаками, інформації про конкретні титули, використаного пристрою та часу перегляду. Сервіс використовує різні технології та моделі машинного навчання, включаючи підсилувальне навчання, нейронні мережі, каузальне моделювання, ймовірнісні графічні моделі та інші [5-6].

Переваги:

- використання складних моделей з алгоритмами глибинного навчання для більш точного прогнозування переваг;
- аналіз змісту фільмів та серіалів (контент-орієнтовані рекомендації) для вдосконалення рекомендацій.

Недоліки:

- рекомендації обмежені контентом, доступним на платформі;
- фільтраційна бульбашка: ризик втрати новизни та різноманітності в рекомендаціях.

### 1.1.3 Spotify

Spotify – це музичний стрімінговий сервіс, який використовує системи рекомендацій для створення персоналізованих плейлистів та рекомендацій для користувачів. Система рекомендацій Spotify включає контент-базову фільтрацію, що аналізує метадані треків, аудіоаналіз та аналіз текстів за допомогою NLP моделей. Вона також включає колаборативну фільтрацію, фокусуючись на схожості організації треків, наприклад, на основі їхньої спільної присутності в плейлистах користувачів. Профілювання користувачів в Spotify ґрунтується на активному та пасивному зворотному зв'язку від користувачів [7-8].

Переваги:

- Постійна адаптація плейлистів до музичних смаків користувачів;
- використання комбінацій колаборативної фільтрації та аналізу аудіо-файлів.

Недоліки:

- точність рекомендацій залежить від активності користувача;
- проблема холодного старту: виклики з наданням рекомендацій новим користувачам.

### 1.1.4 Google

Через свої численні сервіси, включаючи Пошук Google, YouTube та Google Play, компанія використовує складні системи рекомендацій, які допомагають користувачам знаходити релевантний контент. Наприклад YouTube є однією з найбільших та найскладніших промислових систем рекомендацій.

Система використовує глибокі нейронні мережі для генерації кандидатів та окрему глибоку модель ранжування. Система зосереджена на значному впливі користувачів [9].

Переваги:

- розширені алгоритми пошуку, аналіз контенту, персоналізовані рекомендації;
- підвищення задоволеності користувачів та релевантності результатів пошуку.

Недоліки:

- фільтраційна бульбашка: ризик втрати новизни та різноманітності в рекомендаціях.

### **1.1.5 Facebook**

Facebook використовує системи рекомендацій для показу відповідного контенту в стрічці новин, а також для рекомендацій друзів, груп та реклами. Facebook використовує колаборативну фільтрацію та різні алгоритми для рекомендацій, включаючи стохастичний градієнтний спуск та чергуючий найменший квадрат. Facebook стикається з великими наборами даних, що містять понад мільярд користувачів та мільйони предметів, що ставить виклики перед розробкою розподілених алгоритмів [10].

Переваги:

- обробка великих обсягів даних;
- поєднання колаборативної фільтрації з іншими алгоритмами.

Недоліки:

- занепокоєння щодо збору та використання персональних даних користувачів.

### **1.1.6 Goodreads**

А найпопулярнішим сервісом для рекомендацій книг станом на 2023 рік є Goodreads. Це велика соціальна мережа для любителів книг, яка дозволяє користувачам відслідковувати та рецензувати книги, а також отримувати

персоналізовані рекомендації. Сервіс використовує інформацію про книги, які вже прочитав користувач, його оцінки та відгуки, а також переваги інших користувачів зі схожими інтересами для генерації рекомендацій. Goodreads також пропонує різноманітні функції спільноти, такі як обговорення книг, віртуальні книжкові клуби та можливість взаємодії з авторами [11].

Переваги:

- немає проблем з холодним стартом;
- надає досить хороші персоналізовані рекомендації, що точно відповідає смаку користувача.

Недоліки:

- упередженість рекомендацій;
- фільтраційна больбашка;
- є незручності у тому, що користувач постійно повинен сам заносити інформацію про прочитані книги;
- немає можливості купити книгу.

### **1.1.7 Підсумки**

Ці компанії інтенсивно інвестують у дослідження та розробку в галузі штучного інтелекту та машинного навчання, що дозволяє їм створювати все більш точні та ефективні системи рекомендацій. Усі ці сервіси здатні забезпечити значне полегшення життя користувачів, але вони всі різняться своїми послугами та націлені на різні результати. Також кожна з систем мають свої обмеження та проблеми.

Тож для розробки високопродуктивної рекомендаційної системи необхідно не тільки дослідити існуючі методи, а й підібрати підходящі алгоритми для обраної предметної області, провести тестування та оцінити їх ефективність.

Для покращення системи рекомендацій книг можна розширити та удосконалити алгоритми машинного навчання, поєднуючи різні підходи та

спираючись на індивідуальні дані про користувачів, що допоможе краще враховувати інтереси користувачів, знижуючи залежність від популярності книг.

## 1.2 Основні підходи до розробки рекомендаційних систем

Незважаючи на велику різноманітність існуючих алгоритмів їх можна поділити на три групи (рис 1.1).

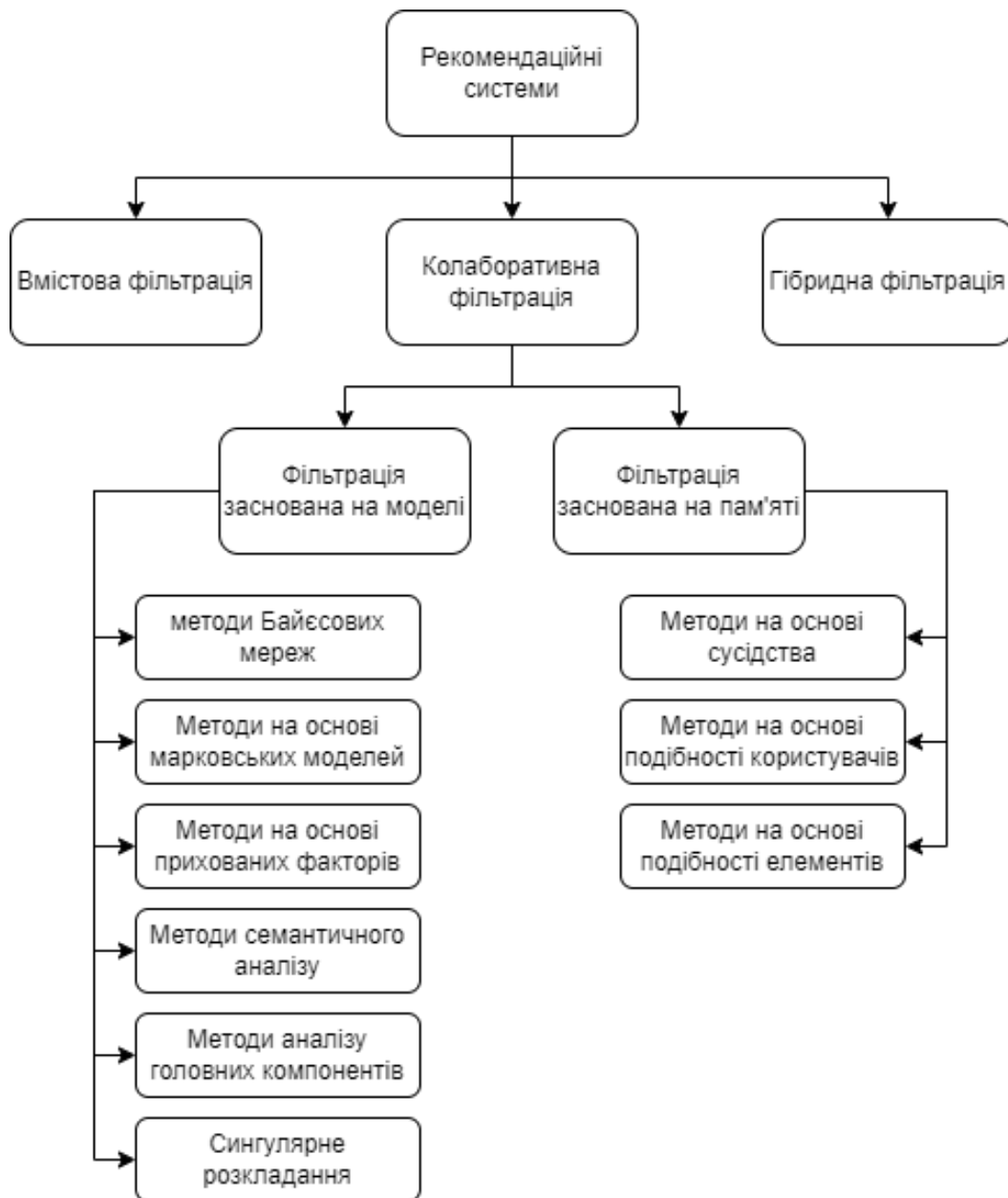


Рисунок 1.1 – Класифікація методів рекомендаційних систем

Однією з ключових складових систем рекомендацій є методи, які використовуються для прогнозування інтересів користувачів та надання рекомендацій. На сьогодні існує чимало підходів, однак для рекомендації книг деякі підходи можуть бути більш ефективними, ніж інші. Тому розглянемо ті, які зазвичай вважаються найбільш підходящими для подібних задач визначимо їхні переваги та недоліки.

### **1.2.1 Колаборативна фільтрація**

Колаборативна фільтрація – це один із найпоширеніших методів рекомендацій, який базується на історії взаємодії користувачів з системою та на спільних інтересах та виборах інших користувачів з подібними уподобаннями. Вона поділяється на дві основні категорії: колаборативна фільтрація за споживачами (user-based) та за об'єктами (item-based) [1, 12]. Дана методика особливо ефективна для книжкових рекомендацій, оскільки вона допомагає виявити книги, які сподобаються користувачам на основі вибору інших зі схожими смаками.

Переваги колаборативної фільтрації:

- ефективність для нових користувачів, що не вимагає великої кількості даних про користувача;
- здатна надавати персоналізовані рекомендації на основі взаємодії користувачів.

Недоліки колаборативної фільтрації:

- проблема холодного старту, важко робити рекомендації новим користувачам та новим об'єктам;
- проблема розрідженості даних, особливо у великих системах дуже важко знайти користувачів, які оцінили однакові товари;
- відсутність рекомендацій для користувачів з унікальним смаком.

### 1.2.2 Змістовна фільтрація

Змістовна фільтрація використовує аналіз змісту об'єктів та порівнює їх з профілем користувача для створення рекомендацій [1, 12]. Така фільтрація ефективна для книг, оскільки може аналізувати атрибути книг, такі як жанр, автор, тематика, стиль написання тощо, для визначення подібності між книгами та інтересами користувачів. Цей метод покладається на описові характеристики об'єктів та інтереси користувача.

Переваги змістовної фільтрації:

- немає проблем з холодним стартом;
- не вимагає великої історії користувача;
- може надавати рекомендацій користувачам з унікальними смаками;
- немає упередженості, може рекомендувати навіть непопулярну товари;
- простота інтерпретації: рекомендації базуються на описових характеристиках об'єктів.

Недоліки змістовної фільтрації:

- проблема селективності фільтрації: надає рекомендації, які дуже схожі на вже переглянуті об'єкти;
- не враховує суб'єктивних вподобань користувачів: не завжди враховує усі смаки користувачів.

### 1.2.3 Гібридні Моделі

Гібридні моделі комбінують колаборативну фільтрацію та змістову фільтрацію для створення більш точних та персоналізованих рекомендацій. Цей підхід дозволяє подолати обмеження обох методів і забезпечити більш широкий спектр рекомендацій [1, 13, 14].

Гібридні системи комбінація колаборативної та контент-орієнтованої фільтрацій може бути особливо ефективною, оскільки вона об'єднує переваги обох підходів, забезпечуючи більш точні та релевантні рекомендації.

Переваги гібридних моделей:

- гібридні моделі можуть забезпечити кращу точність рекомендацій;
- можуть працювати з різноманітними типами даних та об'єктів.

Недоліки гібридних моделей:

- гібридні моделі можуть бути складними для реалізації та оптимізації;
- для гібридних моделей часто потрібні великі обсяги даних.

#### **1.2.4 Глибинне навчання**

Глибинне навчання є новим напрямком у рекомендаційних системах, який використовує нейронні мережі для аналізу даних та створення рекомендацій [15]. Воно відкриває нові можливості для створення більш точних та персоналізованих рекомендацій. Однак, це також несе в собі певні виклики, такі як потреба в значних обчислювальних ресурсах та складності інтерпретації результатів.

Глибинні нейронні мережі використовуються для створення складних моделей рекомендацій. Вони можуть враховувати багато ознак і послідовностей, що поліпшує якість рекомендацій.

Рекурентні нейронні мережі (RNN) допомагають аналізувати послідовності даних. Наприклад, вони можуть дивитися на історію переглядів фільмів та вибирати, що рекомендувати далі.

Згорткові нейронні мережі (CNN) ефективні для роботи з зображеннями та текстом, наприклад, вони можуть допомогти розібратися, які картинки або описи продуктів найбільше подобаються користувачам.

Переваги глибинного навчання:

- моделі здатні до узагальнення, вивчення складних залежностей;

- добре працює на великих наборах даних.

Недоліки глибинного навчання:

- вимагає значних обчислювальних ресурсів та даних.
- глибинні моделі можуть бути важко інтерпретовані.

### **1.2.5 Додаткові підходи**

Окрім основних вже розглянутих підходів існують ще декілька корисних варіантів, поєднуючи з якими можна досягти більш точних результатів. Коротко їх розглянемо [1, 2, 12].

Системи, базовані на знаннях. Ці системи можуть бути корисними для врахування конкретних запитів або вимог користувачів, таких як пошук книг певної тематики або стилю.

Секвенційні рекомендації. Цей підхід може бути корисним для рекомендації серій книг або книг, які слід читати у певному порядку.

Рекомендації на основі графів знань. Можуть виявити складні зв'язки між книгами, авторами, жанрами тощо, що дозволяє створювати більш глибокі та контекстуалізовані рекомендації.

Демографічні рекомендації. Використовують демографічну інформацію користувачів (наприклад, вік, стать, освіту) для рекомендації книг, які можуть сподобатися певній демографічній групі.

Соціальні рекомендації. Враховують вплив соціальних мереж та рекомендації від друзів або інших користувачів, яким довіряє користувач.

Рекомендації на основі візуальних даних. Хоча цей підхід менш традиційний для книг, він може бути корисним для рекомендації книг з особливими обкладинками або ілюстраціями.

Рекомендації на основі афективного обчислення. Використовують аналіз емоційних станів користувачів, наприклад, рекомендуючи розслаблюючі книги, коли користувач під стресом.

Експертні системи рекомендацій. Використовують думки та оцінки літературних експертів для рекомендації книг, особливо корисні для спеціалізованої або академічної літератури.

Мультикритеріальні рекомендації. Враховують декілька факторів одночасно, наприклад, популярність книги, її рейтинг, ціну, жанр тощо, для формування рекомендацій.

### **1.2.6 Підсумки**

Для створення ефективної системи рекомендації книг, поєднання декількох підходів може бути найкращим рішенням, оскільки це дозволить врахувати різні аспекти вибору книг і інтересів читачів. Будемо використовувати гібридний підхід, але додамо ще аналіз демографічних даних. За допомогою такого підходу вирішимо ряд проблем:

Колаборативна фільтрація буде використовувати інформацію про поведінку та вибір інших користувачів, тоді як контент-орієнтована фільтрація аналізуватиме характеристики книг. Їх комбінація дозволить системі рекомендувати книги, які не тільки популярні серед користувачів зі схожими інтересами, але й мають властивості (жанр, стиль, автор), які відповідають індивідуальним перевагам користувача. У свою чергу демографічні рекомендації дозволяють адаптувати вибір книг до конкретної демографічної групи.

Поєднання цих підходів дозволяє створити більш гнучку та ефективну систему рекомендацій, яка може краще задовольнити різноманітні переваги та інтереси читачів.

## **1.3 Опис алгоритмів машинного навчання**

У цьому розділі розглянемо різні алгоритми машинного навчання, які використовуються для створення систем рекомендацій. Кожен з цих

алгоритмів має свої переваги та недоліки, які варто враховувати при виборі підходу для конкретної задачі рекомендацій.

### 1.3.1 Метод найближчих сусідів (k-Nearest Neighbors, KNN)

Метод найближчих сусідів базується на схожості між користувачами або об'єктами [1, 12, 16]. Він визначає  $k$  найближчих сусідів до цільового користувача або об'єкта і рекомендує ті об'єкти, які сподобалися цим сусідам. KNN може використовуватися для колаборативної фільтрації, де система рекомендує продукти, засновані на вподобаннях схожих користувачів (User-based) або на схожості між продуктами (Item-based) (рис. 1.2).

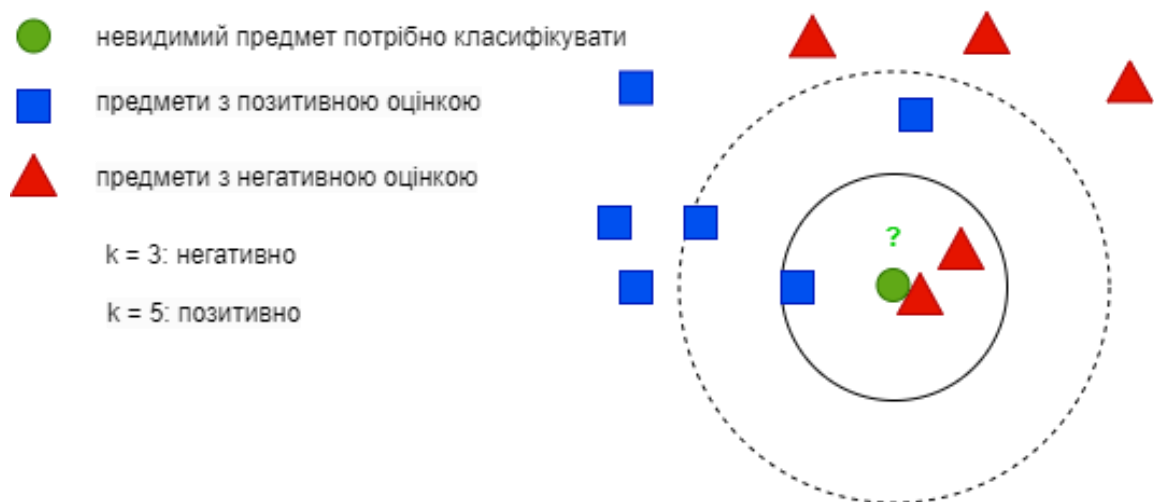


Рисунок 1.2 – Метод найближчих сусідів

Схожість між об'єктами вимірюється за допомогою метрик відстані, таких як Евклідова відстань, косинусна схожість або Манхеттенська відстань. Наприклад, якщо об'єкти — це користувачі, то відстань може визначатися на основі їхніх рейтингів спільних товарів.

Після визначення набору найближчих сусідів рекомендації формуються шляхом агрегування інформації. Наприклад, можна використовувати середні рейтинги товарів, які високо оцінили схожі користувачі.

KNN може бути обчислювально важким, особливо для великих датасетів, тому важливо використовувати методи оптимізації, такі як індексування, зменшення розмірності даних, або паралельні обчислення. А також важливо враховувати фактори, як новизна і різноманітність рекомендацій, щоб уникнути "пастки фільтрації бульбашки" і забезпечити користувачам свіжі та цікаві рекомендації.

Переваги:

- простий в реалізації;
- враховує схожість між об'єктами.

Недоліки:

- велика обчислювальна складність при великій кількості користувачів та об'єктів;
- чутливий до шуму в даних.

### 1.3.2 Метод матричної факторизації

Метод матричної факторизації розкладає матрицю взаємодій користувачів і об'єктів на добуток двох матриць – користувачів і ознак об'єктів (рис. 1.3). Цей процес має на меті виявлення прихованих факторів, які пояснюють спостережувані взаємодії [2, 17].



Рисунок 1.3 – Матрична факторизація

Вихідна матриця, зазвичай, дуже розріджена (з багатьма пропущеними значеннями), оскільки більшість користувачів оцінюють лише невелику кількість продуктів. МФ моделює цю матрицю як добуток двох менших матриць – однієї, що представляє користувачів, і іншої, що представляє продукти. Елементи цих менших матриць відображають приховані фактори. Наприклад, у контексті книг, ці фактори можуть відображати жанри, авторів, стиль тощо. Кількість цих факторів є параметром, який потрібно визначити під час моделювання.

Під час навчання моделі використовується метод оптимізації, такий як стохастичний градієнтний спуск, для мінімізації відстані між фактичними рейтингами та рейтингами, передбаченими моделлю. Зазвичай використовується квадратична функція втрат. Для уникнення перенавчання в моделі часто використовується регуляризація, яка контролює складність моделі.

Після навчання, матрична факторизація може використовуватися для відновлення пропущених значень у вихідній матриці, тобто для передбачення невідомих рейтингів.

Переваги:

- метод добре працює з розрідженими даними;
- здатний враховувати складні взаємодії між користувачами і об'єктами.

Недоліки:

- потребує значної обчислювальної потужності;
- чутливий до перенавчання.

### **1.3.3 Методи ранжування (Ranking)**

Методи ранжування визначають порядок рекомендацій, щоб найкращі об'єкти були вгорі. Це важливо для рекомендаційних систем, які мають велику кількість можливих об'єктів для рекомендації [18].

Weighted Sum Method (WSM): це найпростіший метод, де оцінки різних критеріїв для кожної альтернативи множаться на вагу критерію та сумуються. Використовується, коли всі дані виражені в одиницях виміру.

Weighted Product Method (WPM): схожий на WSM, але замість додавання використовується множення. Це може призвести до проблем із переповненням при обробці великих чисел.

Analytic Hierarchy Process (AHP): заснований на парних порівняннях між альтернативами. Спочатку метод вимагав, щоб відносні значення для кожного критерію давали в сумі одиницю, але пізніше його вдосконалили, пропонуючи ділити кожне відносне значення на максимальне з відносних значень.

Переваги:

- чудово допомагають забезпечити високу релевантність рекомендацій

Недоліки:

- вимагають додаткової обробки результатів.

#### **1.4 Уточнення постановки задачі**

У цій теоретичній частині кваліфікаційної роботи ми провели огляд сучасних підходів та методів у галузі систем рекомендацій. На основі цього огляду можна зробити наступні висновки та уточнення постановки задачі кваліфікаційної роботи.

Метою кваліфікаційної роботи є розробка та імплементація системи рекомендацій з використанням методів машинного навчання та аналізу контенту для покращення персоналізації та релевантності рекомендацій.

Завдання кваліфікаційної роботи:

- зібрати та підготувати вихідні дані для навчання моделі;
- розробити та навчити модель машинного навчання для рекомендацій;
- реалізувати систему рекомендацій, яка враховуватиме демографічні дані, такі як вік, стать, улюблені жанри;

- оцінити точність та ефективність розробленої системи рекомендацій за допомогою відповідних метрик та експериментів;
- провести порівняльний аналіз розробленої системи з існуючими підходами та системами рекомендацій.

Важливість дослідження та розвиток систем рекомендацій є актуальною проблемою в сучасному інформаційному суспільстві. Зростаючий обсяг даних та різноманітність контенту ставлять вимоги до точності та персоналізації рекомендацій. Вирішення цієї проблеми є важливим як для підвищення користувацького досвіду, так і для підтримки бізнес-процесів у різних галузях.

Подальші дослідження можуть включати в себе вдосконалення методів машинного навчання для рекомендацій, розробку нових метрик для оцінки якості рекомендацій, а також вивчення можливостей застосування технологій аналізу контексту та соціальних мереж для покращення рекомендаційних систем.

## 2 ПРОЄКТУВАННЯ СИСТЕМИ РЕКОМЕНДАЦІЙ

У цій кваліфікаційній роботі задумано створити систему рекомендацій для книжкового магазину. Програмний інтерфейс, призначений для взаємодії працівників та покупців із системою, буде розроблений як веб-додаток, що може бути запущений на локальному чи віддаленому сервері.

Система буде побудована на основі клієнт-серверної технології обробки даних, де клієнтські комп'ютери запитують необхідні дані від центрального комп'ютера (сервера), дозволяючи обробку даних як на клієнтському, так і на серверному рівні.

В основу архітектури закладена трирівнева клієнт-серверна структура, що включає клієнтський додаток (часто називаний "тонким клієнтом" або терміналом), сервер застосунків та сервер бази даних, кожен з яких пов'язаний один з одним.

Трирівнева архітектура передбачає три шари програмного забезпечення, що можуть функціонувати на різних платформах: рівень інтерфейсу для кінцевих користувачів, рівень бізнес-логіки та обробки даних, розташований на сервері (іноді відомий як сервер додатка), та рівень бази даних, що може бути реалізований на окремому сервері. На малюнку 2.1 представлено загальний вигляд архітектури системи.



Рисунок 2.1 – трирівнева клієнт-серверна архітектура

## 2.1 Проектування БД

Перший етап проектування полягає у створенні БД, на якому визначаються сутності, їх атрибути та зв'язки між ними. Результатом такого проектування є схема БД, представлена у вигляді ER-діаграми (рис.2.3), на якій відображаються основні сутності предметної області, їх атрибути та зв'язки.

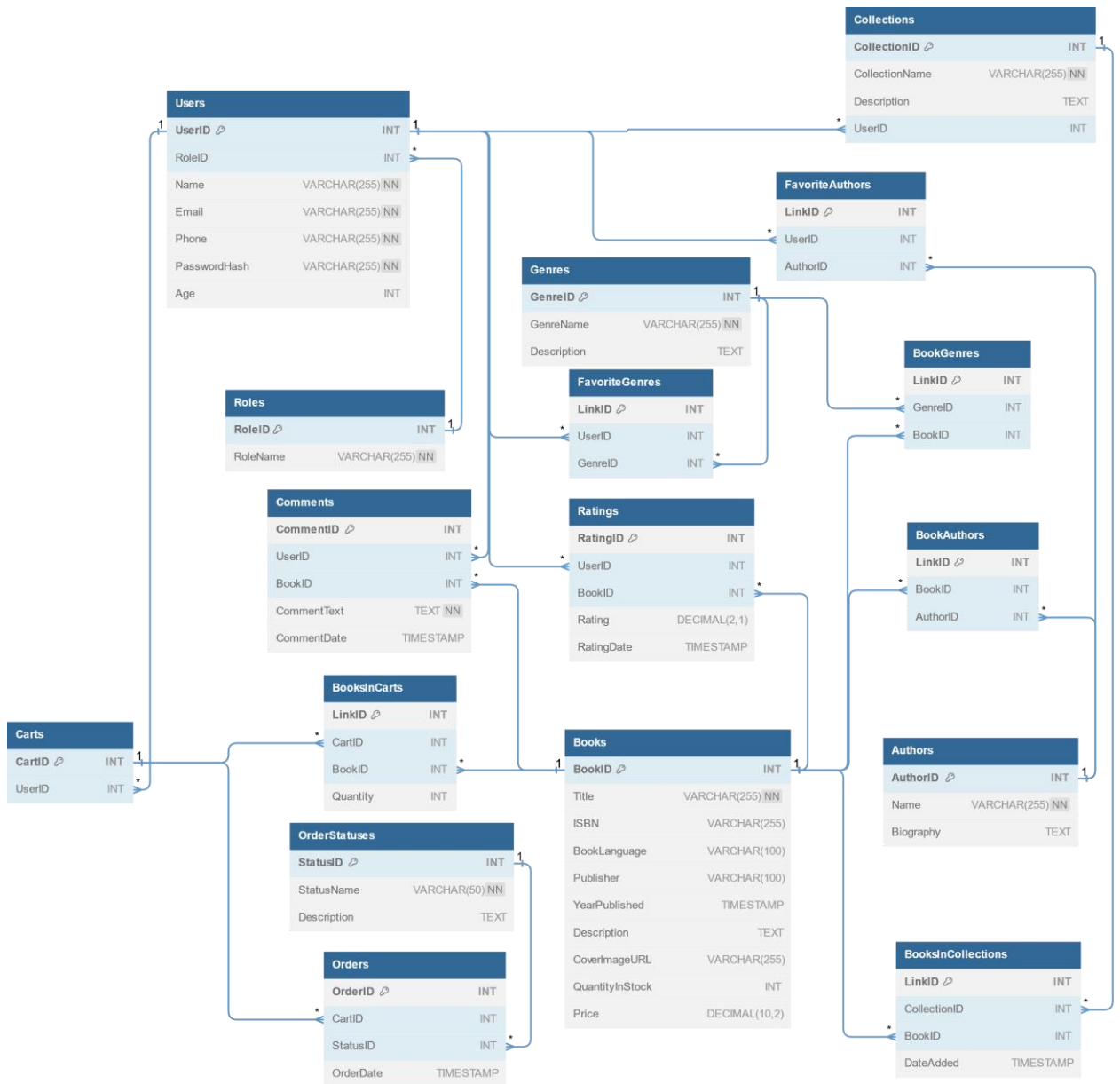


Рисунок 2.2 – ER-діаграма

В інформаційній системі, створеній для книжкового магазину будуть існувати три типи користувачів, кожен з яких матиме свій функціонал:

1) адміністратор:

- додавання книги;
- редагування книги;
- видалення книги;
- перегляд каталогу книг;
- перегляд інформації про конкретну книгу;
- пошук книги;

2) зареєстрований користувач (читач):

- перегляд усього каталогу книг;
- перегляд персональних рекомендацій;
- перегляд інформації про конкретну книгу;
- пошук книги;
- додавання / видалення книги з кошика;
- додавання / видалення книги у колекцію (бажане, в процесі, прочитано);
- покупка книги / перегляд покупок;
- оцінка книги;

3) незареєстрований користувач:

- перегляд каталогу книг;
- перегляд інформації про конкретну книгу;
- пошук книги;
- реєстрація;

У системі варто виділити такі сутності:

- ролі – таблиця з можливими ролями користувачів;
- користувачі – таблиця з інформацією про користувачів;
- книги – таблиця з інформацією про книги;
- автори – таблиця з інформацією про авторів;
- жанри – таблиця з видами жанрів;

- рейтинг – таблиця з оцінками книг від читачів;
- коментарі – таблиця з відгуками про книги від читачів;
- колекції – таблиця з індивідуальними колекціями книг читачів;
- кошик – таблиця з доданими книгами, що планують придбати;
- замовлення – таблиця з усіма замовленнями.

Для забезпечення безпеки даної системи вхід до системи виконується через логін та пароль.

## 2.2 Проектування системи

На рисунку 2.2 представлена схема роботи системи,

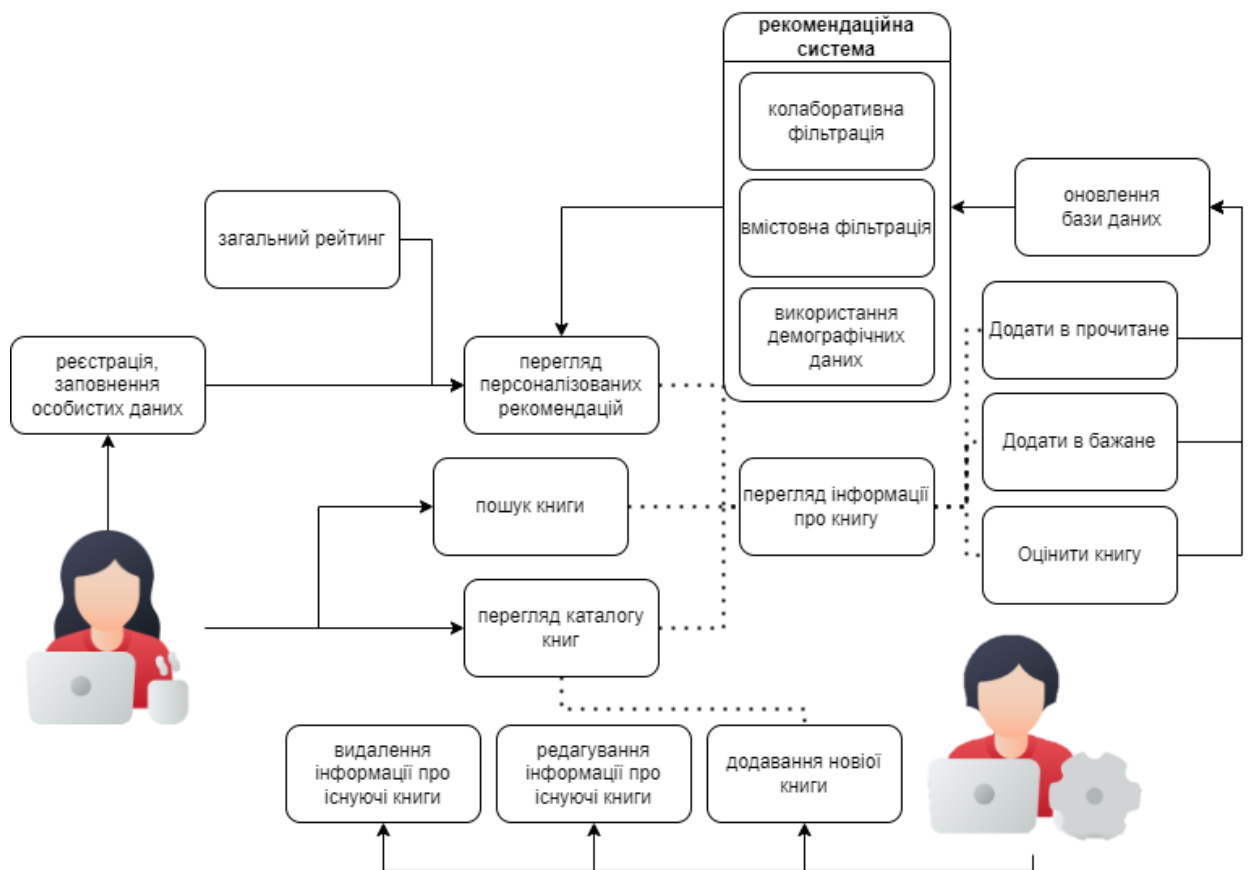


Рисунок 2.3 – схема роботи системи

Розглянемо детальніше етапи роботи системи.

Схема, представляє проект рекомендаційної системи для книжкового магазину. Система працює за таким алгоритмом:

Все починається з обробки файлів даних на якій навчиться модель, вони містять інформацію про книги, їх оцінки та користувачів відповідно. Тому вони послугують загальним рейтингом.

Рекомендаційна система – ядро всього проєкту, яке використовує отримані дані для створення персоналізованих рекомендацій для користувачів. Воно включає кілька підсистем:

- колаборативна фільтрація: використовується для прогнозування інтересів користувача на основі інтересів інших користувачів зі схожими перевагами;
- вмістова фільтрація: для рекомендацій, що базується на атрибутах книг, таких як жанр, автор, або тема;
- використання демографічних даних: це дозволить додавати персоналізацію, використовуючи інформацію, як-от вік, стать, місцеположення користувача, перелік улюблених жанрів чи авторів.

Оновлення бази даних: Регулярне оновлення даних для забезпечення актуальності рекомендацій.

Користувачі взаємодіють з системою через веб-інтерфейс, де вони можуть:

- реєструватися, заповнюючи особистісні дані: це перший крок для входу в систему та отримання персоналізованих рекомендацій;
- переглядати персоналізовані рекомендації: система надає списки рекомендованих книг на основі попередніх дій та переваг користувача;
- пошук книг: можливість шукати книги за назвою, автором, жанром тощо;
- переглядати каталог книг: список усіх доступних книг у магазині (доступно також незареєстрованим користувачам);
- видання інформації про існуючі книги: відображення детальної інформації про кожну книгу;

- редагування інформації про існуючі книги: можливість оновлення даних про книги;
- додавання нової книги: функціонал для розширення асортименту магазину;
- додати в прочитане: можливість користувача позначити книги як прочитані;
- додати в бажане: створення списку бажаних книг для майбутнього придбання;
- оцінити книгу: введення рейтингу книги користувачем, який впливає на майбутні рекомендації.

Ця схема відображає комплексний підхід до розробки рекомендаційної системи, яка інтегрує різні типи фільтрації та персоналізації для надання користувачам найкращого досвіду вибору книг.

### **2.3 Модуль рекомендацій**

Для розробки модуля рекомендацій спочатку необхідно розробити на навчити модель. Для цього був використаний датасет Book-Crossing з порталу Kaggle. Але перед тим, як рекомендаційна система може використовувати дані, їх потрібно підготувати (рис 2.4). Це включає:

- очищення даних: видалення шуму, виправлення помилок, стандартизація форматів;
- трансформація даних: нормалізація, конвертація типів даних, розрахунок додаткових характеристик (feature engineering);
- відбір даних: визначення які дані будуть використовуватись для рекомендацій.

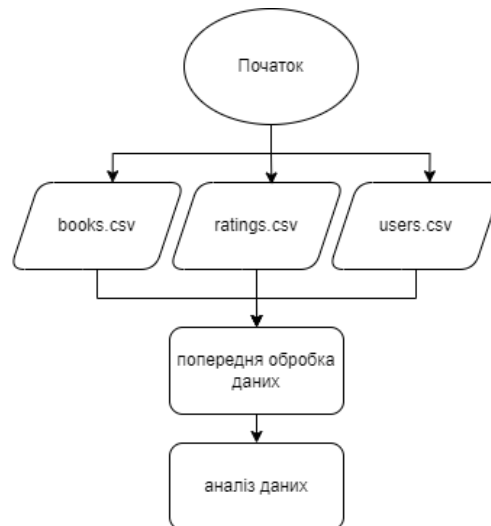


Рисунок 2.4 – Обробка даних для навчання системи

Наступним кроком є розробка самої моделі. У даній роботі пропонується підхід гібридної системи, що буде сполучати у собі колаборативну та вмістову фільтрацію з поєднанням аналізу демографічних даних.

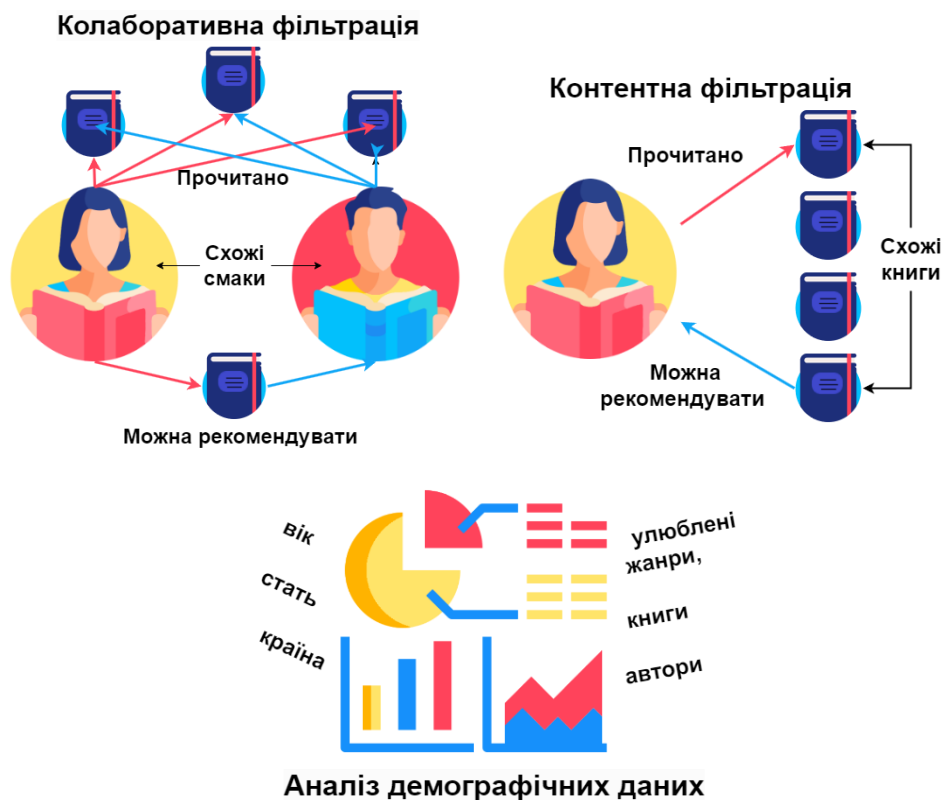


Рисунок 2.5 – Представлення гібридного підходу з поєднання колаборативної та вмістовної фільтрації з використанням аналізу демографічних даних

Колаборативна фільтрація використовує поведінку користувачів, таку як, оцінка книги, покупка книги, додавання у бажане. Це дозволяє генерувати рекомендації за підходом user-based на основі схожості між користувачами. Якщо один користувач полюбляє певні книги, які також подобаються іншому, то книги, які полюбляє другий, можуть бути рекомендовані першому;

Вмістова фільтрація аналізує атрибути продуктів, щоб рекомендувати додаткові продукти зі схожими атрибутами. Наприклад: використання метаданих книги (автор, жанр, рік видання) для знаходження подібних книг;

Використання демографічних даних враховує інформацію, таку як вік, стать, географічне розташування, особисті уподобання, що допоможе у вирішенні проблем з холодним стартом, а також дозволить надавати більш персоналізовані рекомендації.

## 2.4 Засоби реалізації

Для реалізації поставленої задачі були вибрані наступні засоби:

- СУБД PostgreSQL: для зберігання структурованих даних, таких як інформація про книги та користувачів;
- Python: мова програмування для створення серверної частини з алгоритмами машинного навчання;
- Scikit-learn: широкий спектр алгоритмів для машинного навчання в Python,
- Pandas: бібліотека для обробки та аналізу даних;
- NumPy: бібліотека для роботи з багатомірними масивами,
- Flask: фреймворк для створення веб-додатку та інтеграції моделі машинного навчання на Python;
- Bootstrap: для стилізації користувацького інтерфейсу;
- HTML/CSS, JavaScript: для створення зручного та привабливого фронтенду.

Приведемо основні відомості і характеристики кожного з обраних засобів.

### 2.4.1 СУБД PostgreSQL

PostgreSQL – це потужна, відкрита та безкоштовна об'єктно-реляційна система управління базами даних. Вона вирізняється своєю надійністю, гнучкістю та високим рівнем підтримки стандартів SQL. PostgreSQL часто використовується у великих системах, де потрібна складна обробка даних та розширена функціональність [19].

Вибір PostgreSQL для розробки, особливо в контексті складних застосунків, як рекомендаційна система, може бути доцільним через її надійність, розширені можливості SQL та гнучкість у роботі з різними типами даних. Ось кілька ключових переваг PostgreSQL:

- повна підтримка ACID: гарантує надійність транзакцій;
- розширюваність: підтримує різноманітні типи даних, включаючи JSON, XML, HSTORE (для пар ключ-значення), а також підтримує створення користувацьких типів даних;
- підтримка складних запитів SQL: PostgreSQL має потужний механізм запитів, який включає підзапити, об'єднання та віконні функції;
- висока надійність та стабільність: PostgreSQL відомий своєю високою надійністю, що робить його популярним вибором для критичних застосунків;
- підтримка розширень та індексування: підтримує різноманітні способи індексування;
- безпека: надає сильну підтримку безпеки на рівні бази даних, включаючи розширений контроль доступу, шифрування даних тощо;
- активне співтовариство та підтримка: має велике та активне співтовариство, що постійно працює над покращенням та підтримкою продукту.

## 2.4.2 Мова програмування, фреймворки та модулі

Для створення серверної частини веб-додатку та реалізації алгоритмів була вибрана мова Python з модулями Pandas, Numpy та Scikit-learn, так як вони містять повний набір функцій і методів потрібних для створення і навчання моделі рекомендаційної системи.

Python – це високорівнева, інтерпретована мова програмування, створена. Вона відома своєю простотою в читанні та написанні коду, а також за широкий спектр використання – від веб-розробки до наукових досліджень і штучного інтелекту, що ідеально підходить для даної роботи [20].

Ось кілька ключових переваг Python:

- простота коду: легкість у вивченні та читабельність, присутня динамічна типізація, автоматичне керування пам'яттю;
- велика стандартна бібліотека: Python має обширну стандартну бібліотеку, яка охоплює різні області, від веб-розробки до наукових обчислень;
- активне співтовариство: ресурси легко доступні;
- переносимість: Python-програми легко переносити між різними платформами без потреби змінювати код;
- широке використання у науці та аналітиці: Python є популярним вибором у галузях наукових досліджень, обробки даних, машинного навчання та штучного інтелекту завдяки бібліотекам, таким як NumPy, SciPy, pandas, TensorFlow та PyTorch;
- підтримка багатьох парадигм програмування: Python підтримує об'єктно-орієнтоване, процедурне, функціональне програмування;
- використання в розробці веб-додатків: завдяки фреймворкам, таким як Django та Flask, Python ефективний для створення веб-додатків;
- інтеграція з іншими мовами та інструментами: Python можна легко інтегрувати з іншими мовами програмування та системами.

### 2.4.3 Модулі навчання

Scikit-learn: одна з найпопулярніших бібліотек для машинного навчання в Python. Вона містить широкий спектр алгоритмів, включаючи класифікацію, регресію, кластеризацію та зниження розмірності, а також утиліти для попередньої обробки даних, валідації моделі тощо. Підходить для методів рекомендацій, таких як фільтрація на основі контенту або колаборативна фільтрація [21].

Pandas: бібліотека для обробки та аналізу даних. Вона надає швидкі, гнучкі та виразні структури даних, призначені для роботи з "реляційними" або "маркованими" даними. Підходить для обробки великих наборів даних, що є критично важливим для рекомендаційних систем [22].

NumPy: бібліотека для наукових обчислень, що забезпечує підтримку великих масивів та матриць, разом з великим набором математичних функцій для роботи з цими масивами. Використовується для операцій з масивами та матрицями, які є фундаментальними в алгоритмах рекомендацій [23].

Surprise: бібліотека, спеціально призначена для створення та аналізу рекомендаційних систем. Простота використання, забезпечує підтримку різних алгоритмів рекомендації, включаючи колаборативну фільтрацію. Надає потужні інструменти для точної персоналізації рекомендацій [24].

### 2.4.4 Модулі сайту

Flask – це мікрофреймворк для розробки веб-додатків на мові програмування Python. "Мікро" у цьому контексті вказує на простоту його основи: Flask легко розширюється та адаптується до різноманітних потреб розробників [25].

Ось кілька причин, чому Flask може бути хорошим вибором для вашого проекту:

- легкість у використанні: Flask має простий та зрозумілий синтаксис;

- гнучкість: оскільки Flask є мікрофреймворком, він не накладає жорстких обмежень на структуру проекту або бекендові технології, надаючи розробникам свободу у виборі інструментів та бібліотек;
- легкість розширення: Flask може бути легко розширений за допомогою "розширень" (extensions), які додають додаткові функції, такі як ORM, аутентифікація користувачів, форми тощо;
- сумісність з іншими бібліотеками Python: Flask легко інтегрується з іншими бібліотеками та інструментами Python, що дозволяє розробникам використовувати різні бібліотеки для баз даних, машинного навчання, аналітики даних тощо;
- швидкий до розгортання прототипів: Flask дозволяє швидко розгортати прототипи, що є корисним у фазі тестування концепцій та ідей.

Bootstrap: фреймворк для фронтенд-розробки, що використовує HTML, CSS та JavaScript. Він надає готові компоненти та стилі, які можна легко інтегрувати в веб-сайти. Це дозволяє швидко розробляти адаптивні та мобільно-орієнтовані веб-сайти. Він надає універсальну систему сітки, компоненти для навігації, форми, кнопки та інші елементи інтерфейсу, які можна легко налаштувати та адаптувати.

HTML (HyperText Markup Language): є стандартною мовою розмітки для створення веб-сторінок. Він визначає структуру веб-сторінки та слугує фундаментом будь-якого веб-сайту. Він використовується для структурування контенту на веб-сторінці, включаючи текст, зображення, посилання тощо [26].

CSS (Cascading Style Sheets): є мовою стилів, яка використовується для визначення візуального представлення HTML-документів. Вона контролює макет, кольори, шрифти та інші аспекти візуального дизайну та дозволяє створювати більш привабливі та професійно виглядаючі веб-сторінки. Вона дає можливість адаптувати веб-сайти для різних пристроїв та розмірів екрану [27].

### 3 РЕАЛІЗАЦІЯ СИСТЕМИ РЕКОМЕНДАЦІЙ

#### 3.1 Опис класів, системи та моделей

Інтерфейс користувача був реалізований в наступних класах та контролерах:

- `auth_controller.py` – відповідає за авторизацію, вхід/вихід до особистого кабінету веб-сервісу;

- `home_controller.py` – відповідає за головні сторінки, представлення загальної інформації про сервіс, що доступно усім користувачам.

- `book_controller.py` – відповідає за всю дії які можна робити щодо книги, а саме: створення, редагування та видалення книги, перегляд усього каталогу, перегляд рекомендацій, додавання у колекції (бажане, прочитано, в процесі), додавання/видалення книги у кошику, покупка книги;

- `book.py` – відповідає за доступ до модуля моделі рекомендацій.

- `book_store.py` – головний файл, що відповідає за запуск сайту, зв'язує всі контролери та встановлює початкові дані.

- `book_from_api.py` – допоміжний функціонал, що містить методи для додавання книг з API.

- `models` – класи, що відповідають за таблиці сутностей з БД.

- `rescom.py` – клас що відповідає за створення створення та навчання моделі. Тут реалізована вся попередня обробка даних перед подачею їх до потрібних функцій прогнозування рекомендацій;

Для навчання та тестування моделі використано набір даних Book-Crossing [28], що включає 278 858 користувачів (анонімних, але з демографічною інформацією), які надають 1 149 780 оцінок (явних/прихованих) щодо 271 379 книг (рис.3.1).

## 3.2 Опис роботи системи

Для роботи користувача був створений простий, привабливий та інтуїтивно зрозумілий інтерфейс. Сервіс представлений у вигляді веб-застосунку що підтримується більшістю браузерів, а саме Mozilla, Google Chrome, Safari, Microsoft Edge. Головна сторінка показана на рисунку 3.1.

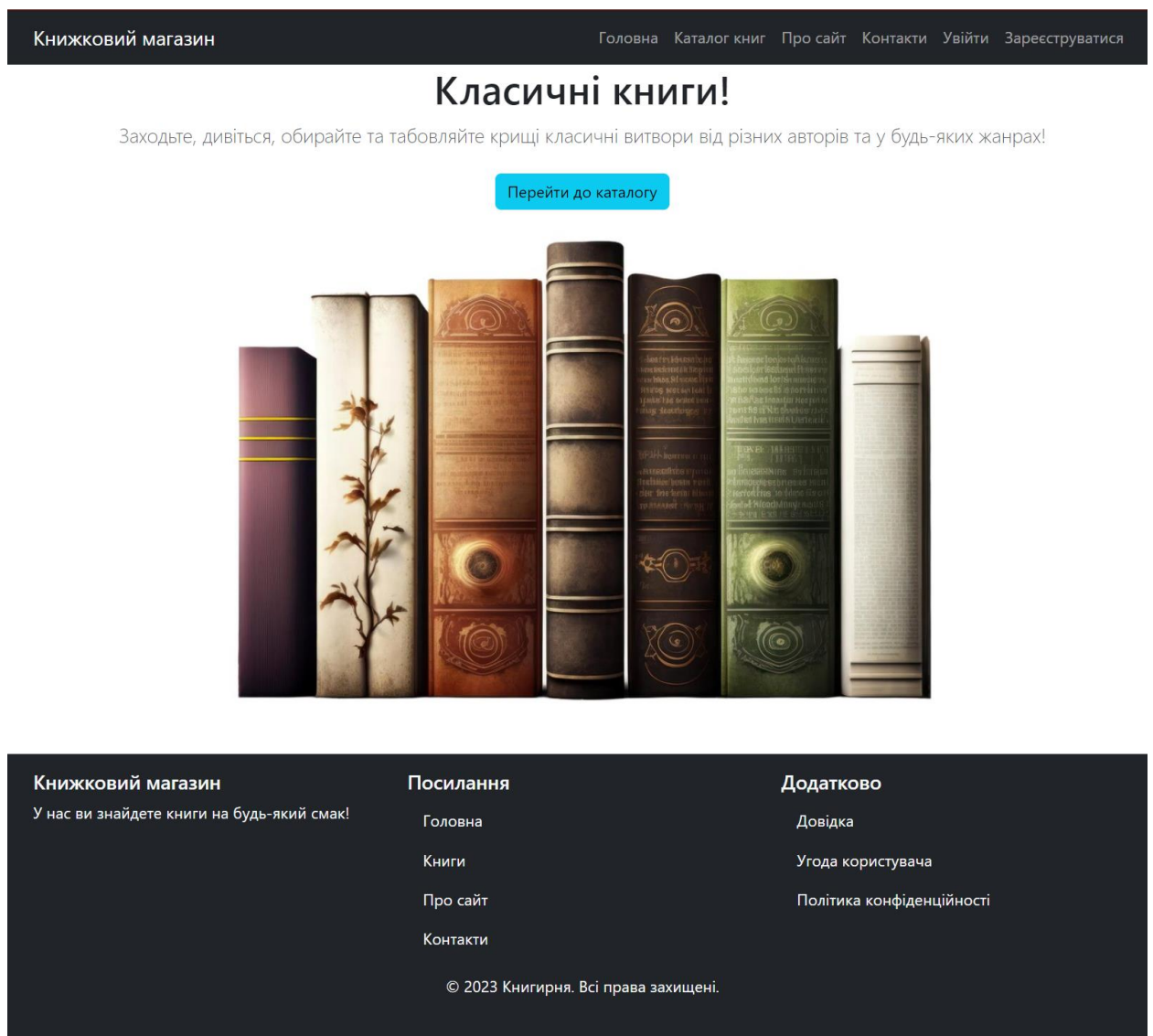


Рисунок 3.1 – Головна сторінка книжкового магазину

Для того щоб користувач зміг отримувати персоналізовані рекомендації, потрібно зареєструватись на сторінці «Реєстрація» та заповнити інформацію про себе, після чого до колекцій додати книг, що подобаються. Створення і

занесення нових даних в систему відбудеться після натискання кнопки «Зареєструватися». Приклад форми реєстрації нового користувача представлений на рисунку 3.2.

Книжковий магазин Головна Каталог книг Про сайт Контакти Увійти Зареєструватися

## Реєстрація

Ім'я

Рік народження

Стать

Країна

Email

Телефон

Пароль

Повторіть пароль

[Зареєструватися](#)

---

<b>Книжковий магазин</b> У нас ви знайдете книги на будь-який смак!	<b>Посилання</b> <a href="#">Головна</a> <a href="#">Книги</a> <a href="#">Про сайт</a> <a href="#">Контакти</a>	<b>Додатково</b> <a href="#">Довідка</a> <a href="#">Угода користувача</a> <a href="#">Політика конфіденційності</a>
--	--	---

© 2023 Книгирня. Всі права захищені.

Рисунок 3.2 – Реєстрація нового користувача

У подальшому після введення логіну та паролю користувачеві будуть доступні всі функції. Сторінка входу показана на рисунку 3.3.

Книжковий магазин

Головна Каталог книг Про сайт Контакти Увійти Зареєструватися

## Вхід до системи

Пошта

Пароль

Залишатись в системі

[Увійти](#)

Новий користувач? [Зареєструватись](#)

**Книжковий магазин**  
У нас ви знайдете книги на будь-який смак!

**Посилання**

- Головна
- Книги
- Про сайт
- Контакти

**Додатково**

- Довідка
- Угода користувача
- Політика конфіденційності

© 2023 Книгирня. Всі права захищені.

Рисунок 3.3 – Вхід до системи

Після авторизації читач матиме розширений доступ до сервісу, що передбачає і список індивідуальних рекомендацій. Для навігації по застосунку застосовується верхня панель управління, на якій розташовані всі необхідні пункти та кнопки. Вигляд панелі зображений на рисунку 3.4.

Книжковий магазин

Головна Каталог книг Про сайт Контакти Особисті рекомендації Кошик Колекції Профіль Вийти

Рисунок 3.4 – Панель навігації сайту для читача

Перегляд всіх книг доданих в БД для пошуку та покупки знаходиться на сторінці «Каталог книг». Тут можна дізнатись назву книги, автора, жанр, ціну та опис. Форма відображення сторінки показана на рисунку 3.5.

## Каталог

**Колекціонер****Авторство:** Джон Фаулз**Жанр:** роман**Мова:** українська**Ціна:** \$21.00

**Опис:** «Колекціонер» (1963) — дебютний роман і перший бестселер Джона Фаулза (1926–2005), стосується найважливіших проблем буття — зв'язку краси і потворності, взаємин людини й суспільства, Творця і творіння. Молодий клерк, який колекціонує метеликів, закохується в юну красуню з мистецьких кіл, яка стає пе...

[Детальніше](#)[У кошик](#)[В колекцію](#)**Зірка над Вифлеємом****Авторство:** Агата Крісті**Жанр:** роман**Мова:** українська**Ціна:** \$27.00

**Опис:** Оповідання до збірки «Зірка над Вифлеємом» Агата Крісті написала уже в поважному віці, маючи великий життєвий досвід і сформований християнський світогляд, а також славу королеви детективного жанру. Втім, це не детективні, але водночас і не традиційні різдвяні історії. Шість оповідань для родинного ...

[Детальніше](#)[У кошик](#)[В колекцію](#)**Джейн Ейр****Авторство:** Шарлотта Бронте**Жанр:** роман**Мова:** українська**Ціна:** \$22.00

**Опис:** Написана в середині дев'ятнадцятого сторіччя, ця історія не втратила популярності й донині. Вона підкорює чимраз більше сердець, захоплює читачів і спонукає їх зазирнути в таємницю кохання. Джейн Ейр — сирота, якої зрелася рідна тітка, віддавши на виховання в притулок. Едвард Рочестер — заможний ан...

[Детальніше](#)[У кошик](#)[В колекцію](#)

1

...

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

...

2713

**Книжковий магазин**

У нас ви знайдете книги на будь-який смак!

**Посилання**[Головна](#)[Книги](#)[Про сайт](#)[Контакти](#)**Додатково**[Довідка](#)[Угода користувача](#)[Політика конфіденційності](#)


© 2023 Книгирня. Всі права захищені.

Для перегляду списку рекомендованих книг, потрібно перейти на сторінку «Особисті рекомендації» (рис. 3.6).

Книжковий магазин

[Головна](#)
[Каталог книг](#)
[Про сайт](#)
[Контакти](#)
[Особисті рекомендації](#)
[Кошик](#)
[Колекції](#)
[Профіль](#)
[Вийти](#)

## Особисті рекомендації



### Ми

**Авторство:** Джордж Орвелл


**Жанр:** роман

**Мова:** українська

**Ціна:** \$23.00

**Опис:** -

[Детальніше](#)
[У кошик](#)
[В колекцію](#)



### Колгосп тварин

**Авторство:** Джордж Орвелл


**Жанр:**

**Мова:** українська

**Ціна:** \$24.00

**Опис:** У казковій повісті "Колгосп тварин" в алегоричній формі розповідається про часи становлення СРСР. Зокрема, тут ідеться про революцію 1917 року, про брехливі обіцянки, що "земля — селянам, а заводи — робітникам", розповідається про розкуркулення, примусові переселення, репресії, розстріли і про мільй...

[Детальніше](#)
[У кошик](#)
[В колекцію](#)



### 1984

**Авторство:** Джордж Орвелл

**Жанр:**

**Мова:** українська

**Ціна:** \$50.00

**Опис:** "1984" — один з найголовніших і вже точно найважливіший роман минулого сторіччя. Важко пригадати якийсь інший літературний твір, який би без зайвої манірності, настільки ж чітко, правдиво і жорстко поставив перед загалом таку ж важливу проблематику. А саме: що таке влада? Яка її природа? Куди прямує...

[Детальніше](#)
[У кошик](#)
[В колекцію](#)

1
2
3

<p><b>Книжковий магазин</b></p> <p>У нас ви знайдете книги на будь-який смак!</p>	<p><b>Посилання</b></p> <p><a href="#">Головна</a></p> <p><a href="#">Книги</a></p> <p><a href="#">Про сайт</a></p> <p><a href="#">Контакти</a></p>	<p><b>Додатково</b></p> <p><a href="#">Довідка</a></p> <p><a href="#">Угода користувача</a></p> <p><a href="#">Політика конфіденційності</a></p>
---	---	--

© 2023 Книжирня. Всі права захищені.


Рисунок 3.6 – Перегляд особистих рекомендацій

Можна переглянути детальну інформацію про книгу (рис. 3.7), додати книгу у кошик (рис. 3.8) або у колекцію (рис. 3.9). Для цього поряд з обраною книгою потрібно натиснути кнопку «Детальніше», «У кошик» або «В колекцію» відповідно.

Книжковий магазин

[Головна](#)
[Каталог книг](#)
[Про сайт](#)
[Контакти](#)
[Особисті рекомендації](#)
[Кошик](#)
[Колекції](#)
[Профіль](#)
[Вийти](#)

[Повернутися до каталогу](#)



Оцінка: 4.0  
★ ★ ★ ★ ☆

## Джейн Ейр

Автор(и): Шарлотта Бронте

ISBN: 9786177914609

Жанр: роман

Мова: українська

Видавництво: Nebo BookLab Publishing

Рік публікації: 2020

**Опис:** Написана в середині дев'ятнадцятого сторіччя, ця історія не втратила популярності й донині. Вона підкорює чимраз більше сердець, захоплює читачів і спонукає їх зазирнути в таємницю кохання. Джейн Ейр — сирота, якої зречлася рідна тітка, віддавши на виховання в притулок. Едвард Рочестер — заможний англійський аристократ, що володіє двома великими, хоч і похмурими на вигляд маєтками — Торнфілд-холлом та Ферндіном. І Джейн, і Едвард шукають свого щастя. Вона — у малому світі, який видається їй безмежно широким, а він — у широкому, проте, на його думку, безмежно малому. Одного дня герої зустрічаються, а згодом між ними спалахує щире, пристрасне почуття, яке змінює їхні долі та змушує стрепенутись і безрадісний Торнфілд-хол, і ще понуріший Ферндін. Але на шляху в закоханих стають минулі помилки й теперішні умовності. Чи здатна їхня незрадлива любов здолати ті всі перешкоди?

Ціна: \$22.00

Залишилось на складі: 24

Бажане
Додати до колекції

Додати в кошик

### Відгуки

Ann

Чудова книга, рекомендую!!!

Дата: 2023-12-25

### Додати коментар

Коментар

Додати коментар

**Книжковий магазин**

У нас ви знайдете книги на будь-який смак!

**Посилання**

[Головна](#)  
[Книги](#)  
[Про сайт](#)  
[Контакти](#)

**Додатково**

[Довідка](#)  
[Угода користувача](#)  
[Політика конфіденційності](#)

© 2023 Книгирня. Всі права захищені.

Рисунок 3.7 – Перегляд детальної інформації про книгу

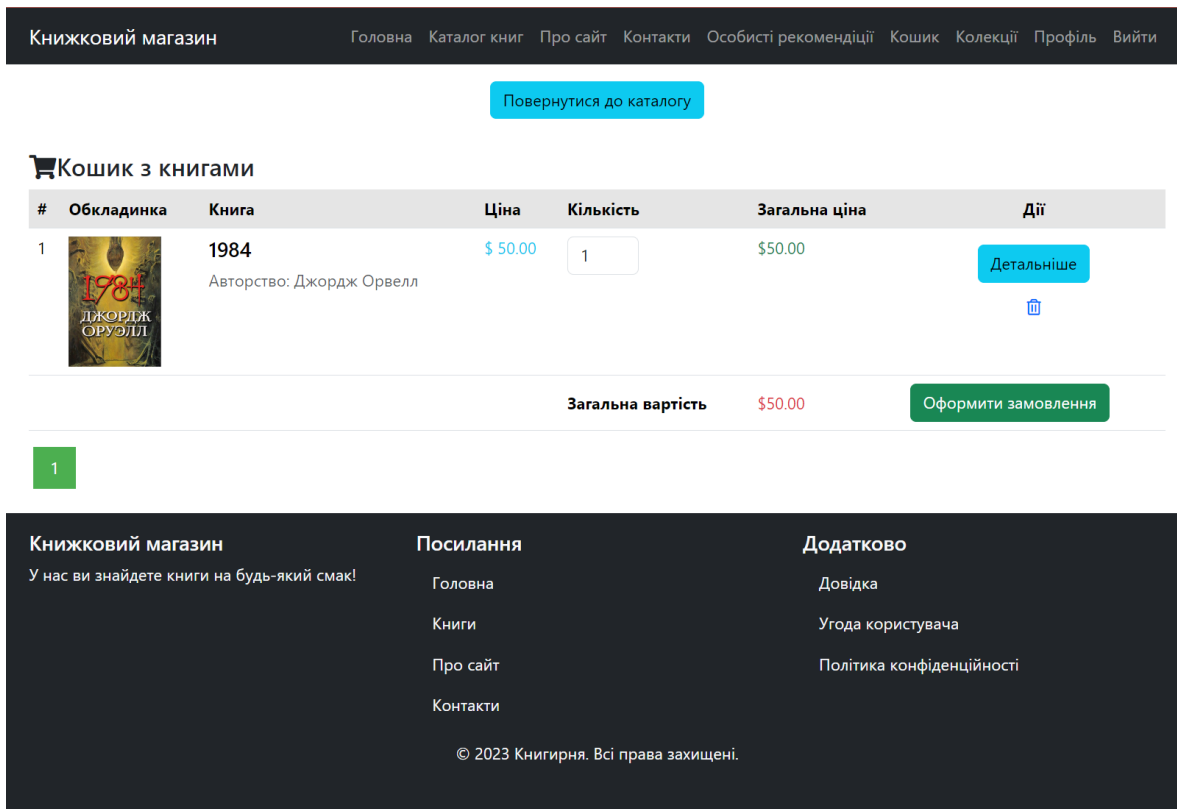


Рисунок 3.8 – Перегляд кошику

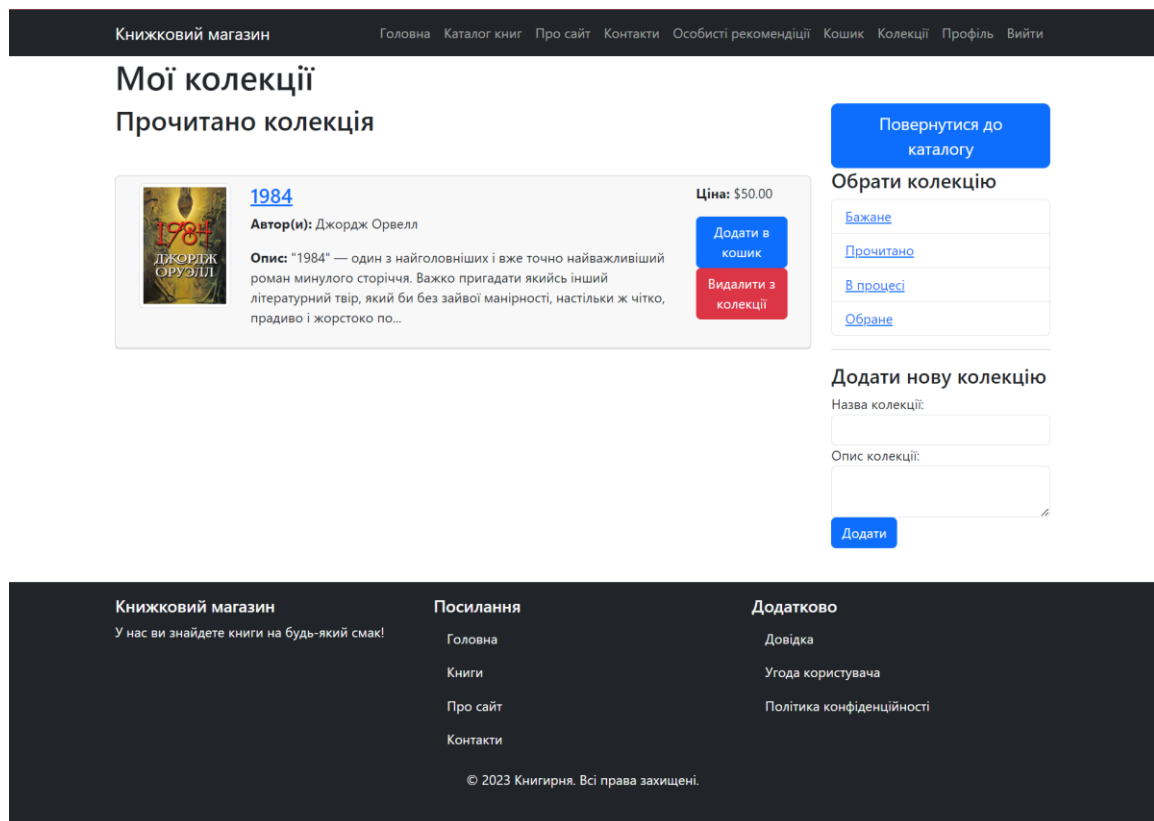


Рисунок 3.9 – Перегляд книг доданих до колекцій

На сторінці з колекціями книг можна переключатись між колекціями, а також створювати власні колекції.

Для адміністраторів панель навігації дещо інакша, є функціонал додавання, редагування та видалення книг, авторів та жанрів (рис. 3.10).



Рисунок 3.10 – Панель навігації для адміністраторів

Сторінка «Каталог книг» теж має трохи інші кнопки (рис. 3.11). Замість кнопок «У кошик» та «В колекцію» у адміністраторів будуть кнопки «Редагувати» та «Видалити» з перенаправленням на відповідні сторінки з можливістю редагувати (рис. 3.12) та видаляти (рис. 3.13) книги.

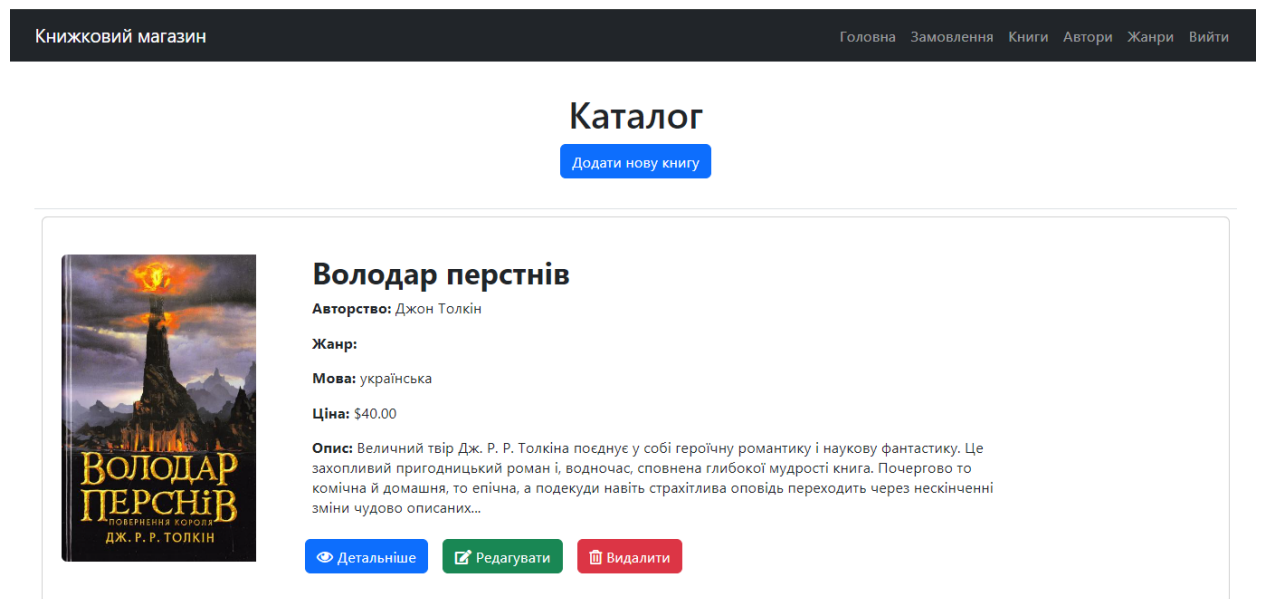


Рисунок 3.11 – Каталог книг для адміністраторів

На сторінці редагування також буде додаткова можливість видалення цієї книги з БД.

## Редагування книги

[Повернутися до каталогу](#)[Перейти до видалення](#)

Назва книги

Володар перстнів

ISBN книги

9786176640349

Мова

українська

Автор

Джон Толкін

Жанр

роман

Кіл-ть екземплярів

70

Рік видавництва

2000

Ціна екземпляру

40,00

Зображення обкладинки

Выберите файл

Файл не выбран

Посилання на обкладинку

<https://book24.ua/upload/iblock/085/0857a6e8c5cf456b038f7562ec7c9d21.jpg>

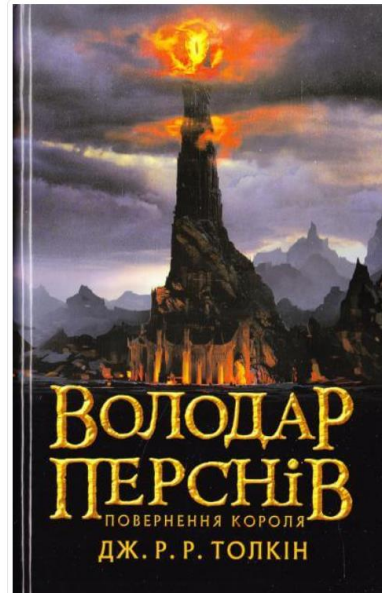
Опис книги

Величний твір Дж. Р. Р. Толкіна поєднує у собі героїчну романтику і наукову

Видавництво

Астролябія

Ок



## Книжковий магазин

У нас ви знайдете книги на будь-який смак!

## Посилання

[Головна](#)[Книги](#)[Про сайт](#)[Контакти](#)

## Додатково

[Довідка](#)[Угода користувача](#)[Політика конфіденційності](#)

© 2023 Книгирня. Всі права захищені.

Рисунок 3.12 – Сторінка редагування книги

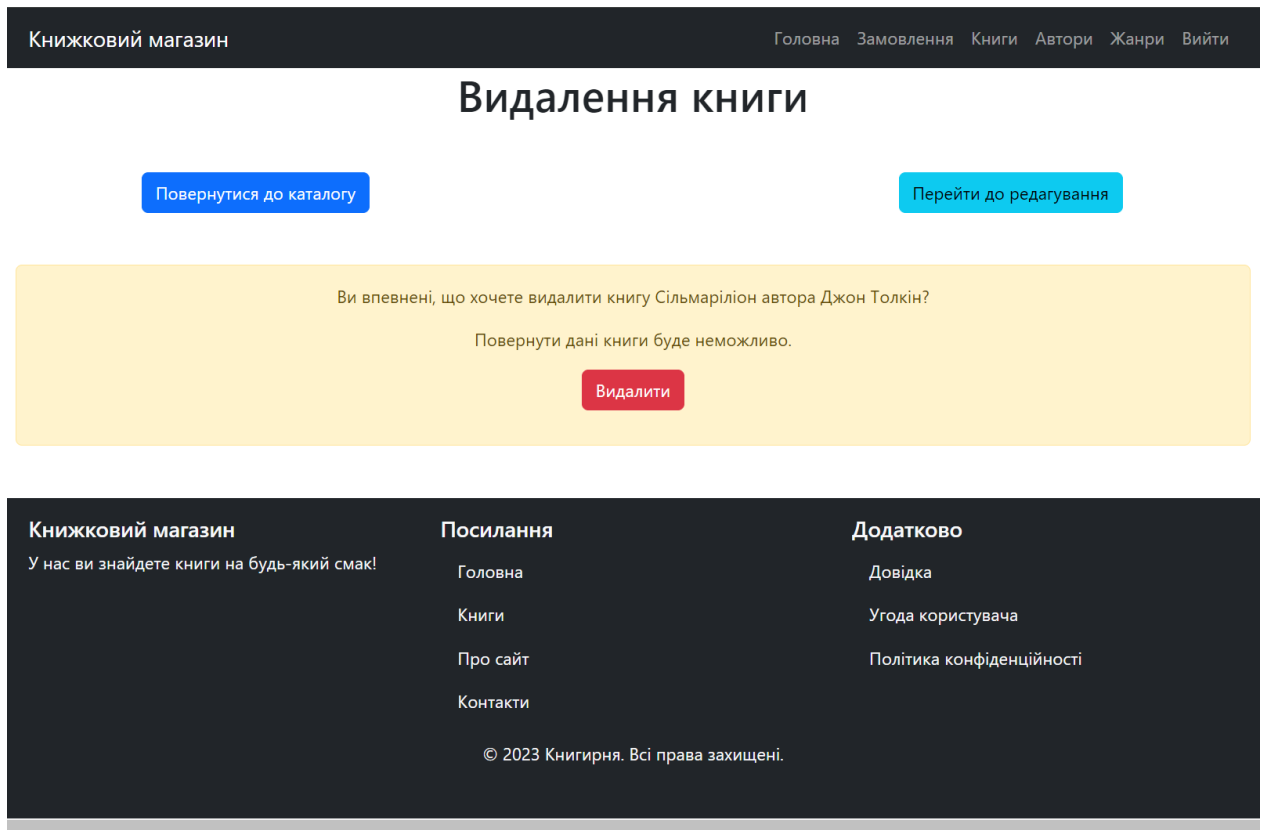


Рисунок 3.13 – Сторінка видалення книги

Для того щоб додати нову книгу, на сторінці «Книги» натиснути на кнопку «Додати книгу», після цього користувача буде перенаправлено на сторінку створення нових книг. На цій новій сторінці, потрібно обов’язково заповнити поля назви та автора, інші поля являються не обов’язковим, та натиснути кнопку «Ок». Приклад форми додавання нової книги для ідентифікації представлений на рисунку 3.14.

## Створення книги

[Повернутися до каталогу](#)

Назва книги

Володар перстнів

ISBN книги

9786176640349

Мова

українська

Автор

Джон Толкін

Жанр

роман

Кіл-ть екземплярів

5

Рік видавництва

2023

Ціна екземпляру

17.49

Зображення обкладинки

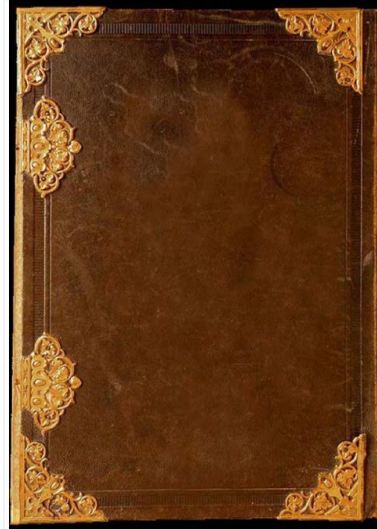
Виберіть файл Файл не вибран

Посилання на обкладинку

Опис книги

Видавництво

Ок



## Книжковий магазин

У нас ви знайдете книги на будь-який смак!

## Посилання

[Головна](#)[Книги](#)[Про сайт](#)[Контакти](#)

## Додатково

[Довідка](#)[Угода користувача](#)[Політика конфіденційності](#)

© 2023 Книгирня. Всі права захищені.

Рисунок 3.14 – Сторінка створення нової книги

Для додавання, редагування та видалення авторів теж створені окремі сторінки.

На рисунку 3.15 наведено приклад сторінки перегляду всіх авторів з кнопками для редагування та видалення. Натиснувши на кнопку «Додати автора» відкривається форма за допомогою якої можна ввести необхідну інформацію про автора (рис. 3.16). Для редагування форма схожа (рис. 3.17).

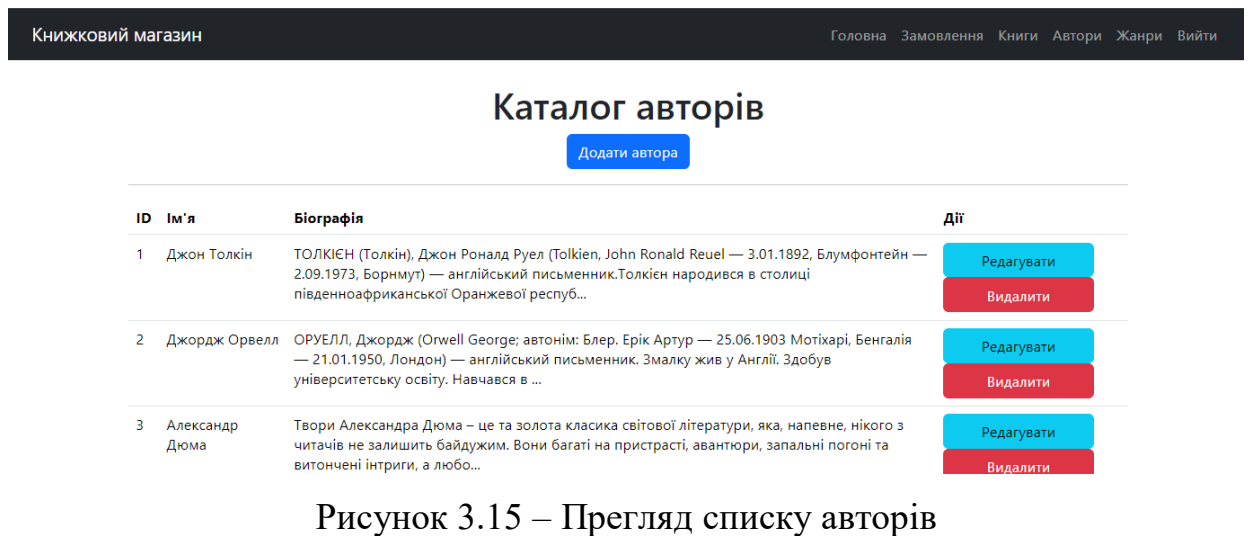


Рисунок 3.15 – Прегляд списку авторів

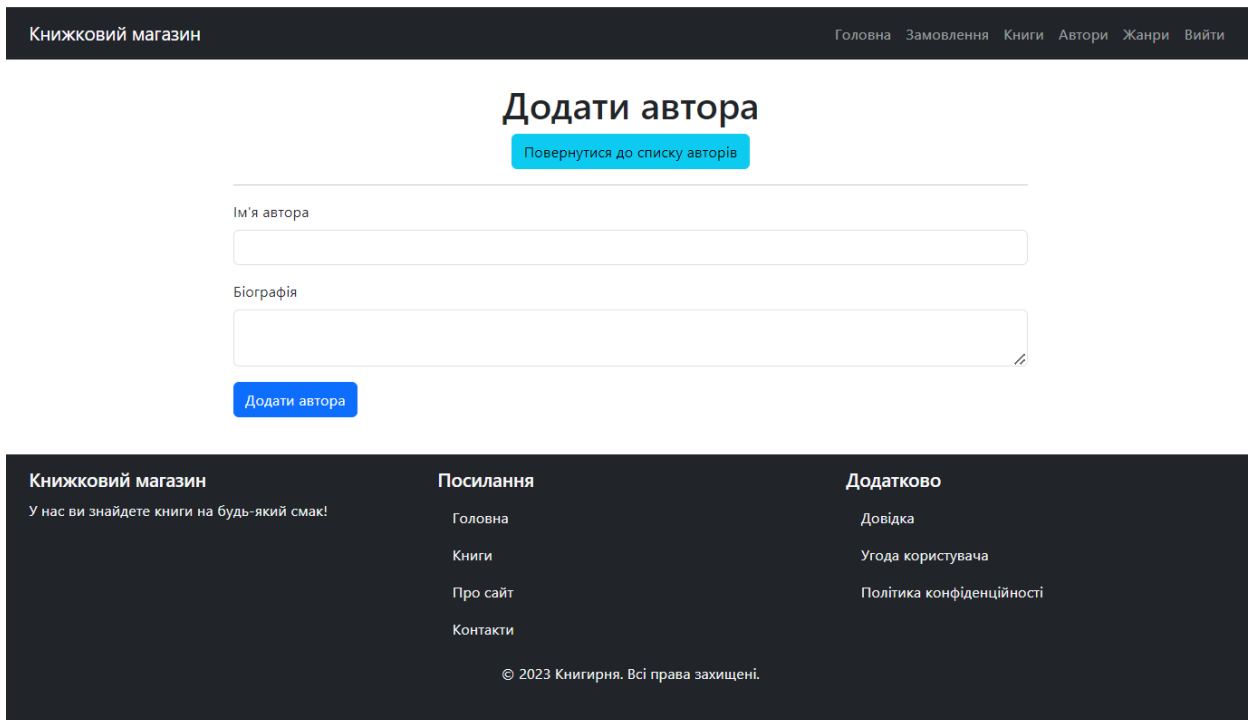


Рисунок 3.16 – Форма для додавання нового автора

Книжковий магазин Головна [Замовлення](#) [Книги](#) [Автори](#) [Жанри](#) [Вийти](#)

## Редагувати автора - Джон Толкін

[Повернутися до списку авторів](#)

---

Ім'я автора

Біографія

[Зберегти зміни](#)

---

<b>Книжковий магазин</b> У нас ви знайдете книги на будь-який смак!	<b>Посилання</b> <a href="#">Головна</a> <a href="#">Книги</a> <a href="#">Про сайт</a> <a href="#">Контакти</a>	<b>Додатково</b> <a href="#">Довідка</a> <a href="#">Угода користувача</a> <a href="#">Політика конфіденційності</a>
--	--	---

© 2023 Книгирня. Всі права захищені.

Рисунок 3.17 – Форма для редагування автора

Для додавання, видалення та редагування списку жанрів існують аналогічні сторінки.

## 4 ТЕСТУВАННЯ СИСТЕМИ

Для навчання та тестування моделі обрано датасет Book-Crossing з платформи Kaggle. Це набір даних містить інформацію про книги та з оцінками читачів, розроблений для навчання моделей рекомендаційних систем. Він включає 278 858 користувачів (анонімних, але з демографічною інформацією), які надають 1 149 780 оцінок (явних/прихованих) щодо 271 379 книг.

В якості метрик визначення точності розпізнавання були взяті: Accuracy, Precision та Recall, а також F1.

Найголовнішим показником якості роботи рекомендаційної системи є точність прогнозування вподобань (Prediction Accuracy) – це оцінка загальної точності, що допомагає визначити, наскільки передбачені рекомендації відповідають фактичним оцінкам користувачів. Для розрахунку точності набір даних було розділено на навчальний та тестовий набори. Після чого використано метрику середню квадратичну помилку (RMSE) або середню абсолютну помилку (MAE) [2, 12].

Середня квадратична помилка (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_i (p_i - q_i)^2}$$

де  $p_i$  – передбачене значення  $i$ -го відношення користувач-книга,  
 $q_i$  – справжнє значення  $i$ -го відношення користувач-книга,  
 $N$  – всі прогнози.

Середня абсолютна помилка (MAE):

$$MAE = \frac{1}{N} \sum_i |p_i - q_i|$$

Значення показників точності модуля рекомендацій зображені на рисунку 4.1. Найбільш високу точність виявлено при використанні запропонованого методу, це 92%.

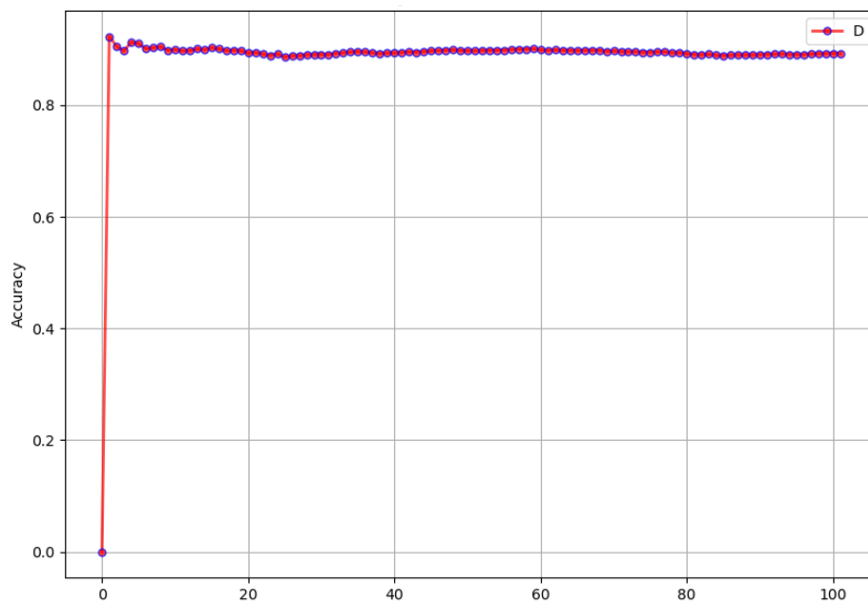


Рисунок 4.1 – Результат тестування точності

Precision вимірює точність, яка частка рекомендованих книг була релевантною, а Recall оцінює повноту, яку частину всіх релевантних книг було рекомендовано. Можливі чотири результати передбачення:

- tp – позитивний прогноз виявився вірним;
- tn – негативний прогноз виявився вірним;
- fp – позитивний прогноз виявився помилковим;
- fn – негативний прогноз виявився помилковим.

Precision є відношенням числа вірно класифікованих прикладів, до загального числа вибірок з міткою розпізнаних правильно і неправильно. То розрахувати його можна за формулою:

$$Precision = \frac{tp}{fp + tp}$$

Коли модель робить багато невірних позитивних класифікацій, це збільшує знаменник та знижує precision. З іншого боку, precision висока, коли:

- 1) модель робить багато коректних позитивних прогнозів;
- 2) модель робить менше неправильних позитивних прогнозів.

Precision – це інтуїтивно зрозуміла здатність класифікатора не позначати як позитивну, вибірку, яка є негативною. Застосування даної метрики до рекомендацій книг показані на рисунку 4.2. Значення Precision як можна побачити майже не коливається і тримається на позначці приблизно 0.96, тобто 96%.

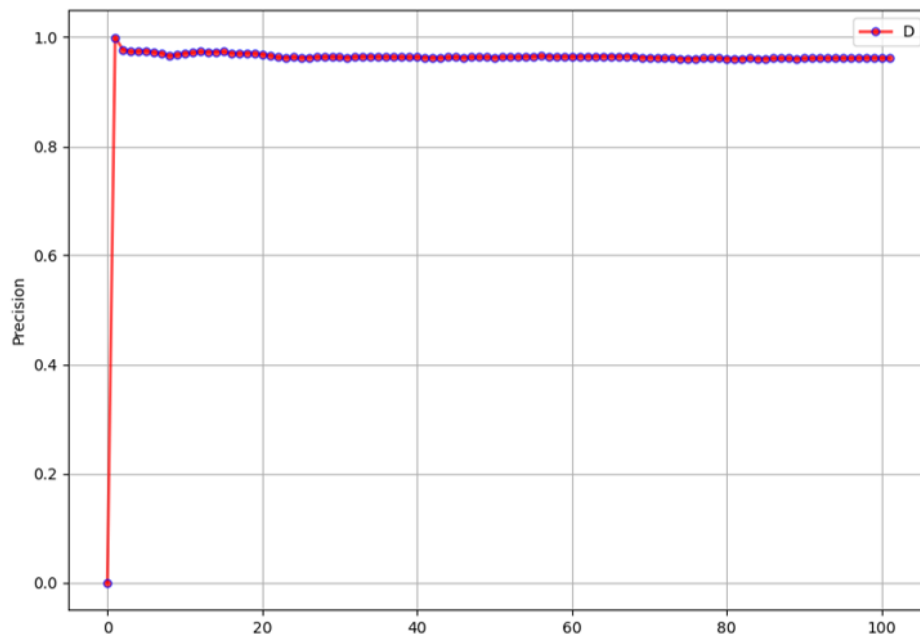


Рисунок 4.2 – Тестування використовуючи метрику Precision

Рівень виявлення Recall, також званий істинно позитивним коефіцієнтом або чутливістю, являє собою частку правильно класифікованих шкідливих екземплярів від загального числа шкідливих екземплярів і розраховується як:

$$Recall = \frac{tp}{fn + tp}$$

Рівень виявлень — це інтуїтивно зрозуміла здатність класифікатора знаходити всі позитивні зразки. На рисунку 4.3 показано результат даної метрики застосованої для рекомендацій. Як можна побачити результат майже не коливається, значення в цих точках приблизно 0.92, що можна виразити як 92%.

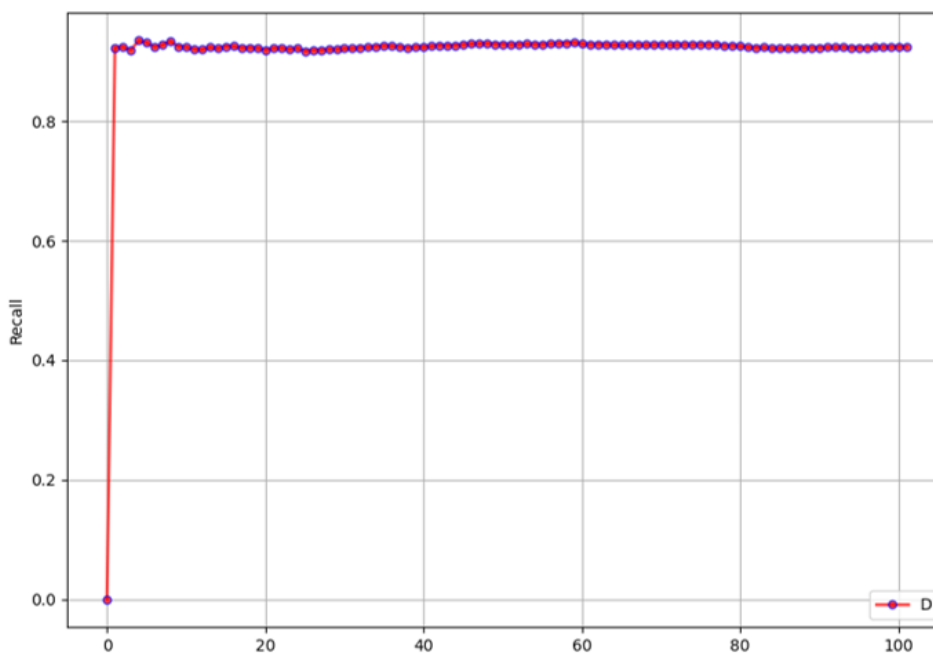


Рисунок 4.3 – Тестування використовуючи метрику Recall

Оцінка F1 – гармонійне середнє точності та повноти, яке забезпечує баланс між цими двома метриками. Вона досягає свого найкращого значення при 1, а найгіршого – при 0. Відносний внесок точності та запам'ятовування в оцінку F1 дорівнює. Формула для оцінки F1:

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

На рисунку 4.4 відображено результат використання даної метрики для даних отриманих після тестування. Загалом коливання в збалансованій оцінці майже не значні і згідно з результатами знаходять біля 0.94 тобто 94%.

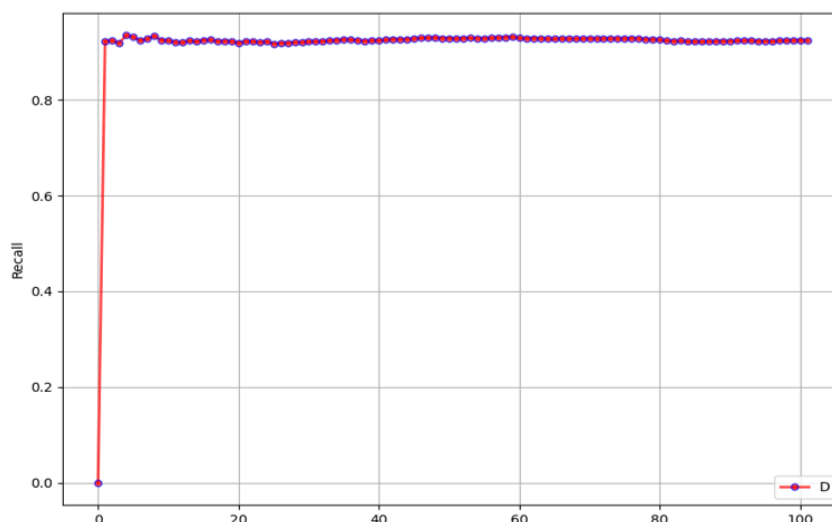


Рисунок 4.4 – Тестування використовуючи метрику F1

Результати тестування даної роботи були оформлені у вигляді статистики в порівнянні з класичними методами рекомендацій (табл. 4.1).

Таблиця 4.1 – Порівняння характеристик першого тестування.

	Запропонований підхід	Колаборативна фільтрація	Вмістова фільтрація	Звичайний гібридний підхід
Показник вірних позитивних передбачень	0.6207947	0.5974478	0.6078454	0.6201596
Показник вірних негативних передбачень	0.6373271	0.6040739	0.6198655	0.63564897
Показник помилкових позитивних передбачень	0.58191427	0.4816137	0.5212365	0.5816137
Показник помилкових негативних передбачень	0.6373271	0.5340739	0.5235499	0.6368998

## ВИСНОВКИ

Ця кваліфікаційна робота зосереджена на створенні та впровадженні інноваційної системи, здатної надавати відповідні та індивідуалізовані рекомендації з використанням технологій машинного навчання.

У дослідженні було здійснено глибокий аналіз наявних рекомендаційних систем та обрано оптимальні алгоритми машинного навчання, які були адаптовані для розробки пропонованої системи. Етап підготовки даних та тренування моделі, що включає відбір та очищення інформації, був ключовим для забезпечення високої точності рекомендацій.

В результаті проведених в роботі досліджень розроблено модель яка використовує гібридний метод, поєднуючи техніки колаборативної та вмістової фільтрації. Цей метод було підсилено за рахунок аналізу демографічних даних, що зробило його ефективнішим у порівнянні з традиційними методами фільтрації.

Для навчання та тестування системи використано дані датасету з платформи Kaggle, що дозволило оцінити ефективність моделі. Реалізацію системи рекомендацій книг виконано у вигляді веб-застосунку з приємним та інтуїтивно зрозумілим інтерфейсом користувача.

Розглянуто результати застосування даної системи які показали, що використання запропонованого методу забезпечує значне покращення у точності та персоналізації рекомендацій. Розроблена система відзначається гнучкістю, масштабованістю та може бути інтегрована у широкий спектр платформ електронної комерції, що робить її високо конкурентоспроможною на ринку рекомендаційних систем.

Створена система має перспективи подальшого розвитку, в тому числі поглиблену модифікацію алгоритмів додавши розширений аналіз вмісту книг, шляхом використання технологій обробки природної мови для аналізу сюжетів, тем, стилів написання тощо, щоб створювати більш точні та відповідні рекомендації.

Завдання кваліфікаційної роботи було досягнуто – створено систему рекомендацій, що використовує методи машинного навчання з підвищеною точністю та персоналізацією пропозицій.

Результати цієї кваліфікаційної роботи було представлено у формі тез доповідей на національній конференції.

**ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ**

1. Ricci, F., Rokach, L., Shapira, B. Recommender Systems Handbook / N.Y: Springer, 2011. – 1008 с.
2. Фальк К. Рекомендовані системи на практиці / М.: ДМК Прес. 2020 – 448 с.
3. Linden, G., Smith, B., York, J. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. / IEEE Internet Computing, 2003. – С. 76-80.
4. The history of Amazon's recommendation algorithm [Електронний ресурс] // Режим доступу: <https://www.amazon.science/the-history-of-amazons-recommendation-algorithm>
5. Gomez-Uribe, C. A., Hunt, N. The Netflix Recommender System: Algorithms, Business Value, and Innovation / ACM Transactions on Management Information Systems (TMIS), 2016. – 130 с.
6. Netflix recommendation system: How it works [Електронний ресурс] // Режим доступу: <https://recoai.net/netflix-recommendation-system-how-it-works/>
7. Oord, A. V. D., Dieleman, S., Schrauwen, B. Deep content-based music recommendation / Neural Information Processing Systems, 2013. – С. 2643-2651.
8. Inside Spotify's Recommender System: A Complete Guide to Spotify Recommendation Algorithms [Електронний ресурс] // Режим доступу: <https://www.music-tomorrow.com/blog/how-spotify-recommendation-system-works-a-complete-guide-2022>
9. Deep Neural Networks for YouTube Recommendations [Електронний ресурс] // Режим доступу: <https://research.google/pubs/pub45530/>
10. Recommending items to more than a billion people [Електронний ресурс] // Режим доступу: <https://engineering.fb.com/2015/06/02/core-infra/recommending-items-to-more-than-a-billion-people/>
11. Goodreads Books Recommendation System [Електронний ресурс] // Режим доступу: <https://monirah-abdulaziz.medium.com/goodreads-books-recommendation-system-619dc190c81>

12. Resnick, P., Varian, H. R. Recommender systems / Communications of the ACM, 1997. – С. 40, 56-58.
13. Adomavicius, G., & Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions / IEEE transactions on knowledge and data engineering, 2005. – С. 17, 734-749.
14. Burke, R. Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction / 2002. – С. 331-370.
15. Zhang, S., Zhao, X., Zhang, L., Liu, T. Deep learning for recommender systems: A concise survey / In Proceedings of the 2019 International Conference on Multimedia, 2019. – С. 1697-1705.
16. Короткий посібник із розуміння алгоритму KNN [Електронний ресурс] // Режим доступу: <https://www.unite.ai/uk/a-quick-guide-to-knn-algorithm/>
17. Koren, Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model / In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008. – С. 426-434.
18. Cremonesi, P., Koren, Y., & Turrin, R. Performance of recommender algorithms on top-n recommendation tasks / In Proceedings of the fourth ACM conference on Recommender systems, 2010. – С. 39-46.
19. Що таке PostgreSQL? - Особливості - Переваги і недоліки [Електронний ресурс] // Режим доступу: <https://uk.education-wiki.com/5154595-what-is-postgresql>
20. What is Python? Executive Summary [Електронний ресурс] // Режим доступу: <https://www.python.org/doc/essays/blurb/>
21. Scikit-Learn [Електронний ресурс] // Режим доступу: [http://cloud-5.bitp.kiev.ua/?page\\_id=640](http://cloud-5.bitp.kiev.ua/?page_id=640)
22. Сучасні бібліотеки машинного навчання [Електронний ресурс] // Режим доступу: <https://travelscode.com/suchasni-biblioteki-mashinnogo-navchannya/>
23. Найпопулярніші бібліотеки ШІ та машинного навчання Python [Електронний ресурс] // Режим доступу: <https://proit.org.ua/naipopuliarnishi-biblioteki-shi-ta-mashinnogho-navchannia-python/>

24. Using Surprise in Python with a recommender system [Електронний ресурс] // Режим доступу: <https://medium.com/@jonahflateman/using-surprise-in-python-with-a-recommender-system-2d6030140926>
25. Flask [Електронний ресурс] // Режим доступу: <https://bizzapps.ru/p/flask/>
26. Що таке HTML [Електронний ресурс] // Режим доступу: <https://mc.today/uk/shho-take-html-ta-yak-za-dopomogoyu-nogo-uvijti-do-it/>
27. Що таке CSS [Електронний ресурс] // Режим доступу: [https://css.in.ua/article/shcho-take-html\\_10](https://css.in.ua/article/shcho-take-html_10)
28. Book Recommendation System [Електронний ресурс] // Режим доступу <https://www.kaggle.com/code/sercanyesiloz/book-recommendation-system/notebook>
29. Пікуш В. А., Савастру О. В. Застосування нейромереж для розв'язання проблем інформаційної безпеки // Тези доповідей 73-ї студентської наукової конференції Одеського національного університету імені І.І. Мечникова. Одеса, 27 квітня 2021 року. – Одеса, 2021. – С. 102-104