ABSTRACT

Today the World Wide Web provides users with a vast array of information, and commercial activity on the Web has increased to the point where hundreds of new companies are adding web pages daily. This has led to the problem of information overload. Recommender systems have been developed to overcome this problem by providing recommendations that help individual users identify content of interest by using the opinions of a community of users and/or the user's preferences.

The aim of this thesis was to design and evaluate different approaches for producing personalised recommendations within the book domain. To achieve this goal, the project first investigated existing recommender systems and profiling techniques.
The next step was to build users' profiles by monitoring users' behaviour, and develop three different approaches for producing recommendations. Finally, an evaluation of the system recommendations' accuracy was done, by first conducting live user experiments and then performing offline analysis to measure the recommendations' accuracy using appropriate methods for testing.

The system evaluation results show that the accuracy of the system recommendations is very good and that a recommender system based on the combination of content-based and collaborative filtering approaches provides more accurate recommendations for the book domain.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This is an experimental project which first, designs…. And second evaluates different approaches for offering recommendations to readers regarding books they may wish to purchase, as part of an online bookshop website.

Today the World Wide Web has provided access to a vast array of information through the web pages, as a result of the Internet growth. Also, commercial activity on the Web has increased to the point where hundreds of new companies are adding Web pages daily. With this increase in the information sources, a problem of information overload occurs, in which the users are trying to deal with an excess of information that is not useful to them as they try to make sensible decisions (Losee, 1989). As a response to this problem, a range of tools to help with retrieving, searching, and filtering have been developed.

The tool most widely used to alleviate the problem of information overload is the search engine. The benefits for the users from search engine technology have decreased as the number of web pages has grown. In addition, the user must first consider the large number of search tools available and decide which one to access. Then the user must interact with each one individually because search engines are typically not personalised to individual users or their prevailing context. Users usually make a choice on the basis of their personal experience or other people's experience. Based on these facts, recommender systems have been developed to provide recommendations that help individual users identify content of interest by using the opinions of a community of users and/or the user's preferences.

Today various recommendation systems play an important role in supporting commercial websites to help users find items that they know they would like to purchase, as well as discover new items about which they had been unaware. The ability to persuade the consumers to buy a suitable item is a significant goal for any recommender system in an ecommerce environment. However, for any recommender system to be successful, the consumer must trust and accept the system’s recommendations. This is done with a clear explanation from the system, presented in a way that is in keeping with the consumer’s preferences. A good recommender system can significantly contribute to achieving the consumer’s acceptance of the system recommendations.

1.1 PROBLEM DEFINITION

This project aims to design and evaluate different approaches for computing recommendations within the book domain to provide personalised recommendations to the users.

1.2 SCOPE AND OBJECTIVES

- An effective solution to the issue of information overload in e-commerce websites is the recommender system.
- This method offers users accurate recommendations.
- most reliable book-related suggestion technology.

Objectives:

- Look into and assess the profiling and recommender systems that are already in use.
- By observing dynamic user behaviours, you can create a user's profile for a recommender system. The user profile needs to change to reflect the user's shifting interests.
- Create a recommender system that uses a variety of computation methods.
- Utilize the right methods to assess the system's recommendations' accuracy.

CHAPTER 2

LITERATURE SURVEY

An idea for Content-Based Book Recommending Using ML Tools for Text Categorization was put forth by Raymond J. Mooney and Loriene Roy. They outline a content-based book recommendation system that classifies text using machine learning and information extraction. Item recommendations are made based on information about the item itself rather than on the preferences of other users. On the other hand, learning customiZed profiles from descriptions of examples enables a system to uniquely characteriZe each customer without the need to match his or her interests to another's. Through the use of automatic text-categorization techniques on semi-structured text downloaded from the web, they have been investigating content-based book recommendations. A database of book data is used by the present prototype system, called LIBRA (Learning Intelligent Book Recommending Agent). After employing a Bayesian learning algorithm to learn about the user's profile, the system generates a ranked list of the most highly suggested additional books from its library. Even when the method provides very modest training sets, the overall results are quite positive. An original content-based book recommender called LIBRA makes use of a straightforward.

A unique book recommendation system was proposed by Binge Cui and Xin Chen. When readers are unable to locate the desired book using the library's bibliographic retrieval system, they are directed to the recommendation pages. It is a web-based system for recommending books to a library's patrons. After logging in, a user can search for books using author names or keywords like book titles. A bibliographic retrieval system will then look for books using the same keywords. If the recommendation system returns any results, submit these keywords to the web books retrieval module. Web Books Retrieval Module allows the librarian or administrator of the online book recommendation system to search the online bookshop using keywords by creating accounts on sites like Amazon. As a result, the web retrieval module searches these online bookshops as the logged-in user when the keyword is presented to it. The user will receive the results from these online booksellers in the form of recommendations. The statistic and analysis module will determine the value of that specific book based on user recommendations. The Auto-Order Module will then generate a book order automatically based on the analysis results according to this value of book. The Short Message and Email Notification Module will get a report from the Book Storage System once the purchased books have been shelved (fig 2.1). Then, utilizing a message and email server, it will inform the readers who have suggested the books that have been acquired.

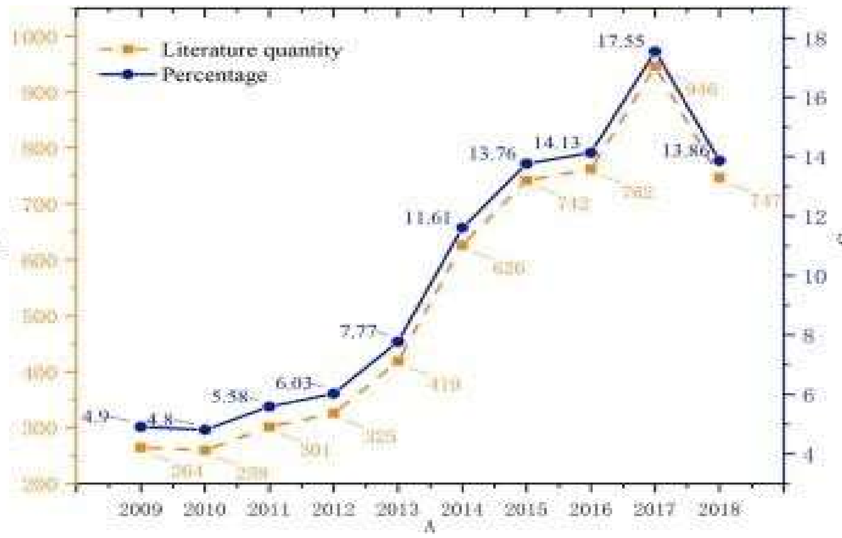

Fig 2.1 Basic Statistics

An effective and best-in-class hybrid recommendation algorithm was proposed by Dharmendra Pathak, Sandeep Matharia, and C. N. S. Murthy , and it provides recommendations that are more in line with user preferences. The hybrid recommendation in this case combines collaborative, content-based, and contextbased algorithms. The primary input for collaborative filtering is rating, or the votes of numerous users, content-based data, which is user-specific information like interests, dates of birth, and priorities, and context-based data, which is behavioural information like time, taste, mood, and weather. The similarity is measured using the cosine similarity. According to the user's prior history, there are subject priority. When they buy a book, do they check to see if the subject priority is changed from what was previously set? If so, topic priority 3 and subject priority 2 should be reset. Priority 1 will remain the same. They came to the conclusion that the proposed Hybrid book recommendation algorithm is superior to the others based on calculations and results.

A book recommendation system based on integrated features of content filtering, collaborative filtering, and association rule mining was proposed by Anand Shanker Tewari, Abhay Kumar, and Asim Gopal Barman. When a customer searches for a book, the search is recorded in the customer's purchase or search history. Recommendation services conduct some filtering when a customer is not online, and the results are saved in the customer's web profile. The recommendations will be created automatically the following time the customer visits the website. Web Usage Mining (WUM) is used in content-based filtering to give customers the pertinent information they need. web server access logs, browser caches, or proxy logs are some examples of historical data that web Usage Mining (WUM) commonly uses to extract knowledge. web use Internet user behavior is recorded by mining, which then processes the information (fig 2.2). The item-based collaborative recommendation algorithm and cosine similarity are both used to measure similarity. Compare the outcomes of collaborative content filtering with association rule mining.

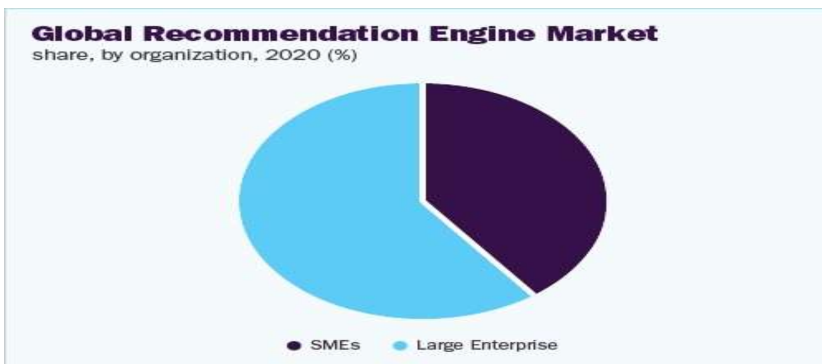

Fig 2.2 Global recommendation system

2.1 INFERENCES FROM LITREATURE SURVEY

For a book recommendation system, Adli Ihsan Hariadi and Dade Nurjanah presented a hybrid-based approach that blends attribute-based and user personality-based methodologies. The MSV-MSL (Most Similar Visited Material to the Most Similar Learner) method is used in this study, and the authors claim that it is the best hybrid attributes-based strategy. When forming neighbourhood ties, the personality trait is used to compare users. The hybrid attribute will use the similarity scores between a target book and its neighbours as well as between the active user and that user's neighbours to generate the recommendation scores of rated books from neighbours. the score for user u's book B, designated as score b. The goal of this is to match up the most similar learner with the most similar visited material. It makes advantage of the collaborative and content values. Utilize the hybrid result as a recommendation after that. That is the webpage that the most similar learner has visited, according to data.

2.2 MOTIVATION

Due to the expansion of the Internet, the World Wide Web now offers access to a wide variety of information via web sites. Additionally, business activity online has grown to the point that every day, hundreds of new businesses add new Web pages. Due to the increase in information sources, a problem known as information overload arises, where users must cope with an abundance of information that is unhelpful to them in order to make sound judgments (Losee, 1989). A variety of tools for accessing, finding, and filtering information have been created as a solution to this issue The search engine is the resource that is most frequently utilised to address the issue of information overload. As the amount of web pages has increased, so too have the benefits for consumers of search engine technology. Additionally, the user must choose which search tool to utilise after carefully weighing the several options. Then, as search engines are often not tailored to specific individuals or their current environment, the user must interact with each one separately. Users typically base their decisions on either their own or other people's experiences. These facts led to the development of recommender systems, which leverage user feedback from a community of users and/or the user's own preferences to generate recommendations that assist individual users in identifying content of interest.

2.3 OPEN PROBLEMS IN EXISTING SYSTEM

The content of the Web is expanding by an estimated 170,000 pages every day, with an Internet growth rate of at least 10% per month (Turban et al., 2000). Due to the abundance of information available to consumers on the World Wide Web, it can be challenging for them to pick exactly what they want. Information overload is a condition when consumers strive to manage more information but are unable to make rational decisions (Losee, 1989).

A variety of retrieval, searching, and filtering techniques have been created to aid with the problem of information overload. The search engine is the most frequently used instrument to help with the issue of information overload. Although search engines are efficient in filtering pages, consumers find it challenging and time-consuming to express their needs in a search query. Due to the exponential growth of web pages, search engine technology's advantages for users have lessened with time.

CHAPTER 3

REQUIREMENT ANALYSIS

3.1 FEASIBILITY STUDIES / RISK ANALYSIS OF THE PROJECT

The project will create and assess a collaborative filtering and content-based recommender system for a real online bookstore. Machine learning methods are typically needed for content-based recommendations in order to identify trends in the products customers like (Middleton, 2003). The experiences of actual users will be reflected in the content-based technology. Users' profiles will be created so that their behaviour may be tracked. Additionally, the system will produce recommendations by comparing the contents of the books in the user's profile with those that the user hasn't reviewed.

3.2 SOFTWARE REQUIREMENTS SPECIFICATION DOCUMENT

3.2.1 Hardware Requirements

- System : Intel® Core™ i5-9300H CPU @ 2.40GHz.
- Monitor : LED.
- Mouse : Logitech.
- Ram : 8.00 GB or above 8.00 GB
- Hard Disk : 1 TB

3.2.2 Software Requirements:

- Operating System : Windows 10, Kali Linux
- Language : Python 3 Framework : Flask 2.2.2

3.2.3 Python:

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

- **Python is Interpreted** − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

- **Python is Interactive** − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

- **Python is Object-Oriented** − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

- **Python is a Beginner's Language** − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

3.2.4 History of Python

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, Small-Talk, and Unix shell and other scripting languages.

Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).

Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

3.2.5 Python Features

Python's features include −

- **Easy-to-learn** − Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.

- **Easy-to-read** − Python code is more clearly defined and visible to the eyes.

- **Easy-to-maintain** − Python's source code is fairly easy-to-maintain.

- **A broad standard library** − Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- **Interactive Mode** − Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- **Portable** − Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- **Extendable** − You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
- **Databases** − Python provides interfaces to all major commercial databases.
- **GUI Programming** − Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
- **Scalable** − Python provides a better structure and support for large programs than shell scripting.

Apart from the above-mentioned features, Python has a big list of good features, few are listed below −

- It supports functional and structured programming methods as well as OOP.
- It can be used as a scripting language or can be compiled to byte-code for building large applications.
- It provides very high-level dynamic data types and supports dynamic type checking.
- It supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

Python is available on a wide variety of platforms including Linux and Mac OS X. Let's understand how to set up our Python environment.

3.2.6 Getting Python

The most up-to-date and current source code, binaries, documentation, news, etc., is available on the official website of Python https://www.python.org.

3.2.7 Windows Installation

Here are the steps to install Python on Windows machine.

- Open a Web browser and go to https://www.python.org/downloads/.
- Follow the link for the Windows installer python-XYZ.msifile where XYZ is the version you need to install.
- To use this installer python-XYZ.msi, the Windows system must support Microsoft Installer 2.0. Save the installer file to your local machine and then run it to find out if your machine supports MSI.
- Run the downloaded file. This brings up the Python install wizard, which is really easy to use. Just accept the default settings, wait until the install is finished, and you are done.

The Python language has many similarities to Perl, C, and Java. However, there are some definite differences between the languages.

3.2.8 First Python Program

Let us execute programs in different modes of programming.

Interactive Mode Programming

Invoking the interpreter without passing a script file as a parameter brings up the following prompt −

```
$ python

Python 2.4.3(#1, Nov112010,13:34:43)

[GCC 4.1.220080704(RedHat4.1.2-48)] on linux2

Type "help", "copyright", "credits" or "license" for more information.

>>>
```

Type the following text at the Python prompt and press the Enter −

```
>>>print "Hello, Python!"
```

If you are running new version of Python, then you would need to use print statement with parenthesis as in **print ("Hello, Python!").** However, in Python version 2.4.3, this produces the following result −

```
Hello, Python!
```

3.2.9 Script Mode Programming

Invoking the interpreter with a script parameter begins execution of the script and continues until the script is finished. When the script is finished, the interpreter is no longer active.

Let us write a simple Python program in a script. Python files have extension **.py.** Type the following source code in a test.py file −

```
Print "Hello, Python!"
```

We assume that you have Python interpreter set in PATH variable. Now, try to run this program as follows −

```
$ python test.py
```

This produces the following result −

```
Hello, Python!
```

CHAPTER 4

DESCRIPTION OF PROPOSED SYSTEM

The application will be developed using the incremental development methodology and will be made up of four increments: Front End, Learning module, Recommendation module and Database increment. The requirements outlined in the Requirements Document will be mapped to manageable increments.

4.1 SELECTED METHODOLOGY OR PROCESS MODEL

4.1.1 Methodology

Recommender systems have been developed to overcome the above mentioned limitations of searching through the massive volume of information available. Recommender systems, in comparison with other filtering tools, require less experience on the part of the user and less effort to specify their interests when querying and operating the system (Resnick and Varian, 1997).

Recommendations systems rely on different technologies for computing recommendations. The most important approaches are content-based filtering and collaborative filtering. Content-based filtering displays users as individuals, while recommender systems employing the collaborative filtering approach display the user as a part of a group (Fasli, 2006). In addition, an advanced recommender system that combines content-based and collaborative filtering to avoid the limitations of each approach, is called a hybrid approach.

4.1.2MODULE DESCRIPTION

Content based filtering approach

The content-based filtering approach identifies the similarity between a user and the new items using the content of the previously evaluated items in the user profile. In addition, each item in a user profile is characterized by a set of attributes which is constructed by extracting a set of features from an item. Such a profile is used to determine if the new item is similar to the item that a user has preferred in the past. For instance, the Newsweeder is a netnewsfiltering system that suggests news articles to the user based on the user's profile (Lang, 1995). Most content-based approaches are performed on textual documents, such as web pages and articles. The textual document can be easily broken down into individual words, unlike video and physical resources, which required sophisticated analysis.

Collaborative filtering approach

Collaborative filtering recommendations are based on the opinions of a community of similar users. The basic idea is that users recommend items to one another. Collaborative filtering makes this possible by asking the users to rate items, which allows the system to recommend new items that similar users have rated highly. For instance, MovieLens is a movie recommender system that uses collaborative filtering to help people find movies they will like in the huge stream of available movies. Collaborative filtering works well for multimedia technology such as music and movies.

Data Set

During the last few decades, with the rise of Youtube, Amazon, Netflix and many other such web services, recommender systems have taken more and more place in our lives. From e-commerce (suggest to buyers articles that could interest them) to online advertisement (suggest to users the right contents, matching their preferences), recommender systems are today unavoidable in our daily online journeys.

In a very general way, recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy or anything else depending on industries).

Recommender systems are really critical in some industries as they can generate a huge amount of income when they are efficient or also be a way to stand out significantly from competitors. As a proof of the importance of recommender systems, we can mention that, a few years ago, Netflix organised a challenges (the "Netflix prize") where the goal was to produce a recommender system that performs better than its own algorithm with a prize of 1 million dollars to win

4.2 DATA SET

Books are identified by their respective ISBN. Invalid ISBNs have already been removed from the dataset. Moreover, some content-based information is given (Book-Title, Book-Author, Year-Of-Publication, Publisher), obtained from Amazon Web Services. Note that in case of several authors, only the first is provided. URLs linking to cover images are also given, appearing in three different flavours (ImageURL-S, Image-URL-M, Image-URL-L), i.e., small, medium, large. These URLs point to the Amazon web site.

4.2.1 Recommender System

Recommender systems intend to provide users with suggestions of items that they may be interested in, based upon their past preferences, history of purchase, or demographic information, as well as the environment of possible items. In addition, a recommender system helps the site adapt itself and provide individual personalisation for each consumer; this increases the sales for the commercial site.

Different forms for providing recommendations have been developed; they can be classified into the following forms: attribute-based recommendations, item-to-item correlation, peopleto- people correlation and non-personalised recommendations (Konstan et al., 2001). For more detailed descriptions.

Recommendations systems rely on different technologies for computing recommendations. The most important approaches are content-based filtering and collaborative filtering. Content-based filtering displays users as individuals, while recommender systems employing the collaborative filtering approach display the user as a part of a group (Fasli, 2006). In addition, an advanced recommender system that combines content-based and collaborative filtering to avoid the limitations of each approach, is called a hybrid approach.

4.2.2 Content based filtering approach

The content-based filtering approach identifies the similarity between a user and the new items using the content of the previously evaluated items in the user profile. In addition, each item in a user profile is characterized by a set of attributes which is constructed by extracting a set of features from an item. Such a profile is used to determine if the new item is similar to the item that a user has preferred in the past. For instance, the Newsweeder is a netnewsfiltering system that suggests news articles to the user based on the user's profile (Lang, 1995). Most content-based approaches are performed on textual documents, such as web pages and articles. The textual document can be easily broken down into individual words, unlike video and physical resources, which required sophisticated analysis.

Content-based filtering has some shortcomings in recommending items. A user's selection is based on the subjective attributes (such as the quality) of the item (Goldberg et al., 1992); in contrast, content based approaches are based on objective attributes (such as the description of an item) about the items. Also, some items the users may be interested in cannot be recommended to them because content-based methods compare new items with the items previously seen by the user, while the user's interests may be beyond the scope of the previously seen items. Finally, multimedia technology such as sound, video or physical items cannot be analysed automatically for relevant attribute information, due to limitations of resources (Jennings et al., 2005).

4.2.3 Collaborative Filtering approach

Collaborative filtering recommendations are based on the opinions of a community of similar users. The basic idea is that users recommend items to one another. Collaborative filtering makes this possible by asking the users to rate items, which allows the system to recommend new items that similar users have rated highly. For instance, MovieLens is a movie recommender system that uses collaborative filtering to help people find movies they will like in the huge stream of available movies. Collaborative filtering works well for multimedia technology such as music and movies. However, it also has some limitations:

New user problem: A new user starts off with a profile of interests from scratch. The system needs to know the user preferences in different items to generate accurate recommendations.

Cold start problem: New items cannot be recommended until more information is obtained when another user either rates an item or provides feedback on the item (Fasli, 2006). As a result, the recommendations generated by the system will not recommend items similar enough to the users' interests.

Scalability: A collaborative filtering algorithm should address the scalability issue as the number of users increase and their collective profile size becomes large (Fasli, 2006).

The schematic diagram of the collaborative filtering process is showed in **Figure 4.1**. As you can see from the figure, there is a list of users denoted by U= {u1, u2,…,um} and a list of items I={i1,i2,….,in}. Each user has a list of items. The collaborative filtering algorithm will generate recommendations(fig 4.1), a list of N items that the active user will mostly like, according to the active user. Also, the process will output a prediction, which is the result prediction on item j for the active user (Sarwar et al., 2001).

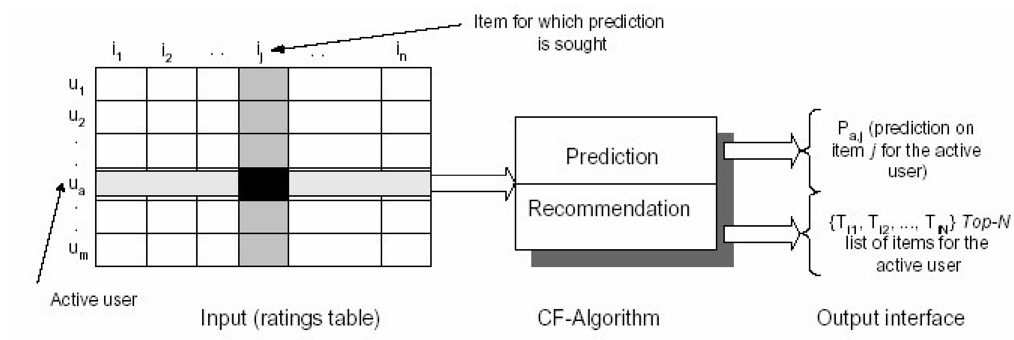

Fig 4.1 Collaborative Filtering Process

4.2.4 Hybrid Approach

Hybrid approach introduced to combine the advantages of both content-based and collaborative filtering techniques help to overcome their limitations. These use thestrength of one set of filtering techniques to overcome the limitation of the other. The hybrid filtering approach is also called “collaborative via content” because content-based profiles are also taken when identifying the similarities among users for collaborative recommendations (Pazzani, 1999).

4.2.5 User - Based collaborative Filtering

User-based algorithm is based on the fact that each user belongs to a larger group of similarly behaving individuals. It uses statistical techniques to find a set of users with similar interests, known as neighbours, in the entire user-item database, to generate a list of recommendation for the active user (Middleton, 2003).

Different measures of similarity that are based on neighbourhood algorithms are used to compute the similarity between the active user and other users in the database, such as the Pearson correlation coefficient and Mean squared differences algorithms (Breese et al., 1998). Moreover, to predict the rating of an item given by the active user, the ratings from the most similar users for the item are averaged and weighted by their similarities to the active user. The Pearson Correlation (fig 4.2) reflects the degree of linear relationship between two variables and ranges from +1 to -1. A positive correlation means that the two users have very similar tastes, while a negative correlation indicates that the users have dissimilar tastes (Fasli, 2006). The Pearson Correlation Coefficient method defines the similarity between two users by:

$$r_{xy} = \frac{\sum_{i=1}^{N} (U_{xi} - \bar{U}_x) * (U_{yi} - \bar{U}_y)}{\sqrt{\sum_{i=1}^{N} (U_{xi} - \bar{U}_x)^2 * \sum_{i=1}^{N} (U_{yi} - \bar{U}_y)^2}}$$

r_{xy}	pearson-r correlation between user x and y
N	number of ratings
U_{xi}	i^{th} rating for user x
$\bar{U}_x$	mean rating for user x

Fig 4.2: Pearson correlation

4.2.6 Item Based collaborative Filtering

The item-based algorithms are developed to overcome the scalability on user-based recommendations. Unlike a user-based approach, the item-based approach identifies the set of items that are similar or related to the item that the active user has evaluated. After that, it computes the similarity between items and then selects the most similar items to the target item within the set of items that the user has rated (Sarwar et al., 2001).

ALGORITHM

- Content based filtering mechanism
- Collaborative based filtering algorithm
- Cosine similarity

4.3 ARCHITECTURE / OVERALL DESIGN OF PROPOSED SYSTEM

Fig 4.3 depicts the architecture of the System

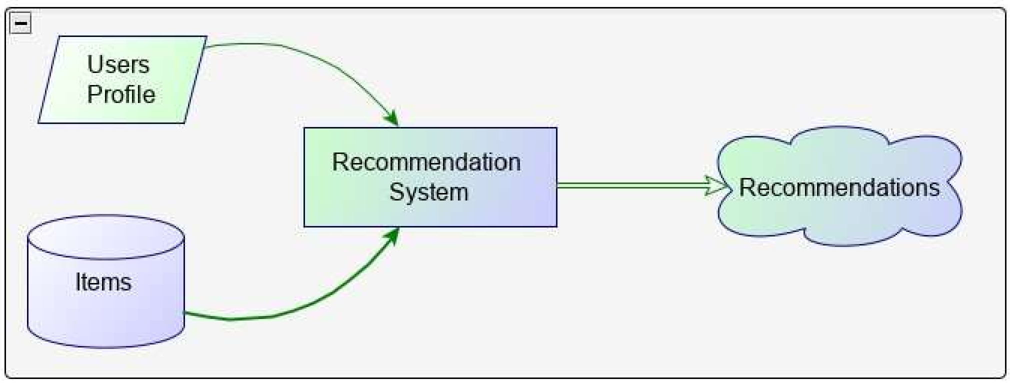

Fig 4.3: System Architecture

4.4 DESCRIPTION OF SOFTWARE FOR IMPLEMENTATION AND TESTING PLAN OF THE PROPOSED MODEL/SYSTEM

Flask Framework:

Flask is a web application framework written in Python. Armin Ronacher, who leads an international group of Python enthusiasts named Pocco, develops it. Flask is based on Werkzeug WSGI toolkit and Jinja2 template engine. Both are Pocco projects.

Http protocol is the foundation of data communication in world wide web. Different methods of data retrieval from specified URL are defined in this protocol.

The following table 4.1 summarizes different http methods –

Table 4.1: Methods and description of the Flask

S.No	Methods & Description
1	**GET** Sends data in unencrypted form to the server. Most common method.
2	**HEAD** Same as GET, but without response body
3	**POST** Used to send HTML form data to server. Data received by POST method is not cached by server.
4	**PUT** Replaces all current representations of the target resource with the uploaded content.

5	**DELETE** Removes all current representations of the target resource given by a URL

By default, the Flask route responds to the **GET** requests. However, this preference can be altered by providing methods argument to **route ()** decorator.

In order to demonstrate the use of **POST** method in URL routing, first let us create an HTML form and use the **POST** method to send form data to a URL.

Save the following script as login.html

```
<html>

<body>

<form action=”http://localhost:5000/login”
method="post">

<p>Enter Name:</p>

<p><input type="text" name="nm"/></p>

<p><input type="submit" value="submit"/></p>

</form>

</body>

</html>
```

Now enter the following script in Python shell.

```
from flask import Flask, redirect, url_for, request app=Flask(__name__)

@app. route('/success/<name>')

def success(name):

return 'welcome %s'% name

@app. route ('/login’,

methods=['POST','GET']) def login (): if

request. method=='POST':

user=request. form['nm']

return redirect(url_for('success’, name= user))

else:

user=request.args.get('nm') return redirect

(url_for ('success’, name= user)) if __name__

=='__main__':

app.run (debug =True)
```

After the development server starts running, open **login.html** in the browser, enter name in the text field and click **Submit**.

Fig 4.4: Local Host Web Server

Form data is Posted to the URL in action clause of form tag.

http://localhost/login is mapped to the **login ()** function. Since the server has received data by **POST** method, value of 'nm' parameter obtained from the form data is obtained by −

```
user = request. form['nm']
```

It is passed to **'/success'** URL as variable part. The browser displays a **welcome** message in the window.

Fig 4.5: Local Host Output Console

Change the method parameter to **'GET'** in **login.html** and open it again in the browser. The data received on server is by the **GET** method. The value of 'nm' parameter is now obtained by −

```
User = request.args.get('nm')
```

Here, **args** is dictionary object containing a list of pairs of form parameter and its corresponding value. The value corresponding to 'nm' parameter is passed on to '/success' URL as before.

4.5 PROJECT MANAGEMENT PLAN

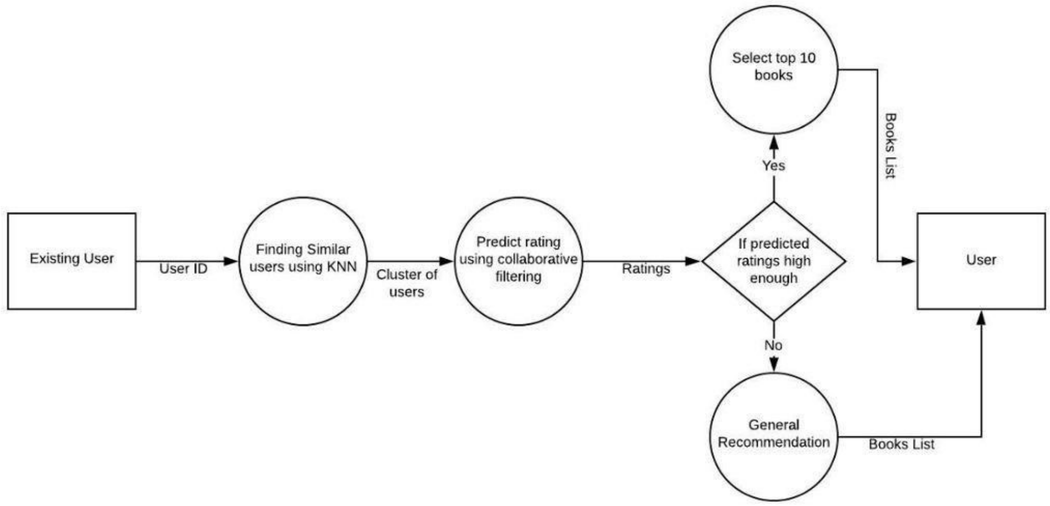

Fig 4.6: Flow Chart

CHAPTER 5

IMPLEMENTATION DETAILS

A recommendation engine is a class of machine learning which offers relevant suggestions to the customer. Before the recommendation system, the major tendency to buy was to take a suggestion from friends. But Now Google knows what news you will read, Youtube knows what type of videos you will watch based on your search history, watch history, or purchase history.

A recommendation system helps an organization to create loyal customers and build trust by them desired products and services for which they came on your site. The recommendation system today are so powerful that they can handle the new customer too who has visited the site for the first time. They recommend the products which are currently trending or highly rated and they can also recommend the products which bring maximum profit to the company.

5.1 DEVELOPMENT AND DEPLOYMENT DETAILS

Machine learning is a process that is widely used for prediction. N number of algorithms are available in various libraries which can be used for prediction. In this article, we are going to build a prediction model on historical data using different machine learning algorithms and classifiers, plot the results, and calculate the accuracy of the model on the testing data.

Building/Training a model using various algorithms on a large dataset is one part of the data. But using these models within the different applications is the second part of deploying machine learning in the real world.

To put it to use in order to predict the new data, we have to deploy it over the internet so that the outside world can use it. In this article, we will talk about how we have trained a machine learning model and created a web application on it usingFlask.

We have to install many required libraries which will be used in this model. Use pip command to install all the libraries.

pip install pandas pip
install numpy pip
install sklearn

5.2 ALGORITHM

Content based Filtering: The algorithm recommends a product that is similar to those which used as watched. In simple words, In this algorithm, we try to find finding item look alike. For example, a person likes to watch Sachin Tendulkar shots, so he may like watching Ricky Ponting shots too because the two videos have similar tags and similar categories.

Collaborative based Filtering: Collaborative based filtering recommender systems are based on past interactions of users and target items. In simple words here, we try to search for the look-alike customers and offer products based on what his or her lookalike has chosen. Let us understand with an example. X and Y are two similar users and X user has watched A, B, and C movie. And Y user has watched B, C, and D movie then we will recommend A movie to Y user and D movie to X user.

Hybrid filtering method: It is basically a combination of both the above methods. It is a too complex model which recommends product based on your history as well based on similar users like you.

There are some organizations that use this method like Facebook which shows news which is important for you and for others also in your network and the same is used by Linkedin too.

Dataset description we have 3 files in our dataset which is extracted from some books selling websites.

- Books – first are about books which contain all the information related to books like an author, title, publication year, etc.
- Users – The second file contains registered user’s information like user id, location.
- ratings – Ratings contain information like which user has given how much rating to which book.

So based on all these three files we can build a powerful collaborative filtering model. let’s get started.

Loading data let us start while importing libraries and load datasets. while loading the file we have some problems like.

- The values in the CSV file are separated by semicolons, not by a comma.
- There are some lines which not work like we cannot import it with pandas and It throws an error because python is Interpreted language.
- Encoding of a file is in Latin

So, while loading data we have to handle these exceptions and after running the below code you will get some warning and it will show which lines have an error that we have skipped while loading.

Preprocessing Data: Now in the books file, we have some extra columns which are not required for our task like image URLs. And we will rename the columns of each file as the name of the column contains space, and uppercase letters so we will correct as to make it easy to use.

The dataset is reliable and can consider as a large dataset. we have 271360 books data and total registered users on the website are approximately 278000 and they have given near about 11 lakh rating. hence we can say that the dataset we have is nice and reliable.

We do not want to find a similarity between users or books. we want to do that If there is user A who has read and liked x and y books, And user B has also liked this two books and now user A has read and liked some z book which is not read by B so we have to recommend z book to user B. This is what collaborative filtering is.

So this is achieved using Matrix Factorization, we will create one matrix where columns will be users and indexes will be books and value will be rating. Like we have to create a Pivot table.

If we take all the books and all the users for modeling, Don't you think will it create a problem? So what we have to do is we have to decrease the number of users and books because we cannot consider a user who has only registered on the website or has only read one or two books. On such a user, we cannot rely to recommend books to others because we have to extract knowledge from data. So what we will limit this number and we will take a user who has rated at least 200 books and also we will limit books and we will take only those books which have received at least 50 ratings from a user.

Exploratory data analysis

The primary goal of EDA is to support the analysis of data prior to making any conclusions. It may aid in the detection of apparent errors, as well as a deeper understanding of data patterns, the detection of outliers or anomalous events, and the discovery of interesting relationships between variables.

Content based filtering /popularity-based filtering

It is a type of Recommendation System which works on the principle of popularity and or anything which is in trend. These systems check about the product or movie which are in trend or are most popular among the users and directly recommend those.

In the content based collaborative filtering we are finding the highest average rating of each book. To find out the highest average rating we merge the ratings dataset with books dataset ratings_with_name=ratings.merge(books,on='ISBN')

Upon merging the rating dataset with book dataset we are grouping the ratings with respect to the name of the book

num_rating_df=ratings_with_name.groupby('Book-Title').count()['Book Rating'].reset_index() num_rating_df.rename(columns={'Book-Rating':'num_ratings'},inplace=True) num_rating_df

Using the mean function calculating the average ratings of the books.

avg_rating_df=ratings_with_name.groupby('Book-Title').mean()['Book Rating'].reset_index() avg_rating_df.rename(columns={'Book-Rating':'avg_rating'},inplace=True) avg_rating_df

By sorting the average rating we can retrive the first 50 highest rating books fro the dataset popular_df=popular_df[popular_df['num_ratings']>=250].sort_values('avg_rating',ascending=False).head(50) popular_df

Collaborative based filtering Approach

Collaborative filtering is used by most recommendation systems to find similar patterns or information of the users, this technique can filter out items that users like on the basis of the ratings or reactions by similar users.

Collaborative filtering is a technique that can filter out items that a user might like on the basis of reactions by similar users.

It works by searching a large group of people and finding a smaller set of users with tastes similar to a particular user. It looks at the items they like and combines them to create a ranked list of suggestions.

There are many ways to decide which users are similar and combine their choices to create a list of recommendations. This article will show you how to do that with Python.

To experiment with recommendation algorithms, you'll need data that contains a set of items and a set of users who have reacted to some of the items.

The reaction can be explicit (rating on a scale of 1 to 5, likes or dislikes) or implicit (viewing an item, adding it to a wish list, the time spent on an article).

While working with such data, you'll mostly see it in the form of a matrix consisting of the reactions given by a set of users to some items from a set of items. Each row would contain the ratings given by a user, and each column would contain the ratings received by an item.

To build a system that can automatically recommend items to users based on the preferences of other users, the first step is to find similar users or items. The second step is to predict the ratings of the items that are not yet rated by a user. x=ratings_with_name.groupby('User-ID').count()['Book-Rating']>200
padhe_likhe_users=x[x].index

filtered_rating=ratings_with_name[ratings_with_name['UserID'].isin(padhe_likhe_u

sers)] y=filtered_rating.groupby('Book-Title').count()['Book-Rating']>=50

famous_books=y[y].index final_ratings=filtered_rating[filtered_rating['Book-

Title'].isin(famous_books)]

pt=final_ratings.pivot_table(index='Book-Title',columns='User-
ID',values='BookRating')

Cosine Similarity

cosine similarity means the similarity between two vectors of inner product space,

It is measured by the cosine of the angle between two vectors. **from**

sklearn.metrics.pairwise import cosine_similarity

similarity_scores=cosine_similarity(pt) similarity_scores.shape def

recommend(book_name):

#index fetch

index=np.where(pt.index==book_name)[0][0]

similar_items=sorted(list(enumerate(similarity_scores[index])),key=lambda

x:x[1],reverse=True)[1:6] data=[] for i in similar_items:

item=[] temp_df=books[books['Book-Title']==pt.index[i[0]]]

item.extend(temp_df.drop_duplicates("Book-Title")['Book-Title'].values)

item.extend(temp_df.drop_duplicates("Book-Title")['Book-Author'].values)

item.extend(temp_df.drop_duplicates("Book-Title")['Image-URL-M'].values)

data.append(item) return data

Accuracy

One of the approaches to measure the accuracy of your result is the Root Mean Square Error (RMSE), in which you predict ratings for a test dataset of user-item pairs whose rating values are already known. The difference between the known value and the predicted value would be the error. Square all the error values for the test set, find the average (or mean), and then take the square root of that average to get the RMSE.

Website Deployment

We are using the pycharm community to deploy the website. By creating the project book recommendation System.

Flask provides configuration and conventions, with sensible defaults, to get started. This section of the documentation explains the different parts of the Flask framework and how they can be used, customized, and extended. Beyond Flask itself, look for community-maintained extensions to add even more functionality.

```
def create_app():     app =
Flask(__name__)
hello.init_app(app)
return app
```

CHAPTER 6

RESULTS AND DISCUSSION

Today the World Wide Web provides users with a vast array of information, and commercial activity on the Web has increased to the point where hundreds of new companies are adding web pages daily. This has led to the problem of information overload. Recommender systems have been developed to overcome this problem by providing recommendations that help individual users identify content of interest by using the opinions of a community of users and/or the user’s preferences.

The aim of this thesis was to design and evaluate different approaches for producing personalised recommendations within the book domain. To achieve this goal, the project first investigated existing recommender systems and profiling techniques. The next step was to build users' profiles by monitoring users' behaviour, and develop three different approaches for producing recommendations. Finally, an evaluation of the system recommendations' accuracy was done, by first conducting live user experiments and then performing offline analysis to measure the recommendations' accuracy using appropriate methods for testing.

The system evaluation results show that the accuracy of the system recommendations is very good and that a recommender system based on the combination of content-based and collaborative filtering approaches provides more accurate recommendations for the book domain.

CHAPTER 7 CONCLUSION

7.1 CONCLUSION

All of our systems– purely content-based, purely collaborative-filtering, and hybrid– performed quite well. Looking back on the project, one thing that we might have chosen to do differently in retrospect would have been to spend more time searching for a dataset of ratings with a higher rating variance per user. Had we been able to find such a dataset, our implementations of algorithms would have been tested on data that would have been more representative of what a typical commercial recommendation system could access in creating its predictions. However, given the data that was available to us, as well as the results our various approaches produced, our systems were largely successful, providing insight into how the different systems we regularly use work and the varying algorithms that make that possible.

7.2 FUTURE WORK

Given more information regarding the books dataset, namely features like Genre, Description etc., we could implement a content-filtering based recommendation system and compare the results with the existing collaborative-filtering based system.

We would like to explore various clustering approaches for clustering the users based on Age, Location etc., and then implement voting algorithms to recommend items to the user depending on the cluster into which it belongs.

7.3 RESEARCH ISSUES

One of the challenges faced by our research was the unavailability of reliable training datasets. In fact, this challenge faces any researcher in the field. However, although plenty of articles about predicting phishing websites using data mining techniques have been disseminated these days, no reliable training dataset has been published publically, maybe because there is no agreement in literature on the definitive features that characterize phishing websites, hence it is difficult to shape a dataset that covers all possible features.

In this article, we shed light on the important features that have proved to be sound and effective in predicting phishing websites. In addition, we proposed some new features, experimentally assign new rules to some well-known features and update some other features.

7.3 IMPLEMENTATION ISSUES

- Handling of sparsity was a major challenge as well since the user interactions were not present for the majority of the books.
- Understanding the metric for evaluation was a challenge as well.

- Since the data consisted of text data, data cleaning was a major challenge in features like Location etc.
- Decision making on missing value imputations and outlier treatment was quite challenging as well.

REFERENCES

[1] Ahuja, Rishabh, Arun Solanki, and Anand Nayyar.” Movie recommender system using K-Means clustering and K-Nearest Neighbor.” In 2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence), pp. 263268. IEEE, 2019.

[2] Badriyah, Tessy, Erry Tri Wijayanto, Iwan Syarif, and Prima Kristalina. ”A hybrid recommendation system for E-commerce based on product description and user profile.” In 2017 Seventh International Conference on Innovative Computing Technology (INTECH), pp. 95-100. IEEE, 2017.

[3] Chen, Junnan, Courtney Miller, and Gaby G. Dagher. ”Product recommendation system for small online retailers using association rules mining.” In Proceedings of the 2014 International Conference on Innovative Design and Manufacturing (ICIDM), pp. 71-77. IEEE, 2014.

[4] Jisha, R. C., Ram Krishnan, and Varun Vikraman. ”Mobile applications recommendation based on user ratings and permissions.” In 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1000-1005.IEEE, 2018.

[5] Keerthana, N. K., Shriram K. Vasudevan, and Nalini Sampath. ”An Effective Approach to Cluster Customers with a Product Recommendation System.” Journal of Computational and Theoretical Nanoscience Vol. 17, No. 1, pp. 347-352.IEEE, 2020.

[6] Kurmashov, Nursultan, Konstantin Latuta, and Abay Nussipbekov. ”Online book recommendation System.” In 2015 Twelve International Conference on Electronics Computer and Computation (ICECCO), pp. 1-4. IEEE, 2015.

[7] Kurup, Ayswarya R., and G. P. Sajeev. ”Task recommendation in rewardbased crowdsourcing systems.” In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1511-1518. IEEE, 2017.

[8] Maya L Pai1,Suchithra M. S2andDhanya M, ”Analysis of Soil Parameters for Proper Fertilizer Recommendation to Increase the Productivity of Paddy Field Cultivation ”,Department of Computer Science and ITSchool of Arts and Sciences,KochiAmrita Vishwa Vidyapeetham, International Journal of Advanced Science and Technology Vol. 29, No. 03, pp. 4681-4696.IEEE, 2020

[9] M¨obius Data Scientist at Healthcare Melbourne, Victoria, Australia” Book Recommendation Dataset Build state-of-the-art models for book recommendation system”, kaggle, 2016.

[10] Mohamed, Marwa Hussien, Mohamed Helmy Khafagy, and Mohamed Hasan

Ibrahim. ” Recommender systems challenges and solutions survey.” In 2019 International Conference on Innovative Trends in Computer Engineering (ITCE), pp. 149-155. IEEE, 2019.

[11] Murali, M. Viswa, T. G. Vishnu, and Nancy Victor. ”A Collaborative Filtering based Recommender System for Suggesting New Trends in Any Domain of Research.” In 2019 5th International Conference on Advanced Computing Communication Systems (ICACCS), pp. 550- 553. IEEE, 2019.

[12] PV Devika, K Jothisree, PV Rahul, S Arjun, Jayasree Narayan. “Book Recommendation System”, 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021

[13] Tewari, Anand Shanker, and Kumari Priyanka. ”Book recommendation system based on collaborative filtering and association rule mining for college students.” In 2014 International Conference on Contemporary Computing and Informatics (IC3I), pp. 135-138. IEEE, 2014.

APPENDIX

A. SOURCE CODE

Fig A.1 Refers to the implementation of the flask for the website deployment.

```
from flask import Flask,render_template,request
import pickle import numpy as np

popular_df=pickle.load(open('popular.pkl','rb'))
pt=pickle.load(open('pt.pkl','rb'))
books=pickle.load(open('books.pkl','rb'))
similarity_scores=pickle.load(open('similarity_scores.pkl','rb'))
app=Flask(__name__) @app.route('/') def index():
    return render_template('index.html',
                           book_name =
list(popular_df['BookTitle'].values),

author=list(popular_df['BookAuthor'].values),
                           image=list(popular_df['Image-URL-
M'].values),

votes=list(popular_df['num_ratings'].values),

rating=list(popular_df['avg_rating'].values)

                           )
@app.route('/recommendation') def
recommendation_ui():
    return render_template('recommendation.html')
@app.route('/recommend_books',methods=['post']) def
recommend():
    user_input=request.form.get('user_input')
index = np.where(pt.index == user_input)[0][0]
    similar_items = sorted(list(enumerate(similarity_scores[0])),
key=lambda x: x[1], reverse=True)[1:6]     data = []     for i in
similar_items:
        item = []
        temp_df = books[books['Book-Title'] == pt.index[i[0]]]
item.extend(temp_df.drop_duplicates("Book-
Title")['BookTitle'].values)
        item.extend(temp_df.drop_duplicates("Book-
Title")['BookAuthor'].values)
        item.extend(temp_df.drop_duplicates("Book-
Title")['ImageURL-M'].values)         data.append(item)
print(data)
    return render_template('recommendation.html',data=data)
 if
__name__=='__main__':
app.run(debug=True)
```

Fig A.1: Flask Sourcecode

B. SCREENSHOTS

First of all we are importing the required libraries and datasets (Fig B.1)

```
In [1]: import numpy as np
        import pandas as pd

In [39]:

In [2]: books=pd.read_csv('Books.csv')
        users=pd.read_csv('Users.csv')
        ratings=pd.read_csv('Ratings.csv')

<ipython-input-2-3b5cc6a4f891>:1: DtypeWarning: Columns (3) have mixed types. Specify dtype option on import or set low_memo
False.
  books=pd.read_csv('Books.csv')

In [3]: books.head()

Out[3]:
```

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-U
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...	http://images.amazon.com/images/P/01951534
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...	http://images.amazon.com/images/P/00020050

Fig B.1 Importing Libraries and Datasets

In the next step data pre-processing is carried out to modify the data as required (FigB.2)

```
Out[8]: ISBN                   0
        Book-Title             0
        Book-Author            1
        Year-Of-Publication    0
        Publisher              2
        Image-URL-S            0
        Image-URL-M            0
        Image-URL-L            3
        dtype: int64

In [9]: users.isnull().sum()

Out[9]: User-ID          0
        Location         0
        Age         110762
        dtype: int64

In [10]: ratings.isnull().sum()

Out[10]: User-ID        0
         ISBN           0
         Book-Rating    0
         dtype: int64

In [11]: books.duplicated().sum()

Out[11]: 0

In [12]: ratings.duplicated().sum()

Out[12]: 0

In [13]: users.duplicated().sum()

Out[13]: 0
```

Fig B.2 Data Pre-processing

In the Fig B.3, the books dataset is merged with ratings dataset to evaluate the highest average rating of the books

##POPULARITY BASED RECOMMENDER SYSTEM

```
In [14]: ratings_with_name=ratings.merge(books,on='ISBN')
```

```
In [15]: ratings_with_name.head()
```

Out[15]:

	User-ID	ISBN	Book-Rating	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-
0	276725	034545104X	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/0345451
1	2313	034545104X	5	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/0345451
2	6543	034545104X	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/0345451
3	8680	034545104X	5	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/0345451
4	10314	034545104X	9	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/0345451

```
In [16]: num_rating_df=ratings_with_name.groupby('Book-Title').count()['Book-Rating'].reset_index()
         num_rating_df.rename(columns={'Book-Rating':'num_ratings'},inplace=True)
         num_rating_df
```

Fig B.3 Popularity based filtering/ Content based filtering Model

Fig B.4 describes the output of the content based filtering model i.e. the top 10 highest average rating books are displayed

```
In [23]: popular_df[popular_df['num_ratings']>=250].sort_values('avg_rating',ascending=False).head(10)
```

Out[23]:

	Book-Title	num_ratings	avg_rating
80434	Harry Potter and the Prisoner of Azkaban (Book 3)	428	5.852804
80422	Harry Potter and the Goblet of Fire (Book 4)	387	5.824289
80441	Harry Potter and the Sorcerer's Stone (Book 1)	278	5.737410
80426	Harry Potter and the Order of the Phoenix (Boo...	347	5.501441
80414	Harry Potter and the Chamber of Secrets (Book 2)	556	5.183453
191612	The Hobbit : The Enchanting Prelude to The Lor...	281	5.007117
187377	The Fellowship of the Ring (The Lord of the Ri...	368	4.948370
80445	Harry Potter and the Sorcerer's Stone (Harry P...	575	4.895652
211384	The Two Towers (The Lord of the Rings, Part 2)	260	4.880769
219741	To Kill a Mockingbird	510	4.700000

Fig B.4 Popularity based filtering / Content based filtering output

Collaborative filtering model recommends the books to users based on the books interacted by the users(fig B.5)

```
In [36]: def recommend(book_name):
             #index fetch
             index=np.where(pt.index==book_name)[0][0]
             similar_items=sorted(list(enumerate(similarity_scores[index])),key=lambda x:x[1],reverse=True)[1:6]
             data=[]
             for i in similar_items:
                 item=[]
                 temp_df=books[books['Book-Title']==pt.index[i[0]]]
                 item.extend(temp_df.drop_duplicates("Book-Title")['Book-Title'].values)
                 item.extend(temp_df.drop_duplicates("Book-Title")['Book-Author'].values)
                 item.extend(temp_df.drop_duplicates("Book-Title")['Image-URL-M'].values)
                 data.append(item)
             return data

In [37]: recommend('To Kill a Mockingbird')

Out[37]: [['The Catcher in the Rye',
           'J.D. Salinger',
           'http://images.amazon.com/images/P/0316769487.01.MZZZZZZZ.jpg'],
          ['Five Quarters of the Orange',
           'Joanne Harris',
           'http://images.amazon.com/images/P/0060958022.01.MZZZZZZZ.jpg'],
          ['Drowning Ruth',
           'Christina Schwarz',
           'http://images.amazon.com/images/P/0385502532.01.MZZZZZZZ.jpg'],
          ['The Bean Trees',
           'Barbara Kingsolver',
           'http://images.amazon.com/images/P/0060915544.01.MZZZZZZZ.jpg'],
          ["The Color of Water: A Black Man's Tribute to His White Mother",
           'James McBride',
           'http://images.amazon.com/images/P/1573225789.01.MZZZZZZZ.jpg']]
```

Fig B.5 Collaborative Based Filtering model Output

RMSE and MAE values are shown in Fig B.6

```
In [44]: pip show pandas

         Note: you may need to restart the kernel to use updated packages.Name: pandas

         Version: 1.5.1
         Summary: Powerful data structures for data analysis, time series, and statistics
         Home-page: https://pandas.pydata.org
         Author: The Pandas Development Team
         Author-email: pandas-dev@python.org
         License: BSD-3-Clause
         Location: c:\users\lenovo\anaconda3\lib\site-packages
         Requires: numpy, python-dateutil, pytz
         Required-by: statsmodels, seaborn

In [40]: print("the mean absolute error is : ",mae)
         print("the RMSE value is : ",rmse)

         the mean absolute error is :  0.9764523415324321
         the RMSE value is :  1.836
```

Fig B.6 RMSE and MAE valuesw

Webpage deployment of content based filtering model is shown in Fig B.7

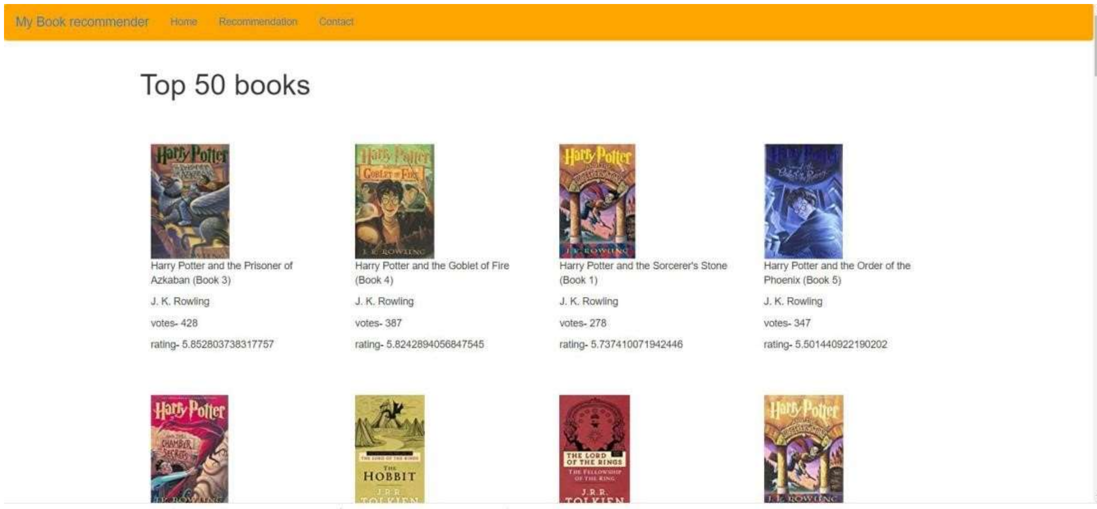

Fig B.7 Webpage for content-based filtering Model

Webpage deployment of collaborative based filtering is shown in the fig B.8

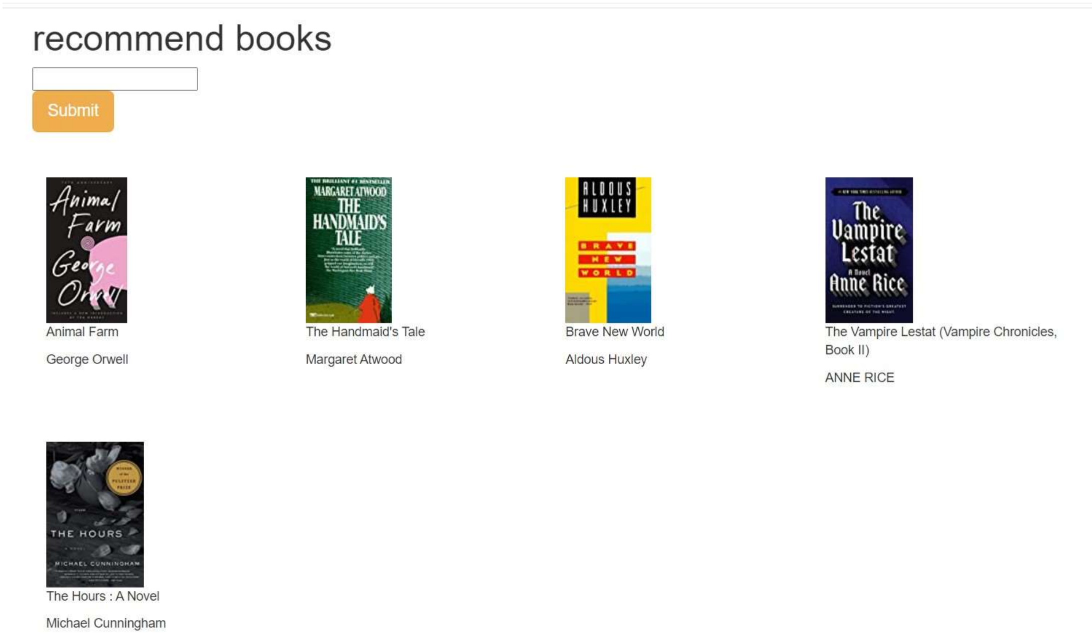

Fig B.8 Webpage for collaborative based filtering model

C. RESEARCH PAPER

Book Suggester System using Machine Learning

Bavisetti Gowtham Ajay
Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, India
bavisettiajay9@gmail.com

Betha Shanmukh Satyadev
Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, India
shanmukhsatyadev@gmail.com

Dr. M.D. Anto Praveena
Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, India
antopraveena.cse@sathyabama.ac.in

Abstract*—As human civilization and web technology become advanced, more and more data is being gathered by online users. This ocean of information provides a wide variety of resources. The question of how to investigate the utilization of these resources has become urgent. These data sets can be analyzed to extract hidden information or finding new information. Data mining is a technique for obtaining usable information from computers and technological methods. A typical tool in today's information age is the search engine. Every day, people utilize search engines like Google, Baidu, Bing, and others. Yet, due to its universality, they are unable to provide consumers with accurate search results that are tailored to their special needs and histories. In response, it suggests the idea of customized built and customized suggestion systems of creating a dual relationship among the user and products of information, utilizing current process of choosing or a familiar connection, quarrying similar suggestions of interest of every user, and carry out customized suggestions. The collaborative recommendation algorithm will be studied in this article along with a kmeans clustering and k-nearest neighbor method that is suggested.

Keywords*—Data mining, collaborative filtering, machine learning, and k-means clustering, k-nearest neighbor.

I. INTRODUCTION

Recommender Systems (RSs) are essential for increasing the e-commerce system's revenue. In recent years, RS has been used in many different fields. Its first concentration was on the entertainment and online retail industries. Its use case is being developed in a variety of fields, including learning online, net banking, booking tickets online, medication, social media, and others.

Customers find it difficult to purchase books from a large selection of books on an e-commerce website. The RS is the best option to reduce the overhead issue. because it enables the user to find an ideal book based on their interests. Our primary goal for research is to develop a suggestion of books for reading books in online.

The twenty-first century is here. Nearly everything in the world uses internet service these days. IoT is also identifying growth, which means that the most important objects in this world are digitalized. eLearning is crucial, particularly in the COVID 19 pandemic. The discipline of eLearning benefits greatly from the use of recommendation systems. The most beneficial is that eLearning offered to end customers was "Response based on the demand" that signifies the appeal of the client will be fulfilled. Because learning online includes variety of portable devices (including laptops, tablets, and smartphones).

The system provides prompt recommendations for any services and goods the customer requests "Whatever and Whenever" thanks to these portable technologies. Having a observation on the scenario when a user wishes to choose an e-book or other reading content but lacks the necessary personal experience. Unfortunately, because there were so many things or e-resources available, they were unable to find the right ones. The system that suggests is what we have suggested to avoid this situation. Fig.1 depicts the flow of the recommender system

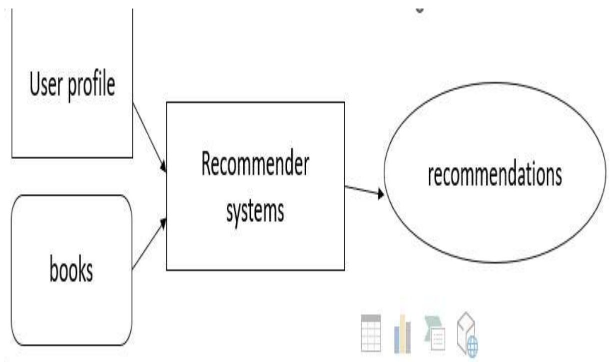

Fig. 1. Flow of a Recommender System

RS was first introduced by Resnick and Varion [1] for developing a collaborative filtering method [2]. Eventually, several researchers utilized the phrase and gave it different definitions. On the internet, numerous recommendation systems are presented and used in different sectors, and each one aims to accurately anticipate in accordance with user interest. Several researchers also looked into the novelty, privacy, scalability, accuracy, usability, and other aspects of RS. Many methods, including collaborative filtering, knowledge-based, demographic, and hybrid methods, can be used to make recommendations.

Collaborative Filtering (CF) is the method for creating recommendations that is most generally used and approved [3]. The CF technique finds comparable products that other users with similar interests favor based on prior interactions (rating, remark) [4]. Based on profile attributes, content-based filtering determines the user's choice based on the items' contents [5]. Knowledge-based works rely on knowledge expertise regarding the user and the item and suggest product based on which items are connected to the user's interests. When suggesting an item, Demographic RS bases its operation on user demographic information from their profiles (such as gender, age, location, etc.).

This article sections are structed as follows : A literature Review on recommender systems is included in Part II. many similarity methods engaged in the RS are presented in Section III. The suggested model of the suggestion of books for online learning is illustrated in Section IV and the words are briefly defined in Section V. The dataset, experimental evaluation, and result analysis are all described in Part VI. The section of the essay that concludes is Section VII.

II. LITERATURE REVIEW

The field of recommendation system has expanded significantly over the past 20 years. Customers have overloaded by the information which is addressed by offering unique, specialized material and repair advice. A few modern cutting-edge techniques are shown in the analysis sections.

Proposal frameworks deal with providing relevant advice in surprisingly intelligent ways. Typically, association rules, content-based filtering, and collaborative filtering are used. Moreover, a straightforward recommendation system for mobile applications is created without sacrificing rating, size, or permission features. There are two models provided, one based on worker incentive features and the other on worker job features. Using real-world datasets, the suggest strategy is compared to several connected methods.

Users decide on a set of data through collaborative filtering. When creating the user want facts set to represent the different facets of the user's profile, special characteristics are taken into consideration. It is a technique for providing users with up-to-date hints from a dataset based on how closely their side interest profiles resemble another user. It is quite tough to side-features for a product or query, and clean that technology cannot be controlled. Customers' information is used by a content-based recommender, either openly or through verification.

A user profile is created using this data, and it is then utilized to provide recommendations to the user. The algorithm recognizes customer's various interests and make suggestions based on the interests of the similar user interests. The capacity of this paradigm to increase the users' present benefits is confined. A huge data set of things can be explored using association rule mining to uncover fascinating associations and interactions between the products. It is a technique for searching for recurrent patterns, parallels, or similarities between datasets from different databases, including relational databases, transactional databases, and various kinds of repositories.

The foremost weakness of the prior methods is the only address one application, namely, an existing client receiving recommendations from the dataset, and do not address usage scenarios like proposing a new book to a new customer or a generic counsel based on the dataset. Our system includes each of these programs, giving users a selection of options for picking the ideal book to read.

The primary shortcoming of the earlier solutions is that they only handle one application, namely, a current customer receiving suggestions from the dataset and do not deal with use scenarios such as recommending a new book to a brand-new customer, or a generic advice based on the dataset. Each of these applications is included in our system, providing customers with a variety of options for selecting the appropriate book to read.

III. METHODOLOGY

Algorithm used in "Book Recommendation System" initiative aims to assist users in selecting the right choice of books that piques their interest and so motivate them to learn more. In this case, we're using Cosine similarity, KNN, and Pearson correlation. Seeing similarities across the books is becoming more and more routine. As previously indicated, the major three use cases for this system are suggestions for current users, recommendations for new users, and ratings for newly uploaded books. Several approaches are used to deal with each of these. The main method used in this project is collaborative filtering based on users. Depending upon the ratings given to the product by another reader who share the target user's preferences, the system predicts what a user will like. Fig 2 depicts the difference between the content based filtering and collaborative based filtering.

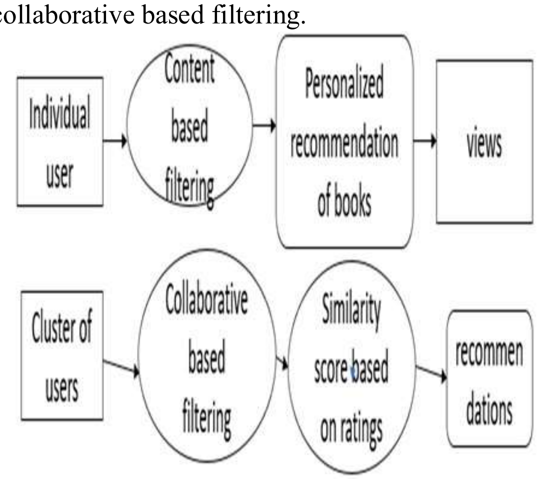

Fig. 2. Content based vs Collaborative based Filtering

A. Dataset Description

For the dataset, three csv files are taken. First one is the books file. The attributes of the books dataset are ISBN, Book title, Book author, year of publication, publisher, image URL small, image URL medium, image URL large. The second dataset is user's dataset. The user's dataset contains the user ID, location, age. The third dataset is the ratings dataset. The ratings dataset contains the user ID, ISBN of the book and the ratings provided by the users for the particular book.

B. Algorithm

1)Pearson correlation: To gauge the linear correlation between variables, co-efficient of Pearson R correlation is utilized. Its purpose is to build a recommendation using rating counts.

depending upon the ratings and the count of rating, a graph of rating distribution is drawn, and most individuals have given ratings of 0.

In order to obtain the count in desc order, first group the rating data frame by ISBN, then take the book rating column. Hence, if more people rate a book, it must be a very well-known book overall. In order to do that, we have created a data frame using all the ISBN numbers, then combined it with the books dataset's most popular titles using the ISBN field. Therefore, in conclusion, we can find the top 5 books based on the number of ratings here. The ratings mean as well as the ratings count were discovered for the correlation. Then, present the data in desc order after generating two fields in data frames: rating of the book and rating count. As a result, if we are using the rating count to determine popularity and my book does not have a decent rating but the greatest number of people have given it a rating, it cannot be a widely read book, so we should not promote it. In order for them to make recommendations, they must take into account the rating average and the amount of ratings. We have some factual importance where users with less than 200 evaluations and books with fewer than 100 evaluations are disallowed in contemplation to create the best recommendation system. At that time, having the opinion to combine user id and ISBN using the pivot table; therefore, by using the indexes as user id on the column features, it would show if the individual has provided any ratings or not. When the pivot is used, the rating table is actually transformed into a 2D matrix if the user has not provided any ratings, which is basically represented as nan. Then a correlation between the ratings and the ratings average is discovered. Fig 3 depicts the output of pearson algorithm.

	Book-Title	num_ratings
0	A Light in the Storm: The Civil War Diary of ...	4
1	Always Have Popsicles	1
2	Apple Magic (The Collector's series)	1
3	Ask Lily (Young Women of Faith: Lily Series, ...	1
4	Beyond IBM: Leadership Marketing and Finance ...	1
5	Clifford Visita El Hospital (Clifford El Gran...	1
6	Dark Justice	1
7	Deceived	2
8	Earth Prayers From around the World: 365 Pray...	10
9	Final Fantasy Anthology: Official Strategy Gu...	4

Fig. 3. Output of the Pearson algorithm

2) KNN: KNN is a machine learning estimator used to find set of alike readers depending upon basic book ratings and provide data using the average rating of the top k nearest neighbors. KNN is a machine learning estimator used to find groups of similar users based on basic book ratings and provide data utilizing the average rating of the highest k nearest neighbors.

In this suggestion system, similar readers are found using similarity of cosine by converting the table to a two-dimensional matrix and then filling in the gaps with NULL. The network data frame's values (ratings) were then converted into a SciPy sparse grid for more accurate computations. Finally, this computation will determine the similarity of the cosine between rating vectors. Unsupervised techniques with sklearn.neighbors are used to get the Closest Neighbors.

To determine how comparable two items are, one can use the cosine similarity measurement. In terms of numbers, it calculates the intersection cosine of two vectors that are visualized into a multidimensional space. The yield value ranges from 0 to 1. (0 denotes the absence of comparability, while 1 suggests that both elements are equally rated at 100.

A collaborative filtering recommendation system based on items is developed using KNN. The five books with the highest similarity scores to the selected book are listed. Since similarity is determined by distance, the higher the value, the less similarity there is. Following that, they are put in ascending order. When a user types in their favorite book from the dataset, the algorithm suggests the user's next five most related novels.

3) Collaborative filtering: Collaborative filtering is one of the methods we use to forecast the book ratings from an already subscribed user by computing the resemblance among readers. The fundamental concept of collaborative filtering is this. Let us say we want to make recommendations for user x. To start, we identify a lot of other users that have user x's likes and dislikes. K Nearest Neighbor is used to accomplish this. An unsupervised learning algorithm is nearest neighbor. It is imported as sklearn.neighbors from the sklearn library. When determining users who are similar, the user's previous ratings of the books are considered.

Three new features were developed. One to find the KNearest Neighbors, second is to forecast the average ratings of a specific book, and the third is to recommend the user the best-rated books. This information is transformed into a cross tabulation with ISBN columns and the id of the user as the index, and all 0 ratings are eliminated. The data is essentially represented as a matrix in the pivot table. It is determined how many books each user has reviewed. The system receives a user's distinctive ID, a distinctive ISBN of book, the number of products the reader has reviewed, and the cross tabulation of ratings.

The standard deviation and the user ratings mean are used to calculate the book ratings. The average of all the user-rated books is represented by the mean rating. The weighted average is calculated using both the mean rate of the mean of the comparable readers of the given book and the measure of their similarity. The distance produced via K Nearest Neighbor is used to determine the similarity value.

Prediction Rate = Rate of the Mean + Standard Deviation.

The user is then shown the top 10 books after the ratings are organized in descending order. If those projected ratings are fewer than 6, a message asking the user to examine some general suggestions determined by Pearson correlation is displayed. Fig 4 depicts the output of collaborative Filtering output

```
[['The Catcher in the Rye',
  'J.D. Salinger',
  'http://images.amazon.com/images/P/0316769487.01.MZZZZZZZ.jpg'],
 ['Five Quarters of the Orange',
  'Joanne Harris',
  'http://images.amazon.com/images/P/0060958022.01.MZZZZZZZ.jpg'],
 ['Drowning Ruth',
  'Christina Schwarz',
  'http://images.amazon.com/images/P/0385502532.01.MZZZZZZZ.jpg'],
 ['The Bean Trees',
  'Barbara Kingsolver',
  'http://images.amazon.com/images/P/0060915544.01.MZZZZZZZ.jpg'],
 ["The Color of Water: A Black Man's Tribute to His White Mother",
  'James McBride',
  'http://images.amazon.com/images/P/1573225789.01.MZZZZZZZ.jpg']]
```

Fig. 4. Collaborative Filtering output

4) Suggestion of books to the New Users: A few books should also be recommended to any new users who need to be added to the data set so that they can rate them. Future recommendations for that person can be enhanced using this rating. The top 10 books, as determined by the total average rating of all the books, are presented to each new user who has been added to this recommendation system. By dividing the total number of ratings for each book by the total number of ratings, the average ratings are determined. Afterwards, they are arranged in the order of decrement. Following the users' ratings of these books, collaborative filtering

	Book-Title	avg_rating
0	A Light in the Storm: The Civil War Diary of ...	2.25
1	Always Have Popsicles	0.00
2	Apple Magic (The Collector's series)	0.00
3	Ask Lily (Young Women of Faith: Lily Series, ...	8.00
4	Beyond IBM: Leadership Marketing and Finance ...	0.00
5	Clifford Visita El Hospital (Clifford El Gran...	0.00
6	Dark Justice	10.00
7	Deceived	0.00
8	Earth Prayers From around the World: 365 Pray...	5.00
9	Final Fantasy Anthology: Official Strategy Gu...	5.00

is used to create the users' subsequent suggestions. Fig 5 depicts the books with highest rating and fig 6 depicts the highest rated author's

Fig. 5. Highest average rated books

	Book-Title	num_ratings	avg_rating
80434	Harry Potter and the Prisoner of Azkaban (Book 3)	428	5.852804
80422	Harry Potter and the Goblet of Fire (Book 4)	387	5.824289
80441	Harry Potter and the Sorcerer's Stone (Book 1)	278	5.737410
80426	Harry Potter and the Order of the Phoenix (Boo...	347	5.501441
80414	Harry Potter and the Chamber of Secrets (Book 2)	556	5.183453
191612	The Hobbit : The Enchanting Prelude to The Lor...	281	5.007117
187377	The Fellowship of the Ring (The Lord of the Ri...	368	4.948370
80445	Harry Potter and the Sorcerer's Stone (Harry P...	575	4.895652
211384	The Two Towers (The Lord of the Rings, Part 2)	260	4.880769
219741	To Kill a Mockingbird	510	4.700000

From the ratings pivot table five readers were chosen randomly. The five users anticipated and saved ratin gs of the previously reviewed books . Also, each of the dataset's actual evaluations was saved. There were 666 total ratings. These served as a basis for calculating the Root Mean Squared Error and Mean Absolute Error values.

From the Fig.7 Root Mean Squared Error value is equal to 1.836 and Mean Absolute Error values is equal to 0.976.

```
the mean absolute error is : 0.9764523415324321
the RMSE value is : 1.836
```

Fig. 7. RMSE and MAE

Fig. 6. Highest-rated authors in terms of average

5) Inserting New Book: To be considered for a user suggestion and to be part of the data set, a new book needs to have a rating. The average ratings of all the author's other books that are already part of the data set are used to compute this score. According to their author, books are sorted. After adding up all of the ratings for each book, a percentage of the total ratings is calculated. It is also possible to see how many people read the book by that author by looking at the overall rating count. The new book that was added to the data collection is then given this rating as its initial rating.

IV. EXPERIMENTAL SETUP AND RESULTS

A. Performance Assessment

Two measures are used to evaluate the collaborative filtering's performance: -

1)RMSE: RMSE is a widely used and accepted approach for determining a model's error. 2) MAE: The only difference between this method and the former is that with MAE, In the case of linear continuous variables, the error is estimated.

V. CONCLUSIONS AND FUTURE WORK

This research suggests a machine learning algorithm named collaborative filtering mechanism for a recommendation of books. By giving book recommendations, people's reading habits improve, which boosts their vocabulary, expertise, and knowledge. The data gathered from reader reviews of completely unrelated novels is utilized. The dataset has an excessive several books and a wide variety of consumers. Our designed system makes the most of the information's unique alternatives to provide a speedy response and high-caliber recommendations. Collaborative Filtering is the algorithm used by the system to generate recommendations. The cosine similarity approach is used to precisely quantify the similarities between the users. Based on the average ratings calculated and gathered from the various users the top rated gooks are suggested for the book readers.

Time constraints are the biggest drawbacks of this machine learning model named collaborative filtering. Collaborative filtering consumes more time to suggest books to the readersbecause of the vastness of the dataset. Usage of the supervised instruction, this can be easily resolved. If the customer seeking for a book to read, our suggestion system may be just what you are looking for. It offers complete satisfaction.

REFERENCES

[1] Kurmashov, Nursultan, Konstantin Latuta, and Abay Nussipbekov. ”Online book recommendation System.” In 2015 Twelve International Conference on Electronics Computer and Computation (ICECCO), pp. 1-4. IEEE, 2015.

[2] Tewari, Anand Shanker, and Kumari Priyanka. ”Book recommendation system based on collaborative filtering and association rule mining for college students.” In 2014 International Conference on Contemporary Computing and Informatics (IC3I), pp. 135-138. IEEE, 2014.

[3] Mohamed, Marwa Hussien, Mohamed Helmy Khafagy, and Mohamed Hasan Ibrahim. ” Recommender systems challenges and solutions survey.” In 2019 International Conference on Innovative Trends in Computer Engineering (ITCE), pp. 149-155. IEEE, 2019.

[4] Badriyah, Tessy, Erry Tri Wijayanto, Iwan Syarif, and Prima Kristalina. ”A hybrid recommendation system for E-commerce based on product description and user profile.” In 2017 Seventh International Conference on Innovative Computing Technology (INTECH), pp. 95-100. IEEE, 2017.

[5] Ahuja, Rishabh, Arun Solanki, and Anand Nayyar.” Movie recommender system using K-Means clustering and K-Nearest Neighbor.” In 2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence), pp. 263-268. IEEE, 2019.

[6] Murali, M. Viswa, T. G. Vishnu, and Nancy Victor. ”A Collaborative Filtering based Recommender System for Suggesting New Trends in Any Domain of Research.” In 2019 5th International Conference on Advanced Computing Communication Systems (ICACCS), pp. 550- 553. IEEE, 2019.

[7] Chen, Junnan, Courtney Miller, and Gaby G. Dagher. ”Product recommendation system for small online retailers using association rules mining.” In Proceedings of the 2014 International Conference on Innovative Design and Manufacturing (ICIDM), pp. 7177. IEEE, 2014.

[8] M¨obius Data Scientist at Healthcare Melbourne, Victoria, Australia” Book Recommendation Dataset Build state-of-the-art models for book recommendation system”, kaggle, 2016.

[9] Kurup, Ayswarya R., and G. P. Sajeev. ”Task recommendation in rewardbased crowdsourcing systems.” In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1511-1518. IEEE, 2017.

[10] Maya L Pai1,Suchithra M. S2andDhanya M, ”Analysis of Soil Parameters for Proper Fertilizer Recommendation to Increase the Productivity of Paddy Field Cultivation ”,Department of Computer Science and ITSchool of Arts and Sciences,KochiAmrita Vishwa Vidyapeetham, International Journal of Advanced Science and Technology Vol. 29, No. 03, pp.

4681-4696.IEEE, 2020

[11] Jisha, R. C., Ram Krishnan, and Varun Vikraman. ”Mobile applications recommendation based on user ratings and permissions.” In 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 10001005.IEEE, 2018.

[12] Keerthana, N. K., Shriram K. Vasudevan, and Nalini Sampath. ”An Effective Approach to Cluster Customers with a Product Recommendation System.” Journal of Computational and Theoretical Nanoscience Vol. 17, No. 1, pp. 347-352.IEEE, 2020.

[13] PV Devika, K Jothisree, PV Rahul, S Arjun, Jayasree Narayan. “Book Recommendation System”, 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021