BOOK RECOMMENDATION SYSTEM USING MACHINE LEARNING

A major Project Report submitted in the partial fulfillment of

the requirements for the award of the degree

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by

G. Siva Sankar Tarun (19471A0522)

G. Anjani Sudheer (19471A0520)

K. Kalyan (19471A0533)

Under the Esteemed Guidance of

Dr. B. Jhansi Vazram M.Tech., Ph.D.

Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NARASARAOPETA ENGINEERING COLLEGE: NARASARAOPET

(AUTONOMOUS)

Accredited by NAAC with A+ Grade and NBA under Tier -1

NIRF rank in the band of 251-320 and an ISO 9001:2015 Certified

Approved by AICTE, New Delhi, Permanently Affiliated to JNTUK, Kakinada

KOTAPPAKONDA ROAD, YALAMANDA VILLAGE, NARASARAOPET- 522601

2022-2023

NARASARAOPETA ENGINEERING COLLEGE: NARASARAOPET

(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the major project entitled " BOOK RECOMMENDATION SYSTEM USING MACHINE LEARNING**" is a bonafide work done by "G. Siva Sankar Tarun 19471A0522, G. Anjani Sudheer 19471A0520, K. Kalyan 19471A0533" in partial fulfilment of the requirements for the award of the degree of BACHELOR OF TECHNOLOGY in the department of COMPUTER SCIENCE AND ENGINEERING** during the academic year **2022-2023** and found worthy of acceptance to requirements of the university.

PROJECT GUIDE

Dr. B. Jhansi Vazram M.Tech., Ph.D.

Professor

PROJECT CO-ORDINATOR

Dr. M. Sireesha M.Tech., Ph.D.

Assoc.Professor

HEAD OF THE DEPARTMENT

Dr. S. N. TirumalaRao M.Tech., Ph.D.

Professor

EXTERNAL EXAMINER

DECLARATION

We declare that this project work titled “BOOK RECOMMENDATION SYSTEM USING MACHINE LEARNING” is composed by ourself that the work contain here is our own except where explicitly stated otherwise in the text and that this work has not been submitted for any other degree or professional qualification except as specified.

G. Siva Sankar Tarun (19471A0522)

G. Anjani Sudheer (19471A0520)

K. Kalyan (19471A0533)

ACKNOWLEDGEMENT

We wish to express our thanks to carious personalities who are responsible for the completion of the project. We are extremely thankful toour beloved chairman sir **M.V. Koteswara Rao,** B.sc who took keen interest in us in every effort throughout this course. We owe out gratitude to our principal **Dr. M. Sreenivasa Kumar** M.Tech, Ph.D(UK),MISTE,FIE(1) for his kind attention and valuable guidance throughout the course.

We express our deep-felt gratitude to **Dr. S. N. Tirumala Rao** M.Tech., Ph.D H.O.D. of CSE department and our guide **Dr. B. Jhansi Vazram** M.Tech., Ph.D. Professor of CSE department whose valuable guidance and unstinting encouragement enable us to accomplish our project successfully in time.

We extend our sincere thanks to **Dr. M. Sireesha** M.Tech., Ph.D. Associate Professor and Coordinator of the project for extending her encouragement. Their profound knowledge and willingness have been a constant source ofinspiration for us throughout this project work.

We extend our sincere thanks to all other teaching and non-teaching staff to department for their cooperation and encouragement during our B.Tech degree. We have no words to acknowledge the warm affection, constant inspiration and encouragement that we receive from our parents. We affectionately acknowledge the encouragement received from our friends thosewho involved in giving valuable suggestions had clarifying out all doubts which had really helped us in successfully completing our project.

By

G. Siva Sankar Tarun (19471A0522)

G. Anjani Sudheer (19471A0520)

K. Kalyan (19471A0533)

ABSTRACT

As the amounts of online books are exponentially increasing due to COVID-19 pandemic, finding relevant books from a vast e-book space becomes a tremendous challenge for online users. Personal recommendation systems have been emerged to conduct effective search which mine related books based on user rating and interest.

Most of these existing systems are user-based ratings where content-based and collaborative learning methods are used. This system proposed an effective system for recommending books for online users that rated a book using the clustering method and then found a similarity of that book to suggest a new book.

A book recommendation system using machine learning is an effective solution to help users discover books they might enjoy based on their past reading habits and preferences. In this system, a collaborative filtering algorithm is used to predict the likelihood of a user enjoying a particular book based on the ratings of similar users. The system uses both implicit and explicit feedback from users to generate personalized recommendations.

The datasets used in the system includes book metadata such as book-title, isbn, author, publisher, and year of publication, as well as user information such as age, location, and reading history and ratings information like user-id, isbn, book-rating. A hybrid approach is used to combine both content-based and collaborative filtering techniques is used to generate recommendations.

Overall, this book recommendation system using machine learning is an effective way to help users discover new books that match their preferences and interests. It can be easily integrated into existing book platforms and can improve user engagement and satisfaction.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE VISION AND MISSION

INSTITUTION VISION

To emerge as a Centre of excellence in technical education with a blend of effective student centric teaching learning practices as well as research for the transformation of lives and community.

INSTITUTION MISSION

M1: Provide the best class infra-structure to explore the field of engineering and research.

M2: Build a passionate and a determined team of faculty with student centric teaching, imbibing experiential, innovative skills.

M3: Imbibe lifelong learning skills, entrepreneurial skills and ethical values in students for addressing societal problems.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VISION OF THE DEPARTMENT

To become a centre of excellence in nurturing the quality Computer Science & Engineering professionals embedded with software knowledge, aptitude for research and ethical values to cater to the needs of industry and society.

MISSION OF THE DEPARTMENT

The department of Computer Science and Engineering is committed to

M1: Mould the students to become Software Professionals, Researchers and Entrepreneurs by providing advanced laboratories.

M2: Impart high quality professional training to get expertise in modern software tools and technologies to cater to the real time requirements of the industry.

M3: Inculcate team work and lifelong learning among students with a sense of societal and ethical responsibilities.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Program Specific Outcomes (PSOs)

PSO1: Apply mathematical and scientific skills in numerous areas of Computer Science and Engineering to design and develop software-based systems.

PSO2: Acquaint module knowledge on emerging trends of the modern era in Computer Science and Engineering.

PSO3: Promote novel applications that meet the needs of entrepreneur, environmental and social issues.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Program Educational Objectives (PEOs)

The graduates of the programmed are able to:

PEO1: Apply the knowledge of Mathematics, Science and Engineering fundamentals to identify and solve Computer Science and Engineering problems.

PEO2: Use various software tools and technologies to solve problems related to academia, industry and society.

PEO3: Work with ethical and moral values in the multi-disciplinary teams and can communicate effectively among team members with continuous learning.

PEO4: Pursue higher studies and develop their career in software industry.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Program Outcomes

1. **Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis:** Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. **Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. **Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. **Life-long learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Project Course Outcomes (COs):

CO425.1: Analyse the System of Examinations and identify the problem.

CO425.2: Identify and classify the requirements.

CO425.3: Review the Related Literature.

CO425.4: Design and Modularize the project.

CO425.5: Construct, Integrate, Test and Implement the Project.

CO425.6: Prepare the project Documentation and present the Report using appropriate method.

Course Outcomes – Program Outcomes mapping

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C425.1		√											√		
C425.2	√		√		√								√		
C425.3				√		√	√	√					√		
C425.4			√			√	√	√					√	√	
C425.5					√	√	√	√	√	√	√	√	√	√	√
C425.6									√	√	√		√	√	

Course Outcomes – Program Outcome correlation

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C425.1	2	3											2		
C425.2			2		3								2		
C425.3				2		2	3	3					2		
C425.4			2			1	1	2					3	2	
C425.5					3	3	3	2	3	2	2	1	3	2	1
C425.6									3	2	1		2	3	

Note: The values in the above table represent the level of correlation between CO's and PO's:

1. Low level
2. Medium level
3. High level

Project mapping with various courses of Curriculum with Attained PO's:

Name of the course from which principles are applied in this project	Description of the device	Attained PO
C3.2.4, C3.2.5	Gathering the requirements and defining the problem, plan to develop a smart bottle for health care using sensors.	PO1, PO3
CC4.2.5	Each and every requirement is critically analyzed, the process model is identified and divided into five modules	PO2, PO3
CC4.2.5	Logical design is done by using the unified modelling language which involves individual team work	PO3, PO5, PO9
CC4.2.5	Each and every module is tested, integrated, and evaluated in our project	PO1, PO5
CC4.2.5	Documentation is done by all our four members in the form of a group	PO10
CC4.2.5	Each and every phase of the work in group is presented periodically	PO10, PO11
CC4.2.5	Implementation is done and the project will be handled by the hospital management and in future updates in our project can be done based on air bubbles occurring in liquid in saline.	PO4, PO7
CC4.2.8 CC4.2.	The physical design includes hardware components like sensors, gsm module, software and Arduino.	PO5, PO6

INDEX

List of figures

1. INTRODUCTION

1.1 Introduction

There is an ever-increasing amount of information available to us in today's recent times, that includes books, music, movies, and more. It can be testing for users to find what they are consider for with this more amount of information and to develop new items that match their interests and satisfaction. Through the vast amount of information this is where recommendation systems come in, helping users to navigate and find what they are looking for more easily.

This highly scalable and trustworthy system for suggesting books can be adjusted for various genres and user preferences. This books recommendation system can be integrated into various types of online stores and digital libraries to offer customers which has individualized books recommendations. A dataset of books and users was used to calculate this system's performance, and the findings demonstrate that the proposed system outperforms current book recommendation systems. With the assistance of this project, users will be able to find new books to enjoy and contribute to the creation of a more sophisticated and individualized.

Book recommendation systems use machine learning algorithms to analyse data from various sources such as users, ratings and books datasets to provide customized recommendations to users. These recommendations can help users to optimize their Books recommends, reduce the amount of time, and mitigate risks associated with users' information and data. This System presents a Book recommendation system that utilizes machine learning algorithms to provide accurate and personalized recommendations to users. This system analyzes data from various sources, and performed various methods such as Popularity-based and collaborative learning methods are used. The system's primary goal is to provides the quality of the recommendations which is very accurate, easy to maintain and simply to use which is using by the user.

The description of the items is used in content-based filtering, which provides suggestions for items that are comparable to the description of the items. Book are recommended using these multiple filtering models depending on multiple the book's content and the user actions. As a result, mine recommendation engine also recommends books to new readers. For clustering the users in this study, we using both techniques as: K-means and Gaussian mixture. To calculate error between absolute numbers and the results, use the Root Mean Square Error formula. RMSE

number used to determine basic accuracy. The similarity of users is determined by the similarity of the ratings given by the users to an item. In this system, we used two methods K-means and Gaussian mixture for clustering the users. The better model is selected based on the silhouette score and used for clustering. So, our recommendation system recommends books to the new users also. In this recommender system, books are recommended based on collaborative filtering technique and similar books are shown using content-based filtering

1.2 Existing System

The primary problem with this is that these systems' recommendations are not transparent. Users may not always understand why a specific book was suggested, which can make it challenging for them to believe in the system and its suggestions. Additionally, the suggestions might not always suit the user's preferences or tastes, which could cause frustration and user engagement.

Disadvantages:

1. Provides single book to choose.
2. Did not provide accurate books to choose.

1.3 Proposed System

Proposed system overcomes the disadvantages of the existing system Overall, these existing systems use a variety of machine learning techniques to generate personalized book recommendations for their users. They take into user's preferences, and interests, as well as book metadata to generate accurate and relevant recommendations.

Advantages:

1. It Provides multiple books to choose based on user rating and interest.
2. users can choose any books among the search books with accurate and quality recommendation .

1.4. System Requirements

1.4.1 Hardware Requirements:

- System type : I3 processor or higher versions
- Cache memory : 4 MB
- RAM : 2 GB or more
- Hard Disc : 8 GB

1.4.2 Software Requirements:

- Operating system : windows 10
- Coding language : Python
- Browser : Any latest browser like chrome
- Python distribution : Anaconda, Flask

1.5 Introduction to Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions inthe future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. A popular saying goes that we are living in an "information age". Terabytes of data are produced every day. Data mining is the process that transforms information into a set of data.

In order to enable the software to independently generate solutions, the prior action of people is necessary. For example, the required algorithms and data must be fed into the systems in advance and the respective analysis rules for the recognition of patterns in the data stock must be defined. Once these two steps have been completed, the system can perform the following tasks by Machine Learning:

- Finding, extracting and summarizing relevant data
- Making predictions based on the analysis data
- Calculating probabilities for specific results

1.6 Some machine learning methods

Machine learning algorithms are often categorized as supervised and unsupervised.

- **Supervised machine learning algorithms** can apply what has been learned in thepast to new data using labeled examples to predict future events. Starting from the analysisof a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any newinput after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors

in order to modify the model

I. Regression : Linear regression is a linear model, e.g., a model that assumes a linear relationship between the input variables (x) and the single output variable (y). More specifically, that y can be calculated from a linear combination of the input variables (x). When there is a single input variable (x), the method is referred to as simple linear regression. When there are multiple input variables, literature from statistics often refers to the method as multiple linear regression.

II. Classification : Classification is a process of categorizing a given set of data into classes, it can be performed on both structured or unstructured data. The process starts with predicting the class of given data points. The classes are often referred to as target, label or categories.

- **unsupervised machine learning algorithms** are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasetsto describe hidden structures from unlabeled data.

  I. Clustering : A clustering problem is where you want to discover the inherent grouping in the data such as grouping customers by purchasing behavior

  II. Association : An Association rule learning problem is where you want to discover rules that describe large portions of your data such as people that buy X also tend to buy Y.

- **Reinforcement machine learning algorithms** is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best. This is known as the reinforcement signal.

1.7 Applications of machine learning

I. Virtual Personal Assistants

II. Predictions while Commuting

III. Videos Surveillance

IV. Social Media Services

V. Email Spam and Malware Filtering

VI. Online Customer Support

VII. Search Engine Result Refining

VIII. Product Recommendations

IX. Online Fraud Detection

2. LITERATURE SURVEY

2.1 Collaborative Filtering with Jaccard Similarity to build a recommendation system

Avi Rana and K. Deeba, et.al. (2019) [1] proposed a paper "Online Book Recommendation System using Collaborative Filtering (With Jaccard Similarity)". In this paper, the author used CF with Jaccard similarity to get more accurate recommendations because general CF difficulties are scalability, sparsity, and cold start. So to overcome these difficulties, they used CF with Jaccard Similarity. JS is based on pair of books index which is a ratio of common users who have rated both books divided by the sum of users who have rated books individually. Books with a high JS index are highly recommended.

2.2 Building a Recommendation System using Keras Deep learning Framework

G. Naveen Kishore, et.al. (2019) [2] proposed a paper "Online Book Recommendation System". The dataset used in this paper was taken from the website "good books-10k dataset" which contains ten thousand unique books. Features are book-id, user-id, and rating. In this paper, the author adopted a Keras deep learning framework model to create neural network embedding.

2.3 Using Quick sort Algorithm approach to design a system

Uko E Okon, et.al. (2018) [3] proposed a paper "An Improved Online Book Recommender System using Collaborative Filtering Algorithm". The authors designed and developed a recommendation model by using a quick sort algorithm, 13collaborative filtering, and object-oriented analysis and design methodology (OOADM). This system produces an accuracy of 90-95%.

2.4 Using UV Decomposition and KNN for building system

Jinny Cho, et.al. (2016) [4] proposed a paper "Book Recommendation System". In this paper, the author uses two approach methods which are Content-based (CB) and Collaborative Filtering (CF). They used two algorithms as UV-Decomposition and K Nearest Neighbors (KNN). They obtained a result with an accuracy of 85%.

2.5 Recommending books through CB and CF approaches

Sushma Rjpurkar, et.al. (2015) [5] proposed a paper "Book Recommendation System". In this paper, the author used Associative Rule Mining to find association and correlation relationships among a dataset of items. They used CB and CF approaches to build a system.

2.6 Detecting patterns, correlations and uses Collaborative Filtering and Associative Rule Mining

Abhay E. Patil, et.al. (2019) [6] proposed a paper "Online Book Recommendation System using Association Rule Mining and Collaborative Filtering". The author detected recurrently occurring patterns, correlations and uses various databases such as relational databases, transactional databases to form associations. They used two approaches i.e., User-based and Item-based Collaborative Filtering, and used the Pearson correlation coefficient to find similarity between the items.

2.7 Uses Demographic, Collaborative Filtering, Content-based to build a Hybrid Recommender System

Suhas Patil, et.al. (2016) [7] proposed a paper "A Proposed Hybrid Book Recommender System". In this paper, the author used techniques such as 14Demographic, Collaborative Filtering, Content-based to build a system and rarely they combined the features of these techniques to make a better recommendation system.

2.8 PHP-based CF, Fuggy logic , Context Engine for recommendation systems

Ankit Khera, et.al. (2008) [8] proposed a paper “Online Recommendation System”. In this paper, the author used the User similarity matrix, Vogoo which is PHP-based CF, Fuggy logic, Context Engine for building recommendation systems. Pearson Correlation is a similarity function in this paper.

2.9 Hybrid Recommender System through Collaborative Filtering

Anagha Vaidya and Dr. Subhash Shinde, et.al. (2019) [9] proposed a paper “Hybrid Book Recommendation System”. In this paper, the author used techniques such as Collaborative Filtering etc. and used the Pearson correlation coefficient. It was published in International Research Journal of Engineering and Technology (IRJET).

2.10 Using Machine Learning Algorithm to build a system

Dhirman Sarma,Tanni Mittra and Mohammad Shahadat Hossain, et.al. (2019) [10] proposed a paper “Personalized Book Recommendation System using Machine Learning Algorithm”. It was published in The Science and Information Organization vol.12.

3. SYSTEM ANALYSIS

3.1 System Requirements

A requirement is a feature that the system must have or a constraint that it must to be accepted by the client. Requirement Engineering aims at defining the requirements of the system under construction. Requirement Engineering include two main activities requirement elicitation which results in the specification of the system that the client understands and analysis which in analysis model that the developer can unambiguously interpret. A requirement is a statement about what the proposed system will do.

Requirements can be divided into two major categories:

- Functional Requirements.
- Non-Functional Requirements.

3.1.1 Functional Requirements :

A Functional Requirement is a description of the service that the software must offer. It describes a software system or its component. A function is nothing but inputs to the software system, its behavior, and outputs. It can be a calculation, data manipulation, business process, user interaction, or any other specific functionality which defines what function a system is likely to perform. Functional Requirements describe the interactions between the system and its environment independent of its application.

- Applying the algorithms on the train data
- Display the recommendations by the model.

3.1.2 Non-Functional Requirements:

Non-Functional Requirements specifies the quality attribute of a software system. They judge the software system based on Responsiveness, Usability, Security, Portability and other non-functional standards that are critical to the success of the software system. Example of

nonfunctional requirement, “how fast does the website load?” Failing to meet non-functional requirements can result in systems that fail to satisfy user needs. Non-functional Requirements allows you to impose constraints or restrictions on the design of the system across the various agile backlogs.

- Accuracy
- Reliability
- Flexibility

3.2. Importance of machine learning in recommendation system

Machine learning is becoming increasingly important in the recommendation sector due to the vast amount of data generated by users, which makes it difficult to provide personalized recommendations manually. By utilizing machine learning algorithms, companies can efficiently process large volumes of data and provide users with relevant recommendations based on their interests and past behaviors.

I. Personalization: Machine learning algorithms can analyze user behavior data, including search history, purchase history, and product interactions, to provide personalized recommendations. This ensures that users receive relevant recommendations that are tailored to their specific needs and interests.

II. Scalability: The recommendation sector generates a large volume of data that can be difficult to process manually. Machine learning algorithms can process this data quickly and efficiently, allowing companies to scale their recommendation engines to handle large volumes of users.

III. Accuracy: Machine learning algorithms can analyze complex data sets and identify patterns and trends that would be difficult to identify manually. This allows recommendation engines to provide highly accurate recommendations, improving user engagement and conversion rates.

IV. Adaptability: Machine learning algorithms can adapt to changes in user behavior and preferences over time. This ensures that recommendations remain relevant and up-to-date,

improving user satisfaction and loyalty.

Overall, machine learning plays a crucial role in the recommendation sector by providing scalable, accurate, and personalized recommendations to users. As the volume of data continues to grow, the importance of machine learning in the recommendation sector is likely to increase even further.

3.3 Implementation of machine learning using Python

Python is a popular programming language. It was created in 1991 by Guido van Rossum.It is used for:

1.web development (server-side),

2.software development,

3. mathematics,

4.system scripting.

The most recent major version of Python is Python 3. However, Python 2, although not being updated with anything other than security updates, is still quite popular.

It is possible to write Python in an Integrated Development Environment, such as Thonny, PyCharm, Net beans or Eclipse, Anaconda which are particularly useful when managing larger collections of Python files.

Python was designed for its readability. Python uses new lines to complete a command, asopposed to other programming languages which often use semicolons or parentheses.

Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose.

In the older days, people used to perform Machine Learning tasks manually by coding all the algorithms and mathematical and statistical formula. This made the process time consuming, tedious and inefficient. But in the modern days, it is become very much easy and efficient compared to the olden days by various python libraries, frameworks, and modules. Today, Python is one of the most popular programming languages for this task and it has replaced many languages in the industry, one of the reason is its vast collection of libraries. Python libraries that used in

Machine Learning are:

I. Numpy

II. SciPy

III. Scikit-learn

IV. Pandas

V. Matplotlib

NumPy is a very popular python library for large multi-dimensional array and matrix processing, with the help of a large collection of high-level mathematical functions. It is very useful for fundamental scientific computations in Machine Learning. It is particularly useful for linear algebra, Fourier transform, and random number capabilities. High-end libraries like TensorFlow uses NumPy internally for manipulation Tensors.

SciPy is a very popular library among Machine Learning enthusiasts as it contains different modules for optimization, linear algebra, integration and statistics. There is a difference between the SciPy library and the SciPy stack. The SciPy is one of the core packages that make up the SciPy stack. SciPy is also very useful for image manipulation.

Skikit-learn is one of the most popular Machine Learning libraries for classical Machine Learning algorithms. It is built on top of two basic Python libraries, NumPy and SciPy. Scikit-learn supports most of the supervised and unsupervised learning algorithms. Scikit learn can also be used for data-mining and data-analysis, which makes it a great tool who is starting out with Machine Learning.

Pandas is a popular Python library for data analysis. It is not directly related to Machine Learning. As we know that the dataset must be prepared before training. In this case, Pandas comes handy as it was developed specifically for data extraction and preparation. It provides high-level data structures and wide variety tools for data analysis. It provides many inbuilt methods for groping, combining and filtering data.

Matplotlib is a very popular Python library for data visualization. Like Pandas, it is not directly related to Machine Learning. It particularly comes in handy when a programmer wants to visualize

the patterns in the data. It is a 2D plotting library used for creating 2D graphs and plots. A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes, etc. It provides various kinds of graphs and plots for data visualization, histogram, error charts, bar chats, etc.

3.4. Scope of the project

The main scope of this project is used to speed up recommendations which is to create such a system, which can provide quality recommendations to their users without the need for long-term registration and maintain high quality recommendation system.

3.5. Analysis

we have 3 files in our dataset which is extracted from the kaggle website.

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.

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/0060973129.0...
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0...
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0...

Fig:3.4.1 Book Dataset.

	User-ID	ISBN	Book-Rating
0	276725	034545104X	0
1	276726	0155061224	5
2	276727	0446520802	0
3	276729	052165615X	3
4	276729	0521795028	6

Fig:3.4.2 Rating Dataset.

	User-ID	Location	Age
0	1	nyc, new york, usa	NaN
1	2	stockton, california, usa	18.0
2	3	moscow, yukon territory, russia	NaN
3	4	porto, v.n.gaia, portugal	17.0
4	5	farnborough, hants, united kingdom	NaN

Fig:3.4.3 User Dataset.

3.6 Data Pre-processing

Before feeding data to an algorithm, we have to apply transformations to our data which is referred as pre-processing. By performing pre-processing, the raw data which is not feasible for analysis is converted into clean data. In-order to achieve better results using a model in Machine Learning, data format has to be in a proper manner. The data should be in a particular format for different algorithms. For example, if we consider Random Forestalgorithm it does not support null values. So that those null values have to be managed using raw data.

Data Pre-processing:

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

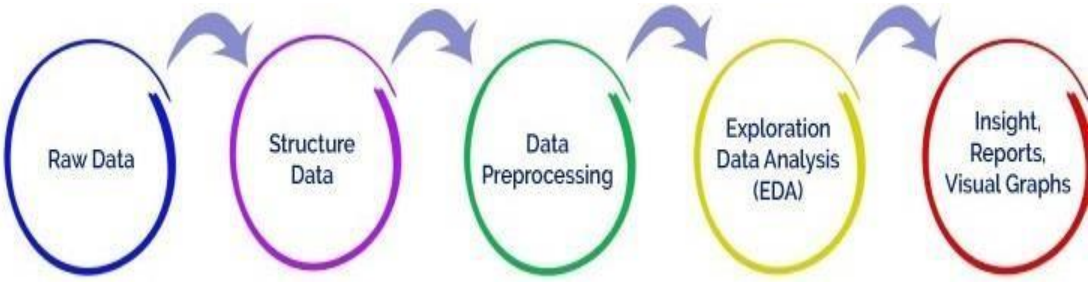

Fig:3.5(1). Data Preprocessing

Need of Data Preprocessing: For achieving better results from the applied model in Machine Learning projects the format of the data has to be in a proper manner. Some specified Machine Learning model needs information in a specified format. For example, Random Forest algorithm does not support null values, therefore to execute random forest algorithm null values have to be managed from the original raw data set. Another aspect isthat data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in one data set, and best out of them is chosen.

3.6.1 Missing values

Filling missing values is one of the pre-processing techniques. The missing values in the dataset is represented as '?' but it a non-standard missing value and it has to be converted into a standard missing value NaN. So that pandas can detect the missing values. In my collected dataset there are no missing values.

3.6.2 Correlation coefficient method

We can find dependency between two attributes p and q using Correlation coefficient method using the formula.

$r_{p,q} = \sum(p_i-p)(q_i-q)/n\sigma p\sigma q = \sum(p_i q_i)-np\, q/\, n\sigma p\sigma q$

n is the total number of patterns, pi and qi are respective values of p and q attributes in patterns i,

p and q are respective mean values of p and q attributes, σp , σq are respective standard deviations values of p and q attributes. Generally, -1≤ rp, q ≤ +1. If rp, q < 0, thenp and q are negatively correlated. If rp, q =0, then p and q are independent attributes and there is no correlation between them. If rp, q > 0, then p and q are positively correlated. Wecan drop the attributes that are having correlation coefficient value as 0 as it indicates that the variables are independent with respect to the prediction attribute. Fig:3.8.2 is the correlation matrix. There are no correlated features in the dataset.

	User-ID	ISBN	Book-Rating	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S
63	278418	0446520802	0	The Notebook	Nicholas Sparks	1996	Warner Books	http://images.amazon.com/images/P/0446520802.0...
65	3363	0446520802	0	The Notebook	Nicholas Sparks	1996	Warner Books	http://images.amazon.com/images/P/0446520802.0...
66	7158	0446520802	10	The Notebook	Nicholas Sparks	1996	Warner Books	http://images.amazon.com/images/P/0446520802.0...
69	11676	0446520802	10	The Notebook	Nicholas Sparks	1996	Warner Books	http://images.amazon.com/images/P/0446520802.0...
74	23768	0446520802	6	The Notebook	Nicholas Sparks	1996	Warner Books	http://images.amazon.com/images/P/0446520802.0...
...	...	...	...	...	...	...	...	...
1026724	266865	0531001725	10	The Catcher in the Rye	Jerome David Salinger	1973	Scholastic Library Pub	http://images.amazon.com/images/P/0531001725.0...
1027923	269566	0670809381	0	Echoes	Maeve Binchy	1986	Penguin USA	http://images.amazon.com/images/P/0670809381.0...
1028777	271284	0440910927	0	The Rainmaker	John Grisham	1995	Island	http://images.amazon.com/images/P/0440910927.0...
1029070	271705	B0001PIOX4	0	Fahrenheit 451	Ray Bradbury	1993	Simon & Schuster	http://images.amazon.com/images/P/B0001PIOX4.0...

Fig:3.6.2.1 Correlation

3.6.3 Cross Validation

Cross-validation is a technique in which we train our model using the subset of the data- set and then evaluate using the complementary subset of the data-set. The three steps involved in cross-validation are as follows :

– Reserve some portion of sample data-set.

– Using the rest data-set train the model.

– Test the model using the reserve portion of the data-set.

3.6.4 Information Gain:

Information gain is a preprocessing technique, which is used to calculate the reduction in entropy. It is commonly used in the construction of decision trees from a training dataset, by evaluating the information gain for each variable, and selecting the variable that maximizes the information gain, which in turn minimizes the entropy and best splits the dataset into groups for effective classification.

3.7 Classification

It is a process of categorizing data into given classes. Its primary goal is to identify the class of our new data.

3.7.1 Machine learning algorithms for classification

Research on data mining has led to the formulation of several data mining algorithms. These algorithms can be directly used on a dataset for creating some models or to draw vital conclusions and inferences from that dataset. Some popular data mining algorithms are Random Forest, Decision tree, Gaussian Naïve Bayes, Support vectormachine etc.

I. **Decision Tree:**

Decision Tree Analysis is a general, predictive modelling tool that has applications spanning a number of different areas. In general, decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions.It is one of the most widely used and practical methods for supervised learning. Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. The decision rules are generally in form of if-then-else statements. The deeper the tree, the more complex the rules and fitter the model.

II. **Gaussian Naive Bayes**:

It is a simple technique for constructing classifiers. It is a probabilistic classifier based on Bayes' theorem. All Naive Bayes classifiers assume that the value of any particular feature

is independent of the value of any other feature, given the class variable. Bayes theorem is given as follows: P(C|X) = P(X|C) * P(C)/P(X), where X is the data tuple and C is the class such that P(X) is constant for all classes. Though it assumes an unrealistic condition that attribute values are conditionally independent, it performs surprisingly well on large datasets where this condition is assumed and holds.

III. **Random Forest:**

A random forest model is a type of ensemble learning algorithm used in machine learning for classification and regression tasks. It consists of a collection of decision tree models that are trained on different subsets of the training data, and then combined to make predictions.

The random forest algorithm works by creating a set of decision trees, where each tree is trained on a randomly selected subset of the training data and a randomly selected subset of the features. This helps to reduce overfitting and increase the generalization of the model.

IV. **Logistic Regression:**

Logistic regression is a statistical method used to analyze and model the relationship between a binary (yes/no) dependent variable and one or more independent variables. It is a type of generalized linear model and is widely used for classification tasks in machine learning.

The logistic regression model uses a logistic function, also known as the sigmoid function, to model the relationship between the independent variables and the dependent variable. The output of the logistic function is a probability score between 0 and 1, which can be interpreted as the likelihood that the dependent variable is equal to 1. The logistic function transforms the linear combination of the independent variables into a probability score, which is then compared to a threshold value to make a binary classification decision.

3.8. Confusion matrix

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. A **true positive** (tp) is a result where the model predicts the positive class correctly. Similarly, a true negative (tn) is an outcome where the model correctly predicts the negative class. A **false positive** (fp) is an outcome where the model incorrectly predicts the positive class. And a false negative (fn) is an outcome where the model incorrectly predicts the negative class.

Sensitivity or recall or hit rate or true positive rate (TPR)

It is the proportion of individuals who actually have the disease were identified as having the disease.

TPR = tp / (tp + fn)

Specificity, selectivity or true negative rate (TNR)

It is the proportion of individuals who actually do not have the disease were identified as not having the disease.

TNR = tn / (tn + fp) =1-FPR

Precision or positive predictive value (PPV)

If the test result is positive what is the probability that the patient actually has the disease.

PPV = tp / (tp + fp)

Negative predictive value (NPV)

If the test result is negative what is the probability that the patient does not have disease.

NPV = tn / (tn + fn)

Miss rate or false negative rate (FNR)

It is the proportion of the individuals with a known positive condition for which the testresult is negative.

FNR = fn / (fp + tn)

Fall-out or false positive rate (FPR)

It is the proportion of all the people who do not have the disease who will be identified as having the disease.

FPR = fp/ (fp + tn)

False discovery rate (FDR)

It is the proportion of all the people identified as having the disease who do not have the disease.

FDR = fp / fp + tp

False omission rate (FOR)

It is the proportion of the individuals with a negative test result for which the truecondition is positive.

FOR = fn / (fn + tn)

Accuracy

The accuracy reflects the total proportion of individuals that are correctly classified.

ACC = (tp + tn) / (tp + tn + fp + fn)

F1 score

It is the harmonic mean of precision and sensitivity

F1 = 2tp / (2tp+ fp + fn)

4. METHODOLOGY

4.1 System Architecture

The methodology for building a book recommendation system using machine learning involves collecting and preprocessing data, extracting relevant features, selecting an appropriate algorithm, training and evaluating the model, and deploying the system for user consumption. Data collection, second is Data preprocessing, third is Feature-extraction, fourth is Algorithm-selection, fifth is Model-training, sixth is Model-evaluation, and the seventh one is Deployment.

System Architecture describes "the overall structure of the system and the ways in which the structure provides conceptual integrity". The system architecture to build a recommendation system involves the following five major steps.

4.1.1 Data Acquisition

4.1.2 Data Pre-processing

4.1.3 Feature Extraction

4.1.4 Training Methods

4.1.5 Testing Data

In Step 4.2.1, Dataset was collected from Good Reads Website in which three datasets are present i.e., Books Dataset, Ratings Dataset, Users Dataset. In Step 4.2.2, Datasets were pre-processed to make suitable for developing the Recommendation system. In Step 4.2.3, Feature extraction is performed in which Truncated-SVD is used to reduce the features of the dataset and Data splitting is done in which training dataset and testing dataset are divided into 80:20 ratio. In Step 4.2.4, Content Based Filtering System is developed in which book description is taken as an input and Collaborative Filtering System is developed by building a model using K-Means Algorithm over Gaussian Mixture after comparing with Silhouette scores. In step 4.2.5, Testing of model with test data is performed.

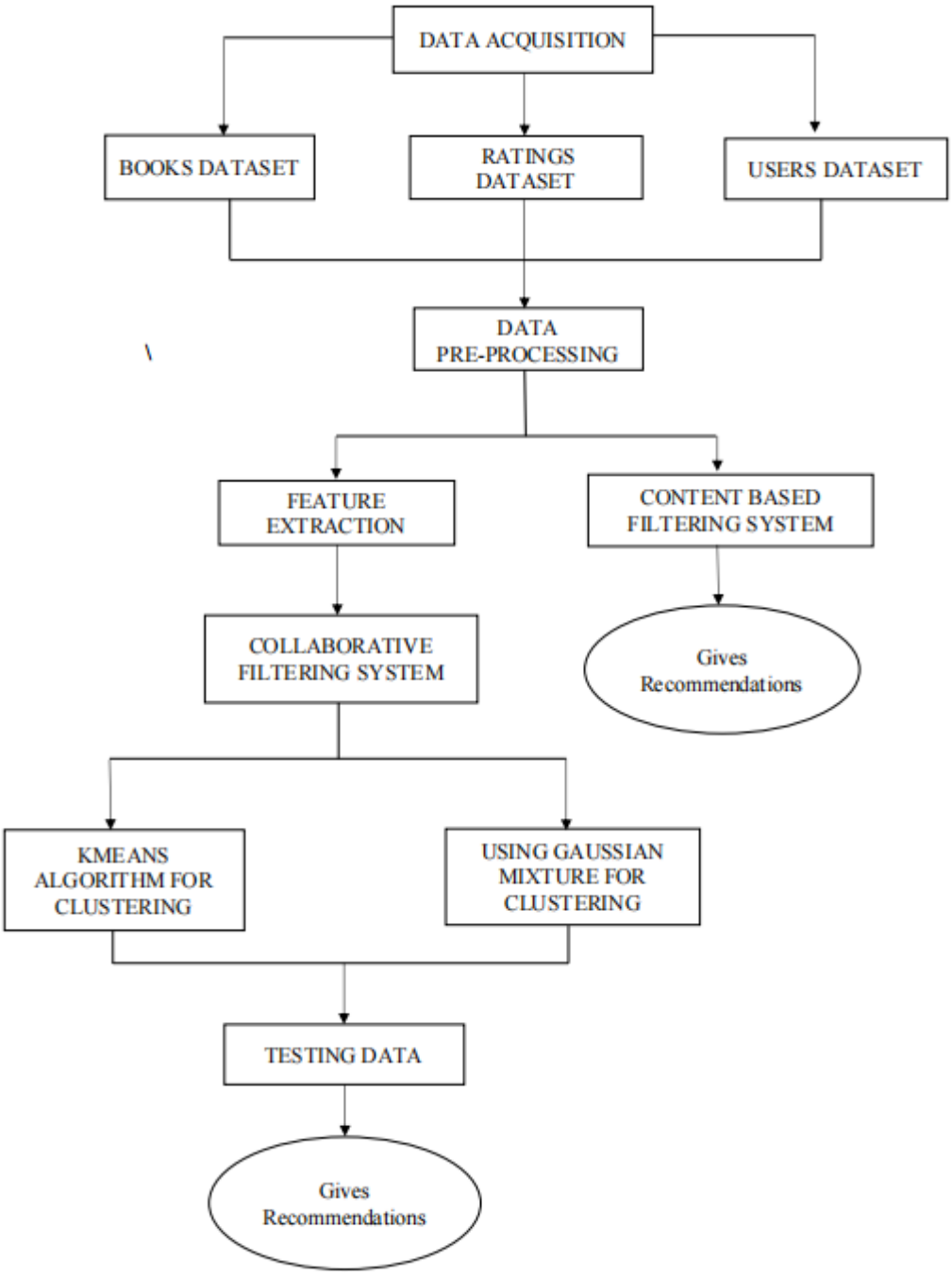

Fig:4.1 System Architecture

4.1.1 Data Acquisition

The goal of this step is to find and acquire all the related datasets or data sources. In this step, the main aim is to identify various available data sources, as data are often collected from various online sources like databases and files. The size and the quality of the data in the collected dataset will determine the efficiency of the model. The Books dataset is collected from the Kaggle website.

4.1.2 Data Pre-Processing

The goal of this step is to study and understand the nature of data that was acquired in the previous step and also to know the quality of data. In this step, we will check for any null values and remove them as they may affect the efficiency. Identifying duplicates in the dataset and removing them is also done in this step.

4.1.3 Feature Extraction

After pre-processing the acquired data, the next step is to reduce the features i.e., Dimensionality reduction. The reduced features should be able to give high efficiency. We used Matrix Factorization technique such as Truncated SVD which takes sparse matrix as input for reduction of features.

4.1.4 Training Methods

Now, we have our training and testing data. The next step is to identify the possible training methods and train our models. We have used two different clustering methods for training models. After that based on the silhouette score of each model, we would decide on which model to use finally.

4.1.4.1 K-Means Clustering: K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters in such a manner that each dataset belongs to only one group that has similar features. Here K defines the number of pre-defined clusters. We have to associate each cluster with a centroid in this algorithm. The sum of distances between the data point and their corresponding clusters should be minimized. The unlabeled dataset is taken as an input and the dataset into k-number of clusters is divided, and

the process is repeated until it does not find the best clusters.

The basic steps involved in K-Means Clustering algorithm is as follows:

Step-1: Select the number K which gives the number of clusters.

Step-2: Select random K number of points or centroids.

Step-3: Each data point to their nearest centroid should be assigned, which forms the predefined K clusters.

Step-4: Calculate the variance and place a new centroid for each cluster.

Step-5: We have to repeat the step-3, each data-point to the new closest centroid of each cluster should be reassigned.

Step-6: If reassignment happens, then go to step-4 or else go to step-7.

Step-7: Stop.

4.1.4.2 Gaussian Mixture: Gaussian Mixture models are powerful clustering algorithms. It assumes that there are a certain number of Gaussian distributions where each distribution represents a cluster. This model groups the data points together into a single distribution. These models used the soft clustering technique for assigning data points to Gaussian distributions.

4.1.5 Testing Data

Once Clustering model has been trained on pre-processed dataset, then the model is tested using different data points. In this testing step, the model is checked for the silhouette score for checking goodness of clustering. All the training methods need to be verified for finding out the best model to be used.

4.2 Collaborative-based Filtering

System recommendations that use collaborative-filtering on prior encounters between users and the goal items. To put simply, attempt to notice customers who give them and who resemble them products recommend on the options made by their lookalikes. Let's use an illustration to clarify.

Users X and Y share some characteristics, and X has seen movies A, B, and C. If user Y has seen user B, user C, and user D, we will suggest user A to user Y and user D to user X.

4.3 Content-based Filtering

The program recommends an item that is comparable that have been used or viewed. To put it simply, this program looks for items that resemble one another. As an illustration, as a result of the two films' similar label and content, if someone likes watching movies, he might also like web series. Only the appearance of the content is identical, and there isn't much attention paid to the viewer. On the basis of prior tastes, it only suggests the product with the highest score.

4.4 Popularity-based Filtering

Popularity-based recommendation systems are simple systems that recommend items based on their overall popularity. However, machine learning can be used to improve the accuracy of such systems by taking into account additional factors that may affect popularity. Here are the steps to build a popularity-based recommendation system using machine learning:

I. Data Collection: The first step is to collect data on user preferences and item popularity. This data can come from various sources, such as user ratings, reviews, purchases, and clicks.

II. Data Preprocessing: The next step is to preprocess the data by cleaning and normalizing it. This involves removing duplicates, handling missing values, and converting the data into a suitable format for machine learning algorithms.

III. Feature Extraction: The next step is to extract features from the data that can be used to train a machine learning model. For example, features such as user demographics, item categories, and item ratings can be extracted from the data

IV. Model Training: The next step is to train a machine learning model using the extracted features. One popular approach is to use collaborative filtering, which predicts a user's rating for an item based on the ratings of similar users and items.

V. Evaluation: The next step is to evaluate the performance of the machine learning model using metrics such as precision, recall, and F1 score. This involves comparing the predicted ratings

to the actual ratings in the test set.

VI. Deployment: The final step is to deploy the popularity-based recommendation system in a production environment. This involves integrating the system with the user interface and backend systems, and continuously monitoring and improving its performance.

Overall, a popularity-based recommendation system using machine learning can improve the accuracy of recommendations and provide a better user experience.

5. IMPLEMENTATION

Index.html:

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <title>Book Recommender System</title>
    <!-- Latest compiled and minified CSS -->
<link rel="stylesheet"
href="https://cdn.jsdelivr.net/npm/bootstrap@3.3.7/dist/css/bootstrap.min.css"
integrity="sha384-
BVYiiSIFeK1dGmJRAkycuHAHRg32OmUcww7on3RYdg4Va+PmSTsz/K68vbdEjh4u"
crossorigin="anonymous">
</head>
<style>
    .text-white{
        color:white
    }
</style>
<body style="background-color:black">

    <nav class="navbar" style="background-color:#00a65a">
        <a class="navbar-brand">My Book recommender</a>
        <ul class="nav navbar-nav">
```

```
        <li><a href="/">Home</a></li>
        <li><a href="/recommend">Recommend</a></li>
        <li><a>Contact</a></li>
    </ul>
</nav>

<div class="container">
    <div class="row">
        <div class="col-md-12">
            <h1 class="text-white" style="font-size:50px">Top 50 Books</h1>
        </div>

        {% for i in range(book_name|length) %}
            <div class="col-md-3" style="margin-top:50px">
                <div class="card">
                    <div class="card-body">
                        <img class="card-img-top" src="{{ image[i] }}">
                        <p class="text-white">{{ book_name[i] }}</p>
                        <h4 class="text-white">{{ author[i] }}</h4>
                        <h4 class="text-white">Votes - {{ votes[i] }}</h4>
                        <h4 class="text-white">Rating - {{ rating[i] }}</h4>
                    </div>
                </div>
            </div>
```

```
            {% endfor %}


        </div>
    </div>

</body>
</html>
```

Main.py:

```
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['Book-Title'].values),
                author=list(popular_df['Book-Author'].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('/recommend')
def recommend_ui():
    return render_template('recommend.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[index])), key=lambda x: x[1],
reverse=True)[1:5]

    data = []
    for i in similar_items:
        item = []
        temp_df = books[books['Book-Title'] == pt.index[i[0]]]
        item.extend(list(temp_df.drop_duplicates('Book-Title')['Book-Title'].values))
        item.extend(list(temp_df.drop_duplicates('Book-Title')['Book-Author'].values))
```

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

        data.append(item)

    print(data)

    return render_template('recommend.html',data=data)
if __name__ == '__main__':
    app.run(debug=True)
```

Home.html

```
<!DOCTYPE html>
<style> h1
{
color:blue;
text-align:center; font-weight:bold;
}
body

{
background-image:url('https://wallpapercave.com/wp/wp8224471.jpg'); background-size:1400px 650px;
background-repeat:no-repeat;
}
div
```

```
{
margin:10px;
}
label

{
background-color:yellow; font-size:20px;
font-weight:bold; width:100%;
}


.a{ width:100%; float:left;
}
</style>
<html lang="en">
<head>
<meta charset="UTF-8">
<H1>BOOK RECOMMENDATION SYSTEM</H1>
</head>
<hr color="orange"><br>
<body>
<form action="/predict" method="post">
<div><label>Enter the book</label>
```

```
<input type="number" min="0" max="150" class="a" name="P" required> <br>
</div><br>
<div><label>Enter the book</label>
<input type="number" min="0" max="150" class="a" name="N" required><br> </div><br>
<div><label>Enter the book</label>
<input type="number" min="0" max="210" class="a" name="K" required><br> </div>

<br>
<div><label>Enter the book</label>
<input type="number" min="0" max="60" class="a" name="temperature" required><br> </div>

<br>
<div><label>Enter the book</label>
<input type="number" min="0" max="100" class="a" name="humidity" required><br>        </div>

<br>
<div><label>Enter the book</label>

<input type="number" min="0" max="14" class="a" name="ph" required><br> </div><br>
<div><label>Enter the book you want</label>
<input type="number" min="0" max="300" class="a" name="rainfall" required><br> </div><br>
<div align="center"><input style="color:blue" type="submit" value="Submit">
</div><br>
</form>
```

```
</body><br>
</html>
```

Recommend.html:

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <title>Book Recommender System</title>
    <!-- Latest compiled and minified CSS -->
<link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap@3.3.7/dist/css/bootstrap.min.css"
integrity="sha384-
BVYiiSIFeK1dGmJRAkycuHAHRg32OmUcww7on3RYdg4Va+PmSTsz/K68vbdEjh4u"
crossorigin="anonymous">
</head>
<style>
    .text-white{
        color:white
    }
</style>
<body style="background-color:black">

    <nav class="navbar" style="background-color:#00a65a">
        <a class="navbar-brand">My Book recommender</a>
        <ul class="nav navbar-nav">
```

```
        <li><a href="/">Home</a></li>
        <li><a href="/recommend">Recommend</a></li>
        <li><a>Contact</a></li>
    </ul>
</nav>

<div class="container">
    <div class="row">
        <div class="col-md-12">
            <h1 class="text-white" style="font-size:50px">Recommend Books</h1>
            <form action="/recommend_books" method="post">
                <input name="user_input" type="text" class="form-control"><br>
                <input type="submit" class="btn btn-lg btn-warning">
            </form>
        </div>

        {% if data %}

         {% for i in data %}
            <div class="col-md-3" style="margin-top:50px">
                <div class="card">
                    <div class="card-body">
                        <img class="card-img-top" src="{{i[2]}}">
                        <p class="text-white">{{i[0]}}</p>
```

```
                    <h4 class="text-white">{{i[1]}}</h4>
                </div>
            </div>
        </div>
        {% endfor %}

        {% endif %}
    </div>
</div>

</body>
</html>
```

db.py:

```
import numpy as np
import pandas as pd
import sklearn

users = pd.read_csv('users.csv')
ratings = pd.read_csv('ratings.csv')
books = pd.read_csv('books.csv')
books['Image-URL-M'][1]
books.isnull().sum()
```

```
books.dropna()

books.isna().sum()

users.isnull().sum()

ratings.isnull().sum()

books.duplicated().sum()

ratings.duplicated().sum()

users.duplicated().sum()

ratings_with_name = ratings.merge(books,on='ISBN')

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)

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)

popular_df = num_rating_df.merge(avg_rating_df,on='Book-Title')

popular_df =
popular_df[popular_df['num_ratings']>=250].sort_values('avg_rating',ascending=False).head(50)

popular_df = popular_df.merge(books,on='Book-Title').drop_duplicates('Book-Title')[['Book-Title','Book-
Author','Image-URL-M','num_ratings','avg_rating']]

popular_df['Image-URL-M'][0]

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['User-ID'].isin(padhe_likhe_users)]

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='Book-Rating')
```

```
pt.fillna(0,inplace=True)


from sklearn.metrics.pairwise import cosine_similarity

similarity_scores = cosine_similarity(pt)

import pickle

pickle.dump(popular_df,open('popular.pkl','wb'))
```

6. RESULT ANALYSIS

After created all the four models by using the training data ,we test the models by using the test data and calculates the accuracies of the models. The following table shows the accuracies of different models:

	K-Means	Gaussian mixture
Silhouette Score	0.0433968332584411	0.016778872313688933

Fig:6.1 Two models' silhouette ratings

The above table shows the accuracies of different models which are created by using the mentioned machine learning algorithms. Among all above models ,the model which is created by using gaussian naive bias algorithm got good accurance. so, we consider it as the final model.

It is determined the differences between the both the typical mean ratings of books for test users and the typical mean ratings of books in the cluster.

Mean rating for 10 random books	3.8876949740034736
Mean rating for 10 books of cluster's favourites	4.3735008665511135

Fig:6.2 Mean-scores comparison

By using RMSE, one can calculate the difference between data values and values predicted by a model.

RMSE	0.5957791790493179
Accuracy	1.167727190936663

Fig:6.3 Values for RMSE and Accuracy

7. OUTPUT SCREENS

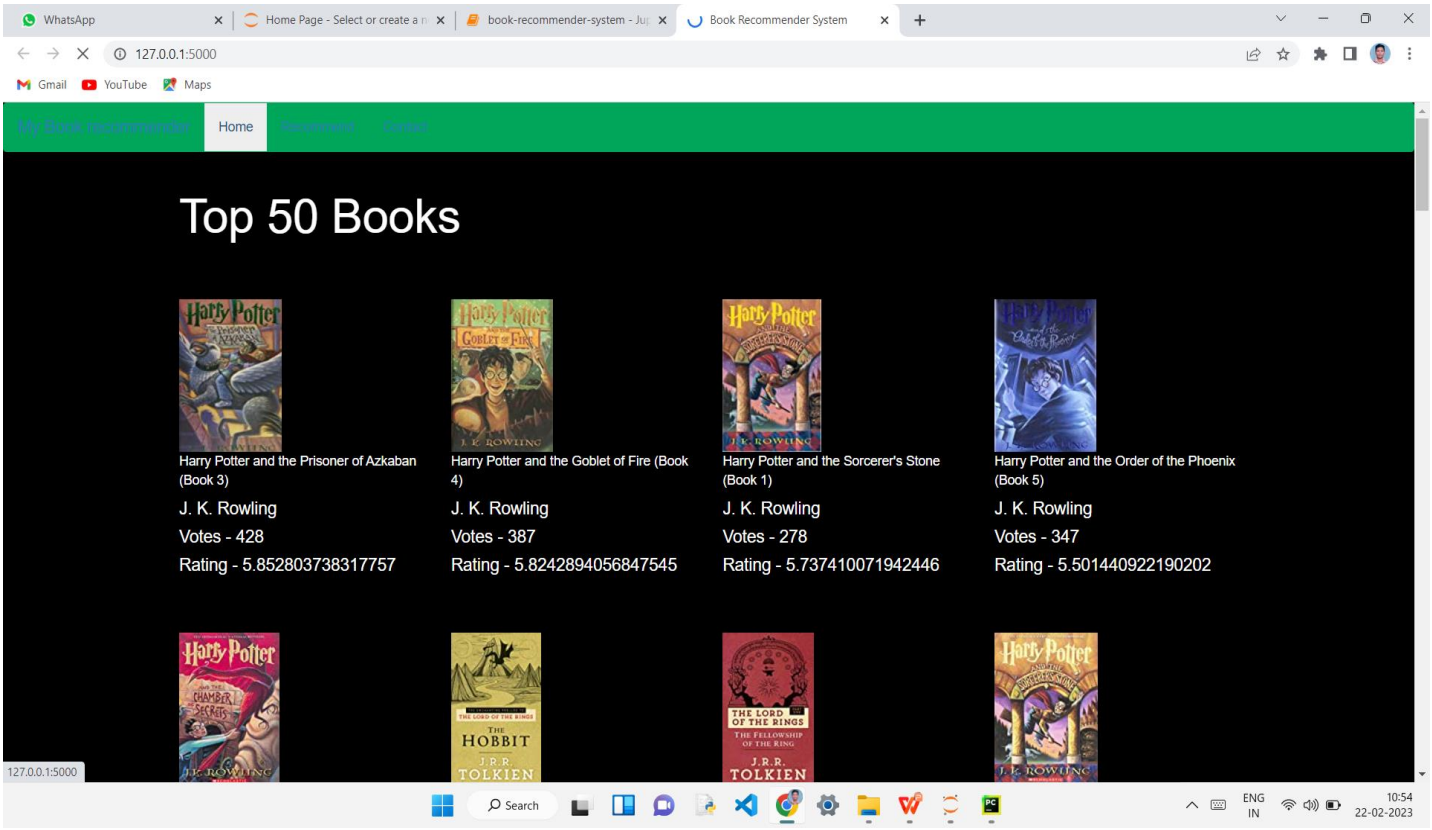

Fig:7.1. Home Page

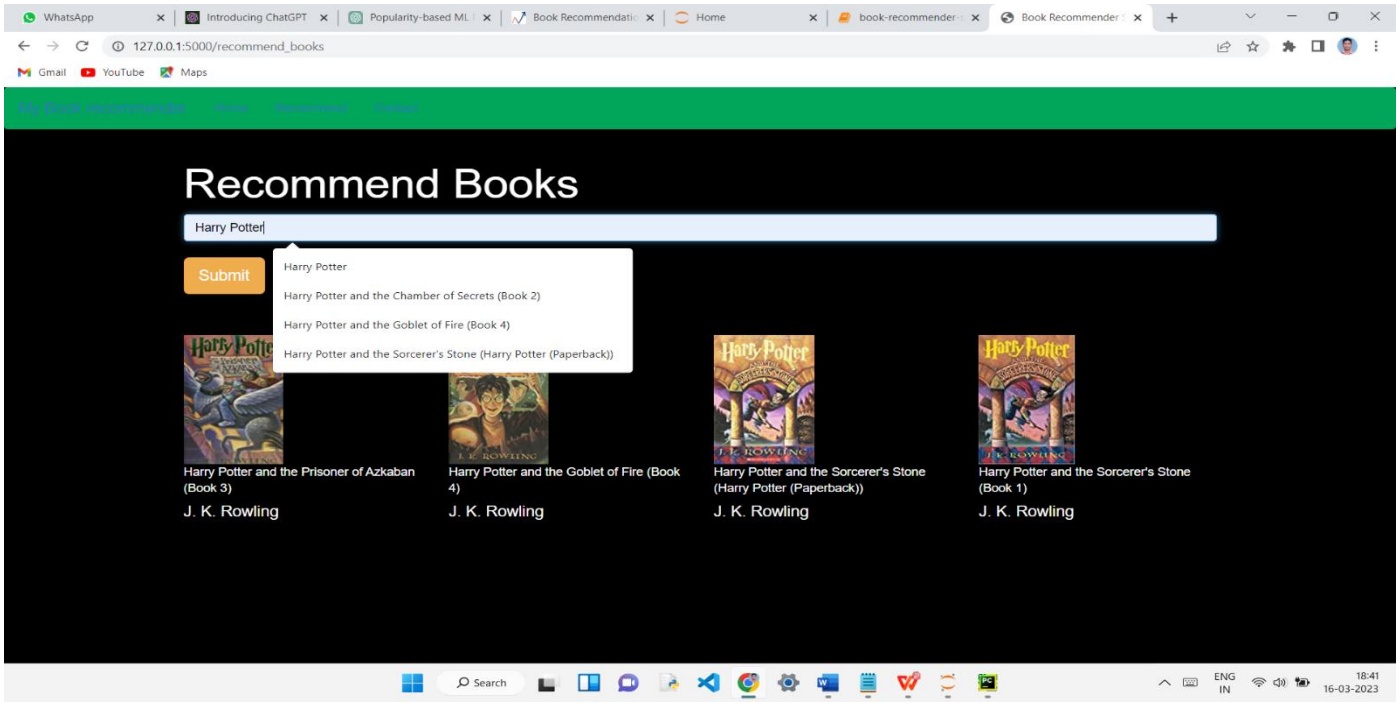

Fig:7.2. Recommend Page

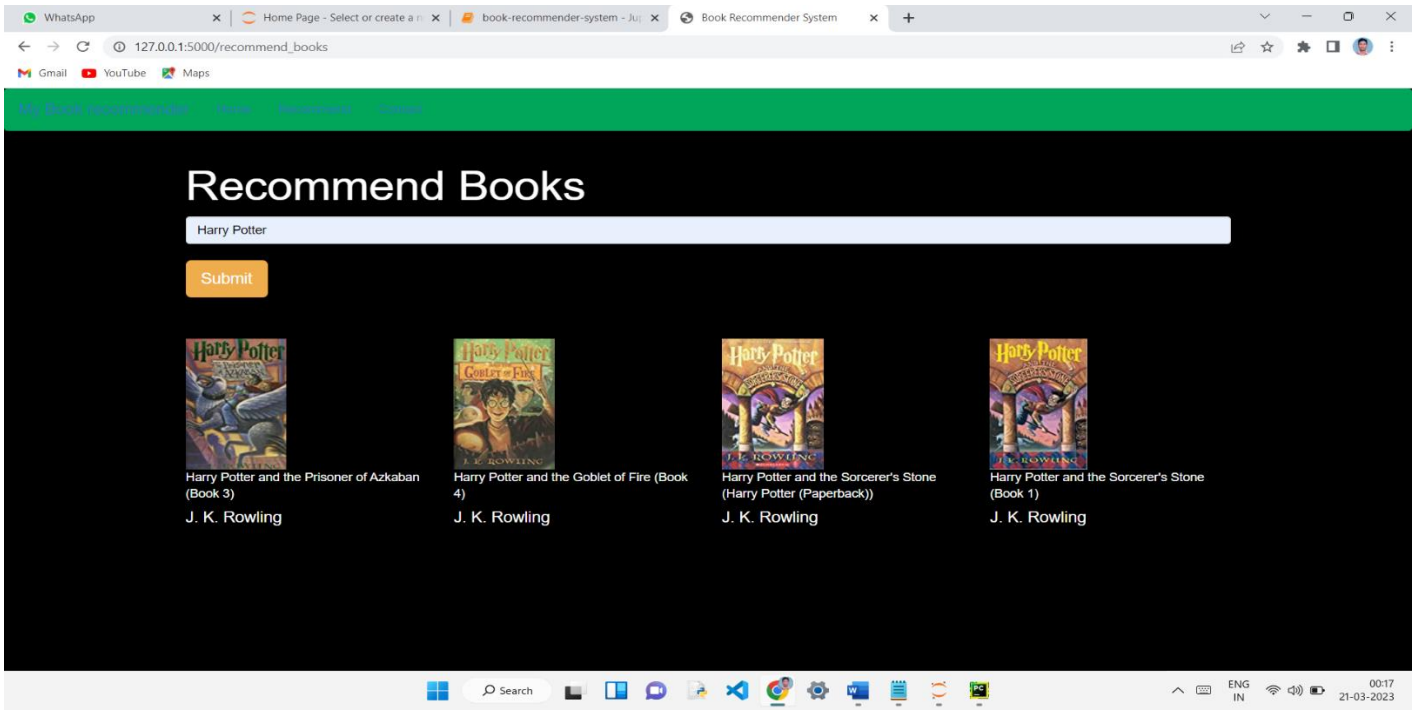

Fig:7.3 Result Page

8. TEST CASES

Test case1:

We have examined the book with the title Animal Farm in this testcase1. therefore, it must deliver accurate findings and produce exact book.

Result:

Although it was unable to produce an accurate result, it did suggest books that were related to that book based on the content and user rating.

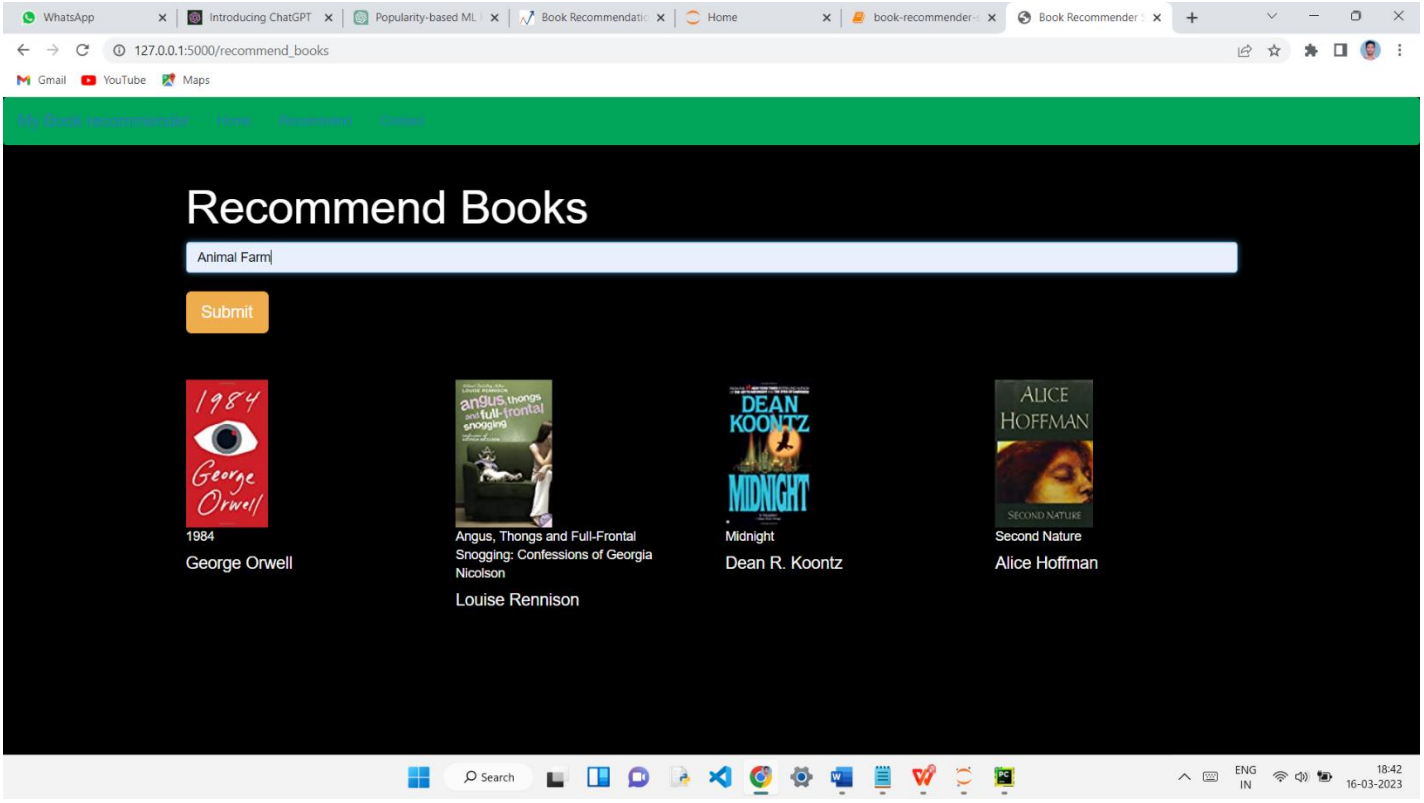

Fig:8.1 Testcase1

Test case2:

This testcase 2 examines the book 1984 and books from the same category or genre and related books are used as recommendations.

Result:

It producing the accurate results and same kind of category books are recommending from the above testcase1.

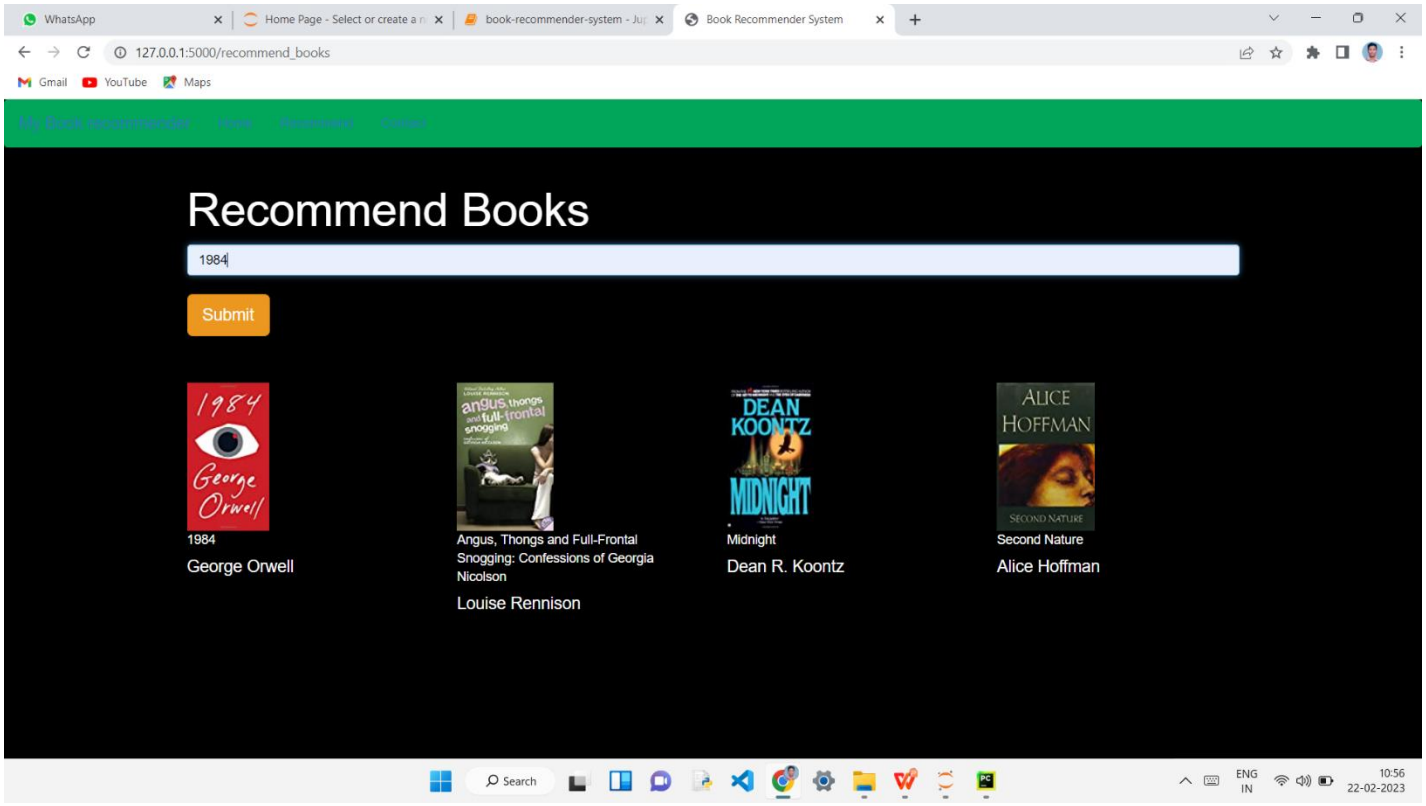

Fig:8.2 Testcase2

9. CONCLUSION

In this project we present a recommendation system based on a collaborative filtering approach. The main goal was to speed up recommendations which is to create such a system, which can provide quality recommendations to their users without the need for long-term registration and have a great profile experience, browsing history etc. Test results indicate that the proposed approach provides appropriate recommendations. The proposed activity can be used in other domains to promote such things as movies, music and other products.

In conclusion, a book recommendation system using machine learning can provide users with personalized recommendations based on their preferences and reading habits. The system can be built by collecting data on user behavior, preprocessing the data, extracting features, training a machine learning model, evaluating its performance, and deploying it in a production environment. Different machine learning algorithms can be used, such as collaborative filtering, content-based filtering, and hybrid filtering, depending on the available data and user requirements. The benefits of such a system include increased user engagement, improved customer satisfaction, and increased revenue for book retailers. However, it is important to address ethical concerns such as user privacy, algorithmic bias, and transparency in the system's decision-making process.

10. FUTURE SCOPE

Expanding to Other Media Types: While the current book recommendation system is focused on books, there is scope to expand it to other media types such as movies, music, and podcasts. This can provide a more comprehensive recommendation experience for users.

Personalization for Groups: The current book recommendation system focuses on individual preferences, but there is scope to personalize recommendations for groups of users, such as families or book clubs. This can provide a more social and interactive recommendation experience for users.

11. BIBLIOGRAPHY

[1] Avi Rana and K. Deeba et.al, “Online Book Recommendation System using Collaborative Filtering (With Jaccard Similarity)” in IOP ebooks 1362, 2019.

[2] G. Naveen Kishore, V. Dhiraj, Sk Hasane Ahammad, Sivaramireddy Gudise, Balaji Kummaraa and Likhita Ravuru Akkala, “Online Book Recommendation System” International Journal of Scientific & Technology Research vol.8, issue 12, Dec 2019.

[3] Ms. Sushma Rjpurkar, Ms. Darshana Bhatt and Ms. Pooja Malhotra, “Book Recommendation System” International Journal for Innovative Research in Science & Technology vol.1, issue 11, April 2015.

[4] Ms. Praveena Mathew, Ms. Bincy Kuriakose, Mr.Vinayak Hegde, “Book Recommendation System through Content Based and Collaborative Filtering Method”, IEEE International Conference on Data Mining and Advanced Computing (SAPIENCE),2016.

[5] Raghuwanshi, Sandeep K., and R. K. Patreiya, “Collaborative Filtering Techniques in Recommendation Systems” , International Journal of Scientific & Technology Research(2019).

[6] A. Gazdar and L. Hidri, "A new similarity measure for collaborative filtering-based recommender systems," Knowledge-Based Systems, vol.188, p. 105058, 2020.

[7] S. S. Sohail, J. Siddiqui and R. Ali, "Book recommendation system using opinion mining technique," 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI),Mysore, 2013, pp. 1609-1614, doi: 10.1109/ICACCI.2013.6637421.

[8] P. Mathew, B. Kuriakose and V. Hegde, "Book Recommendation System through content based and collaborative filtering method," 2016International Conference on Data Mining and Advanced Computing(SAPIENCE), Ernakulam, 2016, pp. 47-52, doi:10.1109/SAPIENCE.2016.7684166.

[9] J. Qi, S. Liu, Y. Song and X. Liu, "Research on Personalized Book Recommendation Model for New Readers," 2018 3rd International Conference on Information Systems Engineering (ICISE), Shanghai, China, 2018, pp. 78-81, doi: 10.1109/ICISE.2018.00022.

[10] E. Cho and M. Han, "AI Powered Book Recommendation System," in Proceedings of the 2019 ACM Southeast Conference, 2019, pp. 230-232.

[11] S. Parvatikar and B. Joshi, "Online book recommendation system by using collaborative filtering and association mining," 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, 2015, pp. 1-4, doi:10.1109/ICCIC.2015.7435717.

[12] Y. Gao, C. Huang, M. Hu, J. Feng, and X. Yang, "Research on Book Personalized Recommendation Method Based on Collaborative Filtering Algorithm," in IOP Conference Series: Earth and Environmental Science, 2019, vol. 252, no. 5: IOP Publishing, p. 052099

Book recommendation system using Machine Learning

Goli Siva Sankar Tarun
Student
Department of Computer Science and Engineering
Narasaraopet, India
golisivasankartarun225@gmail.com

Galam Sudheer
Student
Department of Computer Science and Engineering
Narasaraopet, India
sudheergalam@gmail.com

Kunala Kalyan
Student
Department of Computer Science and Engineering
Narasaraopet, India
kalyankunala@gmail.com

Bolla Jhansi Vazram
Professor
Department of Computer Science and Engineering
Narasaraopet, India
jhansi.bolla@gmail.com

Abstract – Due to COVID-19 pandemic the usage of online books is rapidly increasing, from a huge e-book space finding particular books becomes a immense challenge for users. To perform effective search which mine connected books based on user interest and rating using Book recommendation system. popularity-based and collaborative based methods are used in these systems and based on user ratings and interest. This System for recommending books for users that rating a book using the clustering and above methods and then finding a same of that book to recommend a new book.

Keywords — Books recommendation, Truncated-SVD, Clustering , Popularity-based, Collaborative-based, Root Mean-Square Error.

I. INTRODUCTION

There is an ever-increasing amount of information available to us in today's recent times, that includes books, music, movies, and more. It can be testing for users to find what they are consider for with this more amount of information and to develop new items that match their interests and satisfaction. Through the vast amount of information this is where recommendation systems come in, helping users to navigate and find what they are looking for more easily.

This highly scalable and trustworthy system for suggesting books can be adjusted for various genres and user preferences. This books recommendation system can be integrated into various types of online stores and digital libraries to offer customers which has individualized books recommendations. A dataset of books and users was used to calculate this system's performance, and the findings demonstrate that the proposed system outperforms current book recommendation systems. With the assistance of this project, users will be able to find new books to enjoy and contribute to the creation of a more sophisticated and individualized.

Book recommendation systems use machine learning algorithms to analyse data from various sources such as users, ratings and books datasets to provide customizedrecommendations to users. These recommendations can help users to optimize their Books recommends, reduce the amount of time, and mitigate risks associated with users' information and data.

This System presents a Book recommendation system that utilizes machine learning algorithms to provide accurate and personalized recommendations to users. This system analyzes data from various sources, and performed various methods such as Popularity-based and collaborative learning methods are used. The system's primary goal is to provides the quality of the recommendations which is very accurate, easy to maintain and simply to use which is using by the user.

The description of the items is used in content-based filtering, which provides suggestions for items that are comparable to the description of the items. Book are recommended using these multiple filtering models depending on multiple the book's content and the user actions. As a result, mine recommendation engine also recommends books to new readers. For clustering the users in this study, we using both techniques as: K-means and Gaussian mixture. To calculate error between absolute numbers and the results, use the Root Mean Square Error formula.

RMSE number used to determine basic accuracy.

II. LITERATURE SURVEY

The majority of analyzers used the Pearson's Coefficient function to determine how comparable book-rating were in order to make book recommendations:

1. "Collaborative Filtering for the Book Recommendation System" by D. R. K. Srinivas and S. S. Suresh Kumar, published in the International Journal of Computer Science and Mobile Computing in 2017. This paper explores the use of collaborative filtering for book recommendations, using a dataset which of th

1. "Content-based Recommendation Systems" by Pazzani, M. J., & Billsus, D. (2007)- This survey focuses on book recommendation systems using machine learning. It discusses the importance of author, and keywords and the challenges associated with it, such aslimited data availability and high variability in the data. The survey also covers the different types of machine learning models used for recommendation, including decision trees, support vector machines.

2. "A survey on Book recommendation systems using machine learning" by Bell & Volinsky (2009) - This survey provides a comprehensive overview of Book-recommendation- system which is use Machine learning, including their architecture, and data sources, also performance evaluation metrics. It covers different types of methods for machine learning used in recommendations, such as rule-based systems, collaborative-filtering, and content-based-filtering. The survey alsodiscusses the challenges and limitations of book recommendation systems, such as data sparsity.

Overall, these surveys demonstrate the growing interest and potential of Book recommendation systems using machine learning in the learning sector, and highlight the need for further research to overcome the challenges associated with these systems.

III. EXISTING SYSTEM

Machine learning algorithms are used in a book recommendation system to analyze user data and offer personalized book recommendations based on reading tastes and history. The primary problem with this is that these systems' recommendations are not transparent. Users may not always understand why a specific book was suggested, which can make it challenging for them to believe in the system and its suggestions. Additionally, the suggestions might not always suit the user's preferences or tastes, which could cause frustration and user engagement. Finally, systems for recommending books that heavily depend on machine learning might need a lot of data to properly train the algorithms. For smaller businesses or groups with less access to data, this can be difficult. Furthermore, the process of gathering and storing user data may give rise to privacy issues, which may discourage some users from using the system.

IV. METHODOLOGY

The methodology for building a book recommendation system using machine learning involves collecting and preprocessing data, extracting relevant features, selecting an appropriate algorithm, training and evaluating the model, and deploying the system for user consumption. Data collection, second is Data preprocessing, third is Feature-extraction, fourth is Algorithm-selection, fifth is Model-training, sixth is Model-evaluation, and the seventh one is Deployment.

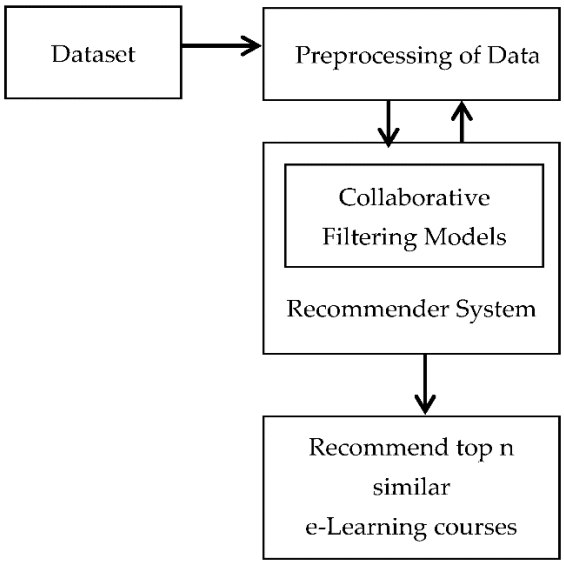

Fig. 1: proposed system task flow

A. DATASET COLLECTION

The Dataset is collect from the Kaggle website: https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset.

we have 3 files in our dataset which is extracted from above kaggle website.

Books - first, there are about books, which includes all the details about books, including an author, title, year of release, etc.

Users: Information about registered users, such as user id and address, is contained in the second file.

Ratings – Ratings contain information like which user has given how much rating to which book.

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/0060973129.0...
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0...
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0...

Fig. 2: Dataset diagram

	User-ID	ISBN	Book-Rating
0	276725	034545104X	0
1	276726	0155061224	5
2	276727	0446520802	0
3	276729	052165615X	3
4	276729	0521795028	6

Fig. 3: Dataset diagram

	User-ID	Location	Age
0	1	nyc, new york, usa	NaN
1	2	stockton, california, usa	18.0
2	3	moscow, yukon territory, russia	NaN
3	4	porto, v.n.gaia, portugal	17.0
4	5	farnborough, hants, united kingdom	NaN

Fig.4: Dataset diagram

A. Data Cleaning and Feature Extraction

The data must be cleaned before being given to any machine learning programs. In the data cleaning procedure, outliers and null numbers are eliminated. There are no null values in the dataset that was gathered. Features are extracted from the data collection after data cleaning. We must examine the correlation between the six distinct features in the dataset that have an impact on the model's output. Correlation is useful for removing characteristics that have little impact on the model's output.

B. Model Architecture.

In this step dataset is splitted in 80% and 20% for training and testing. During training, machine learning algorithms finds the relation between the input and output features. By using this relation, the model able to predict the outputs to the new input values. In this we used the following four machine learning models:

1. content-based-Filtering: The program recommends an item that is comparable that have been used or viewed. To put it simply, this program looks for items that resemble one another. As an illustration, As a result of the two films' similar label and content, if someone likes watching movies, he might also like web series. Only the appearance of the content is identical, and there isn't much attention paid to the viewer. On the basis of prior tastes, it only suggests the product with the highest score.

2. collaborative-based-Filtering: System recommendations that use collaborative-filtering on prior encounters between users and the goal items. To put simply, attempt to notice customers who give them and who resemble them products recomend on the options made by their lookalikes. Let's use an illustration to clarify. Users X and Y share some characteristics, and X has seen movies A, B, and C. If user Y has seen user B, user C, and user D, we will suggest user A to user Y and user D to user X.

3. K-Means: The unlabeled dataset is divided into different clusters by the unsupervised learning algorithm K-Means Clustering. so each collection is only goes to single group shares common feature. In this case, K designates the quantity of predefined groups. In this method, each cluster must be connected to a centroid. A data point's total distance from each of its associated clusters should be as small as possible.

Using the unlabelled dataset as input, k groups are created from the unlabelled dataset. This procedure repeated until no better clusters are found. The k number for this algorithm must be decided in advance.

4. Gaussian Mixture: The clustering methods used by Gaussian Mixture models are effective. The model presupposes that there are a predetermined number of Gaussian distributions, each of which represents a cluster. The data elements are combined in this model to form a single distribution. To assign data points to Gaussian distributions, these models employed the soft clustering method. In a one-dimensional space, the probability density function of the Gaussian distribution.

V. RESULT AND ANALYSIS

After created all the four models by using the training data , we test the models by using the test data and calculates the accuracies of the models. The following table shows the accuracies of different models:

	K-Means	Gaussian mixture
Silhouette Score	0.0433968332584411	0.016778872313688933

Table1: Two models' silhouette ratings

It is determined the differences between the both the typical mean ratings of books for test users and the typical mean ratings of books in the cluster.

Mean rating for 10 random books	3.8876949740034736
Mean rating for 10 books of cluster's favourites	4.3735008665511135

Table2: Mean-scores comparison

By using RMSE, one can calculate the difference between data values and values predicted by a model.

RMSE	0.5957791790493179
Accuracy	1.167727190936663

Table3: Values for RMSE and Accuracy

VI. DEPLOYMENT

The final model is deployed into an application by using flask module of python by creating the user interface which can be accessible easily by the users. It is possible to design a system have to that provides users with high-quality of book recommendations and fosters long-term maintenance and adherence. The application simply takes the input and output

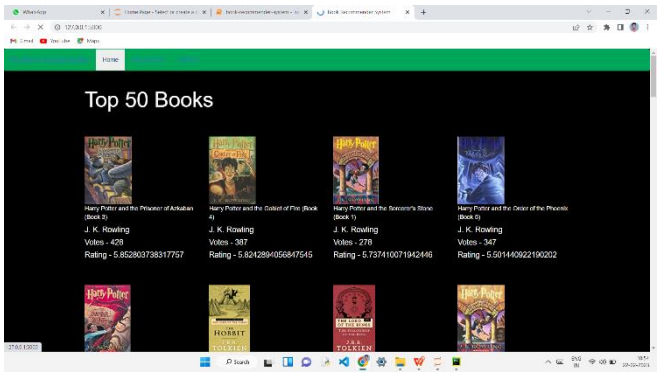

Fig.5: User interface

The following is the recommendations based on user ratings and after typing, select "Search" to have the suggested book, based on user ratings, displayed on the screen. The screenshot of the novels with user ratings is shown in Figure 6.

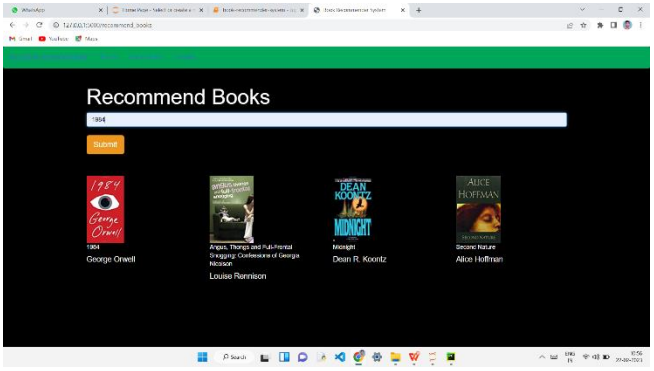

Fig.6: Recommended books

VII. CONCLUSION.

In this paper, we show a collaborative filtering-based recommendation system. The primary objective was to speed up recommendations by developing a system that can give users good recommendations without requiring them to register for an extended period of time.

This project has highlighted the benefits of Machine learning in recommendation system, and how it can help improves the efficiency and the achieves of the book recommendation. By automating the recommendation process, users can save time and effort, while also reducing the risk of making costly mistakes.

VIII. REFERENCES

1. Avi Rana and K. Deeba et.al, “Online Book Recommendation System using Collaborative Filtering (With Jaccard Similarity)” in IOP ebooks 1362, 2019.

2. G. Naveen Kishore, V. Dhiraj, Sk Hasane Ahammad, Sivaramireddy Gudise, Balaji Kummaraa and Likhita Ravuru Akkala, “Online Book Recommendation System” International Journal of Scientific & Technology Research vol.8, issue 12, Dec 2019.

3. Ms. Sushma Rjpurkar, Ms. Darshana Bhatt and Ms. Pooja Malhotra, “Book Recommendation System” International Journal for Innovative Research in Science & Technology vol.1, issue 11, April 2015.

4. Ms. Praveena Mathew, Ms. Bincy Kuriakose, Mr.Vinayak Hegde, “Book Recommendation System through Content Based and Collaborative Filtering Method”, IEEE International Conference on Data Mining and Advanced Computing (SAPIENCE),2016.

5. Raghuwanshi, Sandeep K., and R. K. Patreiya, “Collaborative Filtering Techniques in Recommendation Systems” , International Journal of Scientific & Technology Research(2019).

AB5-1

by Vamshikrishna Namani

Submission date: 10-Mar-2023 02:33PM (UTC+1000)
Submission ID: 2033624034
File name: AB5-1.docx (498.97K)
Word count: 2086
Character count: 11857

Book recommendation system using Machine Learning

Goli Siva Sankar Tarun
Student
Department of Computer Science and Engineering
Narasaraopet, India
golisivasankartarun225@gmail.com

Galam Sudheer
Student
Department of Computer Science and Engineering
Narasaraopet, India
sudheergalam@gmail.com

Kunala Kalyan
Student
Department of Computer Science and Engineering
Narasaraopet, India
kalyankunala@gmail.com

Bolla Jhansi Vazram
Professor
Department of Computer Science and Engineering
Narasaraopet, India
jhansi.bolla@gmail.com

Abstract – Due to COVID-19 pandemic the usage of online books is rapidly increasing, from a huge e-book space finding particular books becomes a immense challenge for users. To perform effective search which mine connected books based on user interest and rating using Book recommendation system. popularity-based and collaborative based methods are used in these systems and based on user ratings and interest. This System for recommending books for users that rating a book using the clustering and above methods and then finding a same of that book to recommend a new book.

Keywords — Books recommendation, Truncated-SVD, Clustering , Popularity-based, Collaborative-based, Root Mean-Square Error.

I. INTRODUCTION

There is an ever-increasing amount of information available to us in today's recent times, that includes books, music, movies, and more. It can be testing for users to find what they are consider for with this more amount of information and to develop new items that match their interests and satisfaction. Through the vast amount of information this is where recommendation systems come in, helping users to navigate and find what they are looking for more easily.
This highly scalable and trustworthy system for suggesting books can be adjusted for various genres and user preferences. This books recommendation system can be integrated into various types of online stores and digital libraries to offer customers which has individualized books recommendations. A dataset of books and users was used to calculate this system's performance, and the findings demonstrate that the proposed system outperforms current book recommendation systems. With the assistance of this project, users will be able to find new books to enjoy and contribute to the creation of a more sophisticated and individualized.

Book recommendation systems use machine learning algorithms to analyse data from various sources such as users, ratings and books datasets to provide customizedrecommendations to users. These recommendations can help users to optimize their Books recommends, reduce the amount of time, and mitigate risks associated with users' information and data.

This System presents a Book recommendation system that utilizes machine learning algorithms to provide accurate and personalized recommendations to users. This system analyzes data from various sources, and performed various methods such as Popularity-based and collaborative learning methods are used. The system's primary goal is to provides the quality of the recommendations which is very accurate, easy to maintain and simply to use which is using by the user.

The description of the items is used in content-based filtering, which provides suggestions for items that are comparable to the description of the items. Book are recommended using these multiple filtering models depending on multiple the book's content and the user actions. As a result, mine recommendation engine also recommends books to new readers. For clustering the users in this study, we using both techniques as: K-means and Gaussian mixture. To calculate error between absolute numbers and the results, use the Root Mean Square Error formula.

RMSE number used to determine basic accuracy.

II. LITERATURE SURVEY

The majority of analyzers used the Pearson's Coefficient function to determine how comparable book-rating were in order to make book recommendations:

1. "Collaborative Filtering for the Book Recommendation System" by D. R. K. Srinivas and S. S. Suresh Kumar, published in the International Journal of Computer Science and Mobile Computing in 2017. This paper explores the use of collaborative filtering for book recommendations, using a dataset which of the user ratings and book metadata in system.

1. "Content-based Recommendation Systems" by Pazzani, M. J., & Billsus, D. (2007)- This survey focuses on book recommendation systems using machine learning. It discusses the importance of author, and keywords and the challenges associated with it, such as limited data availability and high variability in the data. The survey also covers the different types of machine learning models used for recommendation, including decision trees, support vector machines.

2. "A survey on Book recommendation systems using machine learning" by Bell & Volinsky (2009) - This survey provides a comprehensive overview of Book-recommendation-system which is use Machine learning, including their architecture, and data sources, also performance evaluation metrics. It covers different types of methods for machine learning used in recommendations, such as rule-based systems, collaborative-filtering, and content-based-filtering. The survey also discusses the challenges and limitations of book recommendation systems, such as data sparsity.

Overall, these surveys demonstrate the growing interest and potential of Book recommendation systems using machine learning in the learning sector, and highlight the need for further research to overcome the challenges associated with these systems.

III. EXISTING SYSTEM

Machine learning algorithms are used in a book recommendation system to analyze user data and offer personalized book recommendations based on reading tastes and history. The primary problem with this is that these systems' recommendations are not transparent. Users may not always understand why a specific book was suggested, which can make it challenging for them to believe in the system and its suggestions. Additionally, the suggestions might not always suit the user's preferences or tastes, which could cause frustration and user engagement. Finally, systems for recommending books that heavily depend on machine learning might need a lot of data to properly train the algorithms. For smaller businesses or groups with less access to data, this can be difficult. Furthermore, the process of gathering and storing user data may give rise to privacy issues, which may discourage some users from using the system.

IV. METHODOLOGY

The methodology for building a book recommendation system using machine learning involves collecting and preprocessing data, extracting relevant features, selecting an appropriate algorithm, training and evaluating the model, and deploying the system for user consumption. Data collection, second is Data preprocessing, third is Feature-extraction, fourth is Algorithm-selection, fifth is Model-training, sixth is Model-evaluation, and the seventh one is Deployment.

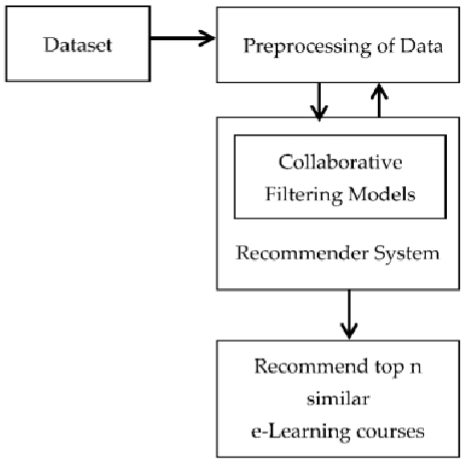

Fig. 1: proposed system task flow

A. DATASET COLLECTION

The Dataset is collect from the Kaggle website: https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset.

we have 3 files in our dataset which is extracted from above kaggle website.

Books - first, there are about books, which includes all the details about books, including an author, title, year of release, etc.

Users: Information about registered users, such as user id and address, is contained in the second file.

Ratings – Ratings contain information like which user has given how much rating to which book.

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/0060973129.0...
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0...
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0...

Fig. 2: Dataset diagram

	User-ID	ISBN	Book-Rating
0	276725	034545104X	0
1	276726	0155061224	5
2	276727	0446520802	0
3	276729	052165615X	3
4	276729	0521795028	6

Fig. 3: Dataset diagram

	User-ID	Location	Age
0	1	nyc, new york, usa	NaN
1	2	stockton, california, usa	18.0
2	3	moscow, yukon territory, russia	NaN
3	4	porto, v.n.gaia, portugal	17.0
4	5	farnborough, hants, united kingdom	NaN

Fig.4: Dataset diagram

A. Data Cleaning and Feature Extraction

The data must be cleaned before being given to any machine learning programs. In the data cleaning procedure, outliers and null numbers are eliminated. There are no null values in the dataset that was gathered. Features are extracted from the data collection after data cleaning. We must examine the correlation between the six distinct features in the dataset that have an impact on the model's output. Correlation is useful for removing characteristics that have little impact on the model's output.

B. Model Architecture.

In this step dataset is splitted in 80% and 20% for training and testing. During training, machine learning algorithms finds the relation between the input and output features. By using this relation, the model able to predict the outputs to the new input values. In this we used the following four machine learning models:

1. content-based-Filtering: The program recommends an item that is comparable that have been used or viewed. To put it simply, this program looks for items that resemble one another. As an illustration, As a result of the two films' similar label and content, if someone likes watching movies, he might also like web series. Only the appearance of the content is identical, and there isn't much attention paid to the viewer. On the basis of prior tastes, it only suggests the product with the highest score.

2. collaborative-based-Filtering: System recommendations that use collaborative-filtering on prior encounters between users and the goal items. To put simply, attempt to notice customers who give them and who resemble them products recomend on the options made by their lookalikes. Let's use an illustration to clarify. Users X and Y share some characteristics, and X has seen movies A, B, and C. If user Y has seen user B, user C, and user D, we will suggest user A to user Y and user D to user X.

3. K-Means: The unlabeled dataset is divided into different clusters by the unsupervised learning algorithm K-Means Clustering. so each collection is only goes to single group shares common feature. In this case, K designates the quantity of predefined groups. In this method, each cluster must be connected to a centroid. A data point's total distance from each of its associated clusters should be as small as possible.

Using the unlabelled dataset as input, k groups are created from the unlabelled dataset. This procedure repeated until no better clusters are found. The k number for this algorithm must be decided in advance.

4. Gaussian Mixture: The clustering methods used by Gaussian Mixture models are effective. The model presupposes that there are a predetermined number of Gaussian distributions, each of which represents a cluster. The data elements are combined in this model to form a single distribution. To assign data points to Gaussian distributions, these models employed the soft clustering method. In a one-dimensional space, the probability density function of the Gaussian distribution.

V. RESULT AND ANALYSIS

After created all the four models by using the training data , we test the models by using the test data and calculates the accuracies of the models. The following table shows the accuracies of different models:

	K-Means	Gaussian mixture
Silhouette Score	0.0433968332584411	0.016778872313688933

Table1: Two models' silhouette ratings

It is determined the differences between the both the typical mean ratings of books for test users and the typical mean ratings of books in the cluster.

Mean rating for 10 random books	3.8876949740034736
Mean rating for 10 books of cluster's favourites	4.3735008665511135

Table2: Mean-scores comparison

By using RMSE, one can calculate the difference between data values and values predicted by a model.

RMSE	0.5957791790493179
Accuracy	1.167727190936663

Table3: Values for RMSE and Accuracy

VI. DEPLOYMENT

The final model is deployed into an application by using flask module of python by creating the user interface which can be accessible easily by the users. It is possible to design a system have to that provides users with high-quality of book recommendations and fosters long-term maintenance and adherence. The application simply takes the input and output provides.

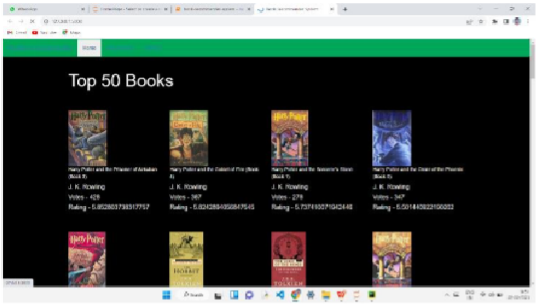

Fig.5: User interface

The following is the recommendations based on user ratings and After typing, select "Search" to have the suggested book, based on user ratings, displayed on the screen. The screenshot of the novels with user ratings is shown in Figure 6.

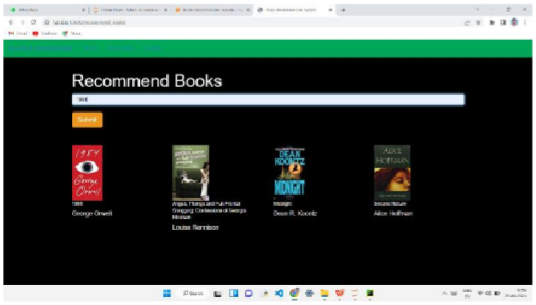

Fig.6: Recommended books

VII. CONCLUSION.

In this project, we show a collaborative filtering-based recommendation system. The primary objective was to speed up recommendations by developing a system that can give users good recommendations without requiring them to register for an extended period of time.

This project has highlighted the benefits of Machine learning in recommendation system, and how it can help improves the efficiency and the achieves of the book recommendation. By automating the recommendation process, users can save time and effort, while also reducing the risk of making costly mistakes.

VIII. REFERENCES

1. Avi Rana and K. Deeba et.al, "Online Book Recommendation System using Collaborative Filtering (With Jaccard Similarity)" in IOP ebooks 1362, 2019.

2. G. Naveen Kishore, V. Dhiraj, Sk Hasane Ahammad, Sivaramireddy Gudise, Balaji Kummaraa and Likhita Ravuru Akkala, "Online Book Recommendation System" International Journal of Scientific & Technology Research vol.8, issue 12, Dec 2019.

3. Ms. Sushma Rjpurkar, Ms. Darshana Bhatt and Ms. Pooja Malhotra, "Book Recommendation System" International Journal for Innovative Research in Science & Technology vol.1, issue 11, April 2015.

4. "Collaborative Filtering for Book Recommendation System using Matrix Factorization" by Anand Rathi. This paper presents a matrix factorization approach to building a book recommendation system and evaluates the system using a dataset of book ratings.

5. "Recommender Systems Handbook" curated by Lior Rokach, Bracha Shapira, and Francesco Ricci. This comprehensive handbook provides an in-depth exploration of the theory and practice of recommender systems, including chapters on collaborative filtering and content-based filtering approaches.

6. "A Hybrid Book Recommendation System using Content-Based Filtering and Collaborative Filtering Techniques" by Prasanna Kumar and J. V. R. Murthy. This paper presents a hybrid approach to building a book recommendation system that combines content-based and collaborative filtering techniques.

AB5-1

ORIGINALITY REPORT

10%	7%	5%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	cse.anits.edu.in Internet Source	2%
2	www.researchgate.net Internet Source	2%
3	ijircce.com Internet Source	1%
4	Himani Shivraman, Umang Garg, Amrita Panth, Ashok Kandpal, Ashish Gupta. "A Model Frame Work To Segregate Clusters Through K-Means Method", 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), 2022 Publication	1%
5	Submitted to Saveetha Dental College and Hospital, Chennai Student Paper	1%
6	Aleide Hoeijmakers, Giovanni Licitra, Kim	1%

classification using passively collected smartphone-based keystroke dynamics within multiple sclerosis", Scientific Reports, 2023
Publication

7	"Advanced Technologies, Systems, and Applications VI", Springer Science and Business Media LLC, 2022 Publication	1%
8	isindexing.com Internet Source	1%
9	digiresearch.vut.ac.za Internet Source	<1%
10	Ahlem Aboud, Nizar Rokbani, Seyedali Mirjalili, Abdulrahman M. Qahtani et al. "A Quantum Beta Distributed Multi-Objective Particle Swarm Optimization Algorithm for Twitter Fake Accounts Detection", Institute of Electrical and Electronics Engineers (IEEE), 2022 Publication	<1%
11	Praveena Mathew, Bincy Kuriakose, Vinayak Hegde. "Book Recommendation System through content based and collaborative filtering method", 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), 2016 Publication	<1

FIRST AUTHOR CERTIFICATE

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

International Conference on

PAPER ID
NECICAIEA2K23091

Artificial Intelligence and Its Emerging Areas
NEC-ICAIEA-2K23
17th & 18th March, 2023

Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that Goli Siva Sankar Tarun , Narasaraopeta Engineering College has presented the paper title Book Recommendation System Using Machine Learning in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar

SECOND AUTHOR CERTIFICATE

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

PAPER ID
NECICAIEA2K23091

International Conference on
Artificial Intelligence and Its Emerging Areas
NEC-ICAIEA-2K23
17th & 18th March, 2023
Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that Galam Anjani Sudheer , Narasaraopeta Engineering College has presented the paper title Book Recommendation System Using Machine Learning in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar

THIRD AUTHOR CERTIFICATE

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

International Conference on

PAPER ID
NECICAIEA2K23091

Artificial Intelligence and Its Emerging Areas

NEC-ICAIEA-2K23

17th & 18th March, 2023

Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that Kunala Kalyan , Narasaraopeta Engineering College has presented the paper title Book Recommendation System Using Machine Learning in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar

FOURTH AUTHOR CERTIFICATE

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

International Conference on

PAPER ID
NECICAIEA2K23091

Artificial Intelligence and Its Emerging Areas

NEC-ICAIEA-2K23

17th & 18th March, 2023

Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that Dr B.Jhansi Vazram, Narasaraopeta Engineering College has presented the paper title Book Recommendation System Using Machine Learning in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar