



---

## **Machine Learning based Efficient Recommendation System for Book Selection using User based Collaborative Filtering Algorithm**

***SURYA P S<sup>1</sup>, Prof. Miriam Thomas<sup>2</sup>, Dr.T.Mahalekshmi<sup>3</sup>***

<sup>1</sup>Final year Student, Sree Narayana Institute of Technology, Kollam, Kerala

<sup>2</sup>Assistant Professor, Sree Narayana Institute of Technology, Kollam, Kerala

<sup>3</sup>Principal, Assistant Professor, Sree Narayana Institute of Technology, Kollam, Kerala

---

### **ABSTRACT:**

Recommender systems are a new breed of online utility that allows users to use the internet while also receiving information about their preferences. Purchasing things through an internet recommender is a relatively simple and quick process. In order to assist users in identifying the appropriate goods, ecommerce websites require a recommendation system. A referral procedure is one of the most effective ways to boost profitability and attract new clients. Existing techniques allow computers to acquire useless data, resulting in a decrease in user attraction and completion of job in a timely and reliable manner. This article presents an overview of the current Recommendation Systems in use in the online book purchase arena. A basic, easy-to-understand approach for book suggestion that assists readers in recommending the appropriate book. In recent years, the administration's recommendation system has concentrated on the information analysis difficulty. Network assets are entirely linked and built swiftly for clients. The proposed method examined the performance of similarity measures in recommending books to a user using a User Based Collaborative Filtering (UBCF) methodology

Keywords: Datamining, Fuzzy Algorithm, Hashing Algorithm, Fingerprint Hashing Algorithm.

---

### **INTRODUCTION**

A recommender system is a type of information filtering system that helps forecast a user's preference for a certain object. We mostly exploited it for commercial purposes. Supported systems are used in a variety of applications and are frequently used to create playlists for music and video services such as Netflix and YouTube. There are also widespread recommender systems for specific themes such as restaurants and online dating. Recommendation systems have been designed to analyze research articles and experts, partners, financial services, and life insurance.

Recommenders often employ collaborative filtering, content-based filtering, or a combination of the two. Collaborative sorting approaches build a model based on a user's previous experience (products that have been liked, rated, or purchased) as well as comparable judgments made by others. A content-based filtering method recommends other items with similar attributes based on a list of discrete, pre-tagged item properties. Present recommendation systems incorporate one or more methodologies in a hybrid system.

Content-based filtering proposes items to users based on a correlation between the user's profile and the item's content. A collection of descriptor terms, usually the words in a text, is used to define an item's content. The user is a Profiles are established by analyzing the meaning of items that the user has seen and defining them using the same terms. The items with the highest positive ratings are recommended to the user. Recommendation Engines are data-intensive systems that use sophisticated patterns to match a set of predetermined parameters and become more efficient as the size of the content they receive grows. CBF employs a variety of strategies to detect similarities in papers in order to generate concrete suggestions.

Furthermore, if the consumer's profile changes, the CBF technique may readily adjust its recommendations. Collaborative recommender systems attempt to forecast item preferences based on how other users have categorised goods for a certain target user. Web search portals are starting to appreciate the relevance of news subject recommendations, despite the two challenges described above. They intend to combine the content-based approach to filtering and the collaborative approach to filtering in order to provide suggestions. Data is filtered using suggestions from others in collaborative filtering. The notion here is that those who enjoyed something in the past would most likely like it again in the future.

Implementing collaborative filtering can be done in two ways: The item-based or user-based method is used in the model-based approach. In a user-based strategy, users play the most important role. When a group of customers shares similar tastes, they form a group. Users are recommended things based on the opinions of people in the same group who share similar likes. The similarity between numerous items is known while adopting a methodology by employing the most appropriate similarity measure to find the user ratings to the objects that he/she has not previously rated.

The anticipated rating value is based on the weighted average of numerous people's suggestions, rather than on the one who is most similar. The weight assigned to a person's ratings is determined by comparing the individual to the target user for whom the forecast should be produced. As a correlation measure, the Pearson correlation coefficient can be utilized.

Using machine learning methods, we create collaborative filtering models in the memory-based approach. This is where we locate the user's ratings of an item.

---

## LITERATURE SURVEY

Different types of methodologies for recommendation systems exist to deliver specialized recommendations. Traditionally, content-based (CBF) and collaborative filtering (CF) have been used. The CBF method is aware of the content of an element, i.e. the material, and may adjust it appropriately based on the user profile. Collaborative filtering, on the other hand, is not content-based and matches to ideas based on specific criteria. In the past, some users agreed to these ideas, and there is a potential that some users will agree to these ideas in the future. Based on the ratings you provide to the products, you can collect data about your tastes.

[1-2] A high-accuracy hybrid recommendation system in Java that combines the benefits of collaborative, content-based, and demographic techniques. This newly introduced book recommendation engine is a professional tool for recommending e-user books. The recommender feature will undoubtedly be a fantastic Java language web application. For today's high-demand websites for online purchases, this form of web application should be advantageous. Because it integrates the qualities of many recommendation approaches, this hybrid recommender system is more accurate and efficient. The book's recommendation engine should reduce the time and effort required to select the best book from a sea of options.

[11-12] For movie suggestions, it employs the svd algorithm on a movie lens data set, and it proves to be the most effective.

[15-16] Proposed a technique for news suggestions on Bing using CCF and the SVD model, which proved to be more efficient than existing algorithms. By utilizing some of the rich contexts and focusing on long-tail users, the proposed CCF combines the benefits of the Content-based (CBF) Filtering approach with the features of the Collaborative (CF) Filtering approach. This CCF is intended for situations such as the subject recommendation on Bing News, when a piece of news could be recommended.

Rich contexts, such as query results, are used to interpret the data.[9-10] For social voting, it employs closest neighbors algorithms and matrix factorization, concluding that affiliations play a crucial role in improving accuracy.[20]

[3-4] The paper discusses some of the issues and obstacles that currently exist in recommender systems and recommends that new techniques to implementing recommender systems should be used. To search the books list based on their rating and content, I used cooperative filtering and content-based filtering. The book suggestion method is mostly based on the book value and rating offered by current users.

[19-20] Proposed recommendations based on the selected artist's resemblance, top artists in a genre, using a hybrid model, and artists in a genre, using a hybrid model, and artists listened to by the user's friends.

[7-8] Comparing Content Based (CBF) and Collaborative Filtering (CF) in Recommendation Systems provides an overview of recommendation systems as well as the differences between the two approaches. The collaborative algorithm in this work employs "User Behavior" to suggest items. In terms of transaction history, ratings, and information on selection and purchase, they take advantage of the actions of other users and objects. Items are recommended to new users based on the experiences and expectations of other users. In content-based filtering, we need to know the content of both the person and the item. Typically, we use the common attribute space content to create user and item profiles.

[5-6] Implements an online recommender system that uses rapid sort and cosine similarity via the CF algorithm. Proposed a technique based on OOADM (Object-Oriented Analysis and Design Methodology), an enhanced collaborative filtering algorithm, and a fast sort algorithm, as well as Django and NOSQL, which proved to be efficient and scalable. This study discusses that recommendation systems based on collaborative filtering methods, content, and hybrid recommendation methods have been proposed, however they all have a number of difficulties that pose research obstacles. Working on this study topic is necessary in order to discover and provide new strategies for overcoming difficulties and reducing costs.

provide guidelines for working together to filter a wide range of applications while keeping concerns of trustworthiness and confidentiality in mind [17-18] For online books, a hybrid recommendation approach is available. In order to acquire book recommendations for a user, Portal used associative rule mining and KNN [13-14]. It discusses numerous ways for evaluating similarity and concludes that item-based methods are more accurate than user-based methods.

---

## DATASET

The data for this was obtained from Kaggle Goodreads- books data, where you can get information about books, authors, and titles, as well as ratings. On the basis of cosine similarity, we recommended the books to users. Many variables, such as goodreads book id, user id, best book id, tag id, tag name, ratings, received ratings, total ratings, books count, and similarity measure, influence a book's recommendations. This system makes use of a data set with good readings. There are seven tables in this data set.

Books.csv, Genres.csv, to read.csv, tags.csv, book tags.csv, max ratings.csv, and my ratings.csv are all examples of CSV files.

- 1.Books.csv- contains a total of 10K books with attributes such as book isbn number, author, rating, and so on.
- 2.Geners.csv - This file contains many book categories. It has the tag name as an attribute, as well as the index.
- 3.Book tags.csv - This file contains the columns tags name given by each user, as well as the number of times the tag name has been applied to a book.
- 4.Max rating.csv- provides a user's rating and includes columns such as book id and rating for the book.
- 5.Ratings.csv- This file contains information about the ratings provided to books by various users and includes columns such as user id, book id, and rating. It contains up to 9,00,000 records.
- 6.to read.csv- includes columns user id and book id and reveals which user marked a book as to be read.
- 7.Tags.csv -Different tags provided by users are kept as tags in this file, which includes the tag name attribute.



Fig1: Recommendations based on similarity

## METHODOLOGY

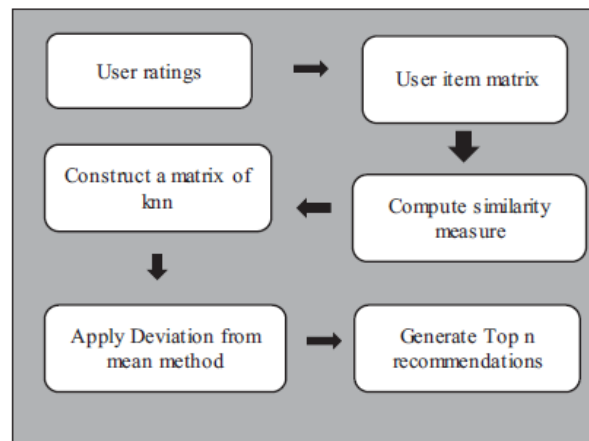


Fig 2: System Architecture

The phases of the proposed method are summarized as follows:

**User rating:** The system gathers feedback from goodreads-book dataset

**User-item matrix:** User related data and device objects are entered as a list of numerical ratings in a User-Item matrix.

**Compute similarity matrix:** Using the similarity methods (e.g., PCC, CPCC, Cosine, Jaccard) we compute similarity matrix for each and every user. The similarity measure ranges from 0, 1.

**Construct a matrix of knn:** Get the k nearest neighbours of target user and generate a matrix with the candidate books and users. The books that are rated by the neighbours and not rated by the target user are candidate books.

**Apply Deviation from mean method:** The predictive scores are calculated from the above generated matrix by applying aggregate method-deviation from mean method. The predictive scores might be binary (0/1) or numerical values which depict the score that target user might give for all books based on the ratings given by neighbours and similarities. Sort the scores of all books to get an order.

**Generate Top n recommendations:** According to the above sorted books based on nearest neighbours and scores. Top n recommendations are given to the target user.

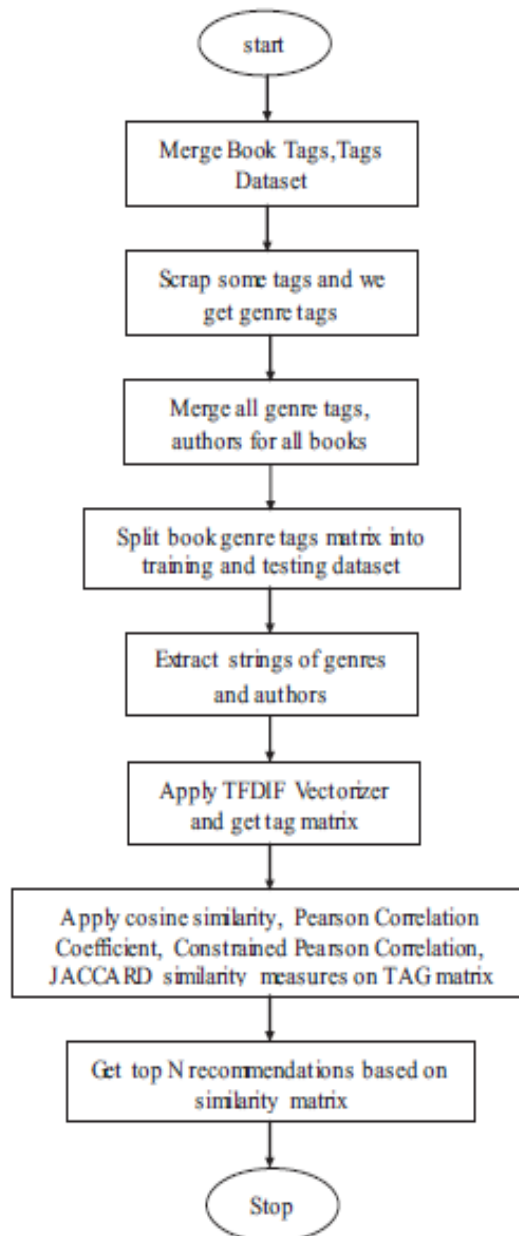


Figure: 3 Flow of system design

#### Similarity Measures:

##### Pearson coefficient correlation:

Pearson coefficient correlation function calculates pairwise correlation. Always diagonal in matrix is correlated perfectly. Here Pearson correlation gives linear correlation between 2 variables.

It ranges between -1 to +1.

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \rightarrow 1$$

##### Constrained Pearson Coefficient:

In PCC the average of the values. But in constrained pearson coefficient the median is considered. In a range [1,5] the median value 3 is considered for calculation. It is almost same as PCC but with more improved results.

$$r = \frac{\sum(x - x_m)(y - y_m)}{\sqrt{\sum(x - x_m)^2 \sum(y - y_m)^2}} \rightarrow 2$$

**Cosine Similarity:**

It is the dot product of the two data points. It measures the cosine angle between the objects. If angle between them is 0 degrees then the similarity is 1. If angle is 90 degrees then similarity score is 0. It is a very popular matrix in finding similarity because it is easy to evaluate specially for sparse vectors.

$$\text{sim}(x,y) = \cos((\vec{x}), (\vec{y})) = \frac{(\vec{x}) \cdot (\vec{y})}{\|\vec{x}\|_2 * \|\vec{y}\|_2} = \frac{\sum_{s \in S_{xy}} r_{x,s} r_{y,s}}{\sqrt{\sum_{s \in S_{xy}} r_{x,s}^2} \sqrt{\sum_{s \in S_{xy}} r_{y,s}^2}} \rightarrow 3$$

**Jaccard Similarity:**

The Jaccard method was proposed by Koutrika and Bercovitz to calculate correlations between pairs of users. The Jaccard approach only takes into account the number of co-ratings for each user pair to describe their partnership. Two users will have a strong relationship if they have different rating trends, and vice versa. It does not consider absolute values of rating.

$$s(x, y)^{Jaccard} = \frac{|I_{xy}|}{|I_x \cup I_y|} \rightarrow 4$$

**Pseudocode:**

- ❖ Start
- ❖ function Top 10 Recommendations
- ❖ Initialize
- ❖ Give counter for all similarity scores of books in index row in similarity matrix
- ❖ Sort all similarity score by sorted ()
- ❖ Get indices of all top 10 recommendations
- ❖ Return book titles

**Algorithm: User based collaborative filtering for Book recommendations**

	Rated	Unrated
Recommended	TP	FP
Not Recommended	FN	TN

**Table 1.Recommendation Confusion Matrix**

**Evaluation Metrics:**

**TP:** the number of books that are recommended to user and rated

**TN:** the number of books that are not recommended and not rated by users.

**FP:**the number of books that are recommended and but are not rated by users.

**FN:**the number of books that are not recommended but rated by users.

**Recall:**It is the ratio of number of correctly predicted positive observations to the total number of positive observations.

$$\text{Recall} = \frac{TP}{TP+FN}$$

**Precision:** Precision is a random error description, a measure of statistical variability. Positive observations predicted for all actual class observations

$$\rightarrow \text{Precision} = \frac{TP}{TP+FP}$$

**F-score:** F-score is the ratio of precision and recall more the Fscore is more the accuracy

$$\text{F-Measure} = \frac{(2 * \text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \rightarrow$$

**MAP [Mean absolute Precision]:** The standard single number metric for evaluating search algorithms is Mean Average Precision (MAP).

Average accuracy (AP) is the average of accuracy values (uninterpolated) at all ranges where relevant document s are located.

---

## BACKGROUND

Technologies used in this project:

### a. FRONT-END DEVELOPMENT

#### 1. HTML

HTML (Hypertext Markup Language) is the industry standard markup language for documents that are intended to be viewed in a web browser. Technologies such as Cascading Style Sheets (CSS) and programming languages like JavaScript can help. Web browsers receive HTML documents from a web server or locally stored files and convert them to multimedia web pages. HTML originally featured cues for the document's look and described the structure of a web page logically.

HTML elements are the fundamental components of HTML pages. Images and other objects, such as interactive forms, can be embedded in the produced page using HTML techniques. HTML allows you to create organized documents by indicating structural semantics for text elements like headers, paragraphs, lists, links, quotations, and other elements. Tags, which are written in angle brackets, separate HTML elements. Tags like `image />` and `input />` insert content into the page immediately. Other tags, such as `p>`, surround and offer information about document text, and may comprise sub-elements such as other tags. The HTML tags are not displayed by browsers, but they are used to read the page's content.

HTML allows scripting languages like JavaScript to insert programs that influence the behavior and content of online pages. CSS determines the appearance and layout of material. Since 1997, the World Wide Web Consortium (W3C), which used to maintain the HTML standards and now maintains the CSS standards, has pushed the usage of CSS over explicit presentational HTML.

#### 2. CSS

CSS (Cascading Style Sheets) is a style sheet language for describing the appearance of a document authored in a markup language like HTML. Along with HTML and JavaScript, CSS is a key component of the World Wide Web. CSS is a style sheet that allows you to separate presentation from content, including layout, colors, and fonts.

This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, and enable multiple web pages to share formatting by specifying the relevant CSS in a separate .css file, reducing complexity and repetition in the structural content and allowing the .css file to be cached to improve page load speed between the pages that share the file and its formatting.

The ability to offer the same markup page in different styles for distinct rendering techniques, such as on-screen, in print, by voice (through speech-based browser or screen reader), and on Braille-based tactile devices, is also made possible by the separation of formatting and content. If the information is accessible on a mobile device, CSS contains rules for different formatting. Cascading gets its name from the priority mechanism used to select which style rule applies when many rules match the same element. This priority-cascading strategy is predictable.

#### 3. BOOTSTRAP

Bootstrap is a free and open-source CSS framework for front-end web development that is responsive and mobile-first. It includes design templates for typography, forms, buttons, navigation, and other interface elements that are based on CSS and (optionally) JavaScript. With almost 142,000 stars, Bootstrap is the second most starred project on GitHub, after only freeCodeCamp (nearly 312,000 stars) and the Vue.js framework.

#### 4. JAVASCRIPT

A scripting language is JavaScript. Learning a scripting language is simple and quick. In run-time, a scripting language is interpreted. It is not compiled like C++, C#, VB.net, and other languages. JavaScript is a client-side scripting language that runs in a web browser. It was created by Netscape and is one of the most well-known scripting languages due to its simplicity. On the other hand, JavaScript can be utilized on the server. JavaScript is supported by the majority of popular browsers. It's simple to use it to interact with HTML components. You may validate text fields, disable buttons, validate forms, and alter the color of your page's backdrop. JavaScript allows you to do all of this. It, too, is a programming language.

features variables, arrays, functions, operators, objects, and much more to aid in the creation of better script for your pages. On the server side, JavaScript can be used to handle database entries, for example. JavaScript code can be directly entered into HTML, or it can be placed in a separate file with the .js extension and linked to the webpage via the .js file.

#### 5. PYTHON 3

Guido van Rossum of the Netherlands' National Research Institute for Mathematics and Computer Science created Python in the late 1980s as a high-level, interpreted scripting language. In 1991, the first version was posted to the alt.sources newsgroup, and in 1994, version 1.0 was released.

Python 2.0 was launched in 2000, and until December 2008, the 2.x versions were the most widely used. The development team decided to release version 3.0 at that time, which included a few minor but important modifications that were not backward compatible with the 2.x versions. Python 2 and 3 have a lot in common, and several Python 3 features have been backported to Python 2. However, they are still incompatible in general.

Python 2 and 3 have both been maintained and developed, with both receiving periodic release updates. The most recent versions accessible as of this writing are 2.7.15 and 3.6.5. Python 2 does, however, have an official End Of Life date of January 1, 2020, after which it will no longer be supported. If you're new to Python, it's best to start with Python 3, which is what this tutorial will cover.

Python is still maintained by the Institute's core development team, and Guido is still in control, with the moniker of BDFL (Benevolent Dictator For Life) bestowed upon him by the Python community. By the way, the name Python is derived from the British comic troupe Monty Python's Flying Circus, of which Guido was (and probably still is) a fan. The Python manual is littered with allusions to Monty Python jokes and films.

.Python is Popular

- Python is Interpreted
- Python is Free
- Python is Portable
- Python is Simple

## 6. PYTHON LIBRARIES

A Python library is a reusable code snippet that you can use in your programs and projects. In contrast to languages such as C++ or C, Python libraries are not tied to any specific context. A 'library' is a vague term that refers to a collection of fundamental components. A library is essentially a collection of modules. A package management, such as `rubygems` or `npm`, can be used to install a library.

.Python Standard Library

Python's Standard Library is a collection of the language's exact syntax, tokens, and semantics. It's included in the standard Python distribution. When we started with an introduction, we mentioned this. It's written in C and takes care of things like I/O and other essential functions. Python is what it is because of all of its functionality. The standard library is built around more than 200 core modules. Python includes this library. You can also use the Python Package Index (PyPI). to access a growing collection of thousands of components in addition to this library

### 1. Matplotlib

### 2. Pandas

### 3. Requests

### 4. NumPy

### 5. SciPy

### 6. Scrapy

### 7. Scikit

## b. BACK-END DEVELOPMENT

### 1. MySQL

MySQL is the most popular open source database software in the world, with over 100 million copies downloaded or distributed over its lifetime. MySQL has become the preferred choice for Web, Web 2.0, SaaS, ISV, Telecom companies, and forward-thinking corporate IT Managers because it eliminates the major problems associated with downtime, maintenance, and administration for modern, online applications due to its superior speed, reliability, and ease of use.

Many of the world's largest and fastest-growing companies, including Yahoo!, Alcatel-Lucent, Google, Nokia, YouTube, Wikipedia, and Booking.com, utilize MySQL to power their high-volume Web sites, vital business processes, and packaged software, saving time and money.

MySQL Enterprise is the company's flagship package, which includes a comprehensive collection of production-tested software, proactive monitoring tools, and premium support services for an affordable annual subscription.

MySQL is an important component of the LAMP (Linux, Apache, MySQL, PHP / Perl / Python) open source corporate software stack. Because of its reduced cost and lack of platform lock-in, LAMP is becoming a more popular alternative to pricey proprietary software stacks.

Sun Microsystems, one of the world's major contributors to open source software, owns, develops, and supports the MySQL database. MySQL was created in Sweden by three Swedes and a Finn: David Axmark, Allan Larsson, and Michael "Monty" Widenius, who had collaborated since the 1980s. The greatest and most widely used database for internet applications in the world.

- Available and affordable for all

- Continuously improved while remaining fast, secure and reliable
- Fun to use and improve
- Free from bugs.

## 2. VS CODE – IDE

Visual Studio Code blends the ease of use of a source code editor with advanced developer features such as IntelliSense code completion and debugging. It is, first and foremost, an editor who stays out of your way. Less time fussing with your environment and more time executing on your ideas thanks to the delightfully frictionless edit-build-debug cycle. Visual Studio Code's core feature is a lightning-fast source code editor that's ideal for day-to-day use. Syntax highlighting, bracket matching, auto indentation, box selection, snippets, and more are all available in VS Code, which supports hundreds of languages. You can traverse your code with ease thanks to intuitive keyboard shortcuts, easy modification, and community-contributed keyboard shortcut mappings. When it comes to serious coding, you'll typically benefit from tools that understand more code than just blocks of text. IntelliSense code completion, sophisticated semantic code understanding and navigation, and code refactoring are all included into Visual Studio Code. When the coding becomes difficult, the tough turn to debugging. We made it happen since debugging is generally the one feature that developers miss the most in a lighter development experience. You may step through source code, analyze variables, view call stacks, and run commands in the console with Visual Studio Code's interactive debugger.

VS Code also connects with build and scripting tools to help you complete typical tasks faster. With Git support in VS Code, you can work with source control without having to leave the editor, including examining pending changes diffs.

## 3. DJANGO

Django is a high-level Python web framework that promotes rapid development and a simple, practical design. It's built by professional developers to take care of a lot of the headaches of web development so you can concentrate on developing your app instead of reinventing the wheel. It's open source and free.

- ❖ Ridiculously fast.
- ❖ Reassuringly secure.
- ❖ Exceedingly scalable.

Django's ability to scale swiftly and flexibly is used by some of the busiest websites on the internet. Django is a free and open-source web framework based on Python that uses the model–template–views (MTV) architectural paradigm. It is maintained by the Django Software Foundation (DSF), a non-profit organization based in the United States. Django's main purpose is to make building complex, database-driven websites easier. The framework promotes component reusability and pluggability, as well as less code, low coupling, rapid development, and the "don't repeat yourself" philosophy. Even for settings, files, and data models, Python is used throughout. Django also has an administrative create, read, update, and delete interface that is produced dynamically and customized using admin models.

### C. TECHNICAL DETAILS

#### i. Artificial Intelligence

Artificial intelligence (AI) is the simulation of human intelligence in robots that have been trained to think and act like humans. The phrase can also refer to any machine that demonstrates human-like characteristics like learning and problem-solving.

Artificial intelligence (AI) is a broad field of computer science concerned with creating intelligent machines that can accomplish activities that would normally need human intelligence. Although AI is a multidisciplinary discipline with many methodologies, advances in machine learning and deep learning are causing a paradigm shift in nearly every sector of the IT industry.

The ability of artificial intelligence to rationalize and execute actions that have the best likelihood of reaching a certain goal is its ideal feature. Machine learning is a subset of artificial intelligence that refers to the idea that computer systems can learn from and adapt to new data without the need for human intervention. This is made possible by deep learning algorithms. Learning is accomplished by absorbing large volumes of unstructured input such as text, photos, and video. Artificial intelligence is founded on the idea that human intelligence may be characterized in such a way that a machine can simply duplicate it and carry out activities ranging from the most basic to the most complicated. Artificial intelligence's goals include simulating human cognitive processes. To the extent that they can be concretely characterized, researchers and developers in the field are making unexpectedly rapid progress in simulating tasks such as learning, reasoning, and perception. Some predict that in the near future, innovators will be able to create systems that can learn and reason about any subject faster than humans can. Others, on the other hand, remain suspicious, claiming that all cognitive activity is loaded with value judgments based on human experience.



Artificial intelligence has a plethora of uses. The technology can be used in a variety of businesses and areas. In the healthcare industry, AI is being tested and used for dosing pharmaceuticals and various treatments in patients, as well as surgical operations in the operating room.

## ii. Machine Learning

Machine learning is a type of data analysis that automates the creation of analytical models. It's a field of artificial intelligence based on the premise that computers can learn from data, recognize patterns, and make judgments with little or no human input. Machine learning is a branch of artificial intelligence (AI) that allows computers to learn and improve on their own without having to be explicitly programmed. Machine learning is concerned with the creation of computer programs that can access data and learn on their own. The learning process starts with observations or data, such as examples, direct experience, or instruction, so that we can seek for patterns in data and make better decisions in the future based on the examples we provide. The fundamental goal is for computers to learn on their own, without the need for human involvement, and to change their behavior accordingly.

### Machine Learning Methods

Machine learning methods are as follows:

#### ❖ Supervised machine learning algorithms

Supervised machine learning algorithms can use labeled examples to apply what has been learnt in the past to fresh data in order to predict future events. The learning algorithm creates an inferred function to generate predictions about the output values based on the examination of a known training dataset. After enough training, the system can provide targets for any new input. The learning algorithm can also compare its output to the correct, intended output and detect faults, allowing the model to be modified as needed.

#### ❖ Unsupervised machine learning algorithms-

It's employed when the data being used for training isn't categorised or labeled. Unsupervised learning investigates how computers might infer a function from unlabeled data to describe a hidden structure. The system doesn't figure out the proper output, but it examines the data and can infer hidden structures from unlabeled data using datasets.

#### ❖ Semi-supervised machine learning algorithms-

Semi-supervised machine learning algorithms straddle the line between supervised and unsupervised learning since they train with both labeled and unlabeled data – often a small amount of labeled data and a big amount of unlabeled data. This strategy can significantly enhance learning accuracy in systems that adopt it. Semi-supervised learning is typically used when the acquired labeled data necessitates the use of skilled and appropriate resources to train / learn from it. Obtaining unlabeled data, on the other hand, usually does not necessitate additional resources..

#### ❖ Reinforcement machine learning algorithms

It's a type of learning that interacts with its surroundings by performing activities and detecting errors or rewards. The most important elements of reinforcement learning are trial and error search and delayed reward. This technology enables machines and software agents to automatically select the best behavior in a given situation in order to improve their efficiency. For the agent to learn which action is better, simple reward feedback is required; this is known as the reinforcement signal.

Machine learning allows for the examination of large amounts of data. While it generally provides faster, more accurate results in identifying profitable possibilities or risky threats, fully training it may take more time and resources. Combining machine learning with AI and cognitive technologies can improve its ability to process massive amounts of data.

.Another categorization of machine learning tasks arises when one considers the desired output of a machine learned system:

- In classification, inputs are divided into two or more classes, and the learner must produce a model that assigns unseen inputs to one (or multi-label classification) or more of these classes. This is typically tackled in a supervised way. Spam filtering is an example of classification, where the inputs are email (or other) messages and the classes are “spam” and “not spam”.
- In regression, also a supervised problem, the outputs are continuous rather than discrete.
- In clustering, a set of inputs is to be divided into groups. Unlike in classification, the groups are not known beforehand, making this typically an unsupervised task.
- Density estimation finds the distribution of inputs in some space.
- Dimensionality reduction simplifies inputs by mapping them into a lower-dimensional space. Topic modeling is a related problem, where a program is given a list of human language documents and is tasked to find out which documents cover similar topics.

## II. EXISTINGSYSTEM

Existing book recommendation engines, such as those used by the best-rated book-buying websites, rely on traditional algorithms to create recommendations. The majority of recommendation systems employ content-based filtering (CBF) techniques, which necessitate a detailed description of objects and a well-organized user profile before making recommendations to users. The system creates suggestions from a source based on the features associated with products and the user's information in the Content-based Recommendation Engine. Content-based recommenders approach suggestion as a user-specific classification problem, with a classifier being learned.

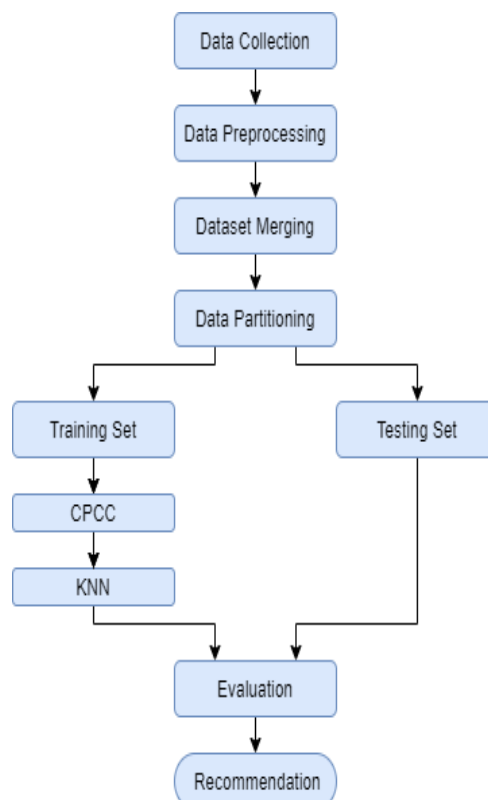
depending on product features, for the user's likes and dislikes A content-based filtering method recommends other items with similar attributes based on a list of discrete, pre-tagged item properties. Present recommendation systems incorporate one or more methodologies in a hybrid system. Content-based filtering proposes items to users based on a correlation between the user's profile and the item's content. A collection of descriptor terms, usually the words in a text, is used to define an item's content. The user profile is formed by analyzing the meaning of items that the users have seen using the same phrases. The items with the highest positive ratings are recommended to the user. Recommendation Engines are data-intensive systems that use sophisticated patterns to match a set of predetermined parameters and become more efficient as the size of the content they receive grows. CBF uses multiple ways to discover similarities in documents, such as cosine similarity or jaccard similarity, to generate concrete suggestions. In addition, if the consumer's profile changes, the CBF technique may readily adjust its recommendations.

## III. PROPOSED SYSTEM

The suggested system employs a user-based collaborative filtering mechanism in its recommendation system. Collaborative recommender systems combine item ratings or suggestions, identify commonalities among users based on their ratings, and generate new recommendations based on inter-user comparisons. Collaborative recommender systems attempt to forecast item preferences based on how other users have categorised goods for a certain target user. Data is filtered using suggestions from others in collaborative filtering. The notion here is that those who enjoyed something in the past would most likely like it again in the future.

The K-Nearest Neighbours machine learning classification technique is employed in the proposed system to locate clusters of similar users. The closeness of instances is determined using a distance measure. The constrained pearson correlation coefficient (CPCC) similarity measure is employed as the distance measure to determine user similarity.

## IV. SYSTEMDESIGN



V. RESULTS AND DISCUSSIONS

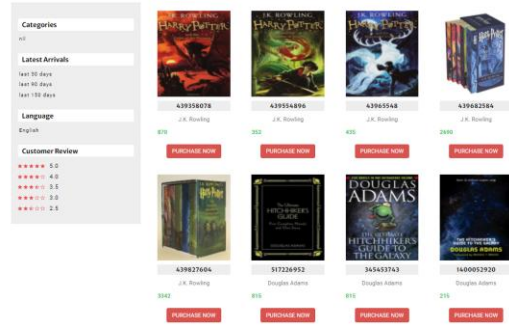


Fig 1: Admin Page



Fig 2: Manage Category Page

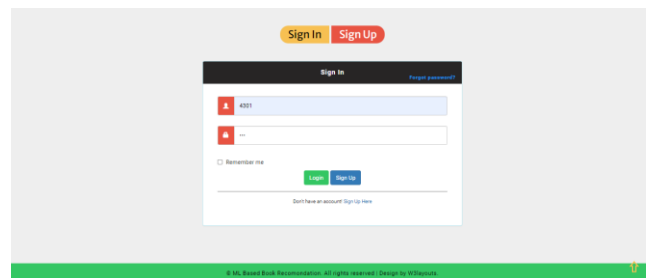


Fig 3: User's Sign Up page

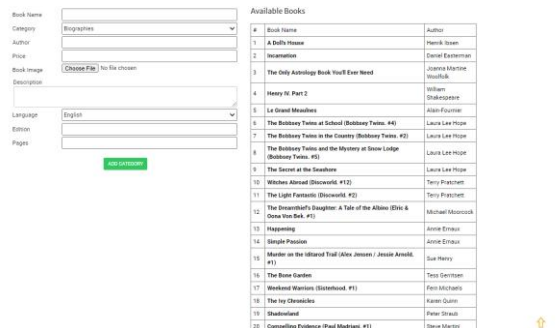


Fig 4: Add Book



Fig 5: User Home Page



Fig 6: User Selection Page



Fig 7: User Recommendation Page

## VI. CONCLUSION

Introduce a new technique to recommend books using user based collaborative filtering for assisting the users in an efficient way. K-Nearest Neighbours supervised machine learning algorithm is used to build the recommendation system based on user-user similarity matrix. The user based collaborative filtering method builds a user-user similarity matrix by using some similarity measure to present the similarity between users in order to utilize it for further processing. For calculating the similarity among users, constrained Pearson correlation coefficient (CPC) similarity measure is used. This similarity score is used by the KNN to find the similar user. The system makes recommendation based on the interests of similar users, i.e., purchase history of similar users.

**VII. REFERENCES**

- [1] AbhilashaSase, KritikaVarun, SanyuktaRathod, Prof. DeepaliPatil(2015),“A Proposed Book Recommender System”, *International Journal of Advanced Research in Computer and Communication Engineering*.
- [2] Murthy ,K.V.S.S.R., &Satyanarayana ,K.V.V (2018),”Intusion detection mechanism with machine learning process “ *International journal of Engineering and Technology*.
- [3] ZhongqiLu, ZhichengDou, JianxunLian, XingXie and QiangYang(2015),” Content-Based Collaborative Filtering for News Topic Recommendation”, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- [4] Narsinga rao,M.R.,Venkatesh Prasad ,V.,SaiTeja, P.,Zindavali, M .&phanindraReddy,O.(2018),”A survey on Preventio of Overfit ting in convoluntional neura networks using machine learning techninques”*International journal of engineering and Technology*.
- [5] E. UkoOkon, B. O. Eke, P. O. Asagba(2018),” An Improved Online Book Recommender System using Collaborative Filtering Algorithm”, *International Journal of Computer Applications*.
- [6] LavanyaReddy,L.Srisai Krishna,J,&vaishnavi,G.V.S.S(2018),”Survey on software reliability prediction using machine learning techniques”, *Journal of Advaced research in dynamical and Control Systems”*
- [7] ManiMadhukar(2014),” Challenges & Limitation in Recommender Systems”, *International Journal of Latest Trends in Engineering and Technology*.
- [8] Lakshmi ,C.r.,rao, D.T.,&Rao,G.V.S(2018),”Fog detection ad Visibility enhancement under partial machine learning approach”*International Conference on Power,control,Signals and instrumentation Engineering*.
- [9] Dharna Patel, Dr. Harish Patidar(2018),” Hybrid Recommendation Solut ion for Online Book Portal”, *International Journal for Research in Applied Science & Engineering Technology*
- [10] Krishna Mohan ,G.,Yoshitha ,N.,Lavanya ,M.L.N,&KrishnaPriya ,A.(2018) ,”Assesment and Analysisb of software reliability usinfmahine learning techniques”*Internationaljournal of Enginerring and technology*
- [11] ParulAggarwal, Vishal Tomar, AdityaKathuria(2017),” Comparing Content Based and Collaborative Filtering in Recommender Systems”,*International Journal of New Technology and Research*.
- [12] Surarchita,V.,VenkateswaraRao ,P.,battacharyya ,D.,Kim,T.(2016) ,”Classification of penaeid prawn species using radial basis probabilistic neural networks and supportvectormachines”*InternationalJournal of Bio-Science and Bio-technology*
- [13] Dr. Mohammed Ismail 1 ,Dr. K. BhanuPrakash ,Dr. M. NagabhushanaRao (2018)”Collaborative filtering-based recommendation of online social voting” *International Journal of Engineering & Technology*.
- [14] Vidhyullatha ,p.,&RajeswaraRao,D.(2016),”Machine learning techniques on mult i dimensional curve fitting data based on r-square and chi-square methods”*InternationalJournal of Electrical and Computer Engineering*
- [15] Immidi Kali Pradeep, M.JayaBhaskar(2018)”Comparative analysis of recommender systems and its Enhancements” *International Journal of Engineering & Technology*.