**Book-Recommender system**

**Rohit Kumar sharma, Data science trainee,**

**Alma Better, Bangalore**

**Abstract:**

Finding relevant books from a huge e-book space has become a tremendous difficulty for internet users as even the volume of online books grows tremendously as a result of the COVID-19 epidemic. Personal recommendation systems have arisen as a means of conducting efficient searches that mine related publications based on user ratings and interests. The majority of these existing systems are based on user ratings and employ content-based and collaborative learning methodologies. The rating technique used by these systems is irrational, as it counts users who have already been unsubscribed from the services and no longer rate books. This document developed an effective system for recommending books to internet users that used the clustering approach to score a book and then determined a book's similarity to suggest a new book. To measure distance and determine similarity between book groups, the suggested system uses the K-means Cosine Distance function and the Cosine Similarity function. Specificity, Sensitivity, and for ten separate datasets, F- Scores were calculated. On averaged, the specificity of the test was higher than the sensitivity, indicating that the test was more accurate. Boring books could be removed from the reader's list using a classifier. A receiver operating characteristic curve (ROC) was also plotted.  find a graphical representation of the accuracy of the classifiers The majority of the datasets fell between the ideal diagonal classifier line and the worst classifier line. According to the findings, recommendations based on a specific book are more accurate than a user-based recommendation system.

***Keywords: - machine learning, Book Recommendation, clustering***

**1. Problem Statement**

A recommender system will help users who do not have enough individual knowledge to pursue through the different types of options offered by a website. It will provide the users with information to assist them to make a decision as to which items to purchase. The challenge with the recommendation system is predicting the users' opinions on various substances and being able to recommend the best goods to each user.

Problems faced by recommender systems.

Data sparsity, scalability, and grey sheep are some other issues with recommendation systems. The term "data sparsity" refers to the fact that the data is dispersed and contains null and missing values. Scalability indicates that predicting a large number of ratings items is challenging. Gray sheep denotes an issue with time and memory. The main objective is to build a Recommender system, which could help them in Recommending proactively.

**2. Introduction**

When selling things online, most businesses have a recommendation system in place. However, almost all websites are not designed with the buyer in mind; instead, the companies' force add-on sells to customers by recommending things that are superfluous and irrelevant. A customized recommendation system (PRS) assists individual users in locating interesting and relevant products among a large number of options. Consumers now have a plethora of ecommerce product selections because to the internet's expansion. Consumers face a serious challenge in finding the correct products at the right moment. Recommendation algorithms are utilized in a wide range of services, including online shopping, music, and movies. Music services like Pandora, for example, discover up to 450 uniquely identifying features of songs to find music that matches its subscribers' tastes. Other music streaming services, such as Spotify, generate weekly song recommendations and tailored radio stations based on the music tastes of comparable users. Netflix, a major television and movie streaming service, use similar algorithms to suggest films to viewers. We can observe how recommendation systems have a surprising amount of influence on the items customers interact with on a regular basis. We've combined various recommendation strategies to properly forecast how consumers will react to books.

## **3. Data Description**

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

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

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

**4. Steps involved:**

* **Data Understanding**

This is the crucial part of the project, here we are observing the type of data we have. Type of features and values dataset contain. According to our visualization we take necessary steps.

* **Exploratory Data Analysis**

After loading the dataset, we performed this method by comparing our different variable with each other’s. This process helped us figuring out various aspects and relationships among the variables. Plotting different graphs specially bar graph helps us to visualize clearly relationship between different variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

* **Null values Treatment**

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we handle them at the beginning of our project in order to get a better result. Null values can cause lots of error so it needs to be handled at very start.

**5. Popularity based Recommender System**

It's a form of recommendation system that works on the basis of popularity or anything that's currently popular. These algorithms look for products or movies that are currently trending or popular among consumers and then immediately recommend them. For example, if a product is frequently purchased by the majority of people, the system will learn that it is the most popular, and it will recommend that product to every new user who has just joined up, increasing the likelihood that the new user would purchase it as well.

**Merits of popularity-based recommendation system**

1. It does not suffer from cold start problems which means on day 1 of the business also it can recommend products on various different filters.
2. There is no need for the user's historical data.

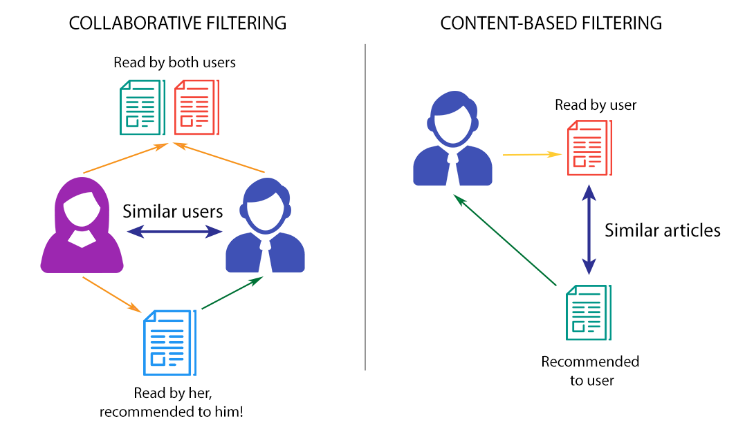
**Demerits of popularity-based recommendation System**

* + Not personalized
  + The system would recommend the same sort of products/movies which are solely based upon popularity to every other user

**Example**

* + Google News: News filtered by trending and most popular news.
  + YouTube: Trending videos
* **Collaborative Filtering Based Recommender System**

The practice of filtering for information or patterns employing strategies incorporating collaboration among numerous agents, viewpoints, data sources, and so on is known as collaborative filtering. Collaborative filtering is often used with very big data sets. Sensing and monitoring data, such as in mineral exploration, environmental sensing over large areas or multiple sensors; financial data, such as financial service institutions that integrate many financial sources; or electronic commerce and web applications where the focus is on user data, etc. have all been subjected to collaborative filtering methods. The rest of this section concentrates on collaborative filtering for user data, while some of the methods and approaches might be applied to other significant applications as well.

****

Mainly, there are two approaches used in collaborative filtering stated below;

1. **User-based nearest-neighbor collaborative filtering**

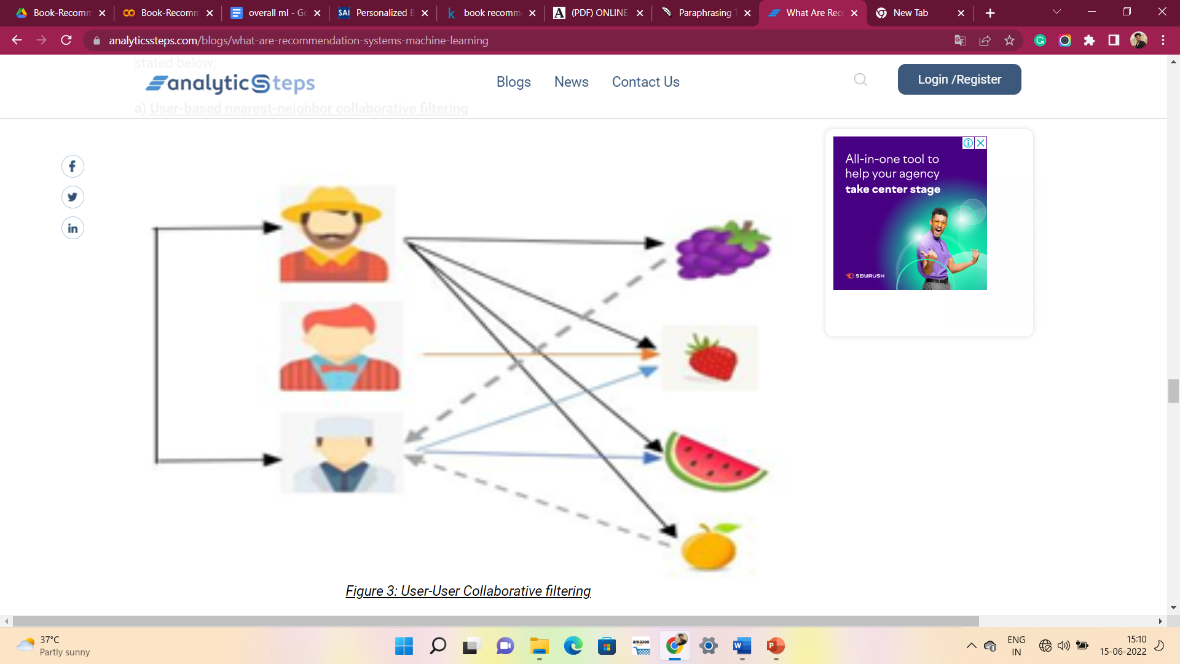
****

Figure shows user-user collaborative filtering where there are three users A, B and C respectively and their interest in fruit. The system finds out the users who have the same sort of taste of purchasing products and similarity between users is computed based upon the purchase behavior. User A and User C are similar because they have purchased similar products.

1. **Item-based nearest-neighbor collaborative filtering**

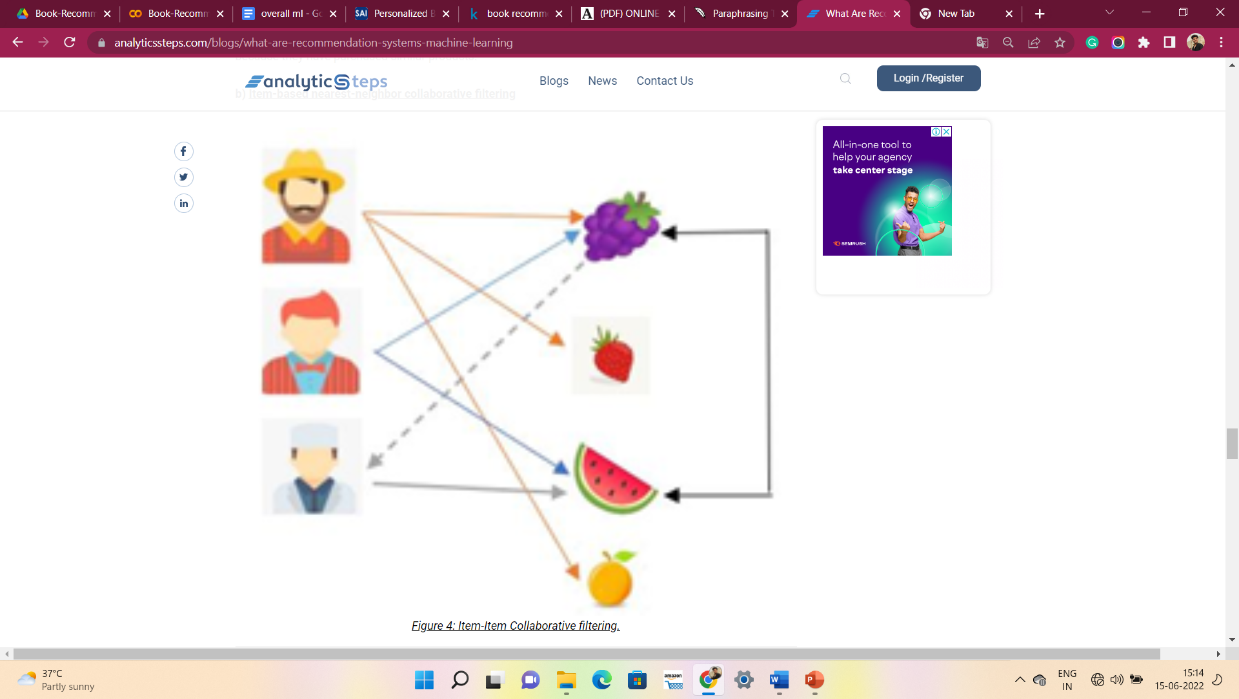
****

Figure shows user X, Y, and Z respectively. The system checks the items that are similar to the items the user bought. The similarity between different items is computed based on the items and not the users for the prediction. Users X and Y both purchased items A and B so they are found to have similar tastes.

**Limitations**

* + Enough users required to find a match. To overcome such cold start problems, often hybrid approaches are made use of between CF and content-based matching.
  + The problem in recommending items to the user due to sparsity problem.

**8. Conclusion:**

The escalating demands of Online Data have led to the invention of new techniques for presenting or rather recommending different products to the users. This project uses Popularity based as well as Collaborative filtering in order to recommend different types of books to the users. Both these techniques get rid of the Data sparsity and Cold Start Problem. Finally, the calculations of both the algorithms give accurate results.

**References-**

1. MachineLearningMastery
2. GeeksforGeeks
3. Analytics Vidhya