Book Recommendation System

# Abstract:

During a last couple of decades, with the upward push of Youtube, Amazon, Netflix, and many different internet services, recommender systems have taken a very important place in our lives.

In a totally trendy way, recommender systems are algorithms aimed toward suggesting applicable objects to users (items being films to watch, textual content to read, merchandise to buy, or whatever else relying on industries).

# Problem Statement

Recommender systems are really critical in some industries as they can generate a huge amount of income when they are efficient or also be a way to stand out significantly from competitors.

The main objective is to create a machine learning model to recommend relevant books to users based on popularity and user interests.

# Introduction

# Recommender systems are truly important in a few industries as they could generate a massive quantity of profits whilst they may be efficient or additionally be a way to stand out notably from competitors.

From e-commerce (propose to consumers articles that might hobby them) to online advertisement (propose to customers the proper contents, matching their preferences), recommender systems are these days unavoidable in our everyday online journey.

**Overview of data**

* Books Dataset

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

* Users Dataset

Contains the users. Note that user IDs (User-ID) have been anonymized and map to integers. Demographic data is provided (Location, Age) if available. Otherwise, these fields contain NULL-values.

* Ratings

Contains the book rating information. Ratings (Book-Rating) are either explicit, expressed on a scale from 1-10 (higher values denoting higher appreciation), or implicit, expressed by 0.

**Steps Involved**

1. **Exploratory Data Analysis**

The number one purpose of EDA is to guide the analysis of facts previous to making any conclusions. It may also useful resource withinside the detection of obvious errors, in addition to a deeper knowledge of facts patterns, the detection of outliers or anomalous events, and the discovery. Followings are the observations from EDA.

* Agatha Christie is the Top authors with highest numbers of book
* Harlequin is the Top Publisher with highest numbers of book
* USA is the Top Country with highest numbers of users
* There are some outlier in the age column. majority of the users are of age 20-35
* Maximum of book have good rating.8 is the most common rating for most number of book

## Feature Engineering

## From the analysis part we get that the Year-Of-Publication was wrongly mentioned for some of the rows.

## Diving deep into the Books dataframe we got to know that for these rows there was actually a column mismatch.

## Recommender System

For building recommendation system, we used three algorithms:

* **Popularity Based Recommendation**
* **Collaborative Based Filtering Using Similarity Score**
* **Collaborative Based Filtering singular value decomposition**

1. **Model Evaluation**

* **For model evaluation we use Mean Average Recall at k for SVD**

**Algorithms:**

### Popularity Based Recommendation:

It is a sort of recommendation system which fits at the precept of popularity and or something that's in fashion. These structures test approximately the books that are in trend or are maximum famous many of the customers and without delay recommend them. For example, if a product is regularly bought with the aid of using maximum people then the system gets to recognizes that that product is maximum famous so for each new consumer who simply signed it, the system will recommend that product to that consumer additionally and probabilities become excessive that the brand new consumer will even buy that. Using this recognition metric we are able to calculate the top books that might be recommended to a consumer.

**In this project we consider book with more than 200 reviews .**

**Our Popularity score formula-**

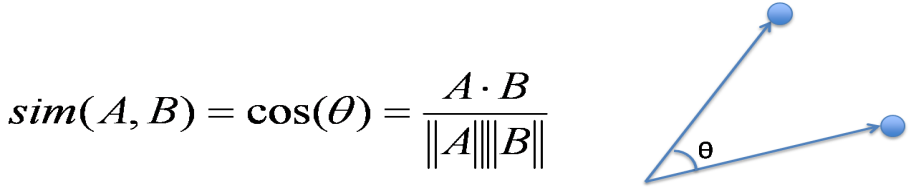
**popularity\_score=0.7\*Avg\_rating+0.3\*Rating\_count**

* **Avg\_rating=Average rating of the book**
* **Rating\_count=Number of rating of the book**

1. **Collaborative Based Filtering:**

In Collaborative Filtering, we generally tend to discover comparable customers and advise what similar customers like. In this sort of recommendation system, we don’t use the features of the item to advise it, as an alternative we classify the customers into clusters of comparable types and recommend every person consistent with the choice of its cluster.

* **Cosine similarity:** Similarity degree refers to distance with dimensions representing features of the data object, in a dataset. If this distance is less, there might be an excessive degree of similarity, however whilst the space is large, there might be a low degree of similarity. In this assignment, I actually have used ‘Cosine Similarity’ that is a famous similarity degree. Cosine similarity is a metric, beneficial in determining, how comparable the information items are regardless of their size. In cosine similarity, information items in a dataset are dealt with as a vector. The formula for cosine similarity-

****

* **singular value decomposition:** The Singular Value Decomposition (SVD) of a matrix is a factorization of that matrix into 3 matrices. It has a few interesting algebraic properties and conveys vital geometrical and theoretical insights approximately linear transformations. It additionally has a few vital packages in data science. In this article, I will attempt to give an explanation for the mathematical intuition at the back of SVD and its geometrical meaning.

#### ****Mathematics behind SVD****

The SVD of  mxn matrix A is given by the formula :

e.JPG

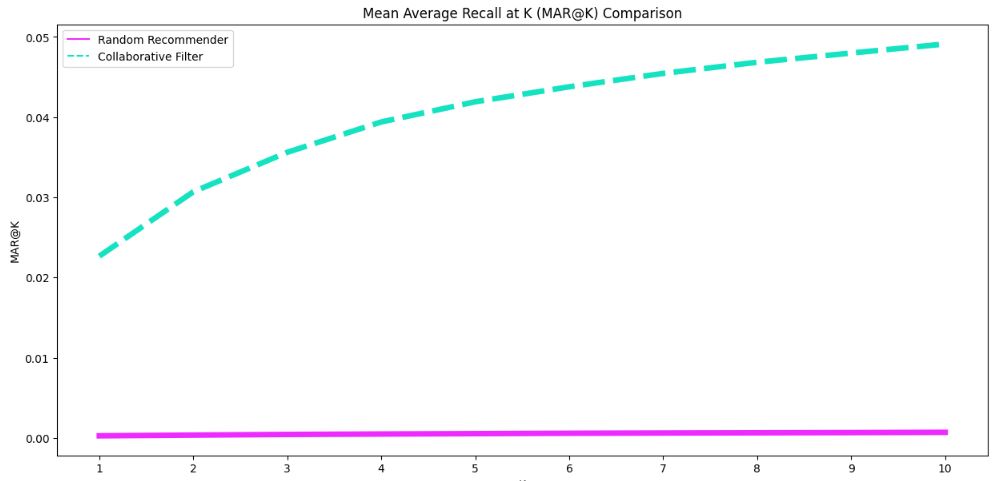
U:  *mxn* matrix of the orthonormal eigenvectors of .

VT: transpose of a *nxn* matrix containing the orthonormal eigenvectors of A^{T}A.

W:  a *nxn* diagonal matrix of the singular values

1. **Evaluation of SVD model using Recall @k**

T MAP@K offers perception into how relevant the listing of recommended items are, while MAR@K offers perception into how nicely the recommender is capable of recoll all of the items the person has rated definitely withinside the test set.



**Conclusion:**

* Majority of the readers were of the age bracket 20–35 and most of them came from North American and European countries namely USA, Canada, UK, Germany and Spain.
* Author of most of the books was Agatha Christie, William Shakespeare, Stephen King
* 8 is the most common rating for most number of book. Rating below 5 are in very few in number.
* MAR@K gives that our SVD recommender is able to recall much more then random Recommender.