Capstone Project Submission

Name, Email and Contribution:

1) **Krushnagopal Brahma**
E-mail: krbrahma54@gmail.com

- Data Preprocessing
- Data analysis.
- Data Visualization
- Feature Engineering
- Frame work of project.
- Model building.
- PPT presentation
- Summary Templet
- Techinical Documentation

Problem definition:

During the last few decades, with the rise of Youtube, Amazon, Netflix, and many other such web services, recommender systems have taken more and more place in our lives. From e-commerce (suggest to buyers articles that could interest them) to online advertisement (suggest to users the right contents, matching their preferences), recommender systems are today unavoidable in our daily online journeys. In a very general way, recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy, or anything else depending on industries). Recommender systems are really critical in some industries as they can generate a huge amount of income when they are efficient or also be a way to stand out significantly from competitors. The main objective is to create a book recommendation system for users.

EDA on given Data set :
Digging into data we understand that

1. The Book-Crossing dataset comprises 3 files.

● Users 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.

● Books 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 the case of several authors, only the first is provided. URLs linking to cover images are also given, appearing in three different flavors (Image-URL-S, Image-URL-M, Image-URL-L), i.e., small,

medium, large. These URLs point to the Amazon website.

● 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 .

Book Recommendation System :

- 1. Popularity Based (Books reviewed by most of the user)

- 2.Popular books in country wise

- 3. Books popular yr wise

- 4. Popular book of same_author

- 5.Popular Book of same Publsher

- 6.Top most rating books.

- 7. Collaborative Filtering (User-Item Filtering)

- 8.Content Based filtering

Conclusion :

In EDA, the Top-10 most rated books were essentially novels. Books like The Lovely Bone and The Secret Life of Bees were very well perceived.

1.Majority of the readers were of the age bracket 30-40 and most of them came from North American and European countries namely USA, Canada, UK, Germany and Spain. Among them USA have highest no book.

2.If we look at the ratings distribution, most of the books have high ratings with maximum books being rated 8. Ratings below 5 are few in number.

3.In book Wild Animus user gives a high no of rating which is 2.5k its so high in compare to other books.

4.Author with the most books was Agatha Christie, William Shakespeare and Stephen King.

5.For modelling, it was observed that for model based collaborative filtering and content based filtering doing best .

6.In other aproach the avg rating The Return of the King and hary poter have most liked books .

Please paste the GitHub Repo link.

Github Link:- https://github.com/brahma102/Book-Recommendation-System.git

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)