Book Recommendation System

ABSTRACT: The main objective of this project is to create a book recommendation system for users. Recommender systems play an important role in some industries as they can generate a huge amount of income when they are used effectively or also be a way to stand out significantly from competitors.

My project includes:

- Data Visualization Machine Learning
- Technologies- Python Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- Data which comprises of three csv file
- 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.
- Ratings: Contains the book rating information. Ratings are either explicit, expressed on a scale from 1-10 (higher values denoting higher appreciation), or implicit, expressed by 0.
- Users: User IDs have been anonymized and mapped to integers. Demographic data is provided (Location, Age) if available else these fields contain NULL values.

Description of the Project:

- Performed exploratory data analysis (EDA) on numerical and categorical data.
- Data Cleaning - Missing value handling and outlies identification
- Feature Selection - Used User-ID, ISBN and Books-Rating for model development.
- Model development - Tried Popularity based model and Collaborative filtering (Model based).
- Data exploration - data processing/cleaning recommendation system developer.

During the last few decades, with the rise of YouTube, Amazon, Ebay, Netflix and many other such web services, recommender systems have taken 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.

Python code implementation

1. After importing the files as data frames, the data has to be cleaned for anomalies.
2. As a part of preprocessing, I created a function to find the missing values in each data frame- Users, Books and Ratings. Then the NULL values are taken care off.
3. All the detailed notes are available in the python file in the form of comments.

At the end of the code we are able to suggest books to the customers based on customer satisfaction rate and based on their personal preference.

We build our model based on K nearest Neighbour. This is a classification algorithm and is non linear.

We used popularity-based filtering to filter the data. Popularity based recommendation system works with the trend. It basically uses the items which are in trend right now.

For example, if any book which is usually bought by every new user then there are chances that it may suggest that book to the user who just signed up.

Book weighted avg formula:

WR = [vR/(v+m)]+[mC/(v+m)]

where,

WR -> Weighted Rating

v -> the number of votes for the books;

m -> the minimum votes required to be listed in the chart;

R -> the average rating of the book; and

C -> the mean vote across the whole report.

Then determined the values of v,m,R,C.

Model based recommender system

Recommender system is used to predict user preference for a set of items based on the past experience. Two the most popular approaches are Content-Based and Collaborative Filtering. Collaborative filtering is a technique used by websites like Amazon, YouTube, and Netflix. It filters out items that a user might like on the basis of reactions of similar users. There are two categories of collaborative filtering algorithms: memory based and model based.

Model based approach involves building machine learning algorithms to predict user's ratings. They involve dimensionality reduction methods that reduce high dimensional matrix containing abundant number of missing values with a much smaller matrix in lower-dimensional space. The goal of this section is to compare SVD and NMF algorithms, try different configurations of parameters and explore obtained results.