Capstone Project Submission

Instructions:
i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Contribution Role :

1. Nikunj Sonule (nikunj.sonule10@gmail.com)

    1. Data Preparation (Data Cleaning and Feature Engineering)

    2. Exploratory Data Analysis

        - Univariate Analysis on numeric and categorical features
        - Analysis from categorical variables

    3. Got top ten books as per ratings

    4. Collaborative Filtering

        - Memory Based approach - Cosine Similarity
        - Model Based approach – Singular Value Decomposition (SVD)

    5. Model Evaluation Metrics

        - RMSE score with each collaborative filtering approach

    6. Getting top recommendation of books and ratings

Please paste the GitHub Repo link.

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

---

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

A book recommendation system is a type of recommendation system where we have to recommend similar books to the reader based on his interest. We've all record and data with three different dataset – Book dataset (ISBN, Book-Title, Book-Author, Year-Of-Publication, Publisher, Image-URL-S, Image-URL-M, Image-URL-L); Users dataset (User-ID, Location, Age); Ratings dataset (User-ID, ISBN, Book-Rating). Providing specific data analysis and prediction to done with this data. The main objective is to build a predictive recommender model, which could help in predicting – how we can predict the best recommendation for users according to their items approach. This would help us in providing better recommendation item to a right specific user.

As the first step, I performed data preparation (i.e data cleaning and feature engineering) and EDA part. The book_data, users_data and ratings_data dataset was an important dataset to gets smaller insights from its. Analysis feature like book_title, book_author, publisher, year_of_publication, age, book_rating were important to get some insights. Some of the null values were present in feature data, we replaced with mean of that particular feature. Deal with mismatch feature like book_title, book_author, year_of_publication, publisher. Only considering age between 5-90 we took users data to analysis and perform recommendation on it.

After data preparation, building recommendation system based on popularity (i.e ratings). These recommendations are usually given to every user irrespective of personal characterization. Merged book_data dataset and ratings_explicit. Considering ISBNs that were explicitly rated for this recommendation system.

The third step was to do Collaborative Filtering - Memory based approach was our first trial on train and test dataset which uses the memory of previous users interactions to compute users similarities based on items they've interacted (i.e user-based approach) or compute items similarities based on the users that have interacted with them (i.e item-based approach). Applying cosine similarity to make item-item similarity need to take transpose of matrix. This matrix would help in manage train-test matrix. After all views predictions based on similarity, we find recommendation on it based on score. Model-based collaborative filtering algorithms provide item recommendation by first developing a model of user ratings. We use Latent Factor Model called Singular Value Decomposition (SVD). SVD made dimensionality reduction technique in machine learning. SVD is a matrix factorization technique. It made us reduces the number of features of a dataset by reducing the space dimension from N-dimension to K-dimension (where K<N). SVD uses a matrix structure where each row represents a user, and each column represents an item. The elements of this matrix are the ratings that are given to items by users.

Model evaluation metrics is important to distinguish the best collaborative filtering – either by memory based or model based approach. The memory based approach – Cosine Similarity shows RMSE score for item based CF is 8.00 and for user based CF it shows 8.00. The score is slightly similar. Model based collaborative filtering made it better score with Latent Factor Model called SVD. The score improved to 1.63 for both SVD RMSE and accuracy score.

The final step was to get recommendation, Collaborative Filtering- Model based approach was among best method to approach recommendation system for this project. So SVD with RMSE score is the best model with 1.63 for this dataset.
That's how I have accomplished my project work in Book Recommendation System.

Drive link :- https://drive.google.com/drive/folders/1gsYarA7WfNL4_FS0xDdWQsi1BS3RE3-h?usp=sharing