Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1). Kunal Gawande

E-mail: gkunal8019@gmail.com

- Data preprocessing
- Perform story generation from visualization
- Collaborative Filtering (User-Item Filtering)
- Collaborative Filtering (Correlation Based)
- Collaborative Filtering (Nearest Neighbor's Based)

2). Bipasha Zade

- Model Deployment
- Feature Importance
- Shapley Additive explanations

3). Deepali Mahajan

E-mail: deepali2062@gmail.com

- Debugging Error
- Data Sorting
- Technical Documentation
- ppt Presentation
- Approach Towards Plan
- Seaborn, matplotlib
- Heatmap
- Linear Model Selection
- Evaluation Matrix

4). Chinmay Rojatkar

E-mail: chinmayrojatkar4@gmail.com

- Data Sorting
- Matplotlib
- ppt Presentation
- Data Visualization
- Technical Documentation
- Approach toward Plan
- Line Plot, Bar plot , Histogram
- Heatmap
- Linear Model Selection
- Data Preparation

5). Nikhil Aggarwal

E-mail: nickagg30899@gmail.com

- Data Cleaning
- Data Analysis
- Error Handling

Please paste the GitHub Repo link.

Kunal Gawande Link:- https://github.com/gkunal8019

Chinmay Rojatkar Link:- https://github.com/ChinmayRojatkar

Bipasha Zade Link:- https://github.com/Bipashazade

Nikhil Aggarwal Link:- https://github.com/Nikhil8815

Deepali Mahajan Link::- https://github.com/deepali2062

Introduction:

Introduction Now-a-days, online rating and reviews are playing an important role in books sales. Readers were buying books depending on the reviews and ratings by the others. Recommender system focuses on the reviews and ratings by the others and filters books. In this paper, the Hybrid recommender system is used to boost our recommendations. The technique used by recommender systems is Collaborative filtering. This technique filters information by collecting data from other users. Collaborative filtering systems apply the similarity index-based technique. The ratings of those items by the users who have rated both items determine the similarity of the items. The similarity of users is determined by the similarity of the ratings given by the users to an item. Content-based filtering uses the description of the items and gives recommendations which are similar to the description of the items. With these two filtering systems, books are recommended not only based on the user's behavior but also with the content of the books. So, our recommendation system recommends books to the new users also. In this recommender system, books are recommended based on collaborative filtering technique and similar books are shown using content based filtering. The required dataset for the training and testing of our model is downloaded from Good Reads website. Matrix Factorization techniques such as Truncated-SVD which takes a sparse matrix of dataset is used for reduction of features. The reduced dataset is used for clustering to build a recommendation system. Clustering is a collaborative filtering technique that is used to build our recommendation system in which data points are grouped into clusters. . In this paper, we used two methods i.e., K-means and Gaussian mixture for clustering the users. The better model is selected based on the silhouette score and used for clustering. Silhouette score or silhouette coefficient is used to calculate how good the clustering is done. Negative value shows that clustering is imperfect whereas positive value shows that clustering was done perfectly. Difference between the mean rating before clustering and after clustering is calculated. Root Mean square Error is used to measure the error between the absolute 2 values and obtained values. That RMSE value is used to find the fundamental accuracy.

Please paste the GitHub Repo link.

GitHub Link:-

https://github.com/gkunal8019/Predicting-sentiment-of-COVID-19-tweets

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

Problem Statement:

During the last few decades, with the rise of YouTube, Amazon, Netflix, and many other such web services, recommender systems have become much more important in our lives in terms of providing highly personalized and relevant content.

The main objective is to create a recommendation system to recommend relevant books to users based on popularity and user interests

Conclusion:

In EDA, the Top-10 most rated books were essentially novels. Books like The Wild Animus and The Lovely Bones: A Novel .
Majority of the readers were of the age bracket 20-50 and most of them came from North American and European countries namely USA, Canada, UK.
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.
Author with the most books was Stephen King, Nora Roberts and James Patterson.
A recommendation system helps an organization to create loyal customers. The recommendation system today are very powerful that they can handle the new customer too who has visited the site for the first time. They recommend the products which are currently trending or highly rated and they can also recommend the products which bring maximum profit to the company.