**Capstone Project Technical Documentation:**

|  |
| --- |
| **Name, Email and Contribution:** |
| 1. Abhishek Kumar Mishra ([Abhishekkumarmishra364@gmail.com](mailto:Abhishekkumarmishra364@gmail.com))  * Objective * Problem Statement * Data Summary * Data Cleaning * Imputing missing value * Difference recommendation system * Challenge * Conclusion |
| **GitHub Repo link.** |
| Github Link:- https://github.com/Abhishek709mis/Book-recommadation-System  Drive Link:- |
| * **Writing a short summary of your Capstone project and its components.** |
| * During the couple of year, with the rise of Youtube, Amazon, Netflix etc., using of recommender systems. In e-commerce feild we suggest (to buyers articles that could interest them) to online advertisement (suggest to users the right contents, matching their preferences), recommender systems are today use in our daily online journeys. * In simple way we can ellaborate, recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read or anything depending on industries). * Recommender systems are really critical in some industries as they can generate a huge amount of income.   **Importing Libraries:**  Some of the libraries like Num-Py for numerical operations, Pandas for data manipulation, matplotlib and seaborn for data visualization were loaded. In addition to these libraries, pycountry library was installed and loaded.  **Reading Data:**  After drive was mounted, data from csv file was read and store in a pandas dataframe.  Data set have three file :   * + - User data set ,     - Book data set and     - Rating data set. * User data set contain :   + - User Id ,     - Location,     - Age * Book data set contain :   + - ISBN,     - Book tittle,     - Book Author ,     - Year of Publication ,     - Publiser      * Rating data set contain :   + - User Id,     - ISBN   **Data Inspection:**  Loading the file of the book data ,user data and rating data.And then exploring the book data set, user data set and rating data set sepreatetly.  **Data Cleaning:**   * Age column has 39.72% missing values   **SYSTEM FLOWCHART:**    **Exploratory Data Analysis:**  **age.png age3.png**    Max age distribution from 20-40  **authors1.png**  Top 10 author  **Publiser1.png**  Top 10 Publishers  **booking.png**  Book number 8 have most rating   **MODELLING:** Popularity Based Recommender.  Content Based Recommendation.  Collaborative Filtering Based Recommendation. **Popularity Based Filtering:** Popularity based recommendation system works basically uses the items which are in trend right now.  Book weighted avg formula:  Weighted Rating(WR)=[vR/(v+m)]+[mC/(v+m)] where,  **v** : is the number of votes for the books;  **m** : is the minimum votes required to be listed in the chart;  **R**  : is the average rating of the book;  and  **C** : is the mean vote across the whole report.  find the values of **v,m,R,C.** **Model Based Collaborative Filtering Recommender:** The recommender system is to predict user preference for a set of items based on the past experience.  Two the most popular approaches are and   1. Content-Based 2. Collaborative Filtering.   Collaborative filtering is a technique used by websites YouTube, Netflix , etc. 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.  Compare SVD and NMF algorithms, try different configurations of parameters and explore obtained results. **Collaborative Filtering based Recommendation System:** Collaborative Filtering is a technique or a method to predict a user’s taste and find the items that a user might prefer on the basis of information collected from various other users having similar tastes or preferences. It takes into consideration the basic fact that if person X and person Y have a certain reaction for some items then they might have the same opinion for other items too.  The two most popular forms of collaborative filtering are:   * **User Based:**Here, we look for the users who have rated various items in the same way and then find the rating of the missing item with the help of these users. * **Item Based:** Here, we explore the relationship between the pair of items (the user who bought Y, also bought Z). We find the missing rating with the help of the ratings given to the other items by the user.  |  |  |  |  |  | | --- | --- | --- | --- | --- | | Model | test\_rmse | test\_mae | fit\_time | test\_time |  |  |  |  |  |  | | --- | --- | --- | --- | --- | | SVD | 1.611699 | 1.24824 | 8.867862 | 8.867862 | | NMF | 2.614838 | 2.226703 | 7.616849 | 0.812754 |  Conclusion  * In EDA, the Top-10 most rated books. * 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. * Ratings distribution, most of the books have high ratings with maximum books being rated 8. * Ratings below 5 are few in number. * The Modelling, it was observed that for model based collaborative filtering SVD technique worked way better than NMF. * Amongst the memory based approach, item-item CF performed better than user-user CF because of lower computation .   . |