**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Rohit Kumar Sharma [rohitraw6@gmail.com](mailto:rohitraw6@gmail.com) |
| **Please paste the GitHub Repo link.** |
| GitHub Link: - https://github.com/Rohitkumar1011/Book-Recommender-system.ipynb |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Summary**  Finding relevant books from a huge e-book space has become a tremendous difficulty for internet users as even the volume of online books grows tremendously as a result of the COVID-19 epidemic. A recommender system will help users who do not have enough individual knowledge to pursue through the different types of options offered by a website. It will provide the users with information to assist them to make a decision as to which items to purchase  This project has three datasets which is extracted from some book selling websites. They are Books – This file contains all the information related to books like an author, title, publication year etc. Users – This file contains registered user’s information like user-id, location. And the file contains Ratings given by users to all the books. Combination of all the dataset is very large and it took bit time to load.  At the start of project, I have imported all the necessary modules including Data visualization modules like Matplotlib, Seaborn. Data understanding at very start with the help of head and tail give us a fair idea about the type of data.  There were lots of missing values especially in user’s dataset, which I have handled by dropping them. Dataset also contain duplicate values which I have dropped in order to improve accuracy.  After that I have done EDA by comparing various variables of books and user’s dataset. Exploratory data analysis is the crucial part of project because it helps in visualizing relation between all the independent and dependent variables. I have plotted book author, book title, Years of publication, books published in a year with the help of graph.  After that I have made popularity-based Recommender system, it’s a form of recommendation system that works on the basis of popularity or anything that's currently popular. In this I have tried to find out most popular books in the dataset with the help of some logic and algorithms. This algorithm can give us top 50 or top 100 books of entire datasets.  After that I have made Collaborative filtering-based Recommender System, here I have made pivot table with users’ columns, books index and ratings values. In this technique basically we are calculating Euclidean distance between two points. Euclidean distance tells how much close values are. I have imported similarity score which categorise similar points. It helps to recommend similar books.  After making a function and applying technique I have checked whether this technique is recommending books or not. I get all the books as a recommendation which are really similar to the book whose recommendation is required.  Finally, I come to the conclusion that both the technique is giving good results with accuracy. |