Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Shubham Srivastava: **E-Mail-** shubham.mach30@gmail.com • Loading Dataset • Data Inspection • Exploratory Data Analysis • Imputing Missing Values • Evaluation using different Recommendation models • Challenges • Conclusion
Please paste the GitHub Repo link.
Github Link:- https://github.com/Shubhamverse/Book-Recommendation-System__Unsupervised-ML **Drive Link:-** https://drive.google.com/drive/folders/1hpxPDbFLnQ1gu-DCNbCLg362-8zhjeA5?usp=sharing
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

In this Unsupervised Learning Project, we were provided with datasets i.e. User_dataset , Books_dataset , Ratings_dataset.

First, we analyze Users_dataset and we found outliers , null values in Age and we imputed it using median value of Age.
Secondly, we analyze Books_dataset and we found errors in year of publication, book title, book author columns and imputed Nan with 'other'.
Thirdly, we analyze Ratings_dataset and we did rating counts based on that we found top rated books.
Then, we merged user and rating dataset on userID and then we merged it with books dataset on ISBN making it Final_dataset.

Next, through Popularity Based Filtering we found Top 20 books where as a cutoff we used 90 percentile. After that we apply Model based collaborative filtering in which we SVD and NMF for dimensionality reduction and we found SVD works better. Then we used Item -Item based filtering to get recommended books. And we also used user Item filtering to get recommended books for user.

Finally, we derive conclusion based on results shown through recommendation lists we obtained is that:
In EDA, the Top-10 most rated books were essentially novels. Books like The Lovely Bone and The Secret Life of Bees were very well perceived.
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.
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 Agatha Christie, William Shakespeare and Stephen King.
For modelling, it was observed that for model based collaborative filtering SVD technique worked way better than NMF with lower Mean Absolute Error (MAE) .
Amongst the memory based approach, item-item CF performed better than user-item CF because of lower computation .