

Project Progress Report – CS410, Fall 2022

Name: Chenfeng Chu; NetID: chu61

Introduction

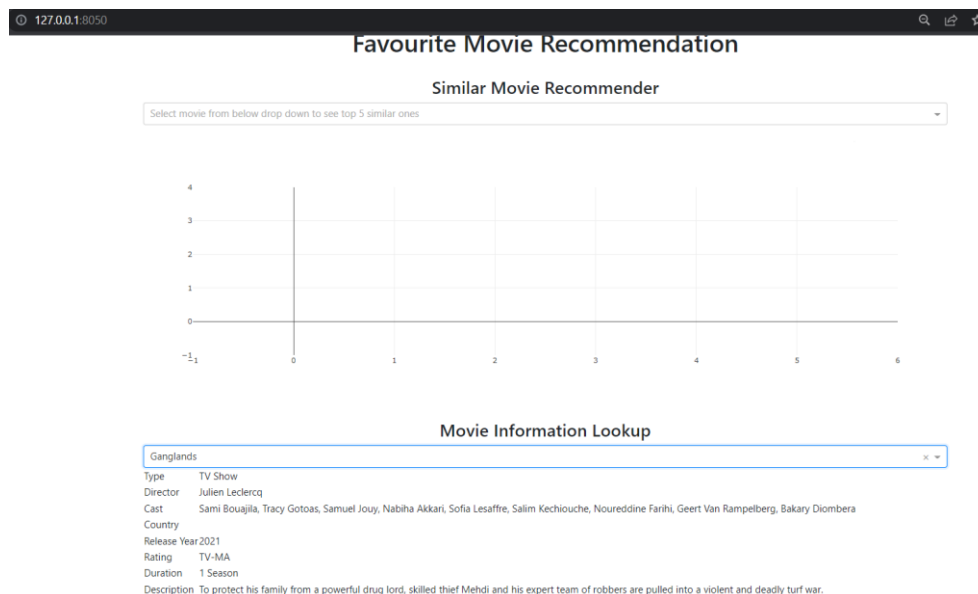
In the project proposal, I have mentioned the following high-level tasks and quick status for each is indicated below.

- Ingest, understand and pre-process dataset – 5 hours **[Completed]**
- Develop effective algorithm to calculate similarity score between two movies – 10 hours **[In progress]**
- Build front end application layout and config application logic to take in user input and return model output – 10 hours **[front end layout is completed; application logic is in progress]**
- Conduct test and debugging – 5 hours **[incorporated into above two tasks]**

In this report, I will cover progress made so far, pending tasks and expected challenges.

Progress made

I have completed data ingestion and explored the dataset to understand what is included and associated data quality issues. I have started implementing the BM25 ranking scoring which will generate the top 5 similar movies based on descriptions of selected one. For the application front end, I have developed the wire frame and completed partial application logic such as fetching basic movie information when selected. Screenshot of current User Interface can be found below:



Pending Tasks

The following tasks are still pending:

- Complete BM25 implementation and incorporate it into the application logic so it will return top 5 similar movies based on description of selected one along with scores

- Complete end to end application logic coding
- Refine User Interface design
- Test, develop final report and demo

Challenges

There are no particular challenges at current stage and I expect to complete the final project before the deadline.