
Capstone Project:

Recommendations Models for Films

Management Report Prepared by:
Win K. Phyo
MS Applied Data Science Candidate '22
Email: wkpwinphyo@gmail.com, win.phyo@mymona.uwi.edu

For:
COMP 6830 - Data Science Capstone Project II
Professor Ricardo Anderson
February 27, 2022

CONTENTS

EXECUTIVE SUMMARY	3
BUSINESS CASE	4
ROADMAP AND CHALLENGES	5
MODELS AND DEPLOYMENT	6
NEXT STEPS	7
REFERENCES	8

EXECUTIVE SUMMARY

We have managed to successfully launch a prototype v1 version of the **Movie Recommendations App** (previously known as **Project Iris**). The initial objective of this project was to build a film recommendations engine that utilized colour and emotion or mood as primary features. Due to lack of available data, however, the colour-based aspects of the model had to be scrapped. The project instead shifted focus towards building out a robust recommendations system using user feedback/preferences data (in the way of Netflix), and also an emotion classification model that can output the predicted emotion of a movie based on that title's description.

The screenshot shows a web-based application interface. At the top, there is a search bar with the text "Harry Potter and the Sorcerer's Stone (a.k.a. Harry Potter and the Philosopher's Stone) (2001)". Below the search bar is a button labeled "Show me recommendations based on what other people are watching 😊". A horizontal scroll bar is visible above a row of movie posters. The posters for "Harry Potter and the Chamber of Secrets", "Harry Potter and the Prisoner of Azkaban", "Harry Potter and the Goblet of Fire", "The Lord of the Rings: The Fellowship of the Ring", and "Harry Potter and the Order of the Phoenix" are displayed. Each poster has a brief description below it. The descriptions are as follows:

- Harry Potter and the Chamber of Secrets:** Cars fly, trees fight back, and a mysterious house-elf comes to warn Harry Potter at the start of his second year at Hogwarts. Adventure and danger await when bloody writing on a wall announces: The Chamber Of Secrets Has Been Opened. To save Hogwarts will require all of Harry, Ron and Hermione's magical abilities and courage. Emotion: anger
- Harry Potter and the Prisoner of Azkaban:** Year three at Hogwarts means new fun and challenges as Harry learns the delicate art of approaching a Hippogriff, transforming shape-shifting Boggarts into hilarity and even turning back time. But the term also brings danger: soul-sucking Dementors hover over the school, an ally of the accused He-Who-Cannot-Be-Named lurks within the castle walls, and fearsome wizard Sirius Black escapes Azkaban. And Harry will confront them all. Emotion: fear
- Harry Potter and the Goblet of Fire:** When Harry Potter's name emerges from the Goblet of Fire, he becomes a competitor in a grueling battle for glory among three wizarding schools—the Triwizard Tournament. But since Harry never submitted his name for the Tournament, who did? Now Harry must confront a deadly dragon, fierce water demons and an enchanted maze only to find himself in the cruel grasp of He Who Must Not Be Named. Emotion: fear
- The Lord of the Rings: The Fellowship of the Ring:** Young hobbit Frodo Baggins, after inheriting a mysterious ring from his uncle Bilbo, must leave his home in order to keep it from falling into the hands of its evil creator. Along the way, a fellowship is formed to protect the ringbearer and make sure that the ring arrives at its final destination: Mt. Doom, the only place where it can be destroyed. Emotion: sadness
- Harry Potter and the Order of the Phoenix:** Returning for his fifth year of study at Hogwarts, Harry is stunned to find that his warnings about the return of Lord Voldemort have been ignored. Left with no choice, Harry takes matters into his own hands, training a small group of students to defend themselves against the dark arts. Emotion: fear

The app currently allows users to find 5 titles similar to any of the 10,000+ in our system by inputting the name of a movie. It also offers a feature for individual users to find titles that we believe they would enjoy most based on the preferences of similar users.

BUSINESS CASE

Content recommendations engine are integral to the products offered by every major content distribution or streaming service today, including Netflix, Spotify, YouTube etc. In the realm of movies, Netflix (\$NFLX), despite recent hits to its stock price, still boasts a market capitalization of around \$100B (albeit declining from \$174B in February, 2022). The ultimate goal of recommendations engine is to keep the user engaged in the product by continuously offering curated, intelligent suggestions for further media consumption based on that particular user's preferences or tastes in content. In today's modern subscription based economy, customer retention becomes especially important in order to maintain steady revenue streams.

Netflix employs a collaborative-based filtering model in its recommendation engine, whereby instead of prioritizing the similarities in content among different titles, they instead build a recommendations engine around what other, similar users tend to consume. In this, and many other ways, data science is core to the success and future growth of companies like Netflix. In fact, according to [levels.fyi](#), the median data scientist compensation package at Netflix is \$505K. To some extent, this signals the value and importance of data science and recommendations systems to the revenue stream of companies built around content.

Despite the relatively large selection of streaming services for movies nowadays, however, the aspect of emotion or mood evoked by a film has been largely ignored in building recommendations systems. It is arguable whether viewers outside of niche film communities would place such a heavy emphasis on the emotion evoked by a film, yet the feelings we experience from watching a movie form a core part of our memories of it - whether we're consciously aware of this or not.

This project aims to build a film recommendations app that presents the user with specific content suggestions based on a robust engine that incorporates both user preference/engagement data, and emotion data as derived from the plot descriptions or overviews of films. The ultimate goal is to provide a platform whether users can submit ratings and reviews for movies they've seen, and to continuously feed this data back into the models to develop a one-of-a-kind content recommendations engine.

ROADMAP AND CHALLENGES

The initial proposed schedule was as follows:

Deliverable	Description	Proposed Date
Build and deploy working pipeline to data warehouse	Build and maintain a pipeline to automatically create and update the necessary datasets in the data warehouse	June 1, 2022
Build colour model	Analyse image data and combine with other data to build model to find titles similar in colour information	July 31, 2022
Build mood model	Analyse data to build model to find titles similar in feeling or emotion	July 31, 2022
Deploy models as recommendations app	Release web app that offers users personalised movie suggestions based on colour, mood, etc.	August 10, 2022

Originally, we had considered using the Microsoft Azure platform to create, update, and manage the various data stores, train the machine learning models, and deploy the final app. As of the release of app version 1, the Azure Data Lake Gen 2 (used as a general purpose data warehouse; currently stores data used to build models) resource and the Azure Synapse workspace (used to manage data pipelines and train machine learning models) have been successfully provisioned. However, due to some level of unfamiliarity with the specific technologies associated with the Azure suite, it was decided that the app would initially be fully built and deployed on local machines.

The colour component of the recommendations engine was completely scrapped due to a lack of accessible data that could be used to form a repository of colour-related movie data. To revisit this feature would require independent analysis of film data using advanced processing techniques to extract colour-based information from films.

We have managed to train a working (and fairly performant) emotion classification model that is currently used to calculate the dominant emotion expressed by a film's overview/description text. The current implementation only presents the user with emotion information related to titles that were already suggested by the recommendations engine, instead of utilizing emotion as a core feature of the model. Future updates intend to build out a more robust content suggestions system that also incorporates this emotion data.

MODELS AND DEPLOYMENT

3 separate models were built for the recommendations system:

- ◆ Ratings-based model that produces suggested titles based on a given movie
- ◆ Ratings-based model that produces suggested titles based on what it thinks a particular user will rate films
- ◆ Emotion classification model that can be used to predict the dominant emotion communicated by texts

All 3 models are currently used by the app to produce recommendations and classify these by emotion. The app prompts users to either select a title from a dropdown list of over 10,000, or to view suggestions based on a given user ID. It then returns a total of 5 content recommendations, their overviews, and predicted dominant mood.

NEXT STEPS

Future updates will see further incorporation of emotion data into the recommendations process, as well as the launch of a registration and ratings submission system, where users can create a profile and submit ratings for movies they've seen. This data will be stored and fed back into the model in order to improve performance.

The Azure platform is intended to be revised as the app grows and scalability becomes important. The core resources necessary for this service have already been provisioned and loaded with some of the necessary data. There needs to be work done, however, to rewrite the current software to an appropriate format for deployment on Azure.

REFERENCES

The Colors of Motion - thecolorsofmotion.com

HappyCoding.io – happyCoding.io

Internet Movie Database (IMDb) – imdb.com

The Movie Database (TMDB) – themoviedb.org

GroupLens MovieLens 25M Dataset - <https://grouplens.org/datasets/movielens/25m/>

GitHub Repository - <https://github.com/wkphyo/capstone-project>

Levels.fyi - <https://www.levels.fyi>