CSC111 Project Report: ScreenSelect

Personalized Movie Recommendation System

Dogyu Lee, Narges Movahedian Nezhad, Sidharth Sawhney, Aastha Sharma

April 3rd, 2023

Part 1: Goal

Due to the advent of streaming services, which provide customers immediate access to a vast collection of information, the entertainment industry has undergone a vital change. Finding the right things to watch, for example, might just be more difficult than ever for viewers. This is due to the overflow of available content. As a solution to this issue, recommendation systems have been created; personalizing and recommending material based on users' viewing patterns and interests. With this in mind, **our project intends to develop a movie customization and recommendation app that takes inspiration from Netflix's recommendation system, leveraging watching patterns, user input, and clever algorithms to provide tailored movie suggestions.**

Our project intends to build movie customization that *varies* from that of others in regard to algorithm, despite the fact that the issue of content overload and the demand for tailored movie recommendations have been thoroughly examined. For instance, we will employ contextual data in addition to the users' viewing history to provide recommendations that are more customized. This goes beyond Netflix's simple recommendations of related content based only on user watch history. Additionally, our project adds a sense of individualism to every user. Unlike other recommendation platforms, we have deliberately chosen not to take into consideration the location of our consumers and users nearby because we understand that everyone has unique taste. Having geographically been near someone in no way guarantees similar preferences when it comes to movies. For this reason, we would go as far as to argue that taking location into account ruins the algorithm and displays less desirable content.

Moreover, ScreenSelect takes into account only the past 10 movies for every user, as opposed to all viewing history—which is the case with other projects in the market. This is largely due to the fact that people's tastes are likely to change over time. By doing so, we are ensuring that our recommendations are updated to appeal to consumers' most recent interests in contrast to the movies they enjoyed in the distant past. This strategy also ensures user happiness and increased engagement with the app. Also, by focusing on the viewer's most recent 10 films rather than their complete viewing history, we may prevent constantly recommending identical films, which can cause user boredom and disengagement. As the user's preferences change over time, our strategy utilizes both the user's viewing history and contextual information to provide recommendations that are more individualized. Overall, our project provides a fresh and original perspective on the issue of information overload. We strive to deliver a highly customized, enjoyable, and useful movie recommendation app to increase user engagement and commitment.

Our project's objective is to boost user engagement by means of providing a highly-customized watching experience that changes depending on each user. We will offer a platform that generates the best recommendations for each user every time by taking into account contextual elements—such as the last 10 watched movies, keywords, genre, and language—going beyond Netflix's straightforward suggestions of related material based on user watch history. We want to provide a highly engaging experience that keeps people riveted to the screen and coming back for more by putting a strong emphasis on tailored suggestions.

Our group has also made several considerations in regard to the ethics of ScreenSelect. We will put in place a number of privacy precautions to make sure that our project doesn't divulge any sensitive information. We never request personally identifiable information when gathering user data. The username we take from our users is the only sort of identification verification input we take from them. This username can never be connected, as it can be anything users choose. This will assist in preserving user privacy and preventing the disclosure of private information. We have also put in place a number of safeguards to make sure that our program doesn't make any incorrect recommendations to vulnerable populations. Our group now has a dedicated email address on the site that we utilize to ask users for feedback, especially from members of vulnerable groups. This will make it easier to spot any offensive ideas or information. The input will help our team improve the app's suggestions and make sure they are appropriate for all users. These precautions will be regularly audited to identify potential privacy and security breaches and ensure that our app complies with privacy regulations.

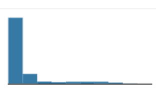
We have concentrated on developing a recommendation system that prioritizes user satisfaction and engagement over other metrics like watch time or income in order to make sure that our project's ideas are valuable to the user as opposed to the Netflix engine. When creating our project, we concentrated on giving user-centered design concepts top priority. Aspects such as reading accessibility (through large fonts) can be seen in all parts of our implementation. This required comprehending the user's requirements and preferences and tailoring the platform to satisfy them. We boosted user engagement and improved the app's suggestions as a result of user testing and feedback. In addition to other helpful measures, we provide users with a wide variety of content, including works that might not be as well-known but are nevertheless well-regarded and pertinent to their interests. This is one of the reasons for our decision to work with a large data set. Users will find fresh material and engage more as a result of this.

In conclusion, the goal of our project is to develop a highly individualized, entertaining, and impactful movie recommendation app. We seek to offer consumers a watching experience that is customized to their interests and keeps them hooked with our platform over time. This is done through machine learning algorithms and contextual elements. Our ultimate purpose is to enhance user engagement and commitment. We will gauge our progress using a variety of test cases that reflect this goal.

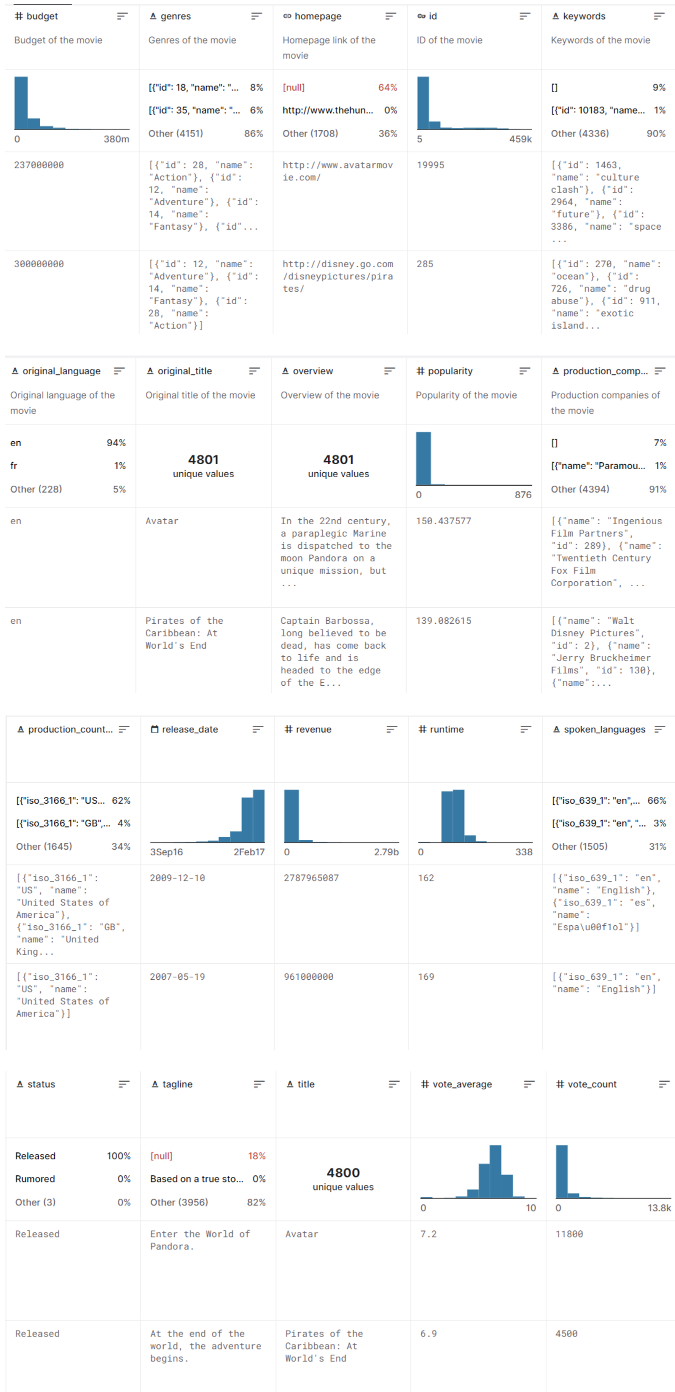
Part 2: Datasets

ScreenSelect used two datasets in the form of CSV files: tmdb_5000_movies.csv (Movies Dataset) and tmdb_5000_credits.csv (Credits Dataset). Both files were downloaded from Kaggle. The Movies Dataset consists of 20 columns with the information and statistics of 4803 movies. Out of the 20 columns in the Movies Dataset, ScreenSelect utilized a subset of nine columns: id, title, genres, keywords, runtime, original_language, overview, vote_average, and release_date. The Credits Dataset consists of four columns with the id, title, cast, and crew for each of 4803 movies. Out of the four columns in the Credits Dataset, ScreenSelect utilized a subset of two columns: id and crew.

CSV File 1

movie_id ID of the movie	title Name of the movie	cast Cast	crew Crew
 5 459k	4800 unique values	4761 unique values	4776 unique values
19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "credit_id": "5602a8a7c3a3685532001c9a", "gender": 2, "...	[{"credit_id": "52fe48009251416c750aca23", "department": "Editing", "gender": 0, "id": 1721, "job": ...
285	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Sparrow", "credit_id": "52fe4232c3a36847f800b50d", "gende...	[{"credit_id": "52fe4232c3a36847f800b579", "department": "Camera", "gender": 2, "id": 120, "job": "D...

CSV File 2



budget	genres	homepage	id	keywords
Budget of the movie	Genres of the movie	Homepage link of the movie	ID of the movie	Keywords of the movie
0 380m	[{"id": 18, "name": "... 8% [{"id": 35, "name": "... 6% Other (4151) 86%	[null] 64% http://www.thehun... 0% Other (1708) 36%	5 459k	[] 9% [{"id": 10183, "name... 1% Other (4336) 90%
237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id"...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": 2964, "name": "future"}, {"id": 3386, "name": "space ...
300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 28, "name": "Action"}]	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "drug abuse"}, {"id": 911, "name": "exotic island...

original_language	original_title	overview	popularity	production_comp...
Original language of the movie	Original title of the movie	Overview of the movie	Popularity of the movie	Production companies of the movie
en 94% fr 1% Other (228) 5%	4801 unique values	4801 unique values	0 876	[] 7% [{"name": "Paramou... 1% Other (4394) 91%
en	Avatar	In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but ...	150.437577	[{"name": "Ingenious Film Partners", "id": 289}, {"name": "Twentieth Century Fox Film Corporation", ...
en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, has come back to life and is headed to the edge of the E...	139.082615	[{"name": "Walt Disney Pictures", "id": 2}, {"name": "Jerry Bruckheimer Films", "id": 130}, {"name":...

production_count...	release_date	revenue	runtime	spoken_languages
[{"iso_3166_1": "US... 62% [{"iso_3166_1": "GB",... 4% Other (1645) 34%	3Sep16 2Feb17	0 2.79b	0 338	[{"iso_639_1": "en",... 66% [{"iso_639_1": "en", "... 3% Other (1505) 31%
[{"iso_3166_1": "US", "name": "United States of America"}, {"iso_3166_1": "GB", "name": "United King...	2009-12-10	2787965087	162	[{"iso_639_1": "en", "name": "English"}, {"iso_639_1": "es", "name": "Espa\u00f1ol"}]
[{"iso_3166_1": "US", "name": "United States of America"}]	2007-05-19	961000000	169	[{"iso_639_1": "en", "name": "English"}]

status	tagline	title	vote_average	vote_count
Released 100% Rumored 0% Other (3) 0%	[null] 18% Based on a true sto... 0% Other (3956) 82%	4800 unique values	0 10	0 13.8k
Released	Enter the World of Pandora.	Avatar	7.2	11800
Released	At the end of the world, the adventure begins.	Pirates of the Caribbean: At World's End	6.9	4500

Part 3: Computational Overview DOROTHY

Part 4: Instructions for Running ScreenSelect DOROTHY

Part 5: Changes to Project Plan DOROTHY

Classes

- The Graph class was modified with five new public methods **verify_vertex**, **add_user_vertex**, **retrieve_item_obj**, **retrieve_vertex_dict**, and **add_movie_vertex**. The preceding methods add user and/or movie vertices to the graph and give access to its vertex objects which are mapped in a private instance attribute dictionary if other functions need it.

- The _Vertex superclass has changed from a class to a data class for cleaner code and not having to add any other abstract methods to it. Its instance attributes have all been changed to have a default None value when initialized. One of its instance attributes **genre**'s type contract changed from str to Optional[set[str] | str] to allow a set of str for the _Movie class which has more than one genre in the Movies dataset. A public instance attribute **neighbours** was removed from the superclass and was added separately to the _Movie and _User subclasses. Representation invariants were added to ensure the instance attributes were non-empty if not the default value.
- _Movie class's private instance attribute **_total_score**'s type contract changed from an int to a dictionary to store users' names and preference-matching scores for itself. The dictionary allowed the system to efficiently access the mapped preference scores of multiple users by their usernames for a specific movie instead of having to compute the score for each user and movie all over again. Representation invariants were added to ensure the instance attributes were not empty strings or 0.
- A new public instance attribute **past_10_neighbours** was added to the _User class to keep track of the recently watched/recommended movies and use them (their scores) to recommend the next movie. Two new public methods **retrieve_top_scores** and **modify_preferences** were added to this class. These methods give access to its private instance attribute dictionary _top_scores if other functions need it and reassign some of its public instance attributes' values. Representation invariants were added to ensure the instance attributes were within the correct range of numbers.

Overall Program

- The plan to use a username and a password for the user login procedure changed to only using the username. Passwords were not used in the program except for the login procedure and each user object has a unique username as its key in the graph's _vertices dictionary. The role of passwords was not much important, so it was excluded from our program.
- The plan to display the recommended movies' posters on the user interface (screen/window) was dropped. It took longer than expected to retrieve the image from the website link and resize it to fit the UI screen.
- The plan to display the overviews of the recommended movies was dropped due to difficulties fitting different length str for each movie on a screen with other existing UI widgets in fixed positions.

Part 6: Discussion

Our goal in this project was to create a movie recommendation app that provides users with tailored recommendations based on their viewing habits and preferences. We understood that such platforms exist in the market, however, we were determined to offer the best suggestions with the highest accuracy for desire. This was done by disregarding our user's geographical location. In order to consistently provide the best suggestions for each user, we also employed a content-based filtering algorithm that takes into account contextual factors including the most recent 10 movies seen, keywords, genre, and language.

We are very proud to say that the results of our computational exploration helped us achieve our goal. Our computational investigation's findings (through user testing) demonstrated that our system was very effective at giving consumers personalized movie selections. Compared to traditional content-based filtering algorithms that simply take into consideration the qualities of each piece of content, our system was able to provide more customized suggestions. This is largely due to the fact that it takes contextual factors and most recently watched into account.

However, we did encounter some limitations during the development of our app. One limitation was the difficulty in loading the movie data into the graph due to the large amount of data and the different formats in which it was presented. For example, in order to access the genre for a movie, we had to access a dictionary value that was in a list which was a string. We used for loops to access each piece and get rid of unnecessary details, but we were not able to typecast the string into different dictionaries of lists. Eventually, we went with our method of normal stripping of brackets, etc. in order to overcome this obstacle.

Another limitation was the difficulty in adding images to the graphic interface of each movie when displaying its title. As the GUI was quite small, we were not able to see how to add images without messing up the layout. Therefore, we chose to take this out and implement a clean version in the future by changing the GUI from the layout import that we have currently used to a fully custom-made file that supports images as well. We believe that this is an achievable goal to set for the future; and it would enhance our current UI even more.

In terms of the next steps for further exploration, we plan on adding strong security to our app by implementing encryption using concepts we learned in CSC110. We will also ask users for their ages and sort movies by age group so that children would never receive 13+/18+ content. Additionally, we could display the user's past top preferences, those that the user has already chosen, to provide more personalized recommendations. This would come in handy when users wish to re-watch past favorites.

To conclude this written report, our project was successful in developing a movie recommendation app that offers tailored suggestions to users based on their watching patterns and interests. While we encountered some limitations, we were able to overcome them and offer a useful tool for movie enthusiasts. Future improvements could be made to enhance security, age-based filtering, and display past user preferences.

Part 7: References—IEEE Format

Course Notes

- D. Liu and M. Badr. "Course Notes for CSC110 and CSC111." Foundations of Computer Science. https://www.teach.cs.toronto.edu/ csc110y/fall/notes/ (accessed Mar. 5, 2023)

Data Set

- A. Dattatray, tmdb_5000_credits, vol.1, Kaggle: Kaggle, 2022. [Dataset]. Available: https://www.kaggle.com/datasets/akshaydattatraykhare/movies-dataset. [Accessed: Mar. 5, 2022].
- A. Dattatray, tmdb_5000_movies, vol.1, Kaggle: Kaggle, 2022. [Dataset]. Available: https://www.kaggle.com/datasets/akshaydattatraykhare/movies-dataset. [Accessed: Mar. 5, 2022].

Module Documentation

- Riverbank Computing and The Qt Company. "Reference Guide — PyQt Documentation v6.4.1." Riverbank Computing. https://www.riverbankcomputing.com/static/Docs/PyQt6/ (Accessed: Mar. 5, 2022).
- J. Bodnar. "Python PyQt6." ZetCode. https://zetcode.com/pyqt6/introduction/ (Accessed Mar. 7, 2023)
- The Qt Company. "Qt for Python" Qt. https://doc.qt.io/qtforpython/index.html (Accessed Mar. 8, 2023)

Idea Research

- M. Gavira. "How Netflix uses AI and Data to conquer the world." LinkedIn. https://www.linkedin.com/pulse/how-netflix-uses-ai-data-conquer-world-mario-gavira/ (Accessed Mar. 7, 2023)
- J. Ciancutti. "How We Determine Product Success." Netflix Technology Blog. https://netflixtechblog.com/how-we-determine-product-success-980f81f0047e (Accessed Mar. 7, 2023)
- D. Chong. "Deep Dive into Netflix's Recommender System." Towards Data Science. https://towardsdatascience.com/deep-dive-into-netflixs-recommender-system-341806ae3b48 (Accessed Mar. 7, 2023)

Inline Citations (Listed in order of use)

- Netflix Help Center. (n.d.). How Netflix's Recommendations System Works. Privacy and Security Help Page. Retrieved March 26, 2023, from https://help.netflix.com/en/node/100639
- Help Center. (n.d.). How to See Viewing History and Device Activity. Privacy and Security Help Page. Retrieved March 28, 2023, from https://help.netflix.com/en/node/101917