Anime Recommendation System

Matt Fournier, Christie Hui

MyAnimeList

Description

The purpose of this project is to create an anime recommendation system for users based on existing user data on myanimelist.net. Existing user data consists of anime titles, genre, types, ratings, and amount of users who have watched a particular anime.

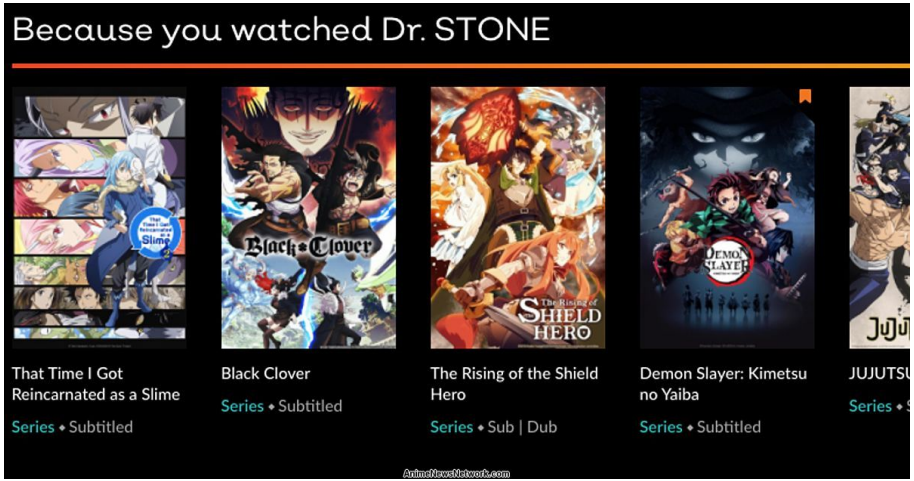
Some interesting questions we intend to answer are:

- Which genres of anime contribute to ambiguous users' decisions?
- How much do aspects such as rating, episode count, and type influence a user's choices?
- Is popularity of an anime or similarity to other users more significant in impacting users' preferences?

Prior Work

Streaming sites such as Crunchyroll have recommendation systems for users based on their previous anime watch history.

Our project will try to incorporate a two-way recommendation system in which user input can be provided in addition to the existing user data in the data set.



Datasets

- Anime Recommendations Database: https://www.kaggle.com/CooperUnion/anime-recommendations-database
- Found on Kaggle
- Downloaded on both team members' machines

Content

Anime.csv

- anime_id - myanimelist.net's unique id identifying an anime.
- name - full name of anime.
- genre - comma separated list of genres for this anime.
- type - movie, TV, OVA, etc.
- episodes - how many episodes in this show. (1 if movie).
- rating - average rating out of 10 for this anime.
- members - number of community members that are in this anime's "group".

Rating.csv

- user_id - non identifiable randomly generated user id.
- anime_id - the anime that this user has rated.
- rating - rating out of 10 this user has assigned (-1 if the user watched it but didn't assign a rating).

Proposed Work

- **Exploratory Data Analysis:** the data provided in the dataset is already pretty clean, but we plan on doing some outlier detection, visualizations, and feature analysis.
- **Data pre-processing:** for example, the dataset consists of the value -1 for user anime entries without ratings. We will figure out how to process these values as we do the EDA and pre-processing section.
- **Model fitting:** Determining a proper classification model for the data set.

List of tool(s)

- Jupyter Notebook
- Pandas, numpy, sklearn, scikitlearn (for calculations, EDA, data processing, analysis)
- matplotlib, seaborn (for visualizations)

Evaluation

- Cross validation
- Precision
- Accuracy
- F1 score
- Recall
- R-squared value