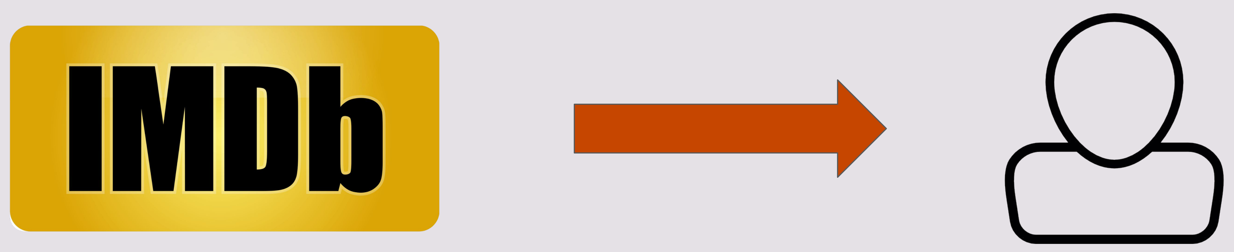
Movie Recommendation Model

By Grant Doan, Prayash Joshi, Reagan Orth, Ved Patel

BACKGROUND

We aimed to create a recommendation system that considers user interests and similarities between users. Due to a lack of user input, existing approaches suffer from single-movie or search-based recommendations, popularity bias, and unreliable suggestions.



Limitations of the current solutions:

Cold Start: Not having any prior knowledge on a user

Data Sparsity: Difficult to find users who have reviewed the same movies

Scalability: Inaccurate results are inevitable with large amounts of data being processed on popular sites.

Filter Bubble: Users are diverged from having different opinions

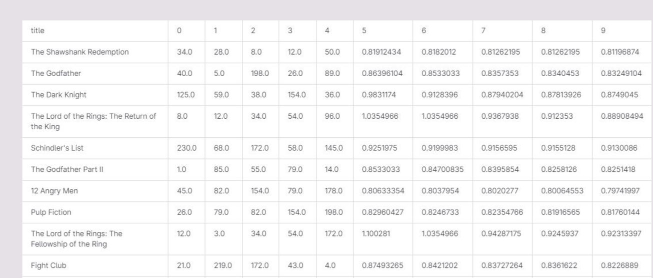
We combat the issues with preexisting models by offering a way for users to enter a rating for movies directly. We will carefully evaluate their ratings through our machine-learning ecosystem to curate a set of movies they will surely enjoy.

OUR PROCESS

Initially, we started by scraping two different datasets from IMDb:
- Dataset 1: Reviews of top movies
- Dataset 2: Title, Description, Genre, Directors, and Actors of top 1000 movies

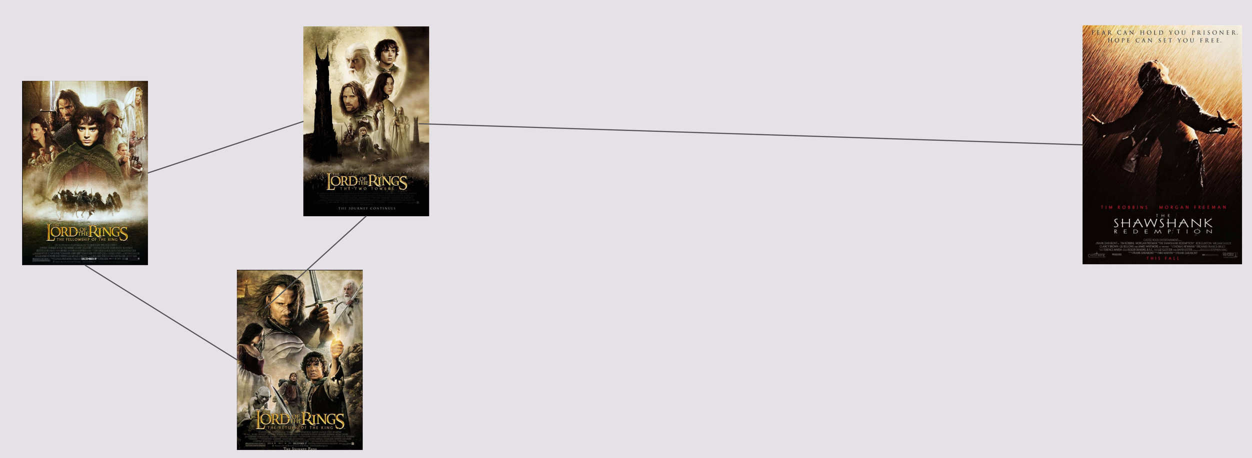
Now we had to perform NLP preprocessing on the first dataset which involved:
- Cleaning up punctuation and creating a sparse matrix on the reviews
- DF/TDF to remove insignificant words (Giving special attention to names mentioned in reviews)
- Pretrained BERT for Aspect Based Sentiment Analysis
- TF/IDF for finding specific film elements that are preferred by users

From the second data set we were able to create a similarity matrix which determines how close different movies are:



Next we created a Graph Neural Network to create a high dimensional mapping for the movies:

- The graph itself is large and unweighted
- Movies are nodes
- Edges represent similarities
- Similar movies will be close together



But how will this help recommend movies?

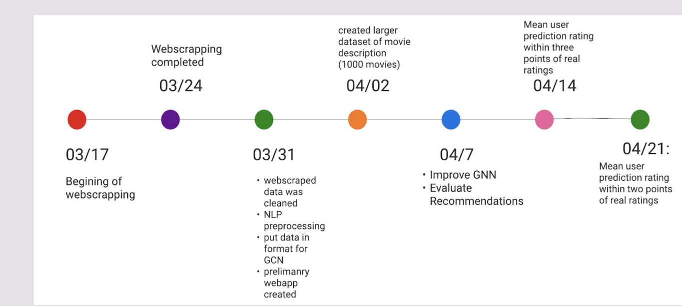
- Check movies most similar to movies the user already likes
- Check descriptive words in any user reviews provided
- Combine BERT pretrained-model and TF/IDF to find what user's find interesting in movies

And finally combine the movie preferences…

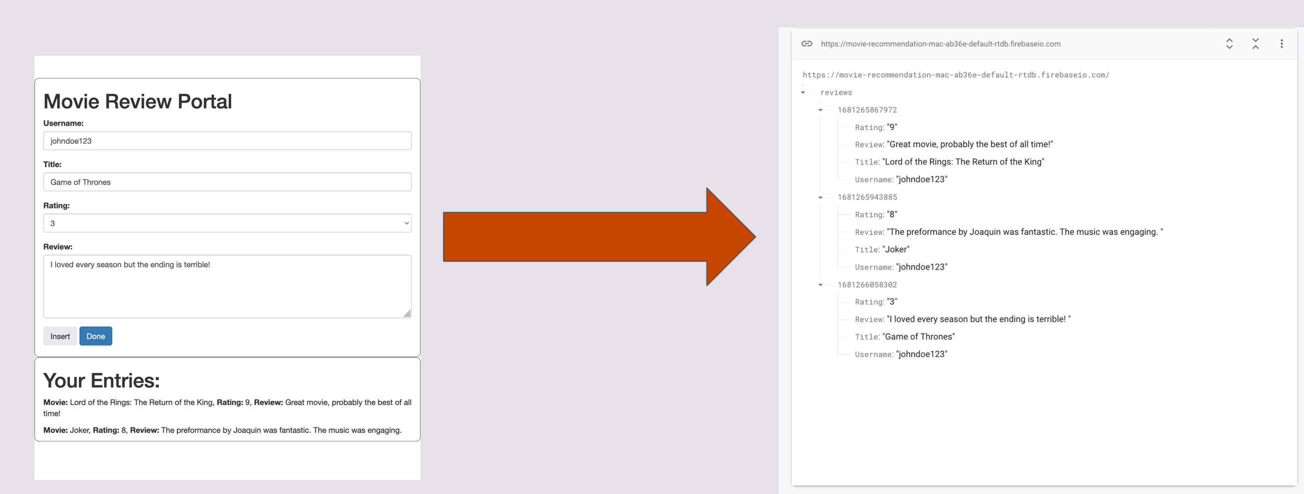
Results and Impact

Graph Neural Network:
- Built similarities graph for unsupervised problem.
- Worked on improvements such as gathering more film data.
- Pretrained BERT model does a good job of evaluating overall sentiment with 90% accuracy.
- Extracting topics from reviews is difficult, but positive/negative classification works well.
- Pursued aspect-based sentiment analysis using PyABSA.

Here is our project timeline:



Webpage:
- Takes in user reviews and provide recommendations based on them.
- First iteration took one review at a time and stored in Firebase real-time database.
- Second iteration allows for input of multiple movies and transitioned to a dynamic website.



Suppose you entered this review for The Lord of the Rings: The Return of the King:

"Rating: 9"
"I love this movie. It has amazing world building and the characters are played very well. I wish more movies were like this."

This review would result in these results: