MOVIE RECOMMENDATION SYSTEM PROJECT

KAZI | SUMAIYA | MAHAM | MUHAMMAD ELBADRI

ABSTRACT

Our goal for this project was to establish a successful recommendation system that analyzed movie content, such as popularity, crew/cast information, and movie content to arrive at the best possible recommendations. We prepared our data, created thresholds based on distributions, used text-mining techniques, and finally created a machine-learning model.

EXPLORATORY DATA ANALYSIS

We determined the data types in each dataset to understand the rows and columns. After examining the 'movies' dataset, the team set a threshold based on the column's popularity, vote count, and vote average. This trimmed the dataset significantly, which later increased the performance of the ML algorithm.

DATA PREPERATION

We decided only to use movies with English as their original language. Additionally, we filled in the null values for "overview" but decided against filling in the null value for "homepage" because it had a limited value to our data set. Finally, we unpacked columns with information on what we would use in our content-based approach. We merged our movies_df and credits_df data into one dataset for easier handling.

MACHINE LEARNING

We cleaned the text data using Python libraries and the Porter stemmer to normalize the text. Using the TFIDF-Vectorizer and Cosine Similarity, the algorithm gave back promising results.

RESULTS

Using the algorithm, the resulting movies were very similar to the input result.

Input: Frozen

Output:
Wreck-It Ralph
The Princess and the Frog Pocahontas
Brave
Snow White and the Seven Dwarfs

Input: The Dark Knight

Output:
Batman Returns
The Dark Knight Rises
Batman
Batman Begins
Man of Steel