Movie Recommendation Application

Ryan S. Shaw

Northwest Missouri State University, Maryville MO 64468, USA
S546850@nwmissouri.edu

Abstract. Keywords: data analytics · movies · Python · web scraping

1 Introduction

I plan to work in the domain of web scraping. Primarily because it is the most interesting to me but also because I feel I did not get to fully explore web scraping in the web scraping course due to time constraints.

I would like to scrape my data from the online website IMDB primarily. I may also source data from Rotten Tomatoes.

I intend to solve the problem of trying to decide which movies to watch. I feel like this is something everyone struggles with at some point. Scrolling through endless movies on Netflix or some other streaming service and not knowing what to choose.

Steps taken would be as follows:

1. Scrape the data from the previously mentioned websites.
2. Clean/organize the data into a usable format.
3. Create a Python script that requests user information (name, birthdate, gender, favorite genre).
4. Create a model that can recommend some movies that the user may like based on their information and the movie rating.

Python will be a huge component for me in this project. I will need it to for just about every aspect of this project.

2 Data Sourcing

I used Python and web scraping (specifically the Beautiful Soup Module) to gather my data from the IMDB website. The data is in HTML format being written intitially for a webpage and collected using Python. The attributes that will be gathered are: Title, Release Year, Rating, Runtime, Genre, Metascore, Movie Description and Votes. IMDB organizes movies by genre, so moving forward I would like to implement a method for users to search movie recommendations by genre.

3 Datasets

4 Data Cleaning

5 Results and Analysis

6 Limitations

7 Conclusion

□

References

1. Imdb: Ratings, reviews, and where to watch the best movies and tv, https://www.imdb.com/
2. Rotten tomatoes: Movies — tv shows, https://www.rottentomatoes.com/