

CS 410 Project Proposal: Letterboxd Film Recommender System

MovieRekomendr

Team Members

Aliva Panigrahi (alivabp2) - Team Captain

Chirag Gupta (chiragg4)

Shreyah Prasad (sprasa20)

Topic Description

Letterboxd is an online social film review platform, where members come together to write short or even detailed reviews about popular films/television. We plan to perform sentiment analysis on Letterboxd's most popular reviews to develop a movie recommender system. Based on the positivity rate, the system returns a ranked list of movies. If time permits, we can implement a personalized recommender system based on users' preference of genre, controversiality, etc... The recommender system will be evaluated by comparing the system's outputs to the popular movies list on Letterbox. Our project relates to the course by creating a push system for a text information system. Additionally, we will be implementing sentimental analysis, a topic discussed in this course.

Movies are an important topic due to how much public interest they garner. They can be very controversial and are arguably one of the largest forms of influential media and discussion. Diving into the world of movies can be intimidating at first, especially with surface-level knowledge in films. Someone unfamiliar with movies may not know a good starting point or what qualities to evaluate a movie on. Most film review sites simply provide ratings and reviews rather than actually recommending movies. These lists tend to be similar across sites, rather than taking in personal input, which is a unique attribute Letterboxd has. Since these film reviews have a very personalized approach, the recommender system would be human-like (ie: friends recommending movies to friends).

Tools

We plan to use Python to develop our system.

Python libraries:

- BeautifulSoup
- Selenium: <https://selenium-python.readthedocs.io>
- NLTK: <https://www.nltk.org>
- SCLKit: <https://scikit-learn.org/stable/>
- NumPy: <https://numpy.org/doc/>
- PyGUI: [Dear PyGui's Documentation — Dear PyGui documentation](#)

Tasks & Timeline

- Learning Selenium **8HRS**
- Evaluating Letterboxd **10HRS**
 - Identifying important page attributes needed to pull information
 - Testing how to pull needed information
- Using Selenium scrape all necessary information **14HRS**
 - ie: most popular film reviews of all time, ratings, text content of reviews, genres, themes, etc...
- Sentiment analysis of reviews (research and implementation) **12HRS**
- Create a ranked list based on scraped information **8HRS**
- Create basic UI **8HRS**