SENTIMENTAL ANALYSIS TO FILM REVIEWS ON LETTERBOXD

Team Member: *Zejia Chen*

Statement of Research Question:

The questions that I would love to answer in the final projects are:

◆ What is the sentiment toward films by a particular director? Are there any splits in opinions? What can we learn from it? How does the sentiment change over time as the director's career progresses?

◆ What are some salient characteristics to the topics of positive and negative reviews?

◆ Can we gauge the performance our analysis by assessing the accuracy of different adopted methodology?

Project Overview:

I have always considered myself a cinephile. Normally on the weekend, I will pick an intriguing film to enjoy with my friends as a weekly ritual. Although my friend and I all have similar film taste, we still spend an incredible amount of time in choosing the film that we ended up watching.

One of the ways that we determine if a film is watchable or not is by looking up its review from various sources. Movie reviews help users decide if the movie is worth their time. A summary of all reviews for a movie can also help users make this decision by not wasting their time reading all reviews.

As a result, I would love to utilize this opportunity to conduct a sentimental analysis on reviews of a series of films. After getting the sentiment score of all film reviews, I will then generalize characteristics of those sentiments to offer an abstract of positive/negative sentiments regarding a particular film.

Sentiment analysis can be very useful in various scenarios. It determines the attitude of critics depending on their reviews. Great Sentiment analysis of a movie review can rate how positive or negative a movie review is and hence the overall rating for a movie.

By evaluating the performance of different methodologies, I wish to offer a model that can accurately reflect film reviewers' sentiment and offer pertinent guidance for people who are struggling to find the right movie to watch.

Summary of Related Works:

1. *Analyzing Movie Reviews using Sentiment Analysis*, Gayathri Devi Nagalapuram
   https://medium.com/nerd-for-tech/analysing-movie-reviews-using-sentimental-analysis-77e28e463b1b
   - **Similarity:** The research topic of this project is the same as mine. It treats as a useful reference to the structure of my entire research, specifying the step that I am going to take to get to the goal.
   - **Difference:** The research primarily leverages machine-learning to conduct its sentimental analysis which is quite different from my approaches. The author also uses a disparate data set compared to mine which I will cover in the next session.

2. *Analyzing Star Wars Movie Scripts,* XAVIER
   https://www.kaggle.com/xvivancos/analyzing-star-wars-movie-scripts
   - **Similarity:** The project is also about conducting a text (sentimental) analysis pertaining a film-related text document, in this case a movie script.
   - **Difference:** Instead of a analyzing the overall sentiment from a large text document like this project does, I will be generalizing sentiments across a corpus of brief film reviews.

3. *Sentiment Analysis for Movie Reviews*, Ankit Goyal & Amey Parulekar
   https://cseweb.ucsd.edu/classes/wi15/cse255-a/reports/fa15/003.pdf
   - **Similarity:** One of this research paper's goal is to explore the performance of multiple NLP models which is something that I am also interest in doing for my project.
   - Difference: The research is also heavily based on machine-learning and making prediction depending on people's past reviews. Mine, on the other hand, is more focus on scrutinizing sentiments of existing film reviews.

Proposed Dataset:

Movie-rating websites are often used by critics to post comments and rate movies which help viewers decide if the movie is worth watching. Instead of finding the data on conventional movie rating website, like IMDB or Rotten Tomatoes, where most of reviews are made by professional critics, I decide to come up with a more specifc dataset which is more tailored to my research direction.

The website that I am going to getting the film reviews is called *Letterboxd* – a social platform focused on sharing opinions about, and love of, film. The platform allows users, many of whom are just film lovers, to rate and record their opinions about films which makes it ideal to conduct my desired sentimental analysis. Plus, I want to keep the scope of my research topic more specific. Hence, I am planning to scrap reviews of film that created by one of my favorite yet controversial director on Letterboxed.

Proposed Methods:

To answer my first two research questions, I will primarily work with the sentimental analysis toolkit (VADER) that we learned in class. I will also use topic modeling as well as vectorization in some ways that I can extract topics to generalize the characteristics of positive/negative sentiment in reviews.

Since we will cover BERT in subsequent weeks, I am also hoping to conduct similar NLP analysis using this method. Eventually, I am going to investigate the performance as well as the tradeoff between these two approaches, and hopefully can have a clear response to the third research question.

Proposed Presentation Format:

In terms of presentation format, I am planning on writing a research paper to describe my initiative, data & method, and findings. In addition, I might also make a PPT to illustrate my findings and insights visually.

Proposed Timeline:

WEEK	GOAL
NOV 9, 11	**Getting the dataset (at least some of it)**
NOV 16, 18	**Testing methods & Conducting intended analysis**

NOV 23, 25	**Continue with my analysis (if it worked) & drafting the research paper**
NOV 30, DEC 2	**Working on the research paper**
DEC 7, 9	**Final review & submission**