Sentiment Analysis for Movie Review

Srimedha Bhavani Chandoo
Department of Data Science
University of Colorado Boulder
Boulder, Colorado, USA
srbh1140@colorado.edu

Jagrati Chauhan
Department of Data Science
University of Colorado Boulder
Boulder, Colorado, USA
jach5937@colorado.edu

Navya Prasad Malur Narasimha Prasad
Department of Data Science
University of Colorado Boulder
Boulder, Colorado, USA
nama5610@colorado.edu

ABSTRACT

Movie reviews are crucial tools for assessing a film's performance, overcoming the limitations of traditional numerical or star ratings. While quantitative evaluations provide a surface-level picture of a film's success or failure, the compilation of textual reviews provides a more in-depth qualitative understanding. Textual reviews, beyond simply categorizing grades, operate as detailed narratives, revealing the intricate tapestry of a film's strengths and shortcomings. These evaluations serve as a platform for critics and spectators to express themselves, highlighting key areas that contribute to a film's overall impact. The data was collected using web scraping techniques and data cleaning and visualisation was done. Three different variations of Naive Bayes model were implemented in this project and were compared with each other. Out of these three models, Bernoulli's Naive Bayes is performing the best in terms of accurate prediction of the sentiment.

1 INTRODUCTION

Diving deeper into a movie review for a more detailed evaluation of whether a movie meets the collective expectations of both its audience and critics can be done using a textual review better. It opens the door to understanding the complex dynamics that shape a film's reception. This research attempts to peel back the layers of textual movie reviews to reveal the nuances that a numerical rating system may miss. Exploring these subtle critiques yields a richer understanding, offering vital insights into the alignment between a film's planned impact and its actual reception by discriminating spectators and critics alike. Movie reviews were web scraped from online platforms such as IMDB for this project[3]. Data cleaning, pre-processing and analyzing was done once it was collected to proceed with the review analysis. The reviews collected were analyzed for different types of sentiments present using the sentiment intensity analyzer available in the nltk toolkit. Once the data was stored and processed, we developed three different types of Naive Bayes model - (Gaussian, Bernoulli and Multinomial) that predicted which of the following sentiments each review expressed: strong positive, positive, neutral, negative, or strong negative. [8]

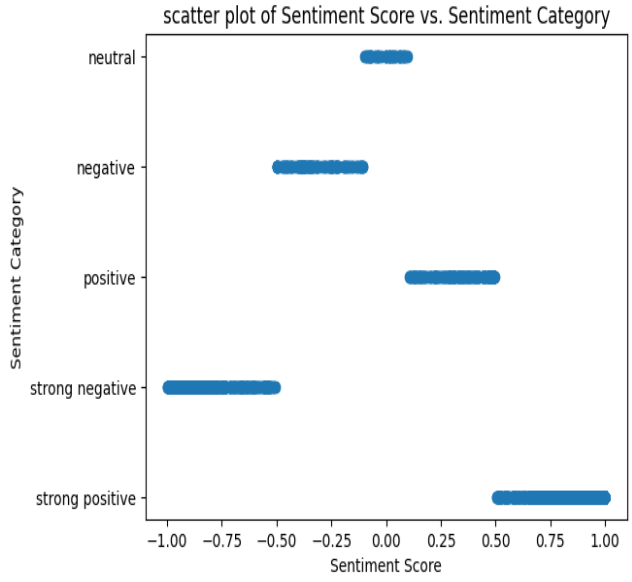
2 DATA EXTRACTING – WEB SCRAPING

We had an interest, in gathering data from real time websites for this project since our main focus was on extracting information. We collected around 5000 reviews for movies from the IMDB website. This dataset provided us with a wealth of details, including the date and time of the reviews

viewers ratings, for each film and the movie titles. The data we obtained was carefully crafted, making sure that there were no missing values or anomalies. It was a wide collection of data. To conduct sentiment analysis we eliminated any columns that were not useful. We proceeded with data analysis and visualization before moving on to data modeling.[1]

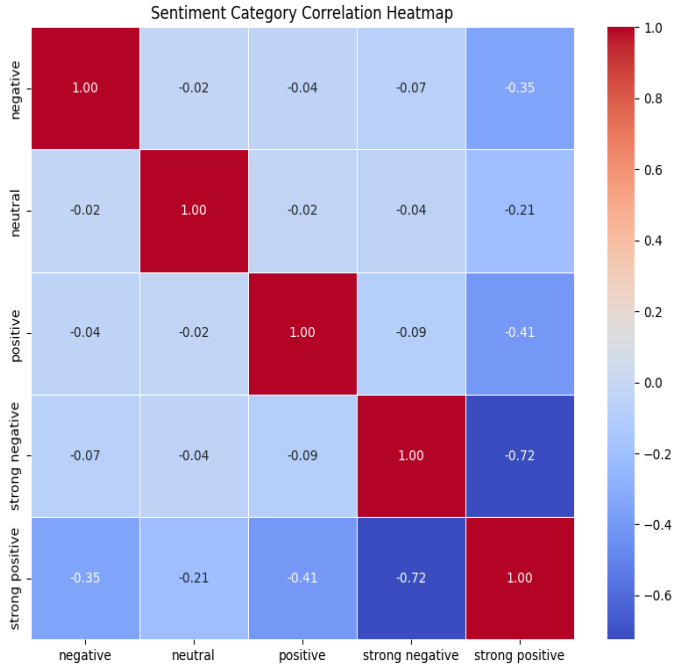
3 DATA ANALYSIS

Extracting insights, from information heavily relies on the understanding and interpretation of data. Making decisions becomes easier when we analyze patterns, trends and connections within data sets. To present information clearly and accessibly visualizing data through charts and graphs is an approach. Whether its for business intelligence or scientific research effective data analysis and visualization are crucial, in improving communication and understanding. They empower individuals and organizations to draw conclusions from their data set [2].

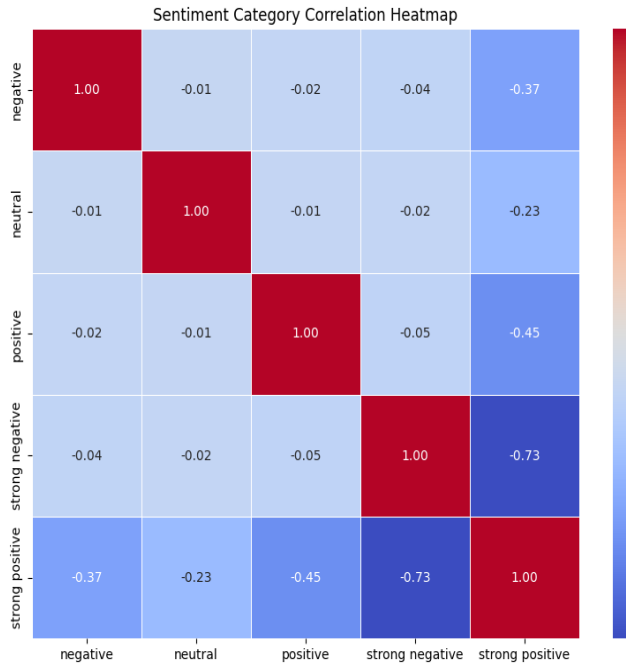


(a) Figure 1: Division of Sentiment Category Based on Sentiment Scores

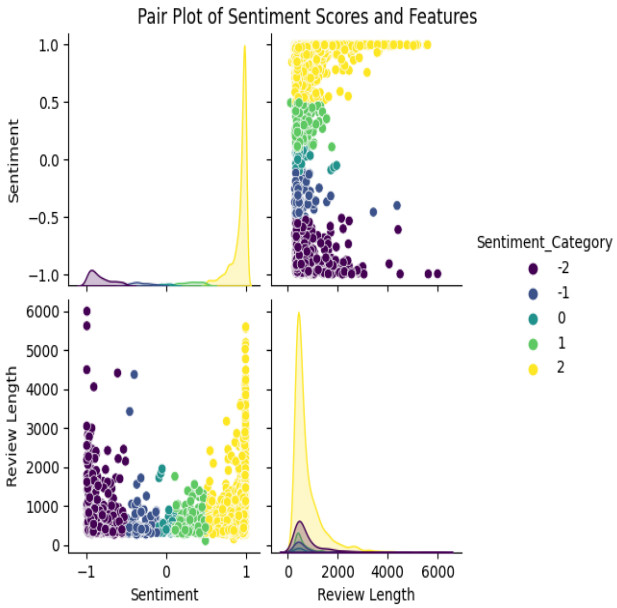
The scatter plot shows how sentiment scores and sentiment categories are related. When we look at the plot it's clear that positive sentiment scores are generally higher, than other ones. We can see this from the cluster of data points in the category which're closer to the top of the y axis compared to the spread of data points, for the negative category.



(a) Heatmap for the reviews of Oppenheimer



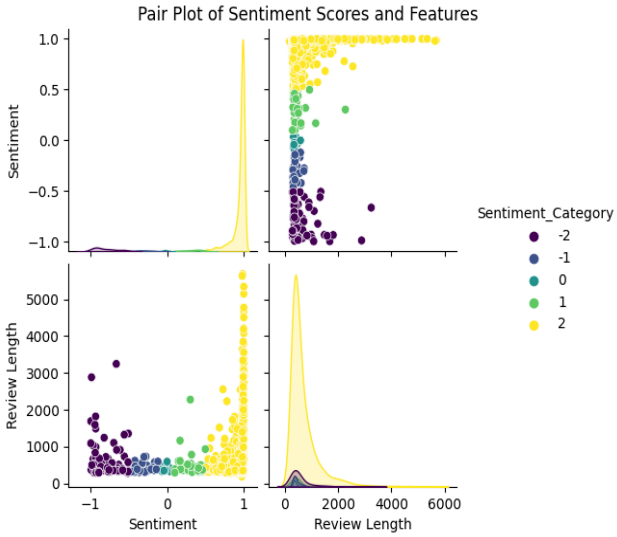
(b) Heatmap for the reviews of Barbie



(a) Pair plot for the reviews of Oppenheimer

Figure 2: Heatmap

This heatmap figure 2 illustrates how people feel towards sentiment categories displaying the percentage of individuals who're positive, negative or neutral. The intensity of the colors, on the heatmap represents the level of correlation, between pairs of sentiment categories. Darker colors indicate stronger correlations while lighter colors suggest weaker or negative correlations.



(b) Pair plot for the reviews of Barbie

Figure 3: Pair Plot

Figure 3 is a pair plot of sentiment scores and features, color-coding different sentiment categories. This pair plot can provide us with the following insights:

3.1 Sentiment Score Distribution:

The diagonal histograms show the distribution of sentiment scores for each sentiment category.

It looks that the sentiment ratings for each sentiment category are focused around specific values.

3.2 Association Between Sentiment Scores and Variables:

The scatter plots off the diagonal demonstrate the association between sentiment scores and other variables. Observing the clusters of points can reveal how sentiment ratings vary in relation to the attributes.

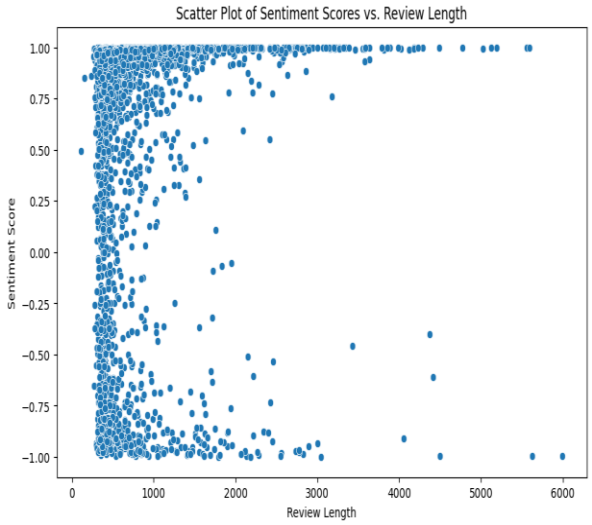
3.3 Sentiment Category Separation:

Using the pair plot, we can see how successfully the sentiment categories are separated depending on the characteristics. Clear distinctions across sentiment categories imply that some characteristics play a substantial role in influencing sentiment.

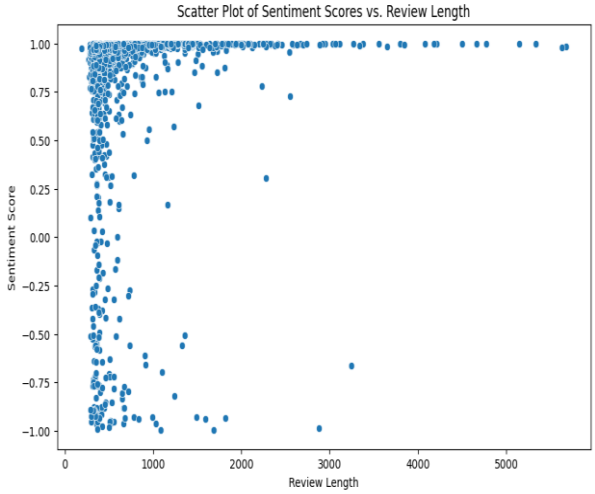
3.4 Potential Patterns or Trends:

Using the pair plot, we can see how successfully the sentiment categories are separated depending on the characteristics.

Clear distinctions across sentiment categories imply that some characteristics play a substantial role in influencing sentiment



(a) Scatter plot for the reviews of Oppenheimer

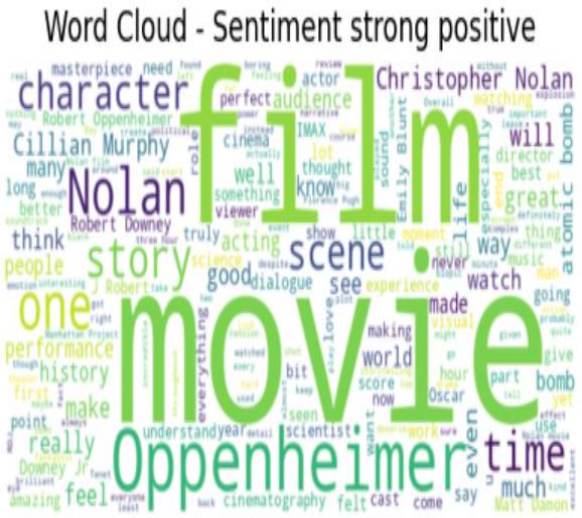


(b) Scatter plot for the reviews of Barbie

Figure 4: Scatter Plot

The graph 4 displays a connection, between the sentiment score and the length of reviews. It implies that as the length of a review increases the sentiment score generally decreases. In words longer reviews are more likely to convey sentiments compared to

shorter ones. However it is important to mention that this correlation is not particularly strong which suggests there are cases that defy this pattern. There are reviews with positive sentiment scores and several short reviews.



(a) Word cloud for the reviews of Oppenheimer

(b) Word cloud for the reviews of Barbie

Figure 5: Word Cloud

The figure 5 depicts a mix of negative opinions, about the movie. The used words are related to the film itself like "Oppenheimer," "movie," "film," "time " "good," "fun," "funny," "world," "made," "feel," "think," and character names like Barbie and Ken. These words are mostly neutral as indicated by their color. However there are also a words such, as "boring" and "watch" that suggest some individuals found the movie slow or uninteresting.

However, there are some blue words, such as "great," "good," and "amazing," indicating that some people loved the film. Overall, the graph reveals that Oppenheimer is a divisive film, with certain admirers and others dislike it and similarly for the movie Barbie. We have included strong positive graph for Oppenheimer and strong negative graph for the Barbie movie.

These visuals are definitely helpful, in giving us an understanding of how the data operates and the presence of any biases within it. Additionally we are gaining an appreciation for the actions taken during the data analysis process. Why these visual representations play a role. Each of the plots mentioned above offers insights, in their distinct manner as shown by the above plot.[7].

4 Data modeling

Data modeling in machine learning generally involves creating mathematical representations of real-world phenomena. Through techniques like classification, regression or neural networks, models learn patterns and relationships within data and help in predicting the patterns. This process enables predictive analysis and decision-making,

forming a crucial step in developing accurate and efficient machine learning systems from both business and social point of views. [6]

Stemming and lemmatization are text normalization techniques in Python. Stemming reduces words to their root or stem form, removing suffixes. The NLTK and SpaCy libraries are the most commonly used tools for stemming. In this project, the NLTK tool has been used. Lemmatization goes a step further, reducing words to their base or dictionary form, considering the word's grammatical meaning for more accurate results in sentiment analysis. For instance, cared, caring, care all stem down to the word care which is considered the root word in this scenario. We made use of these techniques to break down our reviews in the data set obtained.

After breaking down the words and storing them as their root forms, another useful feature of the NLTK library was employed. The sentiment intensity analyzer. The NLTK toolkit offers a tool called "SentimentIntensityAnalyzer" for sentiment analysis, in Python. It calculates sentiment scores for text allowing us to quantify positivity, negativity and neutrality. By utilizing a trained model this analyzer takes into account both polarities and intensities providing a nuanced understanding of sentiment in everyday language. This tool finds its usefulness in applications such as monitoring media analyzing customer feedback and making sentiment based decisions, in natural language processing projects.

In our project, three different variations of Naive Bayes model is used - Gaussian, Bernoulli and Multinomial[5]. These three models are the most commonly used variations for text analysis and text classification in machine learning and data mining problems.

4.1 Bernoulli Naive Bayes:

Bernoulli Naive Bayes is a variant of the Naive Bayes algorithm suitable for binary data. It assumes features are binary (present or absent), making it ideal for text classification tasks like spam detection. The model calculates probabilities using the Bernoulli distribution, distinguishing between occurrences and non-occurrences of feature within a document.

4.2 Gaussian Naive Bayes:

Gaussian Naive Bayes assumes that features follow a Gaussian distribution. It's suitable for continuous data, making it applicable to tasks such as classification in medical diagnostics or image recognition. The model estimates mean and variance parameters for each class, allowing it to handle real-valued features effectively.

4.3 Multinomial Naive Bayes:

Multinomial Naive Bayes is designed for discrete data, commonly used in text classification where features represent word frequencies. It models the likelihood of observing word counts given a class, making it well-suited for tasks like document categorization. The model assumes features follow a multinomial distribution and is widely employed in natural language processing applications. [4]

5 Result

Table 1: Results from the Barbie dataset

Barbie	Gaussian	Bernoulli	Multinomial
Accuracy	97.26 %	94.29 %	97.26 %
Precision	0.9734	0.9455	0.9569
Recall	0.9726	0.9429	0.9550
F1-Score	0.9723	0.9424	0.9554

Table 2: Results from the Oppenheimer dataset

Oppenheimer	Gaussian	Bernoulli	Multinomial
Accuracy	63.09 %	61.97 %	73.34 %
Precision	0.6573	0.6229	0.7308
Recall	0.6309	0.6198	0.7314
F1-Score	0.6202	0.6153	0.7306

After reviewing the information it becomes apparent that the multinomial model outperforms both the Gaussian and Bernoulli models. When applied to the Barbie data set the Multinomial Naive Bayes model achieved an accuracy rate of 97.26%. However when applied to the Oppenheimer data set its accuracy rate dropped to 73.34%. This difference, in accuracy could potentially be attributed to biases in the collected data set.

6 Conclusion

To summarize, in this project, we attempted to mine our own data for two well-known films, Oppenheimer and Barbie, using web scraping techniques. We created a few visualizations after processing and analyzing the scraped data, which truly helped us understand the patterns and trends prevalent in the reviews. We used three distinct variations of the Naive Bayes model after lemmatizing the terms in the reviews and using a sentiment intensity analyzer. As seen in the results section, the Multinomial Naive Bayes model produces the best results. There are certain adjustments that can be made to improve the model's performance. These are covered in the following section, Future Work.

7 Future Work

- The project can be taken forward to include time-dependent lexicon, which can help us to understand how the public perceives the movie over time. Over time a movie can gain more criticism/ acceptance. We can also collect the gender and age of the reviewers to analyse how the movie is perceived by different age groups and men and women. These insights can help us get deeper understanding about what kind of movies will be liked by different subset of people. This can be further extended to build a movie recommendation model.
- The current model does not take into consideration, the inter word meaning dependency and the nuance usage of words. Hence just by going by the literal meaning of words, some of the reviews that are creatively and positively written can be mistaken treated as a negative sentiment review.
- The supervised learning model can be modified to use more advanced deep learning models like current neural networks (RNNs) or transformers, for

sentiment analysis. These models often excel at capturing complex relationships in sequential data.

- The performance and significance of the model can be enhanced by taking into account the number of upvotes and downvotes received for a review. This is one of the improvements that can be implemented in the future. Just because one viewer didn't enjoy the movie and left a review doesn't necessarily mean that every one else who watches it will also have the same bad experience. However if a review receives a big number of downvotes then it becomes easier to understand that most viewers actually appreciated the film in contrast, to one persons negative remark.

REFERENCES

[1] Ayat Abodayeh, Reem Hejazi, Ward Najjar, Leena Shihadeh, and Rabia Latif. Web scraping for data analytics: A beautiful soup implementation. In 2023 Sixth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU), pages 65–69, 2023.

[2] Lavanya Addepalli, Lokhande Gaurav, Sakinam Sindhuja, Saem Hussain, Joydeep Mookerjee, Vamsi Uppalapati, Waqas Ali, and Vidya Sagar. Assessing the performance of python data visualization libraries: A review. International Journal of Computer Engineering in Research Trends, 10:28–39, 01 2023.

[3] Papadopoulos S Frangidis P, Georgiou K. Sentiment analysis on movie scripts and reviews: Utilizing sentiment scores in rating prediction. Springer Nature, 2020 May 6.

[4] Muzammil Khan Yazeed Ghadi Hanan Aljuaid Zubair Nawaz Kifayat Ullah, Anwar Rashad. A deep neural network-based approach for sentiment analysis of movie reviews. In 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), 2022.

[5] Suresh Kumar, Kamlesh Sharma, Dhruv Veragi, and Ankit Juyal. Sentimental analysis of movie reviews using machine learning algorithms. In 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), volume 1, pages 526–529, 2022.

[6] B. Lakshmi Devi, V. Varaswathi bai, Somula Ramasubbareddy, editor – "Venkata Krishna P. Govinda, K." and Mohammad S. Obaidat. Sentiment analysis on movie reviews. In Emerging Research in Data Engineering Systems and Computer Communications, pages 321-329, Singapore, 2020, Springer Singapore.

[7] Barbara Ramalho, Joaquim Jorge, and Sandra Gama. Representing uncertainity through sentiment and stance visualization: A survey. Graphical Models, 129:101191,2023

[8] Asiri Wijesinghe, Sentiment analysis on movie reviews. BOOK, 2015/10/14

Link of the YouTube video

Click here