A 72% Accurate Model that Spotlights Sentiment Bias!

Sentiment Prediction of Rotten Tomatoes Movie Reviews

Hank Dickerson, Erin Moulton, Varun Pavuloori

Group 1: DS 4002 – Data Science Project
September 30, 2024

Project Details

Motivation

- Understand the word choice of movie reviews that influence Rotten Tomato movie classifications as positive or negative
- A simple rating can majorly influence views and overall perceptions of a film

Context

- This relationship is valuable for **filmmakers** to understand genre preferences and trends, **marketers** to make informed decisions before ratings are released and the **audiences** to better recognize potential critic bias

Project Details

Hypothesis:

- The sentiment of a movie review can be accurately predicted by the presence of specific keywords in the review

Research Question:

- What keywords are most frequently used in movie critic reviews that correlate to positive and negative sentiments?

Modeling Approach:

- Sentiment Classification, TF-IDF Analysis [1] [2], and Chi-Squared Feature Collection

Data

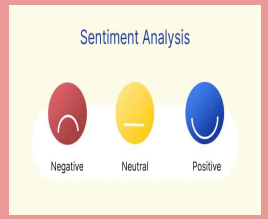
- **Acquisition**: Kaggle. There were no licensing or ethical concerns
- **1,444,963** total movie critic reviews, with each row representing an individual text review
- Cleaned dataset to remove ~16,000 duplicate reviews and ~70,000 values in the review column leaving **1,359,668** usable movie reviews

Data Dictionary

Column	Description	Type
movieTitle	unique identifier for movie each review	Object
year_created	year when the review was created	Datetime
isTopCritic	Boolean flag indicating if the critic is recognized as a top critic.	Boolean
fresh_or_rotten	Review classification state (e.g., “fresh” or “rotten”)	Object
review	full text of the review	Object

Analysis Plan & Justification

-remove duplicate and null review
-remove unnecessary columns



-Use LLM model to classify reviews as either pos/neg

-calculate TF-IDF score for pos/neg reviews
-filter out neutral words

-Chi-squared statistic of the top 10 words associated with pos and neg reviews

-Train a logistic regression model to classify movie reviews as either positive or negative based on their TF-IDF features.

Tricky Analysis Decision

NLTK vs. LLM

- Switched from using manual sentiment analysis with NLTK to using an LLM
- LLMs are far better at understanding language contextually

Bias and Uncertainty Validation

- Dataset difficulties:
  - Sarcasm in reviews
    - “This movie is great if you are a fan of garbage cinema”
    - “If you ever have a few hours of your life you never want to get back, this movie is a terrific choice”
  - Critic Bias
  - Changing society
    - Dataset contains reviews from many years ago, changing views on what constitutes a good movie
      - Corrected by only taking reviews from 20 years to present
      - Still includes Pre-Covid reviews

Results & Conclusions

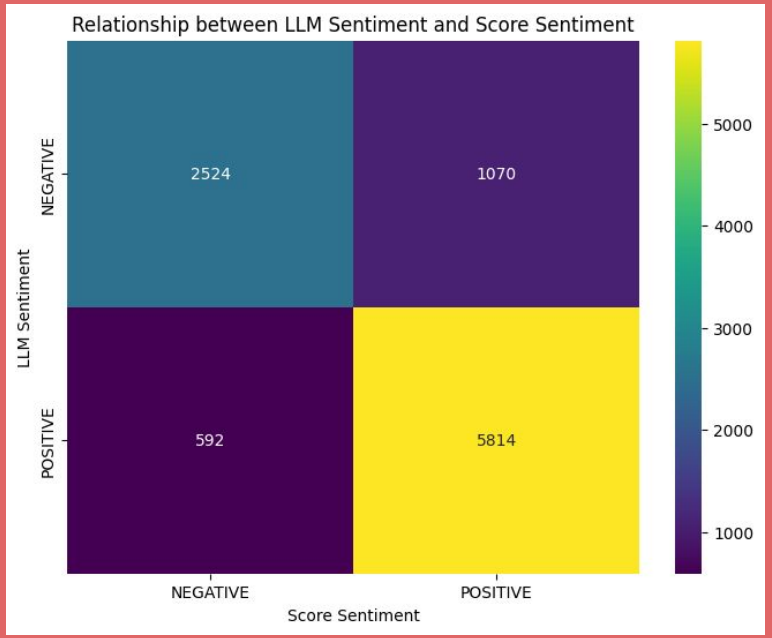
- We found our final model to be 72% accurate
- We reject the null hypothesis (desired 80% accuracy)
- Positive reviews were easier to predict

```
Model Evaluation Metrics:
Accuracy: 0.7210
Precision: 0.7364
Recall: 0.8790
F1 Score: 0.8014

Classification Report:
              precision    recall  f1-score   support

    NEGATIVE       0.67      0.44      0.53       719
    POSITIVE       0.74      0.88      0.80      1281

    accuracy                           0.72      2000
   macro avg       0.70      0.66      0.67      2000
weighted avg       0.71      0.72      0.70      2000
```



Next Steps

- Adjust and implement more rules using an NLTK model
- This would allow for the highest possible accuracy
- Test the new model on different review mediums

References, Resources, Acknowledgements

Jablonski, J. (2023, October 21). *Natural language processing with python's NLTK package*. Real Python. https://realpython.com/nltk-nlp-python/

P. Choksi, "Rotten Tomatoes Movie Reviews - 1.44M rows," *Kaggle.com*, 2024. https://www.kaggle.com/datasets/priyamchoksi/rotten-tomato-movie-reviews-1-44m-rows (accessed Sep. 22, 2024).

Pipelines. Hugging Face Documentation. (n.d.) https://huggingface.co/docs/transformers/en/main_classes/pipelines

Wang, S., & Manning, C. D. (n.d.). *Baselines and Bigrams: Simple, Good Sentiment and Topic Classification*. https://aclanthology.org/P12-2018.pdf

Welcome to Vadersentiment's documentation! VaderSentiment 3.3.1 documentation. (n.d.-b). https://vadersentiment.readthedocs.io/en/latest/

Wu, J. Y., & Pao, Y. (n.d.-b). *Predicting Sentiment from Rotten Tomatoes Movie Reviews*. https://nlp.stanford.edu/courses/cs224n/2012/reports/WuJean_PaoYuanyuan_224nReport.pdf

Github Repository: https://github.com/erinmoulton/Sentiment-Prediction-of-Rotten-Tomatoes-Critic-Movie-Reviews

Questions?


Word Cloud of Most Significant Words in Movie Reviews
fails
bad
worst
dull
doesn
mess
half
entertaining
like
bland