

IMDB Movie Review Sentiment Analysis

Morgan Didjurgis

22 Mar 2022



Outline

- Business Problem
- Data
- Model
- Results
- Conclusions



Business Problem

1

Provide automatic classification of movie reviews as positive or negative.

2

Overall accuracy best metric to evaluate the model's efficacy.

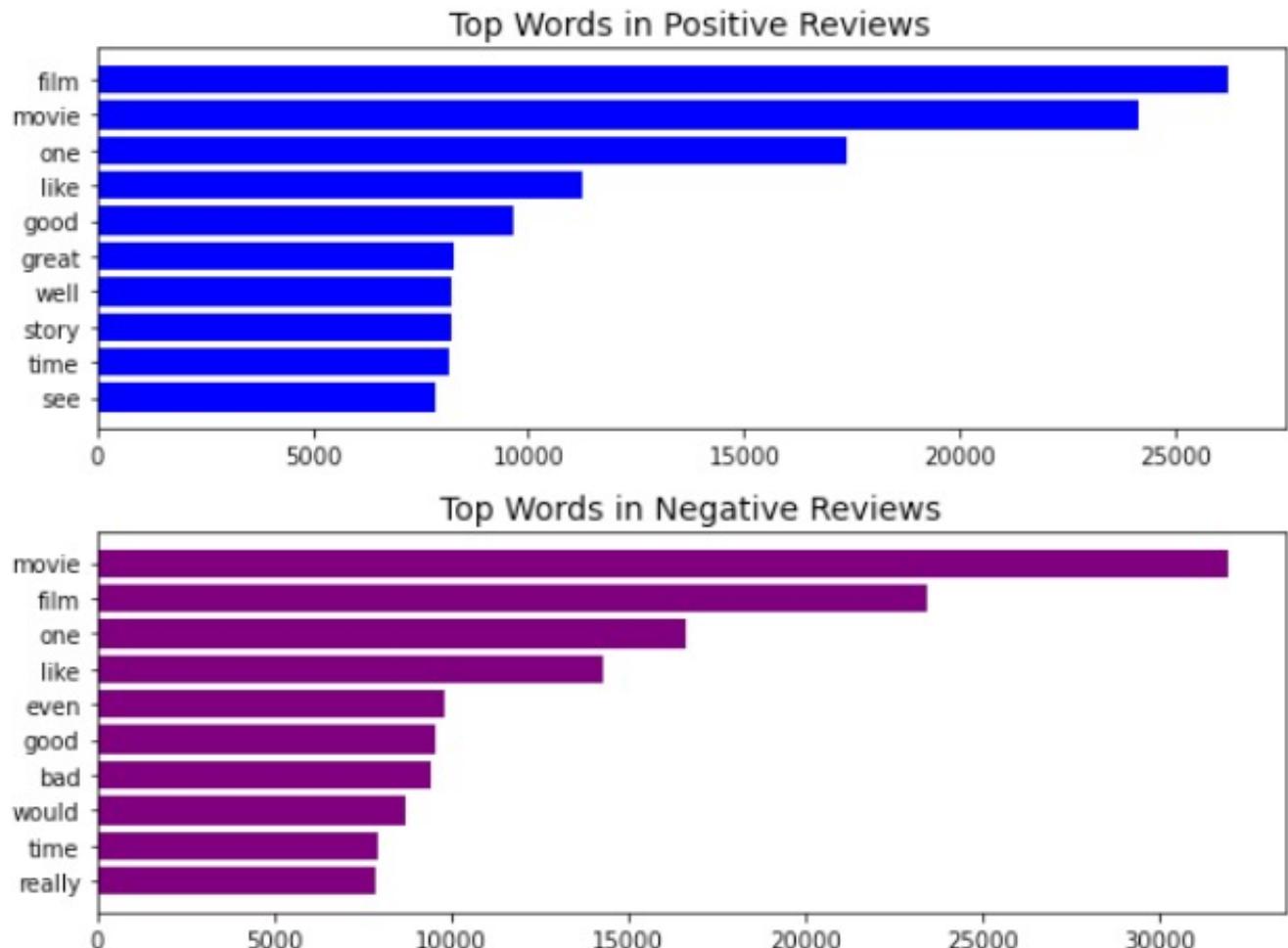
3

Balance accuracy with ease of deployment and interpretability.

Data

- List of 40,000 movie reviews
- Labeled Positive / Negative
- Balanced classes
- Over 90k different words in all reviews
- **kaggle** Dataset

<https://www.kaggle.com/datasets/yasserh/imdb-movie-ratings-sentiment-analysis>



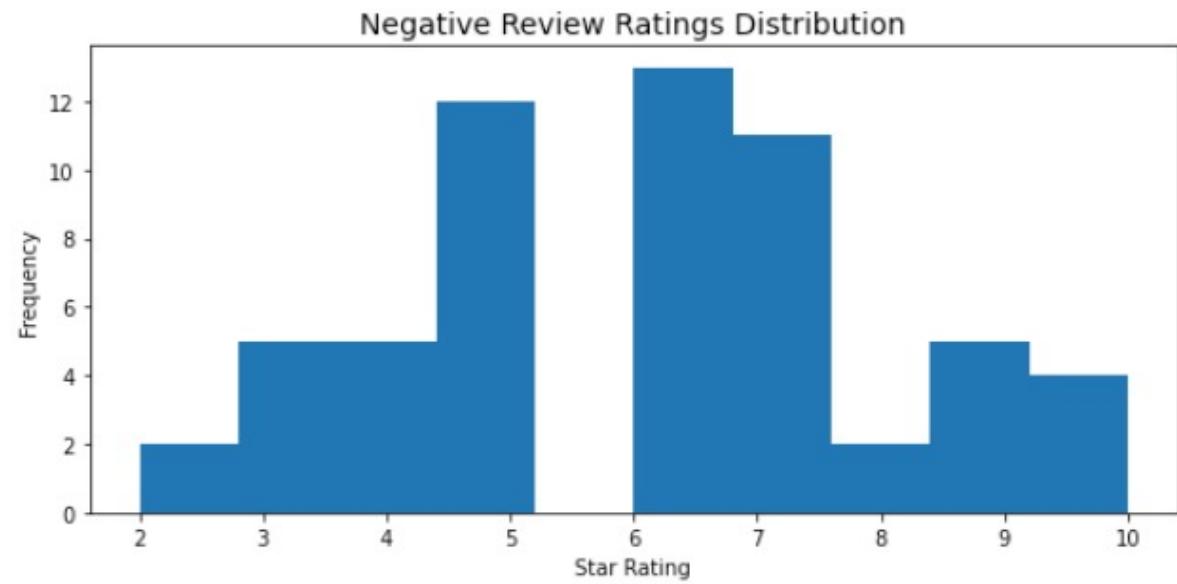
Model

- Logistic Regression
 - 12,000 words included as features
 - Model accounts for frequency of word in each review and number of reviews that contain the word
- Model Performance:
 - 90% accuracy in training and testing
- Limitations of the model:
 - Only provides overall ‘positive’ or ‘negative’ classification
 - Aspect of individual statements not evaluated

Important Features

Results: Example Movie Review Analysis

- Classified over 200 reviews from all nine Star Wars episodes
- Positive Reviews: 8.0 avg rating
- Negative Reviews: 6.0 avg rating
- Would be challenging to use star ratings to denote if the review was positive or negative



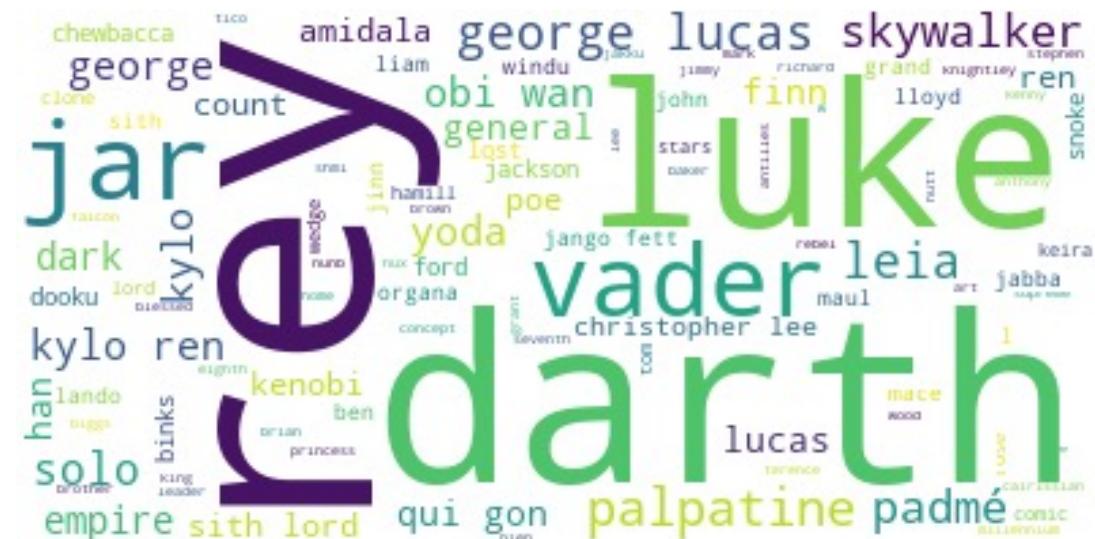
Results: Example Movie Review Analysis

Character mentions in positive vs. negative reviews...

Positive Reviews



Negative Reviews



Conclusions

- Model classifies IMDB movie reviews as positive or negative with an accuracy of 90%
- Can be used to gain further insight into positive/negative sentiment

Next Steps

- Transfer learning to apply more sophisticated models
- Expand number of labels (add neutral)
- Aspect sentiment analysis

