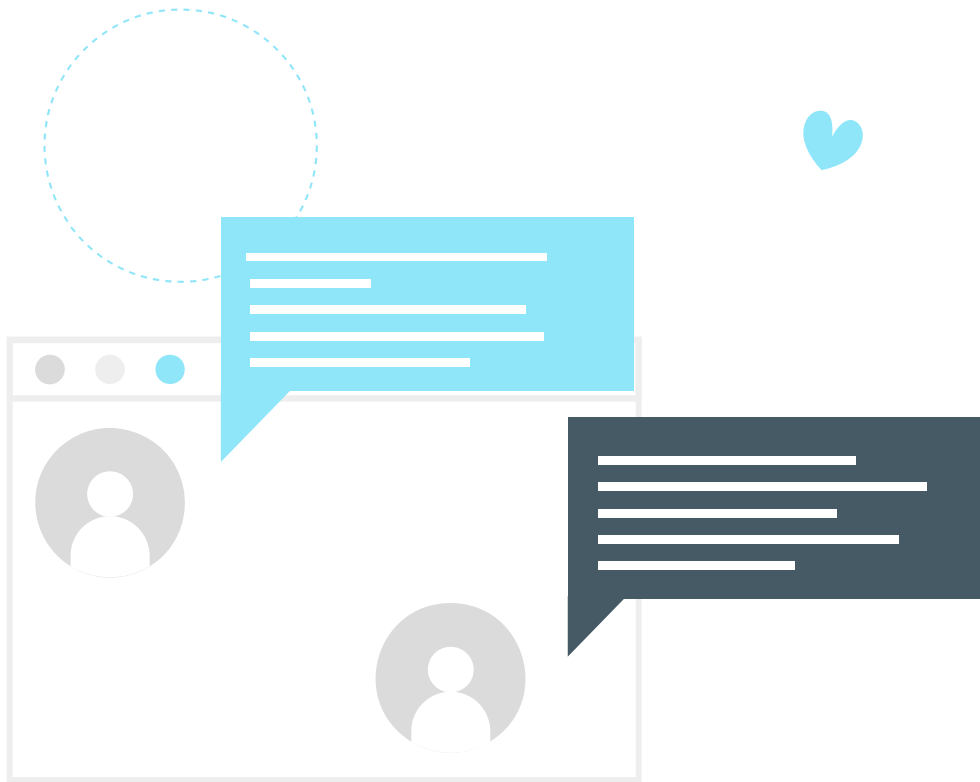


Twitter Sentiment Analysis

Sean Hart September 15, 2022



01



The Dataset



Initial Dataset

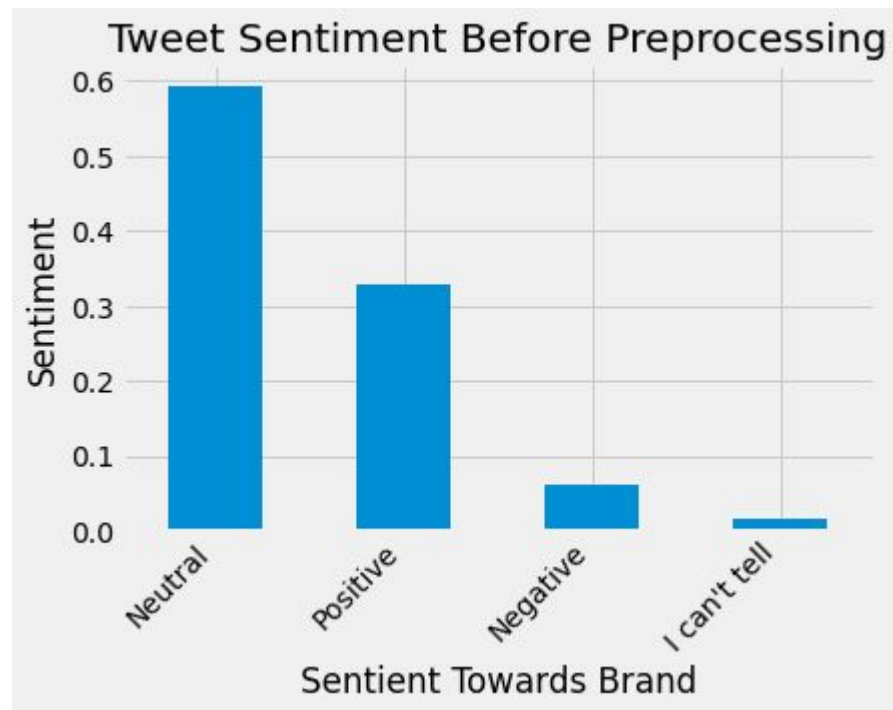
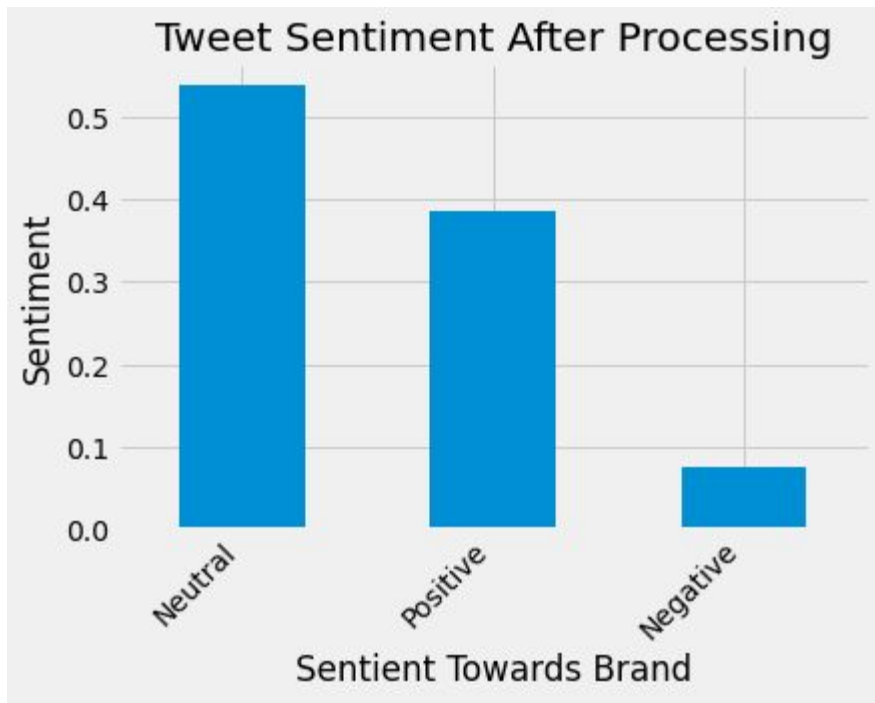


- 9902 Tweets from 2011 South by Southwest Conference
- Dataset included Tweet, Sentiment, and Object of Sentiment
- The dataset was highly imbalanced – very few negative sentiment examples.





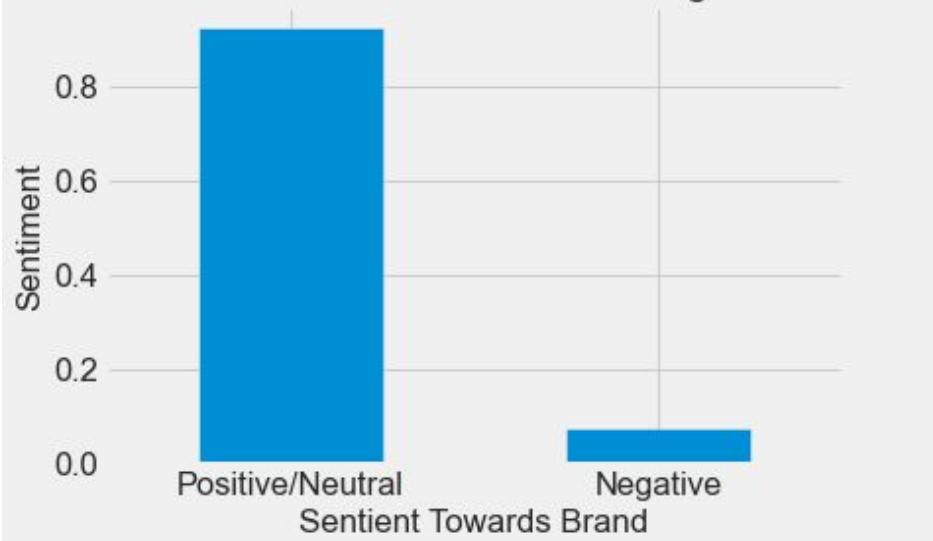
Sentiment Breakdown



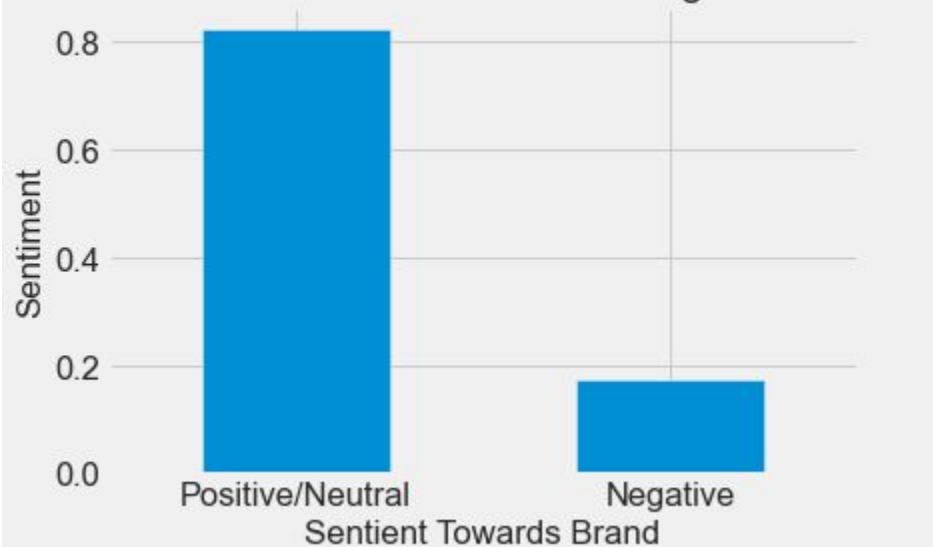


Sentiment - Binary Modeling

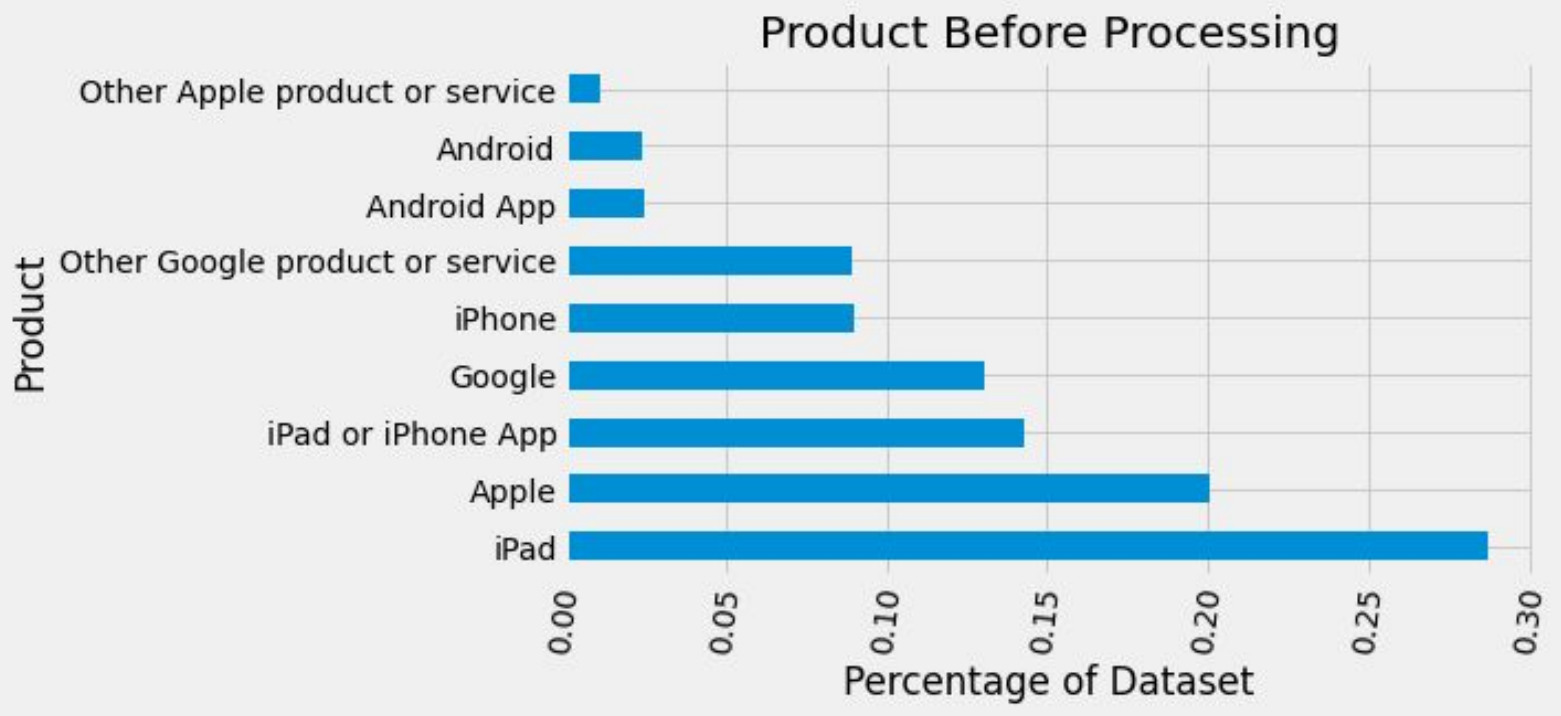
Tweet Sentiment Before Additional Negative Tweets



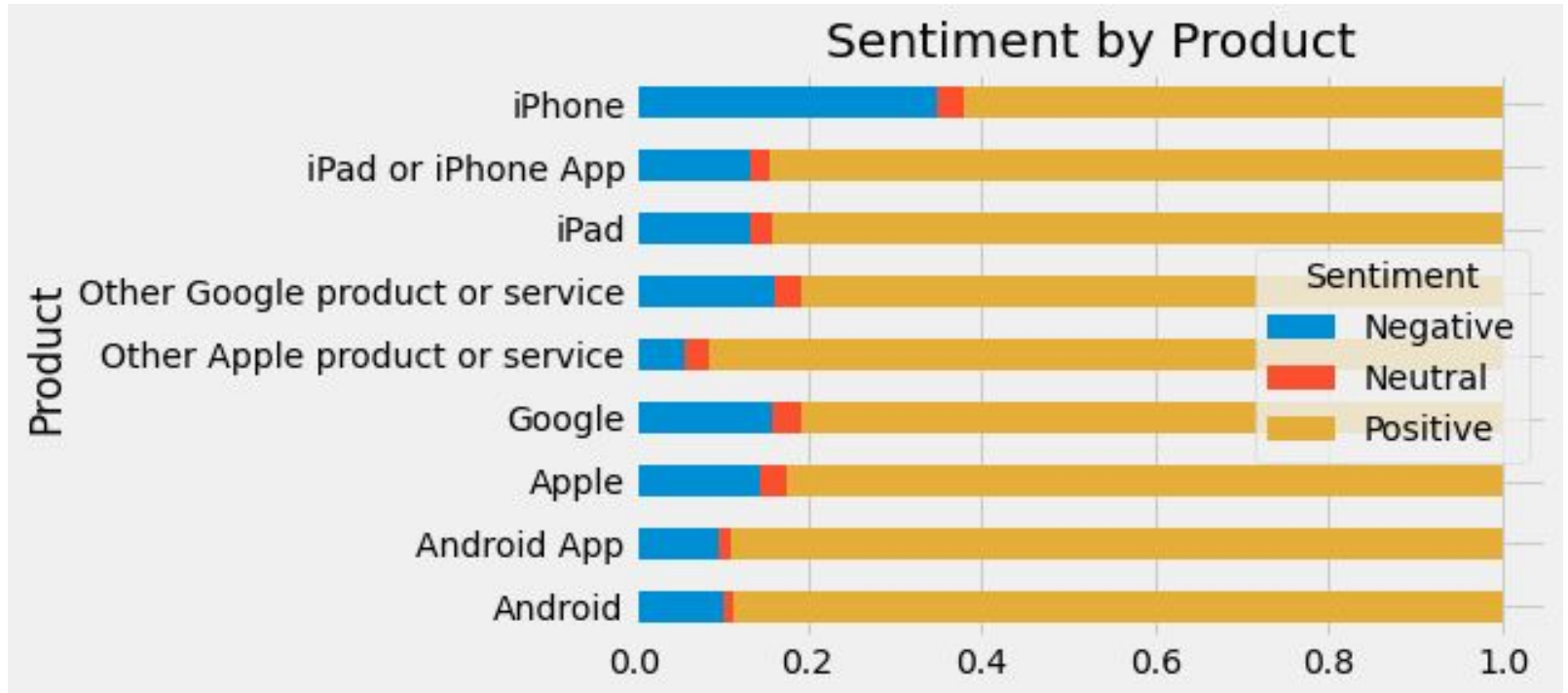
Tweet Sentiment After Additional Negative Tweets



Product Breakdown



Product Breakdown



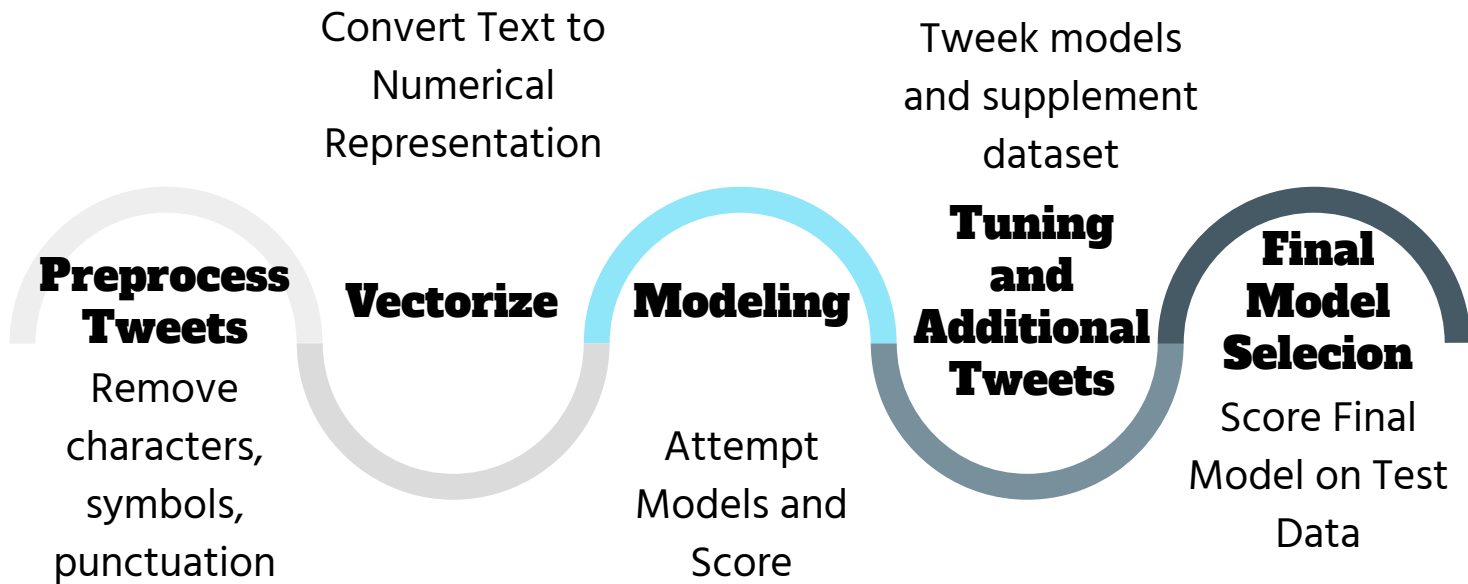


02

Project Metrics & Methods



Project Steps



Project Metrics

Recall - Casting the Net

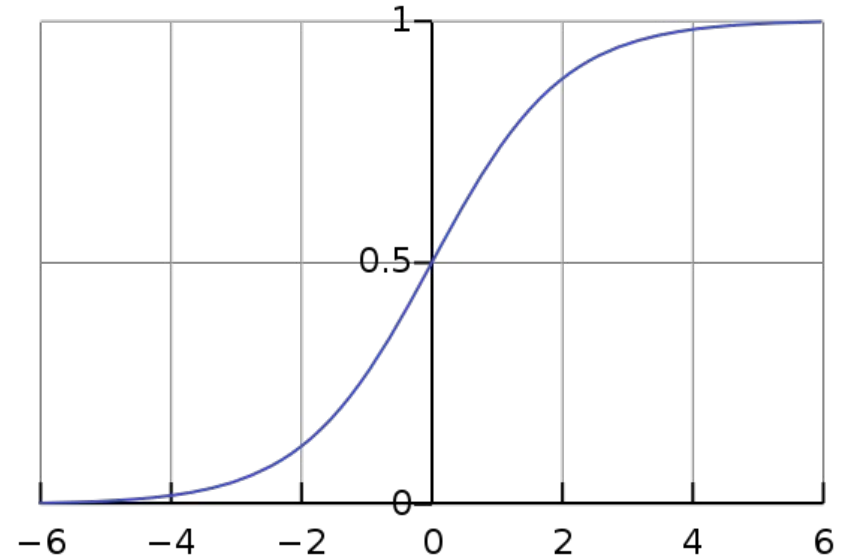
- Recall – What percentage of Negative Tweets Were Captured
- Precision – What percentage of the time were the predicted actually Negative.
- Optimized with F-Beta 2 Scorer. Sought a Balance Between Recall and Precision.



Final Model Selection

Tuned Logistic Regression Model

- Simple Model Nimble Model Using Bag-of-Words Vectors.
- Selected Models Using Cross-Fold-Validation and Grid Search
- Augmented the Training Data with an Additional 1219 Negative Tweets





03

Results & Interpretation

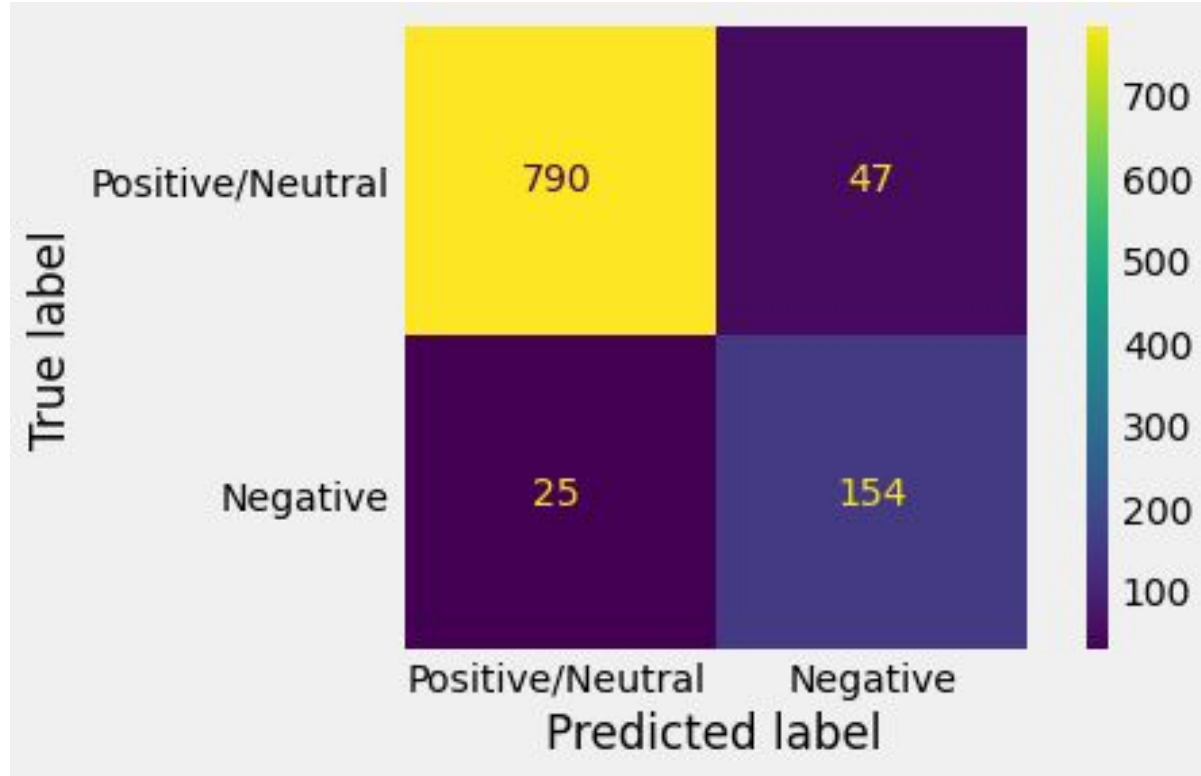
Final Model Results

Recall Score - 86%

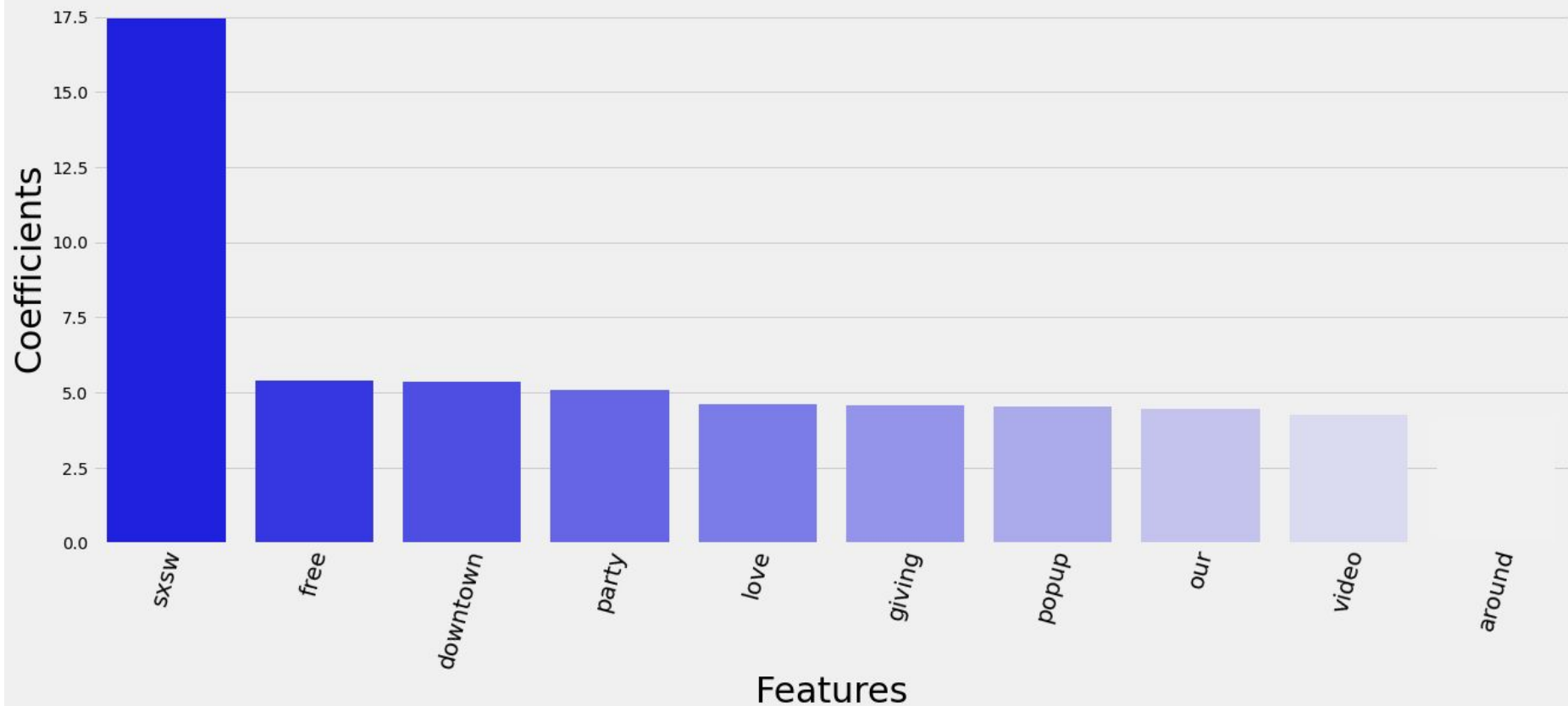
Precision - 77%

F1 Score - 81%

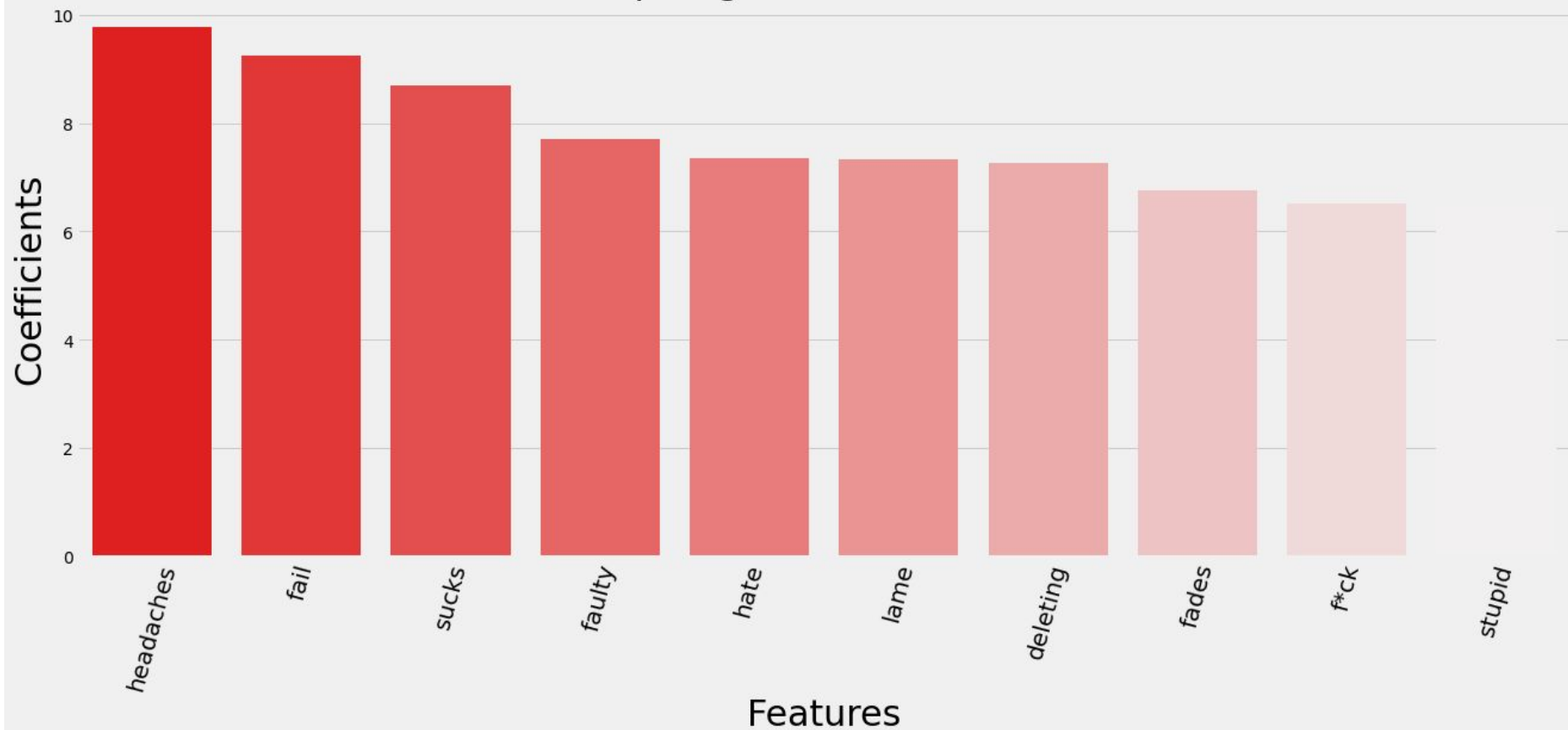
Accuracy - 93%



Top Positive Coefficients



Top Negative Coefficients







04

Conclusion & Future Work



Conclusions



- Attack the battery issue – major source of complaints.
- Throw some parties – people like free stuff, it can turn the tide of sentiment.
- People want to like brands at SXSW – it's a succes!





Future Work



- Analyze the use of emojis and GIFS in tweets.
- Gather location-based data in real-time to respond to negative and positive feedback on-site.
- Experiment with Sentiment forecasting – could any of these reactions be predicted by Twitter chatter in advance of the event?

