Twitter Sentiment Analysis

Sean Hart September 15, 2022

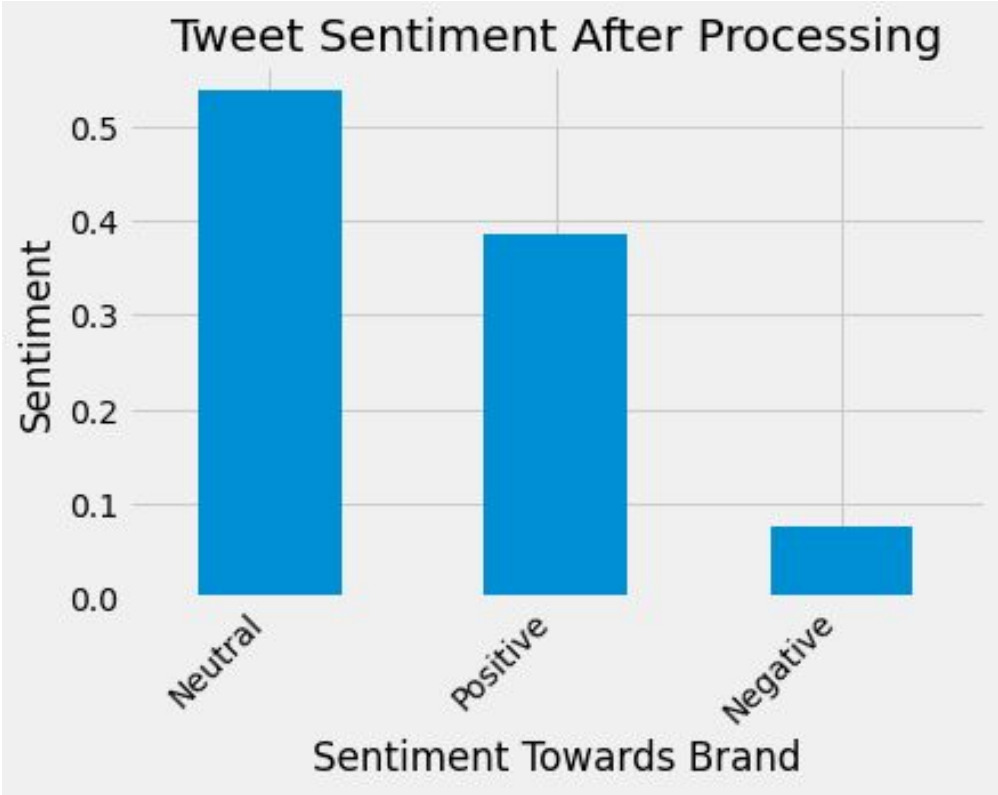
01

The Dataset

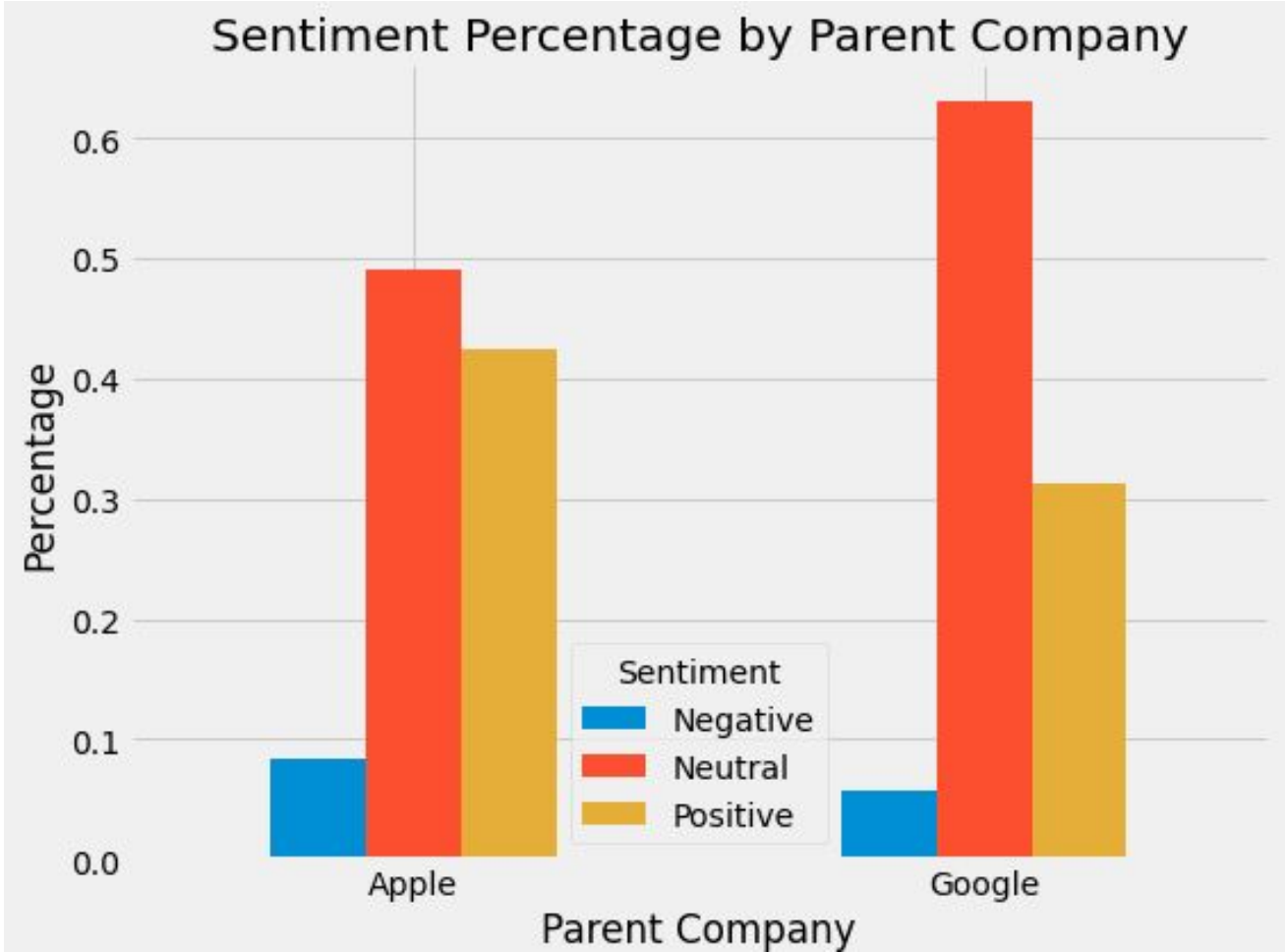
Initial Dataset

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

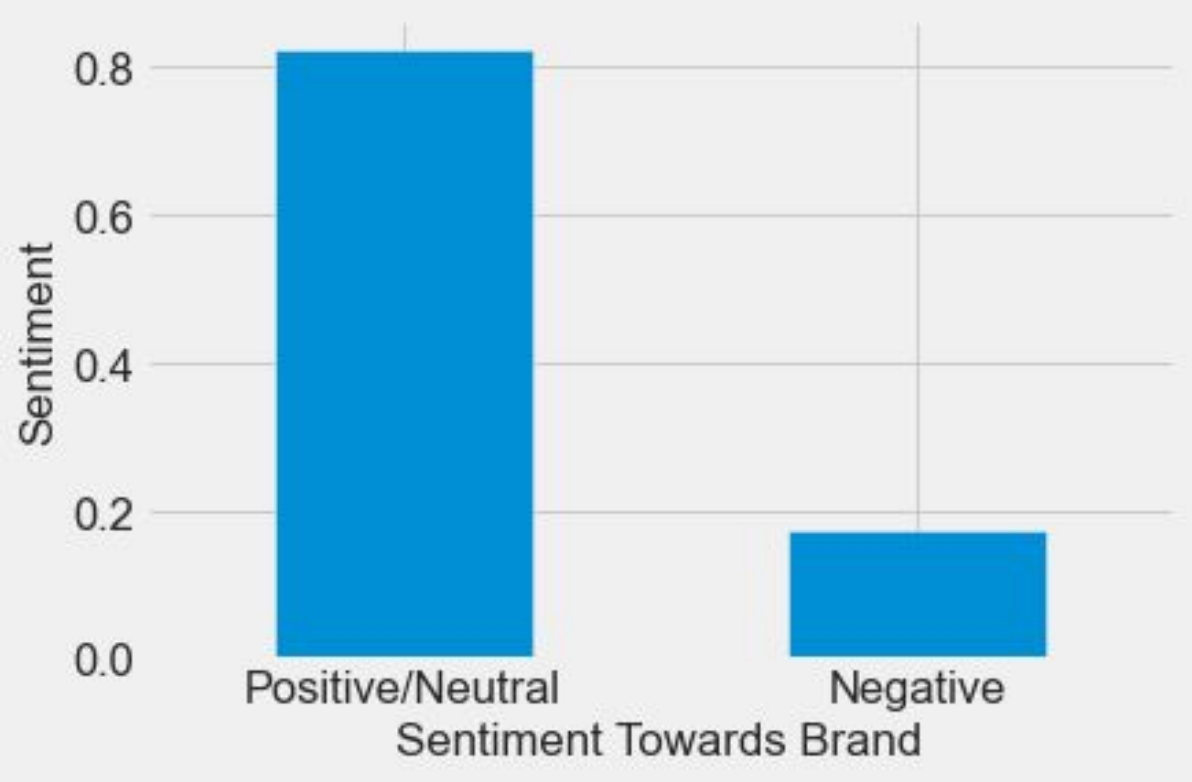
Sentiment Breakdown

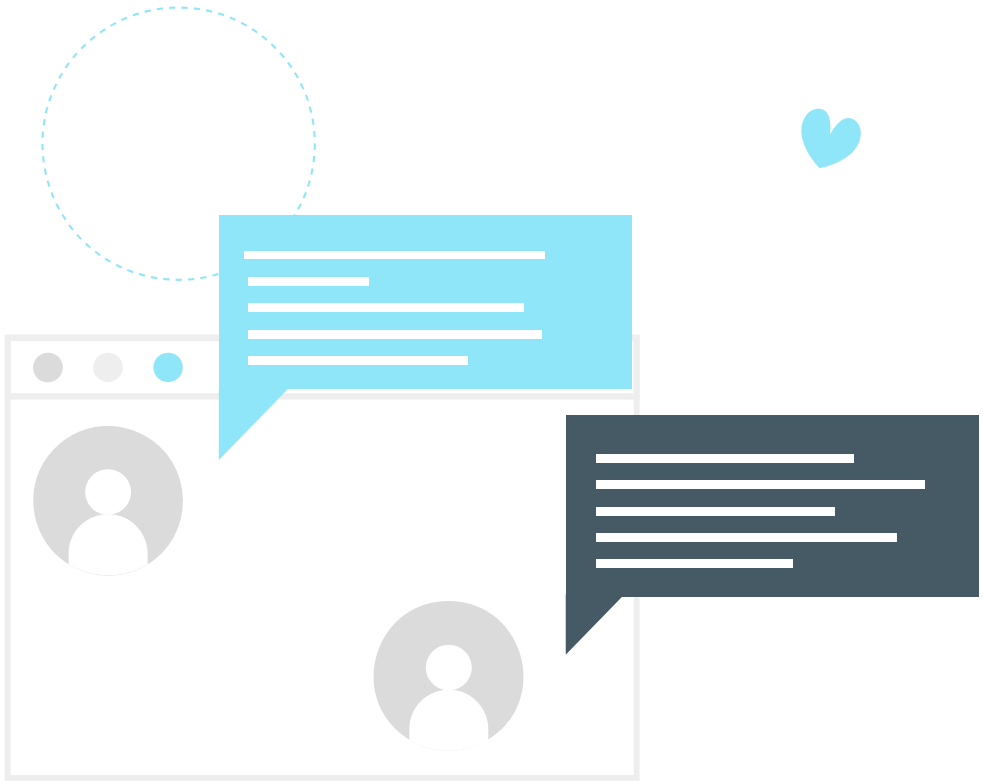


Company Sentiment



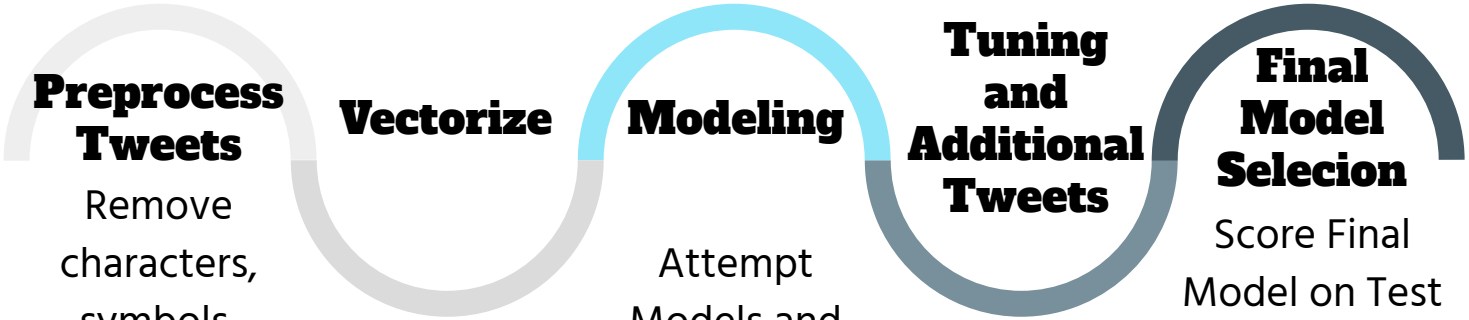
Sentiment - Binary Modeling





02 Project Metrics & Methods

Project Steps



Preprocess Tweets

Remove characters, symbols, punctuation

Vectorize

Convert Text to Numerical Representation

Modeling

Attempt Models and Score

Tuning and Additional Tweets

Tweak models and supplement dataset

Final Model Selecion

Score Final Model on Test Data

Project Metrics

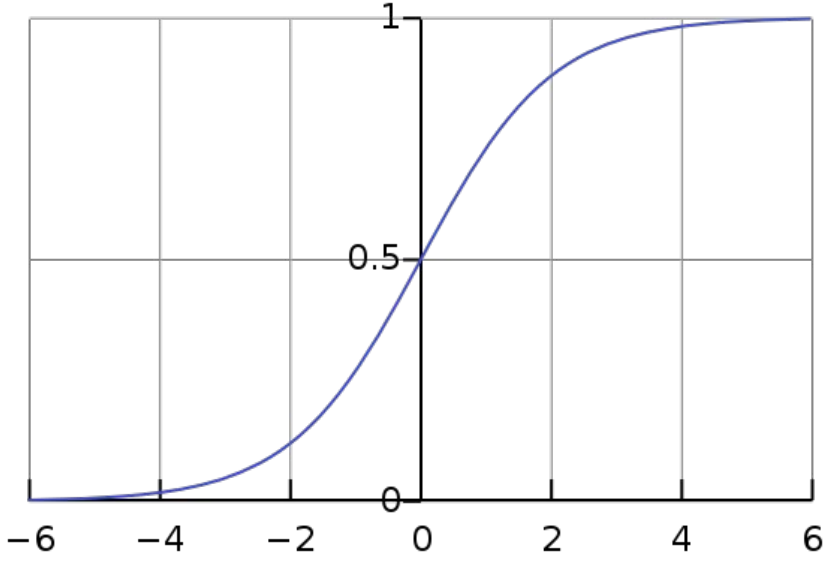
Recall - Casting the Net

- Recall – What percentage of **Negative Tweets** Were Captured
- Precision – What percentage of the captured Tweets **Were Indeed Negative**
- Lean Toward Recall **without** sacrificing Precision.

FInal Model Selection

Tuned Logistic Regression Model

- Simple Model Used a Bag-of-Words Approach
- Split into Training/Validation/Testing Data
- Augmented the Training Data with an Additional 1219 Negative Tweets



03

Results & Interpretation

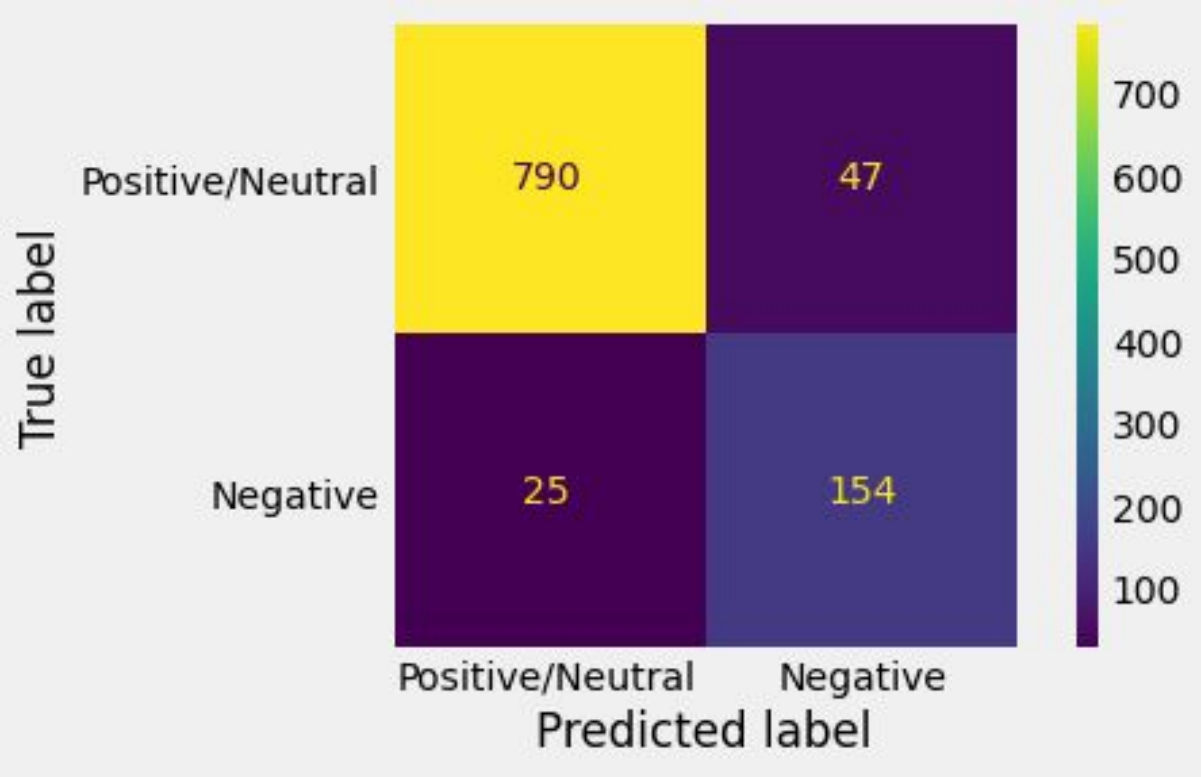
Final Model Results

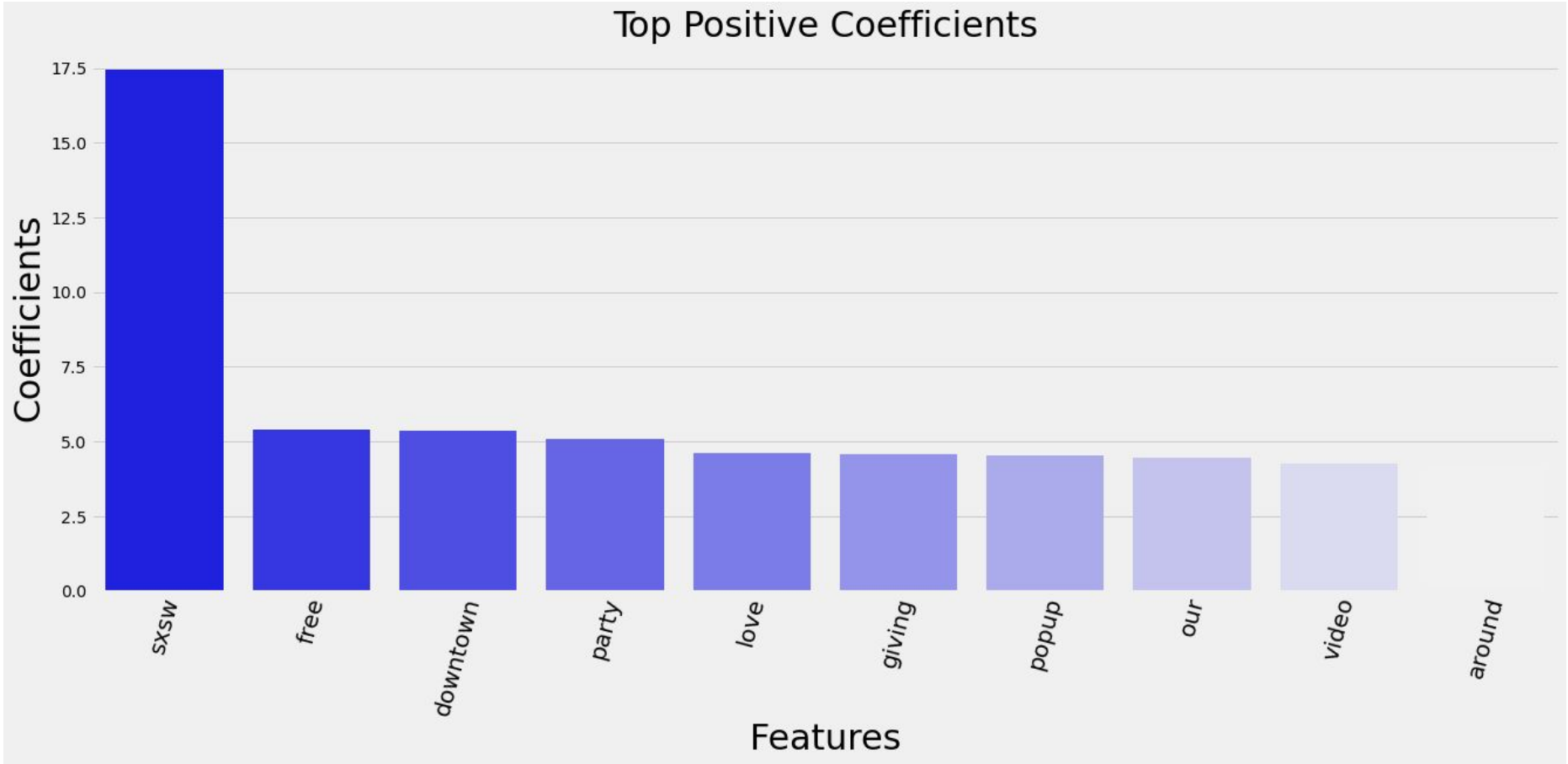
Recall Score - 86%

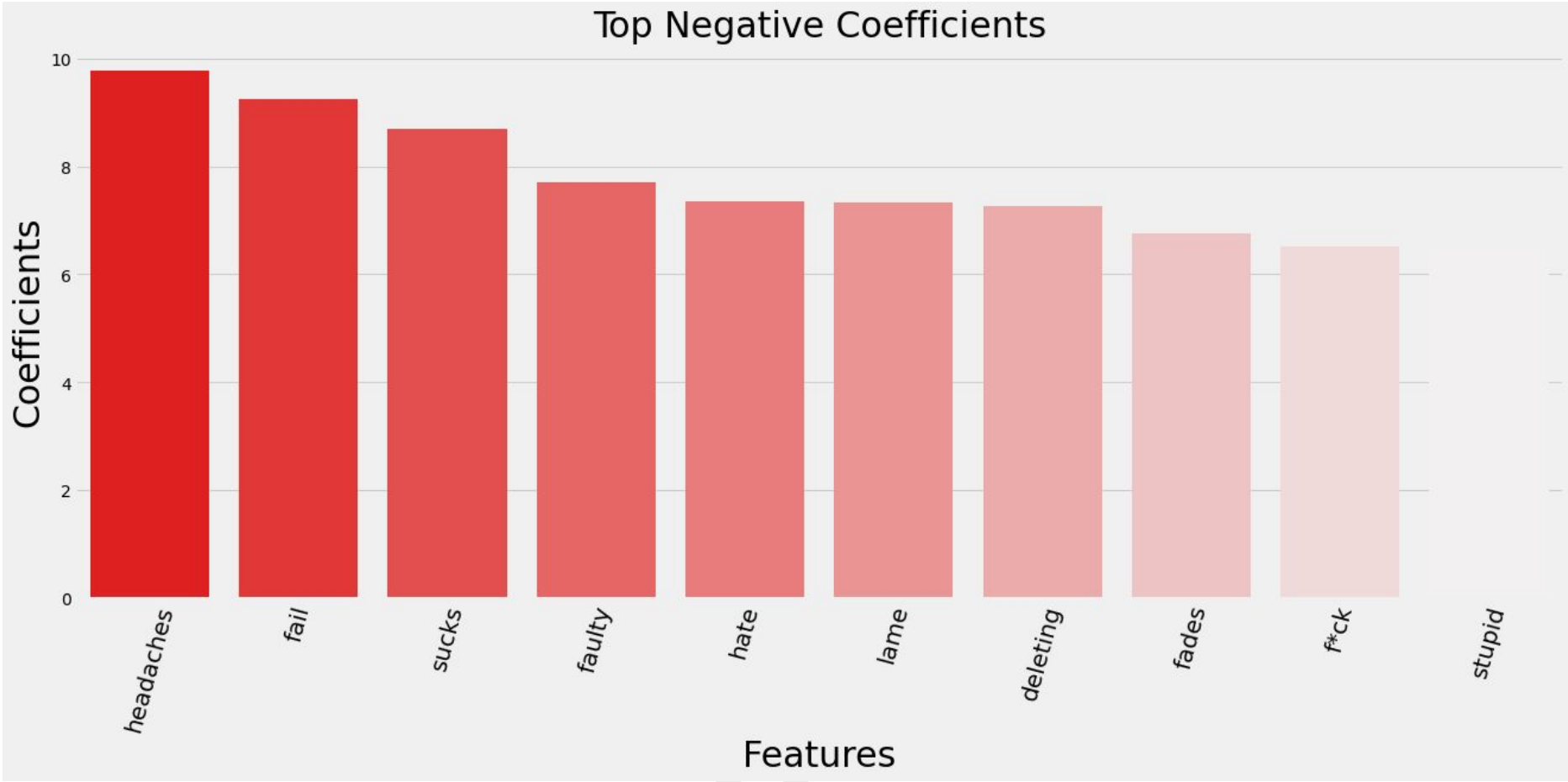
Precision - 77%

F1 Score - 81%

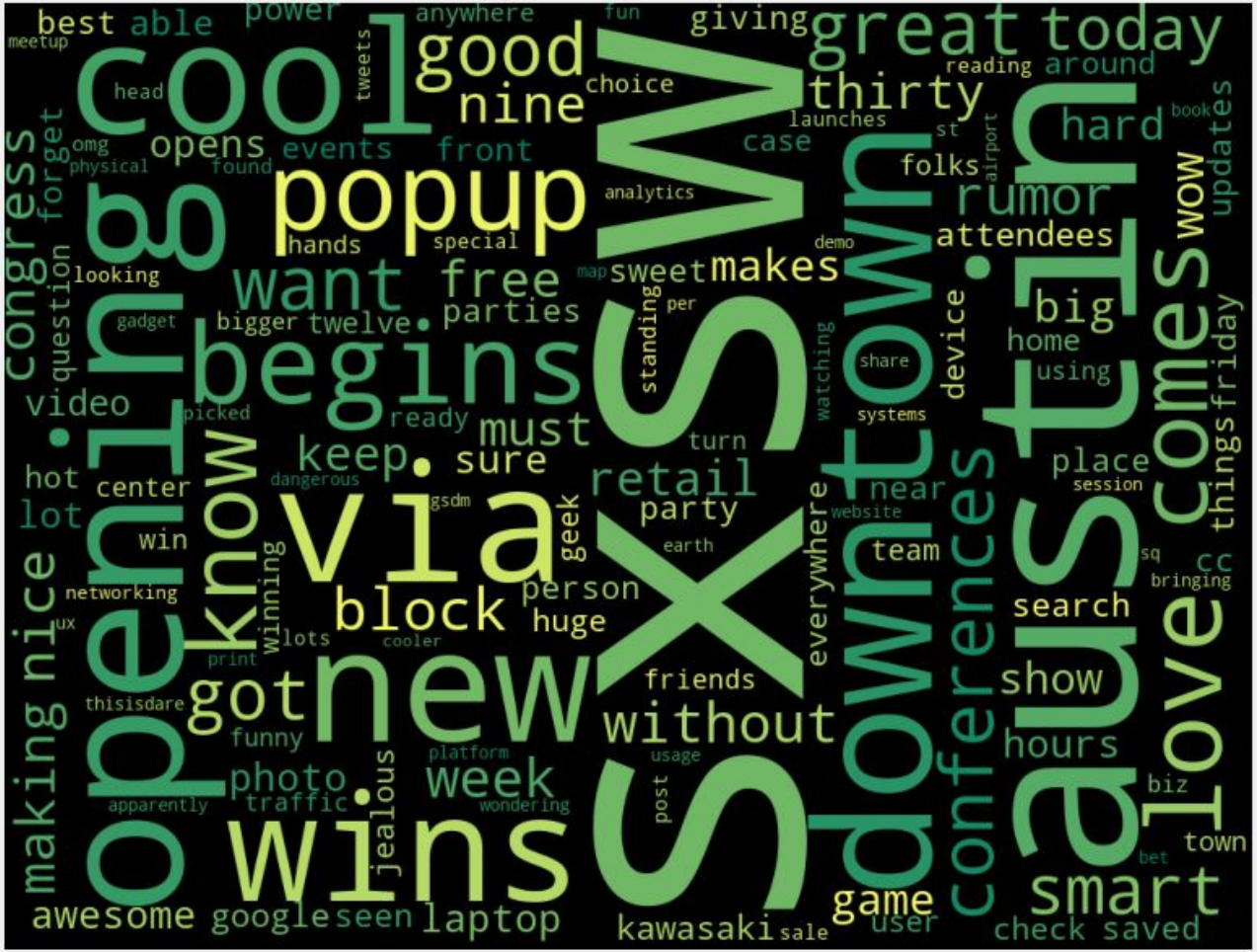
Accuracy - 93%



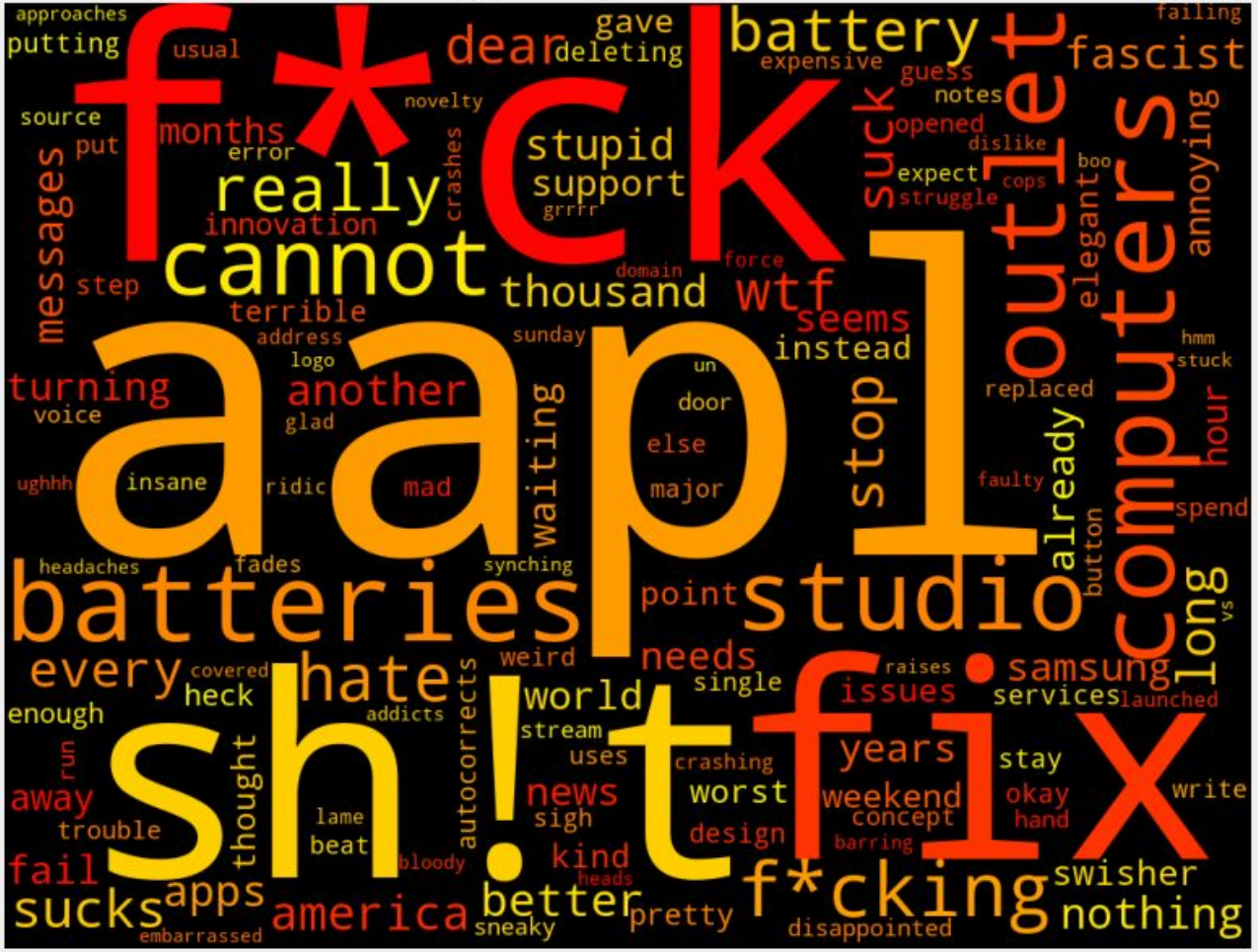

Top Positive Coefficients
Coefficients
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
sxsw
free
downtown
party
love
giving
popup
our
video
around
Features


Top Negative Coefficients
Coefficients
10
8
6
4
2
0
headaches
fail
sucks
faulty
hate
lame
deleting
fades
f*ck
stupid
Features

Apple Positive Words



Apple Negative Words



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 success!

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.
- Can we predict sentiment in advance? It's a difficult task to anticipate future problems.