Twitter Sentiment Analysis

Amanda Potter, Adam Roth

Project Goal

NLP model to identify negative consumer sentiment on Twitter

Data

Dataset of Tweets evaluated and labeled for positive or negative emotion towards a brand or product

Definitions

Tweets:

- Text string of up to 140 characters, generally includes abbreviated ‘words’

Sentiment:

- Did the user express positive or negative emotion in the tweet

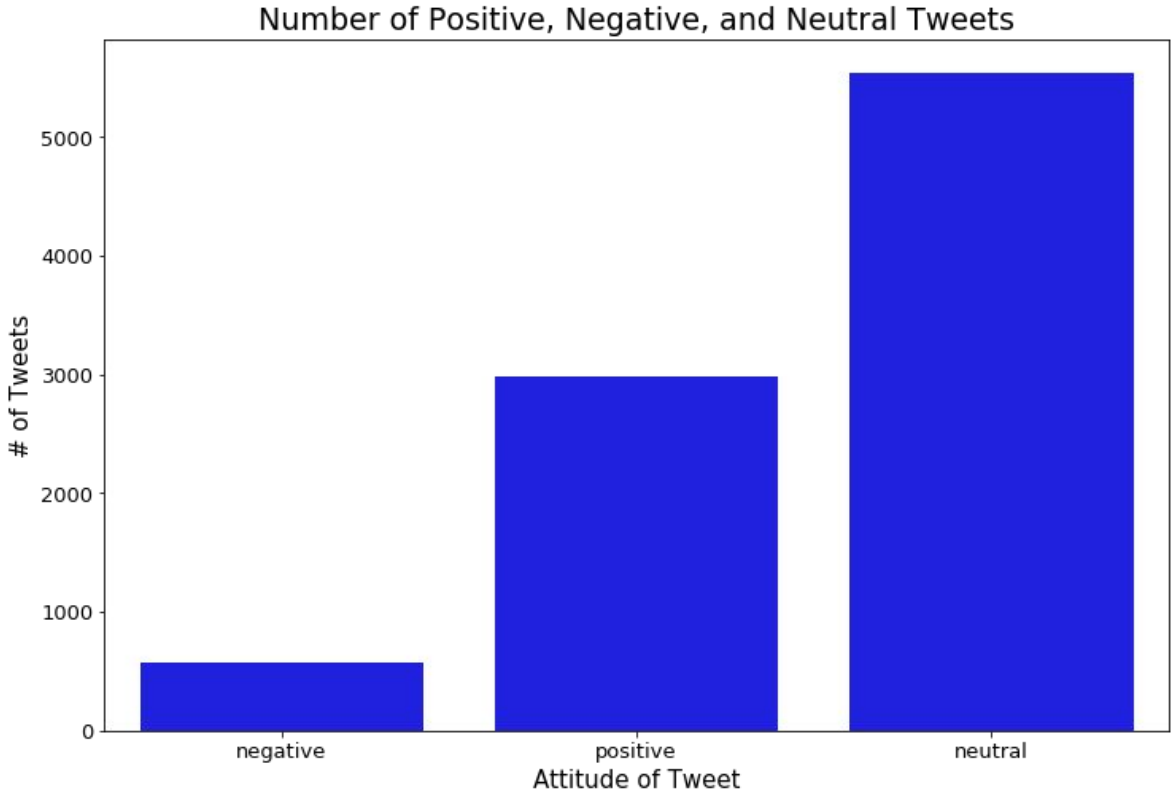
Polarity:

- Scale defining how negative or positive the emotion expressed in tweet

Initial Distribution

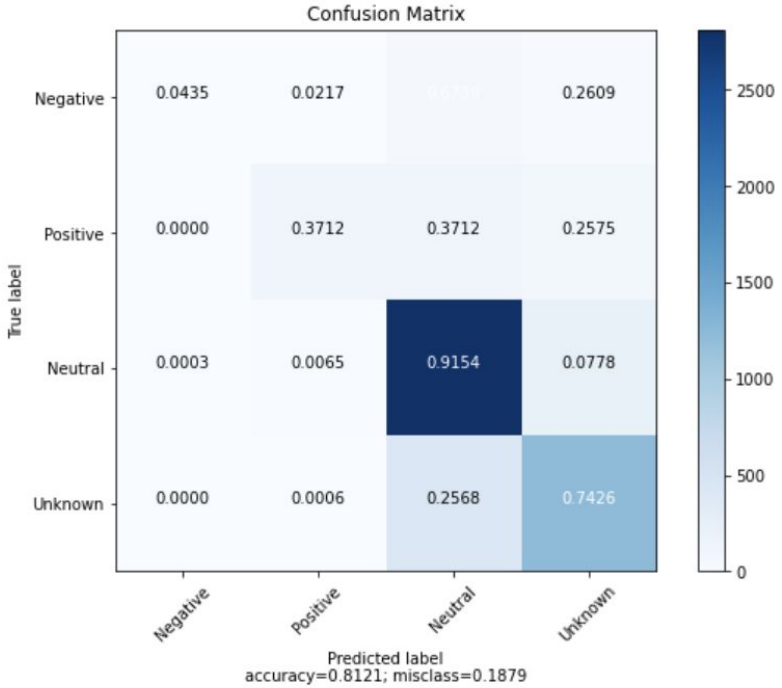
9093 tweets

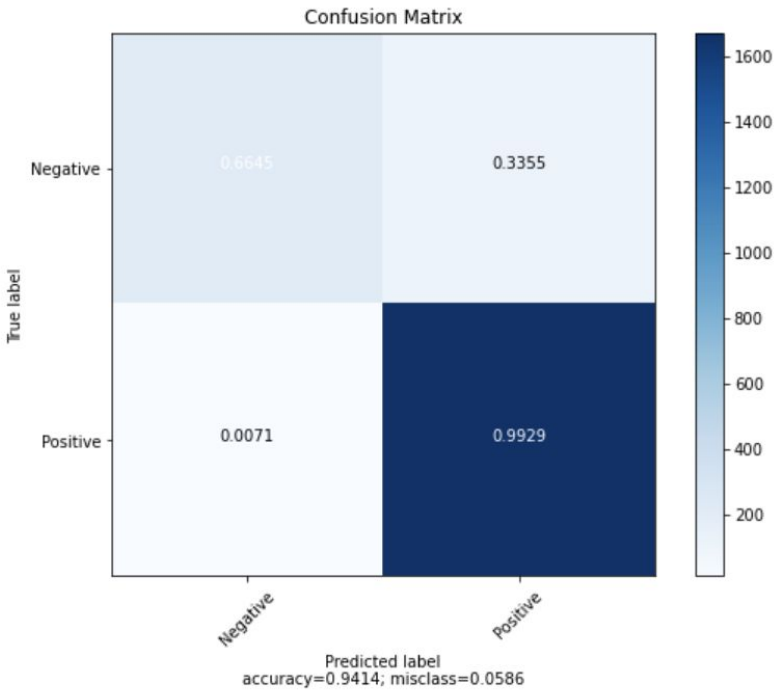
- 61% neutral/ not known
- 33% positive
- 6% negative



Initial Models

Two Initial Models



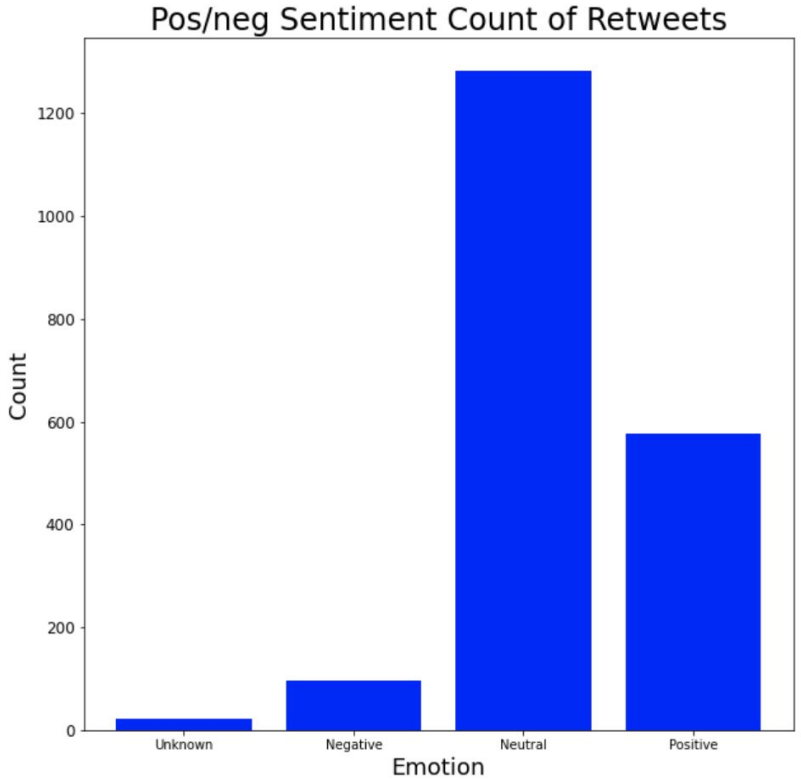


Refining the Model

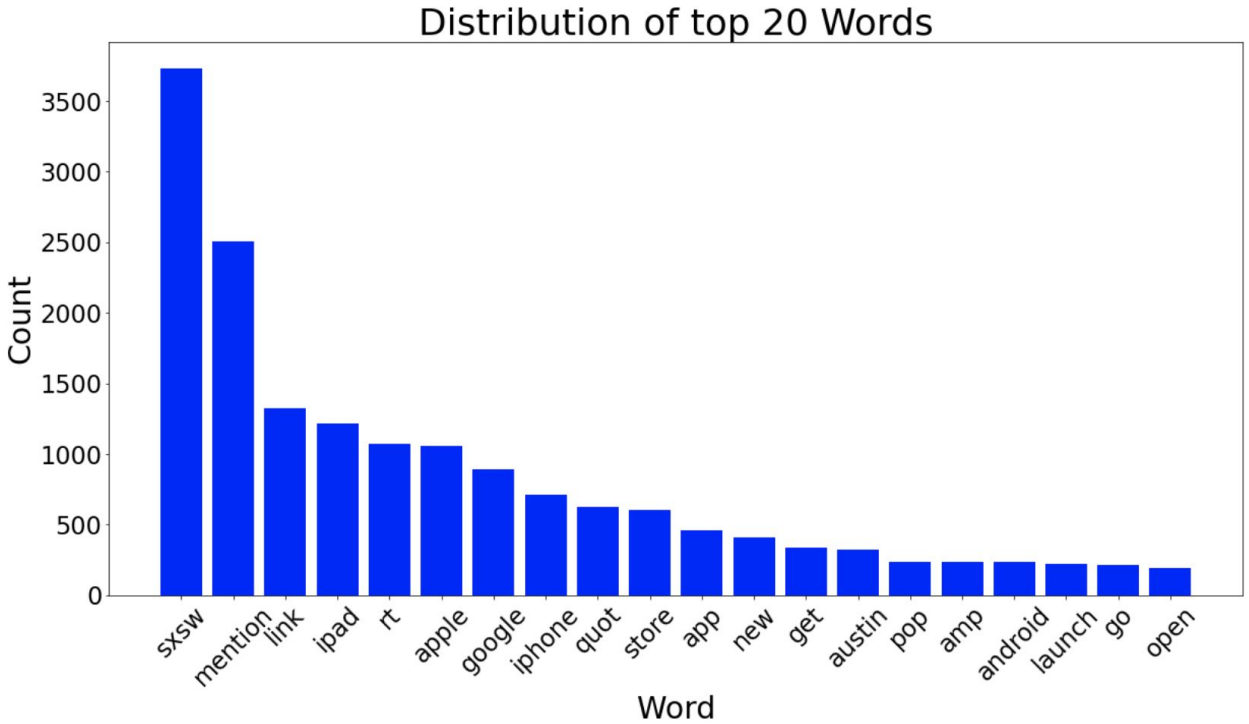
Lots of Retweets

9093 tweets

- 1977 were retweets
- Like the overall dataset, most did not have sentiment



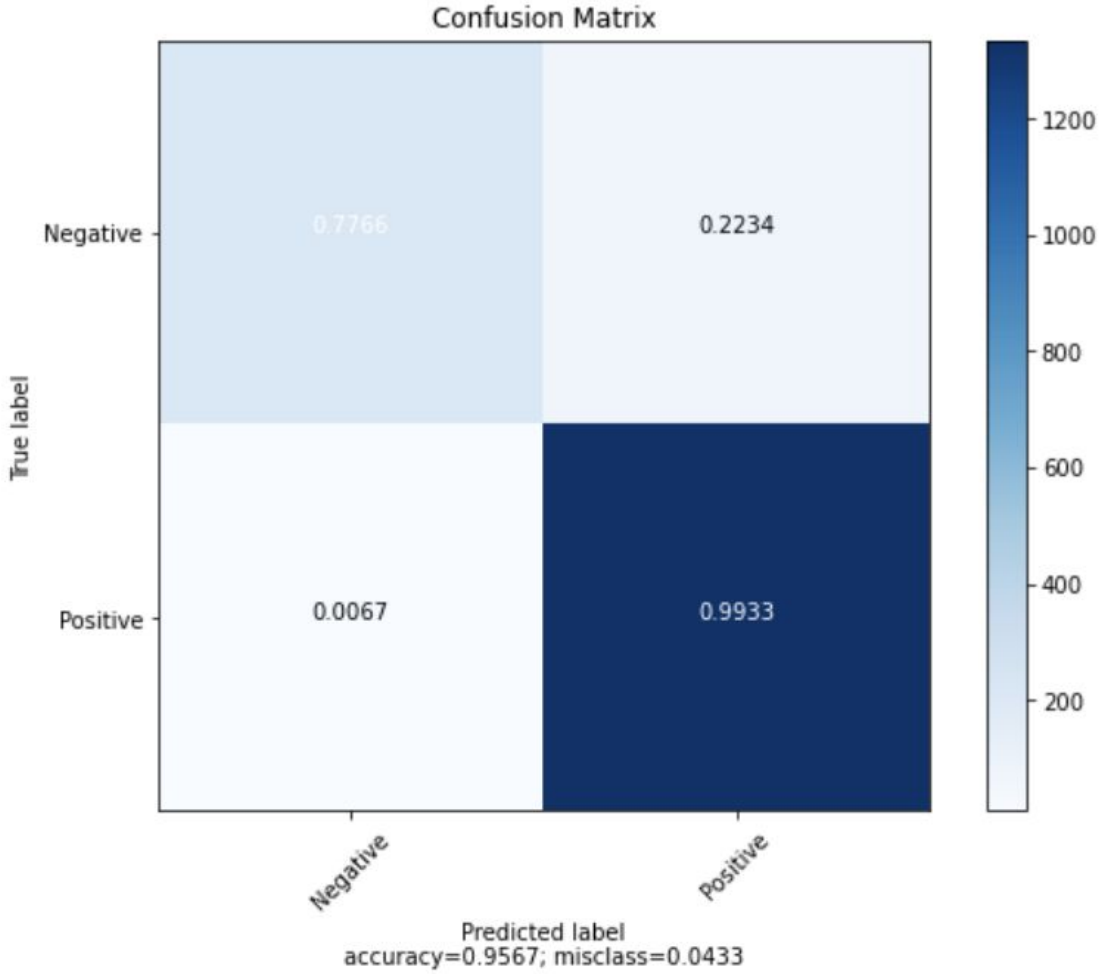
Additional Stop Words



Additional stop words were determined from the frequency distribution

Refined Model

Removal of retweets and additional stop words had marginal impact



Recommendations / Further Questions

Recommendations/Next Steps

- Use this model to direct customer service to tweets that have not been classified as positive
- Investigate use of ‘polarity’ score to improve model
- Identify more recent datasets to refine model
- Investigate use of Google’s BERT library

Contact

Amanda Potter

amandaspotter@gmail.com

Adam Roth

aroth605@gmail.com