

# NATURAL LANGUAGE PROCESSING – TWITTER SENTIMENT CLASSIFICATION

CATHERINE FRITZ – PROJECT 4

7/19/2021



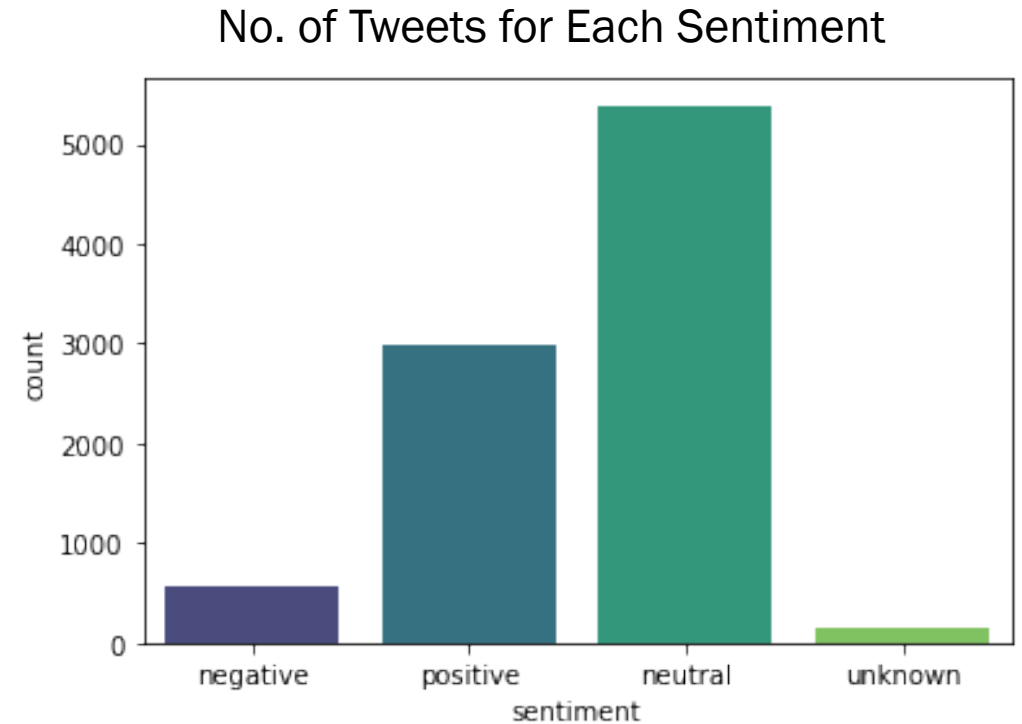
---

## **BUSINESS UNDERSTANDING**

- Businesses need to get feedback on their products
- Product reviews only one source
- Informal reviews on social media
- Predict sentiments based on social media posts

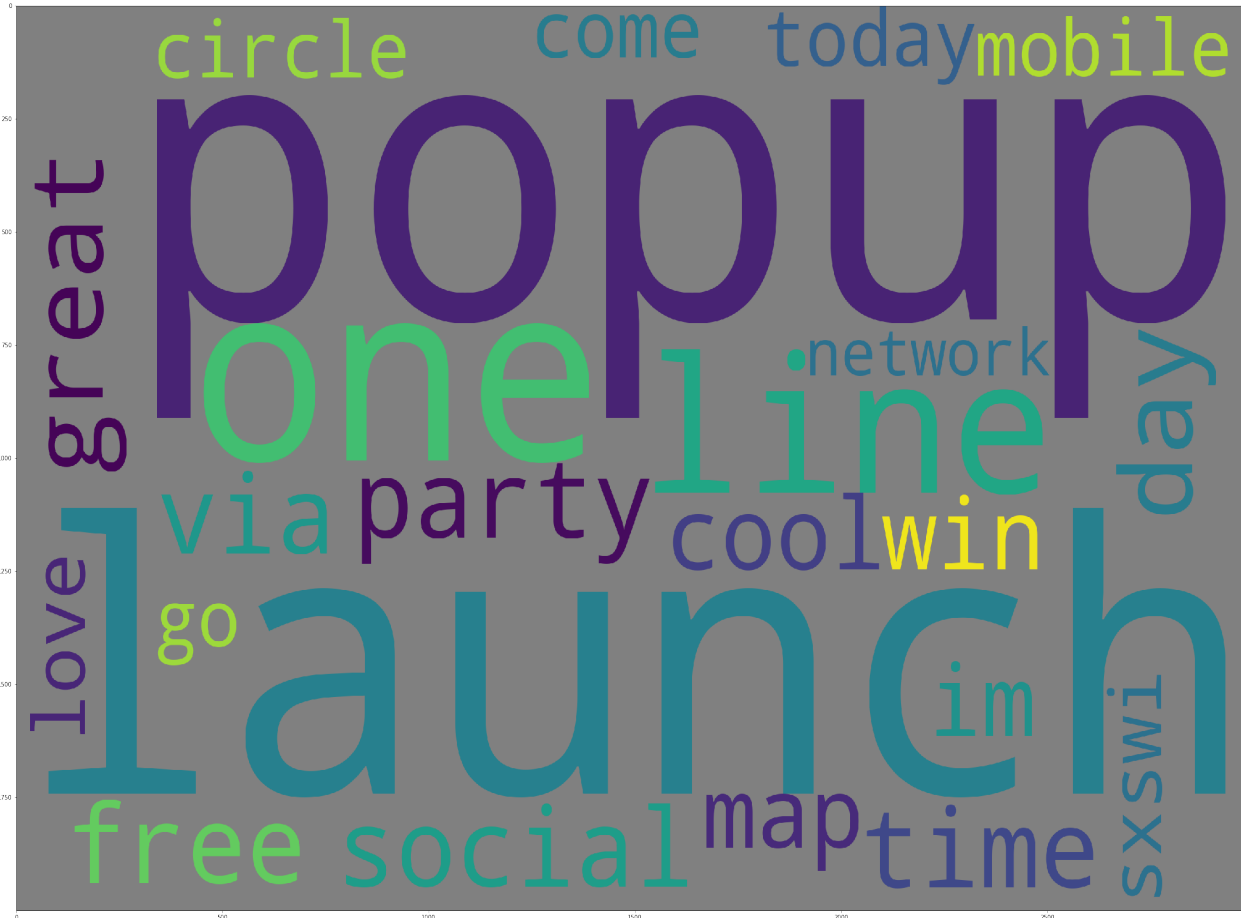
# DATA UNDERSTANDING

- Data classified by a human as positive, negative, neutral, or if the sentiment is unknown
- Subject of the tweets center around Apple or Android products.
- Data imbalance



# DATA UNDERSTANDING

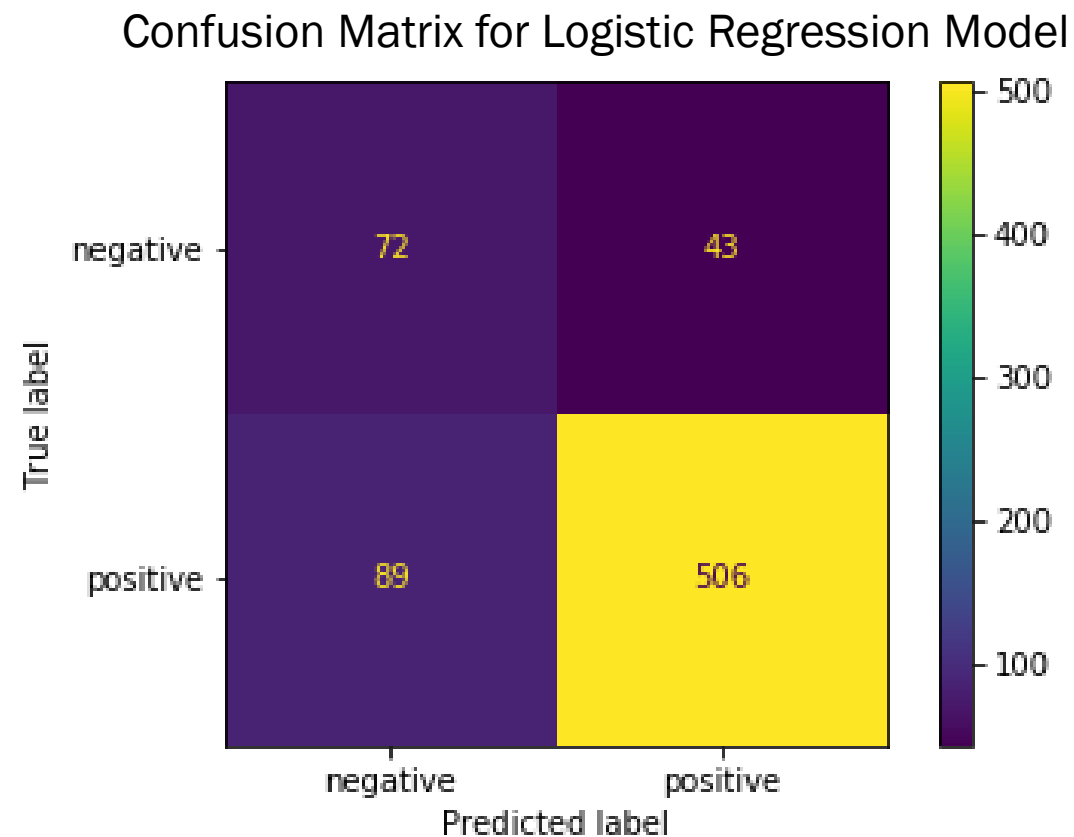
Top Words for Positive Sentiment



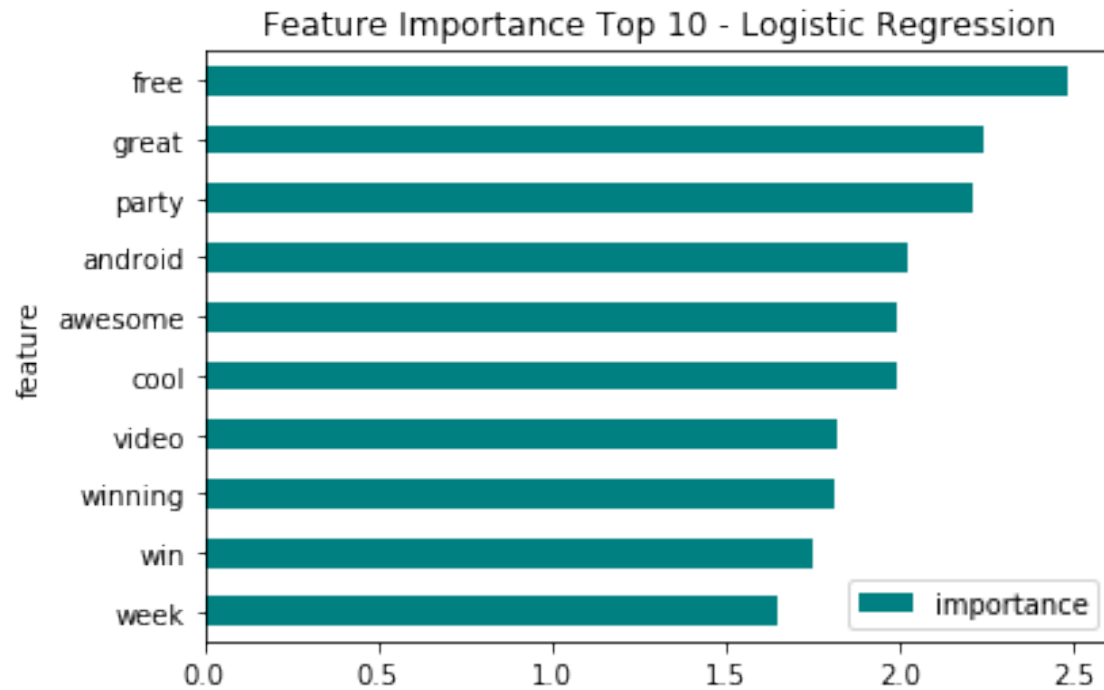
- To make the data suitable for modeling, the following steps were taken:
- Remove undesirable characters
- Remove Twitter specific text like @ tags
- Tokenize the text
- Remove common stop words for the English language and also specific to Twitter (like "RT" or "link").
- Lemmatize the stop words to consolidate similar words

# MODEL & EVALUATION

- Best Model: Logistic Regression
- Captures positive and some negative class data
  - True positive rate: 85.0%
  - True negative rate: 62.6%



# MODEL & EVALUATION



- Feature importance for logistic regression model:
  - Lists the top words the model used to determine classification

---

## CONCLUSION

- Based on the current model and data, can predict if a tweet is positive.
- Reduced ability to predict negative sentiment.

---

## RECOMMENDATIONS

- Tweets can in fact be a source to capture sentiments
- The model should be periodically retrained with new data to improve results, especially for the negative class
- The top words give a variety of insight into why people are posting tweets in the first place. For example "free" could indicate the occasion to tweet about a product, while "great" is how a tweeter feels. One interpretation could be one could encourage reviews by giving out free products.



---

## **FUTURE WORK**

1. Update model with more data from Twitter
2. Incorporate posts from other social media
3. Incorporate deep learning techniques



# THANK YOU

NAME: CATHERINE FRITZ

EMAIL: CMFRITZO@GMAIL.COM

GITHUB: @CMFRITZ

LINKEDIN: LINKEDIN.COM/IN/CATFRITZ

ADDITIONAL INFORMATION CAN BE FOUND AT

[HTTPS://GITHUB.COM/CMFRITZ/PROJECT\\_4\\_NLP](https://github.com/cmfritz/project_4_nlp)