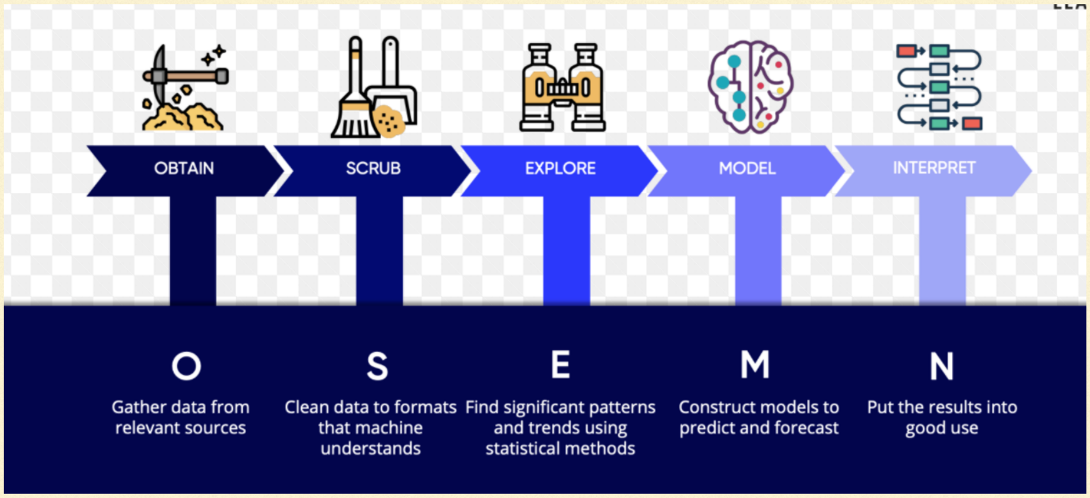
TWITTER SENTIMENT ANALYSIS

We have high emotions regarding different products, brands or services. My anger towards Apple company does not benefit me but it can benefit the Apple company, They can use my experience to improve their customer services.

GOAL

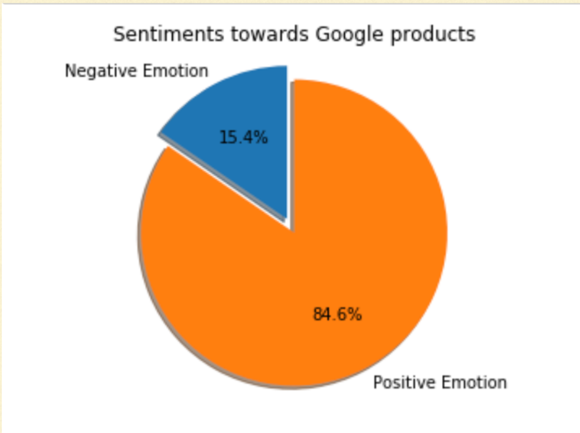
1. Build a model that will predict if a tweet is positive, negative or neutral.

2. Find the best predicting model.



METHODOLOGY USED

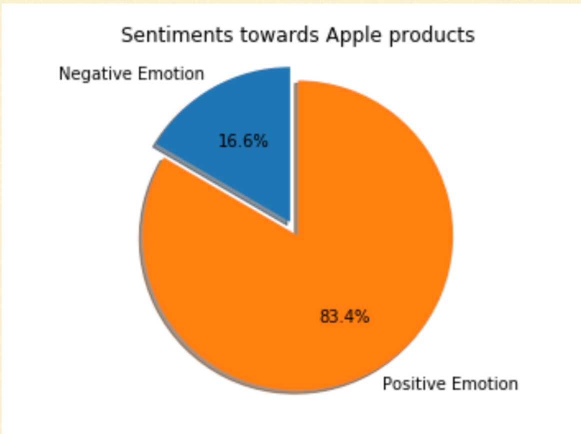
TWEET SENTIMENT TOWARD GOOGLE PRODUCTS AND SERVICES



- **85% POSITIVE EMOTION.**
- **15% NEGATIVE EMOTION**

I learned that you are more likely to talk about your experience if you are either extremely dissatisfied or the opposite. Which makes me think that either the dataset is not balanced or people are mostly happy with google products and services.

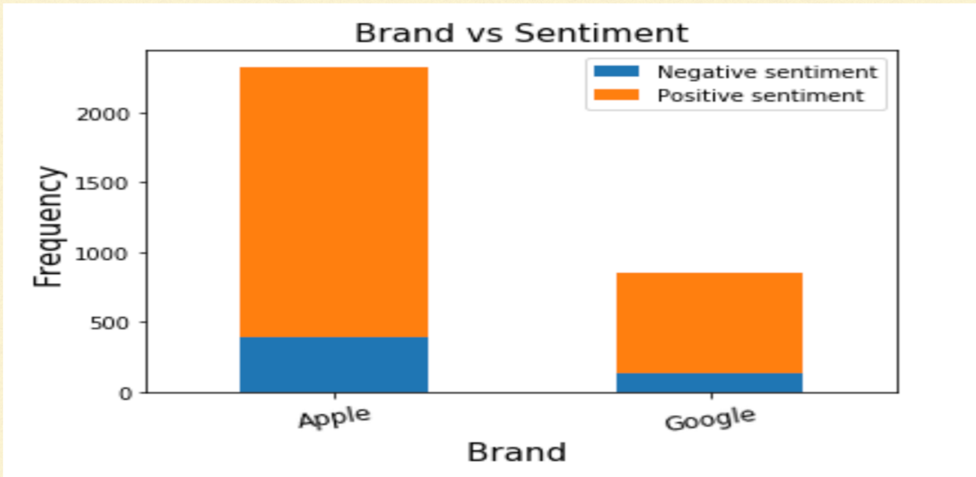
TWEET SENTIMENT TOWARD APPLE PRODUCTS AND SERVICES



- **83% POSITIVE EMOTION.**
- **17% NEGATIVE EMOTION**

I can tell the same thing about Apple products. Happy customers make up the high majority of this group.

SENTIMENT VS BRAND



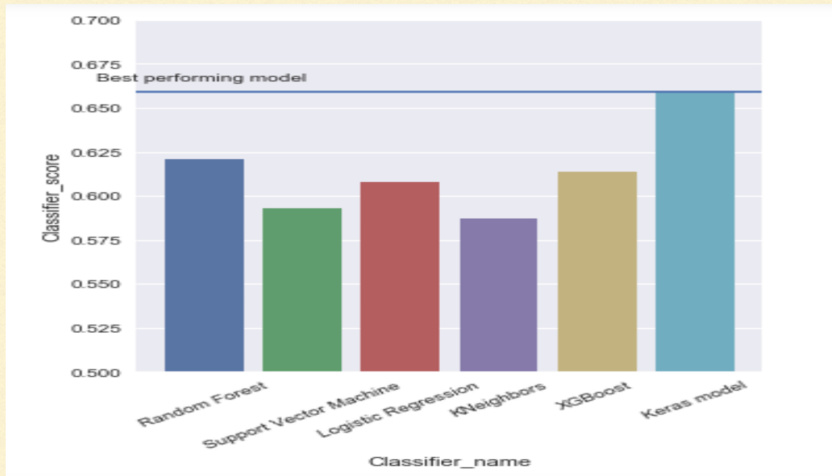
If we compare Apple and Google inspired sentiments, we can see that there are slightly more unhappy people in the Apple customers group than Google group. We also have a lot more people talking about Apple than Google.

After I cleaned the data, I built some models using different classifiers and I compared them using the accuracy metric.
Accuracy is the percentage of examples correctly classified: true samples divided by total samples. In our case accuracy is correctly predicted tweets, positive or negative out of the total number of tweets.

Over 0.85 accuracy means that over 85% of the tweets were correctly labeled as either positive, or negative. Since the random change of getting a correct label is in this case similar to flipping a coin, over 50%, an accuracy of over 85% is a pretty decent one.
Precision would be an even more useful metric since it gives you information about false positives. In this case false positives are negative tweets predicted to be positive. I included using the precision metric in my future work.

MULTI CLASS MODELS

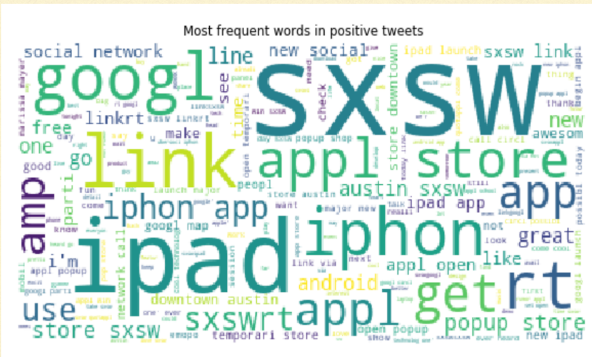
- Keras was the best performing Model with an Accuracy Score over 0.65.

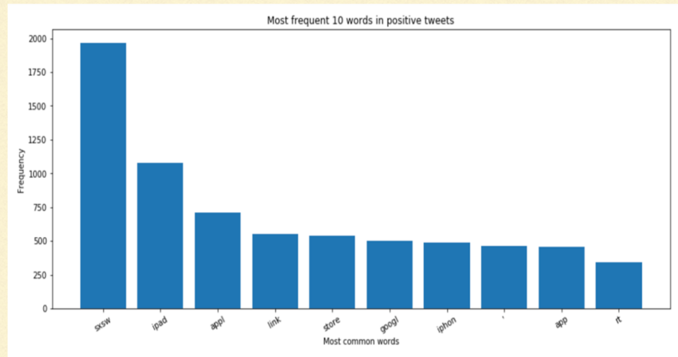


I also build models that will classify tweets into four categories and the accuracy for my best performing model is over 0.65. In this case the probability of randomly guess the label is 0.25 so labeling correctly in 65% of cases looks decent again, but there is room for improvement.

FINDINGS

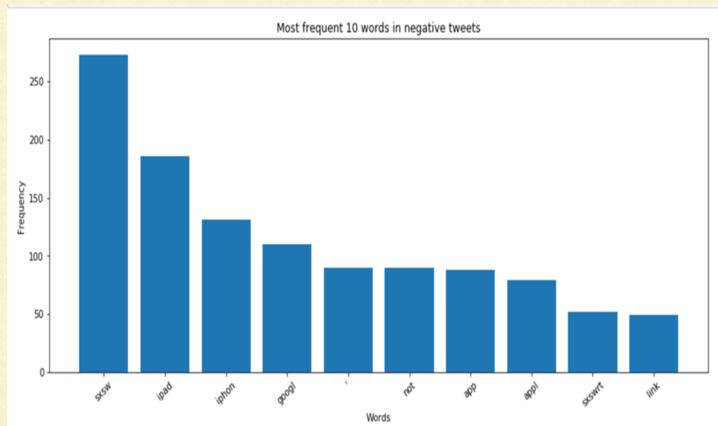
Most common words in positive tweets:





FINDINGS

Most common words in negative tweets:



As you can see the most frequent words in both types of tweets are almost the same. This makes it harder for the model to correctly classify.

Recommendations

TELL ME HOW YOUR CUSTOMER SERVICE IS, AND I'LL TELL YOU IF YOU MAKE MONEY!

A good tweet sentiment predicting model will help customer service identify:

1. Negative tweets in order to fix or to understand better why the negative emotion toward the brand/product.

2. Positive tweets for show casing customer experience and training purposes.

FUTURE WORK

Improve the model:

1. Balance classes on the 2-class models.

2. Find a way to get different metrics like Precision in the Keras model.

Precision offers a good window into understanding false positives. In our case false positives are real negative tweets that are predicted to be positive by our model.

THANK YOU