Natural Language Processing

By: Susanna Han

Purpose of NLP Analysis:

An analysis using a Natural Language Processing model allows us to analyze text data, which makes analyzing the score of the 9,093 product tweets possible. Finding the correlation and important features of the tweets help us to provide insight to what words makes up a positive review or negative review.

The data set we used has three columns which includes the tweet (the review), which product the review is referring to, and emotion, whether or not the review was positive, negative, or neutral.
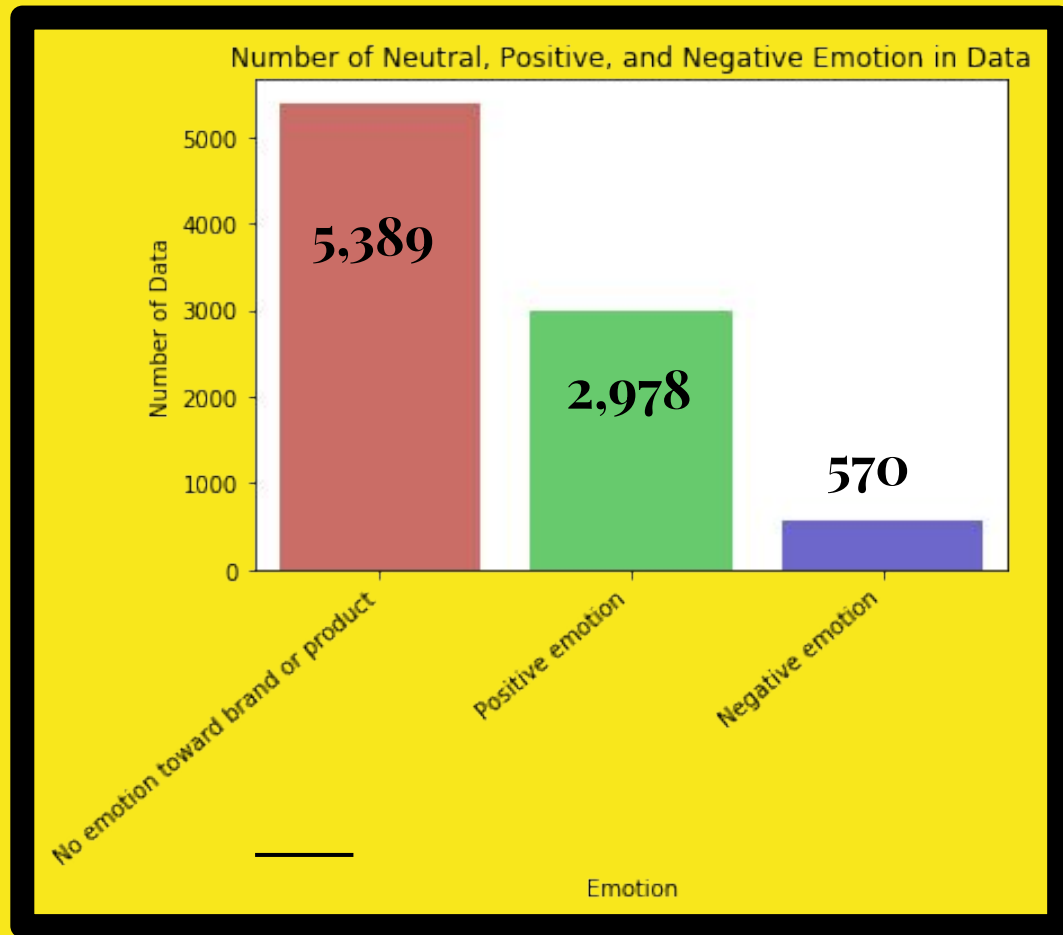
Data collected

No emotion (Neutral)

59%

Positive emotion

33%

Negative emotion

6%

I can't tell (Unknown)

2%

Number of Neutral, Positive, and Negative Emotion in Data

Tokenize → **Stop Words** → **Regular Expression**
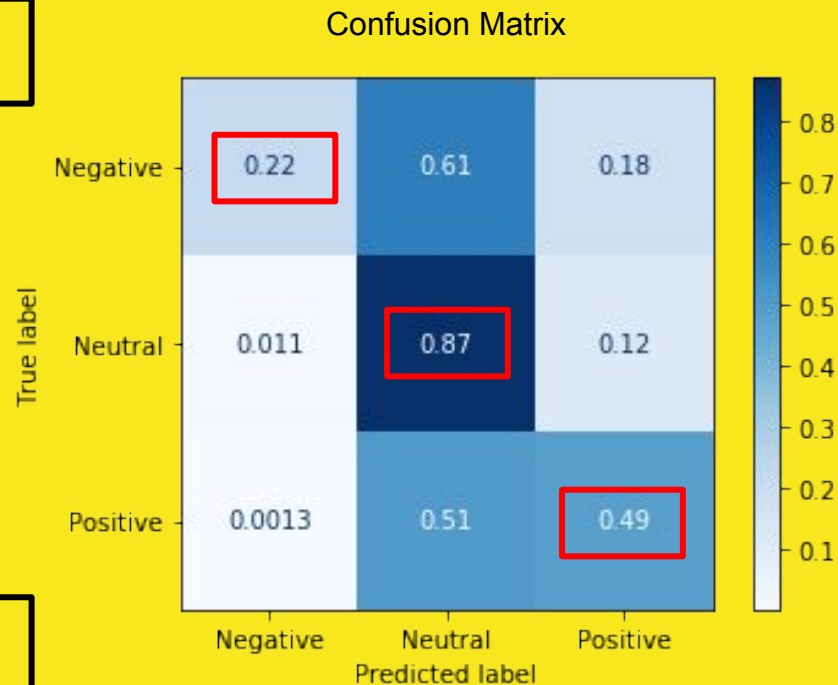
Separates each word and punctuation

A list of common words that do not add meaning to a sentence.

Finds and replace patterns such as hashtags and links

Methods:

Tuned Model

- Average accuracy of 70 %

- Overfit model

- Training text accuracy of 96%

- Imbalanced data shows in model.

Confusion Matrix

Most Frequent Positive Tweet Words

Most Frequent Negative Tweet Words

Word cloud of the common words used in positive tweets and negative tweets

Shows us the vocabulary used in the tweets.

Insight & Recommendations:

1. There are a lot of negative responses correlated with the ipad2. – Looking into the correlation between negative responses and the ipad2.

2. In both the negative and positive responses people mention the design. – The design is important. A lot of people tweeting about apple and google products are interested in the design.

3. There are positive responses towards the ipad/iphone.

Future Work :

- Use different methods such as Linear SVC, Support Vector Classifier.

- Create more vibrant and clear visuals to help understand the data in a deeper way.

- Try different tuning parameter methods such as GridSearchCV

Thank You!