Twitter Sentiment Analysis

Prabagini Thambinathan

Objective

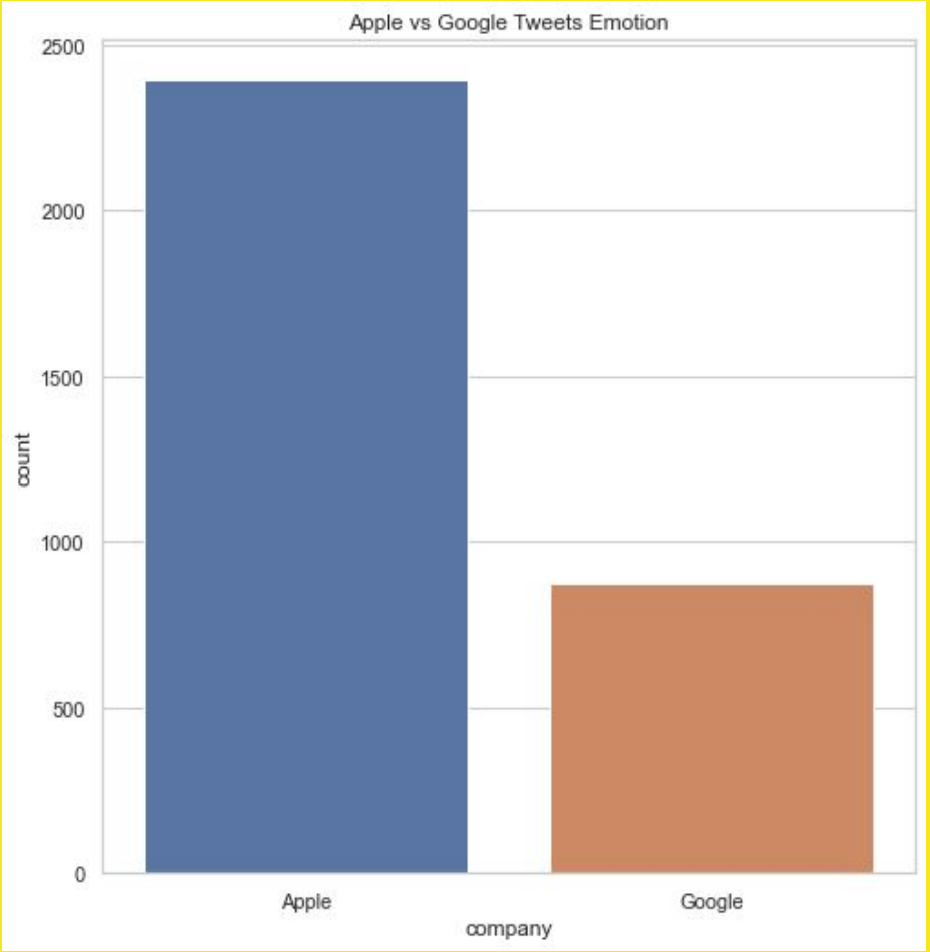
- The objective of this project is to perform supervised sentiment analysis on twitter tweets about Google and Apple products using Natural Language Processing and the model can rate the sentiment of the tweet based on its content.

Data

- Obtained the dataset from CrowdFlower
- Source: https://data.world/crowdflower/brands-and-product-emotions
- This dataset has tweets of human raters rated the sentiment in over 9000 Tweets as Positive, Negative,Neutral

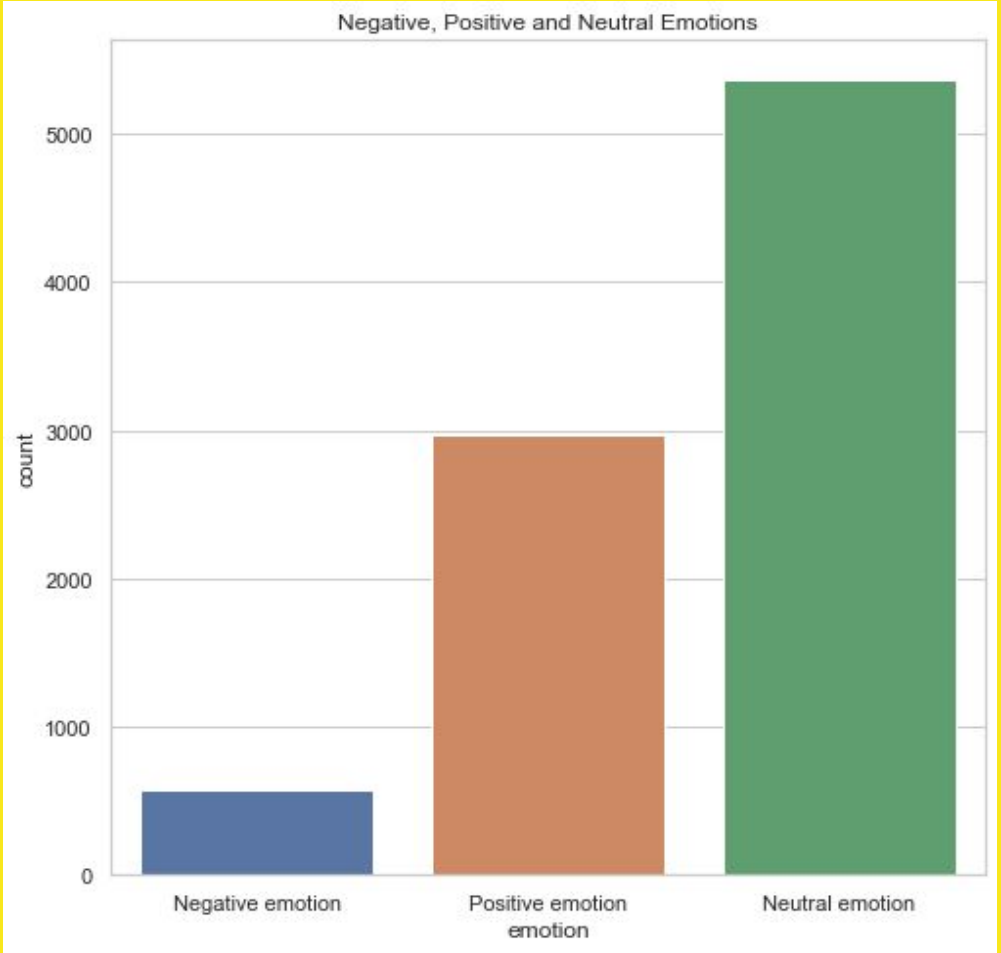
Apple Vs Google Tweets

From the graph it tells us that Apple company products has more tweets than Google products.



Tweet Sentiment

Neutral sentiment has more than 60 % of tweets compared to positive and negative tweets.



Methodologies

- Removal of stop words
- Lemmatization technique
- Vectorization - Term Frequency - Inverse Document Frequency
- Tokenizer

Modelling

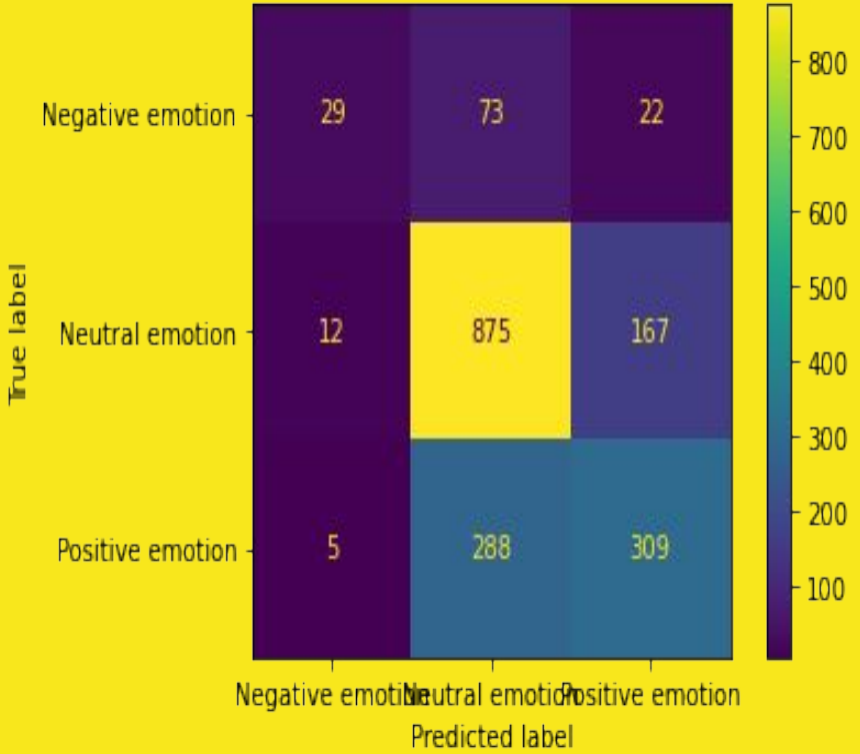
- Logistic Regression
- Multinomial Naive Bayes
- Random Forest
- Decision Tree
- XG Boost

Modelling

- Cross validation
- SMOTE technique for multiclass imbalance
- Performed Deep Neural Networks with different epochs

Results

- Most of the model had trouble predicting positive or negative sentiment because it was overfitted
- Random Forest Model was relatively successful in extracting both positive and negative sentiments



Next Step

It is recommended to gather more data that has a higher percentage of labels that indicate positive or negative sentiment.

Further tuning of models using grid search.

Adding more embedding layers and hyper parameters in neural network model to avoid overfitting.

Thank you

Prabagini Thambinathan

Email:mail4prabagini@gmail.com