SENTIMENT ANALYSIS

A Natural Language Processing Model

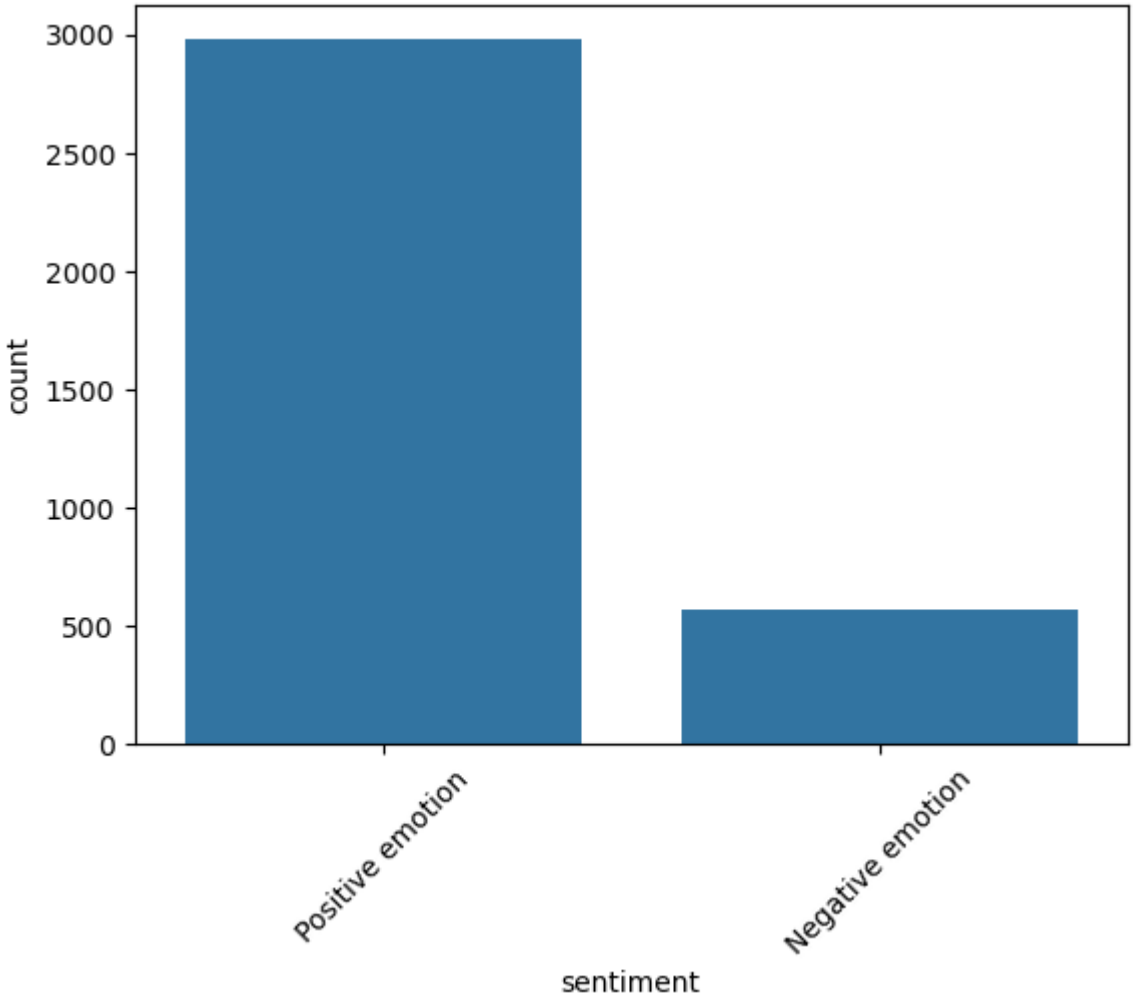
Introduction

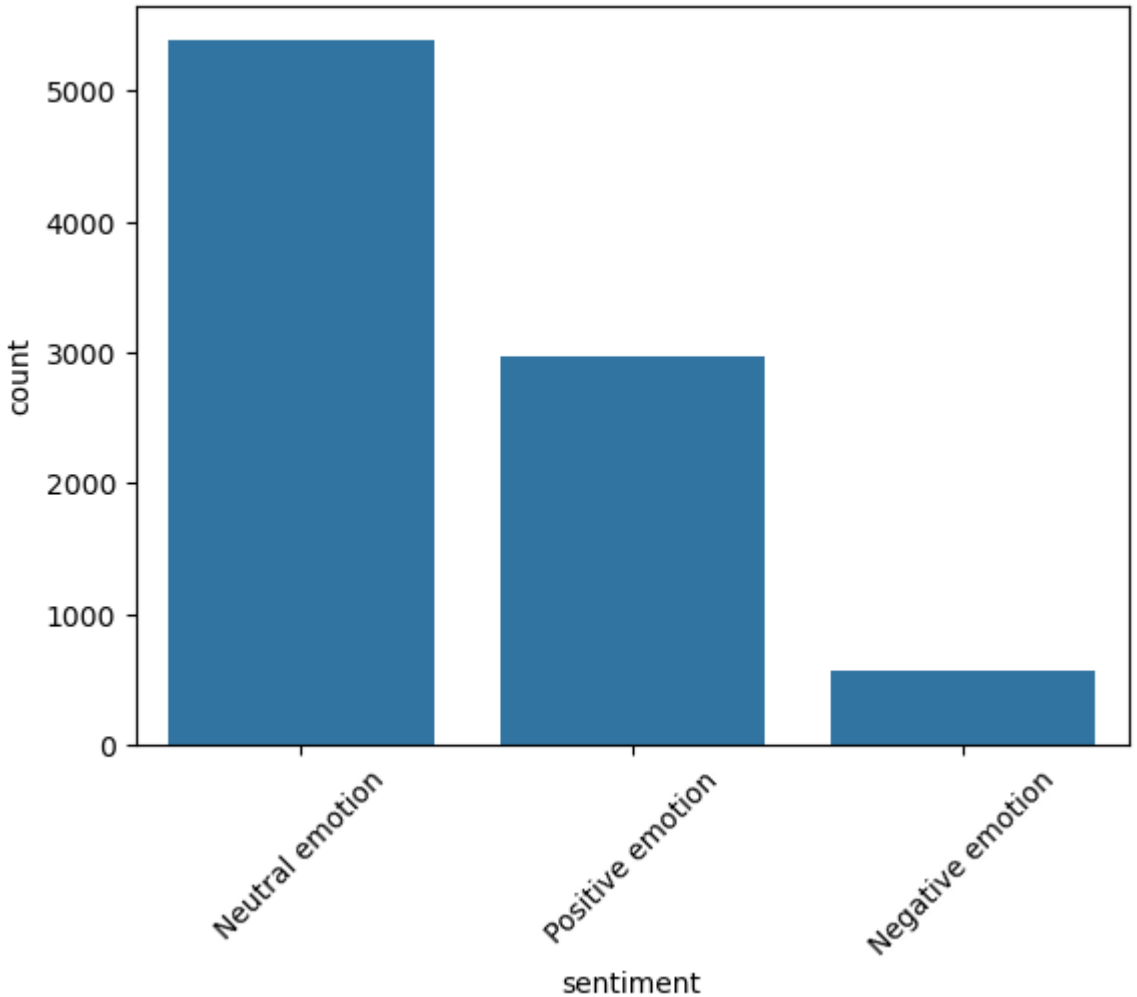
This project seeks to analyze and develop a predictive model of the sentiments from random people's tweets about Google products and Apple. The emotions from the tweets are either positive, negative, or neutral. We seek to create a binary classifier model and a multi-class model. The binary classifier will determine whether a tweet is positive or negative towards a product while the multi-class will include the neutral emotion.

The natural language processing technique enables us easily determine whether a post depicts satisfaction or dissatisfaction without having to go through the entire posts and this makes the process scalable, and that is the end goal for this project.
To develop the model, we used the https://data.world/crowdflower/brands-and-product-emotions from data world collected by CrowdFlower.

Data

The data has three columns the tweet_text column contains the tweets, the emotion_in_tweet_is_directed_at contains the brand to which the tweet was directed and the last column is_there_an_emotion_directed_at_a_brand_or_product has the emotion. There are 9093 rows in the data however there are 5802 null values in the second column and 1 in the first column. There are 4 values under the emotion's column, positive, negative, neutral and unknown.


3000
2500
2000
1500
1000
500
0
count
Positive emotion
Negative emotion
sentiment


5000
4000
3000
2000
1000
0
count
Neutral emotion
Positive emotion
Negative emotion
sentiment

Modeling and Conclusion

After preprocessing the data we head on to the modeling bit, where we will make a binary and multi-class model.
The baseline model is the MulitonomialNB which had a score of 84%.
To improve the prediction, we used the random forest classifier which resulted in a score of 86%. Since the classes had an imbalance the F1 score was the best measure of performance with a score of 83% for the improved binary classifier.

The multi-class classifier had a score of 66% indicating the model easily predicts the positive or negative sentiment compared to when the neutral sentiment is included.

In Conclusion: We can now be able to classify the emotion of a tweet given the text in the tweet.