Module 4 project

By: Phillip Ojo

Business Case

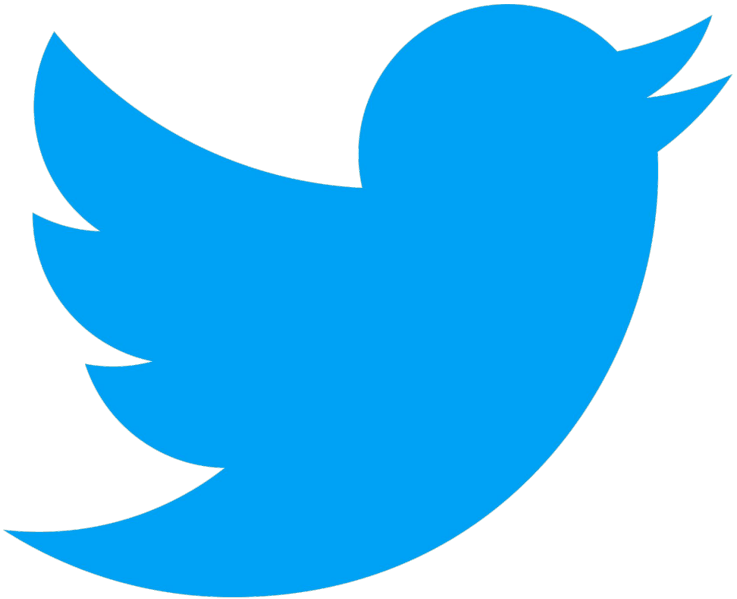
Twitter has over 300 million monthly active users, which allows businesses to reach a broad audience and connect with customers without traditional marketing techniques. On the downside, there's so much information that it's hard for brands to quickly detect negative social mentions that could harm their business.

That's why sentiment analysis, which involves monitoring emotions in conversations on social media platforms, has become a key strategy in social media marketing. Listening to how customers feel on Twitter allows companies to understand their audience, keep on top of what's being said about their brand, and their competitors, and discover new trends in the industry.

For this project, we are tasked with building an NLP model to analyze Tweet sentiment about Apple and Google products.

Twitter Dataset

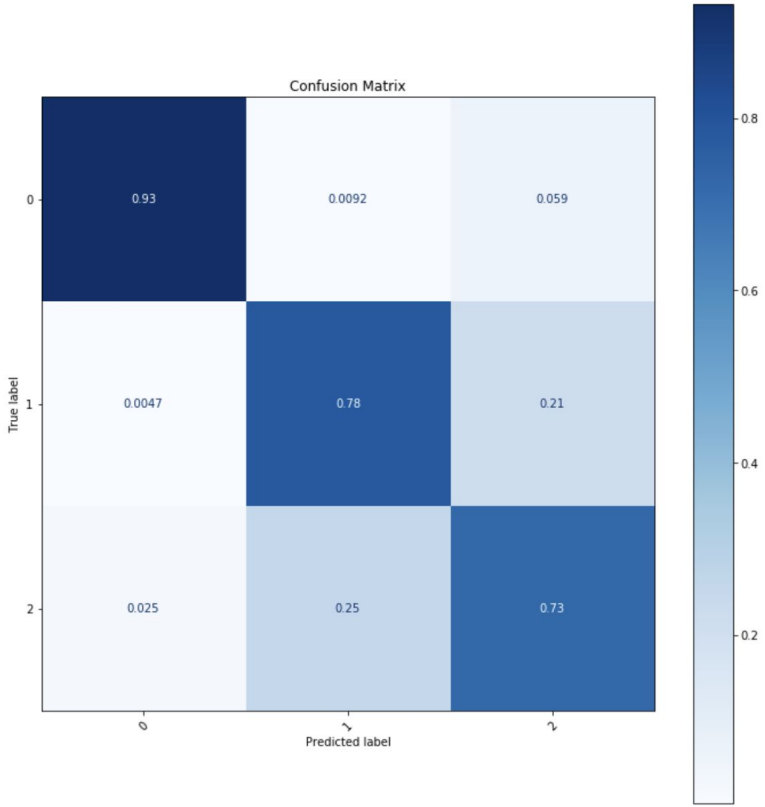
- Collected Tweets from Over 9,000 Users
- Analyzed 7668 Unique Words/Characters.
- Imbalanced Dataset (60% Neutral, 33% Positive, 6% Negative)



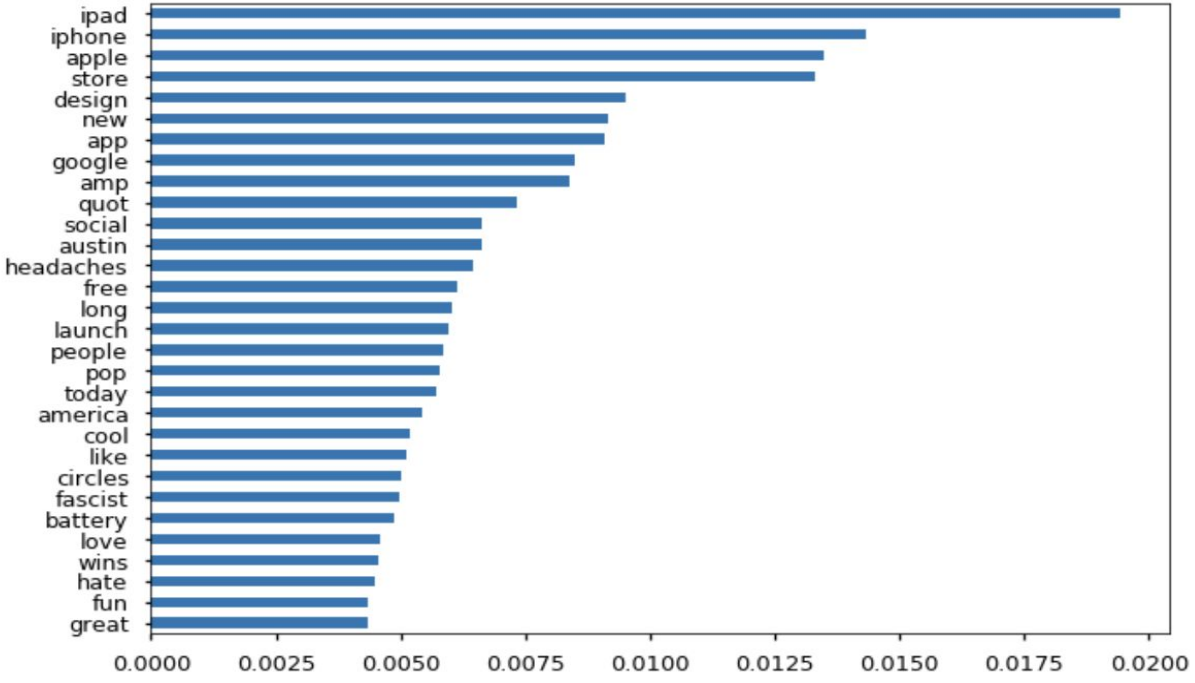
Model

```
------------------------------------------------------------
MODEL ACCURACY
------------------------------------------------------------
TRAINING ACCURACY: 91.56%
TESTING ACCURACY: 81.32%
------------------------------------------------------------
CLASSIFICATION REPORT
------------------------------------------------------------
              precision    recall  f1-score   support

           0       0.97      0.93      0.95      1085
           1       0.75      0.78      0.77      1070
           2       0.73      0.73      0.73      1078

    accuracy                           0.81      3233
   macro avg       0.82      0.81      0.81      3233
weighted avg       0.82      0.81      0.81      3233

------------------------------------------------------------
```

click to expand output; double click to hide output



Important Features



Conclusions / Recommendations

- After using a "upsampling" technique, our model was able to predict Negative tweets by a user 93% of the time which is very impressive to do with the very initial imbalanced data set.

- Words like "ipad", "iphone", "apple" were important in helping the model predict the different classes. We will have to do future analysis to see which classes it specifically impacted in predicting.

Future Work

- Improve the model by removing words in very low or high frequency

- Find better ways to clean the tweets for better pre processing. Alot of misspelled words and repeated letters.

Thank You