SENTIMENT ANALYSIS OF TWITTER DATA

With Natural Language Programming

A comparison of Naïve Bayes & Neural Networks for Tweet Sentiment Analysis

By Brian Amani

Project Overview

- Introduction
- Data Understanding
- Modelling
- Results
- Recommendations and Conclusions

Data Understanding
1. Data Cleaning

	tweet_text	emotion_in_tweet_is_directed_at	is_there_an_emotion_directed_at_a_brand_or_product
0	.@wesley83 I have a 3G iPhone. After 3 hrs twe...	iPhone	Negative emotion
1	@jessedee Know about @fludapp ? Awesome iPad/i...	iPad or iPhone App	Positive emotion
2	@swonderlin Can not wait for #iPad 2 also. The...	iPad	Positive emotion
3	@sxsw I hope this year's festival isn't as cra...	iPad or iPhone App	Negative emotion
4	@sxtxstate great stuff on Fri #SXSW: Marissa M...	Google	Positive emotion

The tweet sentiment data is contained in a dataframe of 9093 rows x 3 columns,. I started off with cleaning and then visualizing the data to have a better sense of the method to approach the project.

Data Cleaning

1. Dropping null values
2. Removing characters that don't make sense
3. Simplifying the data by introducing a new column "Brand"
4. Renaming the columns and rephrasing some of the sentiments

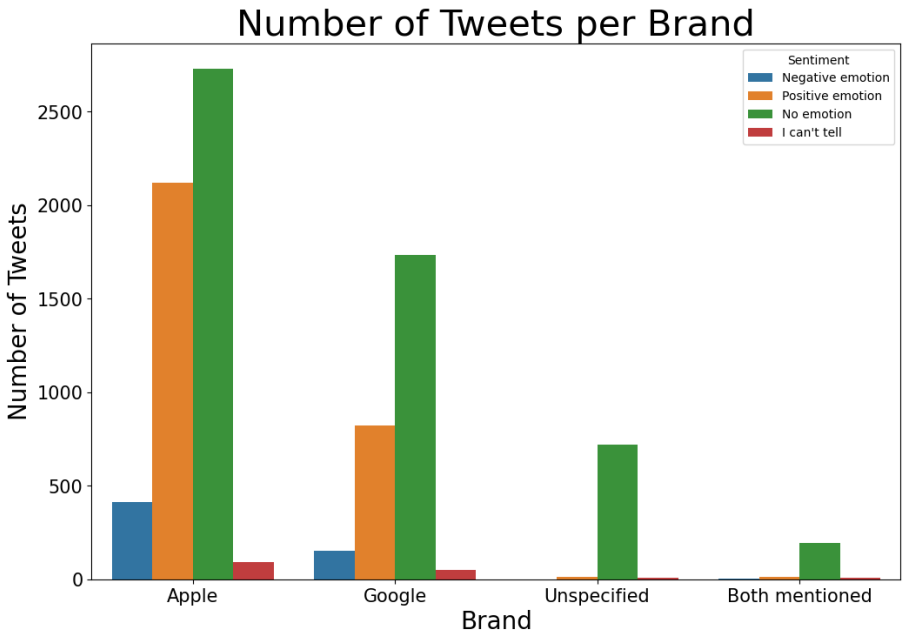
The cleaned dataframe had 9068 rows x 4 columns

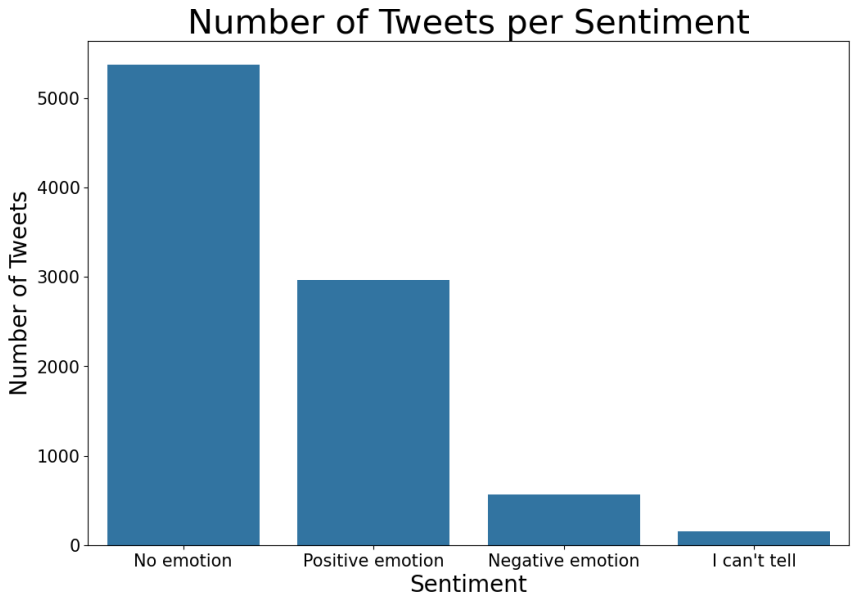
	Tweet	Product	Sentiment	Brand
0	.@wesley83 I have a 3G iPhone. After 3 hrs twe...	iPhone	0	Apple
1	@swonderlin Can not wait for #iPad 2 also. The...	iPad	1	Apple
2	@sxsw I hope this year's festival isn't as cra...	iPad or iPhone App	0	Apple
3	@sxtxstate great stuff on Fri #SXSW: Marissa M...	Google	1	Google
4	@teachntech00 New iPad Apps For #SpeechTherapy...	Unspecified	0	Apple
...	...	...	...	...
9063	@mention Yup, but I don't have a third app yet...	Unspecified	0	Google
9064	Ipad everywhere. #SXSW {link}	iPad	1	Apple
9065	Wave, buzz... RT @mention We interrupt your re...	Unspecified	0	Google
9066	Google's Zeiger, a physician never reported po...	Unspecified	0	Google
9067	Some Verizon iPhone customers complained their...	Unspecified	0	Apple

9068 rows × 4 columns

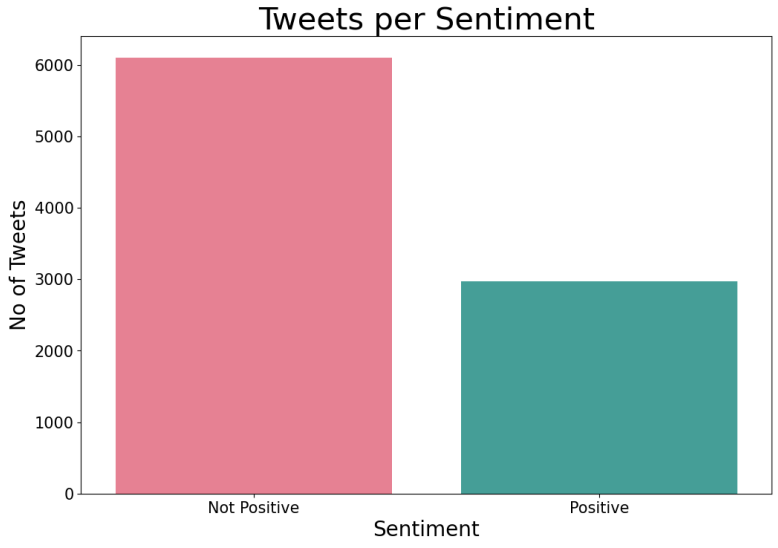
Data Understanding

2. Data Visualization



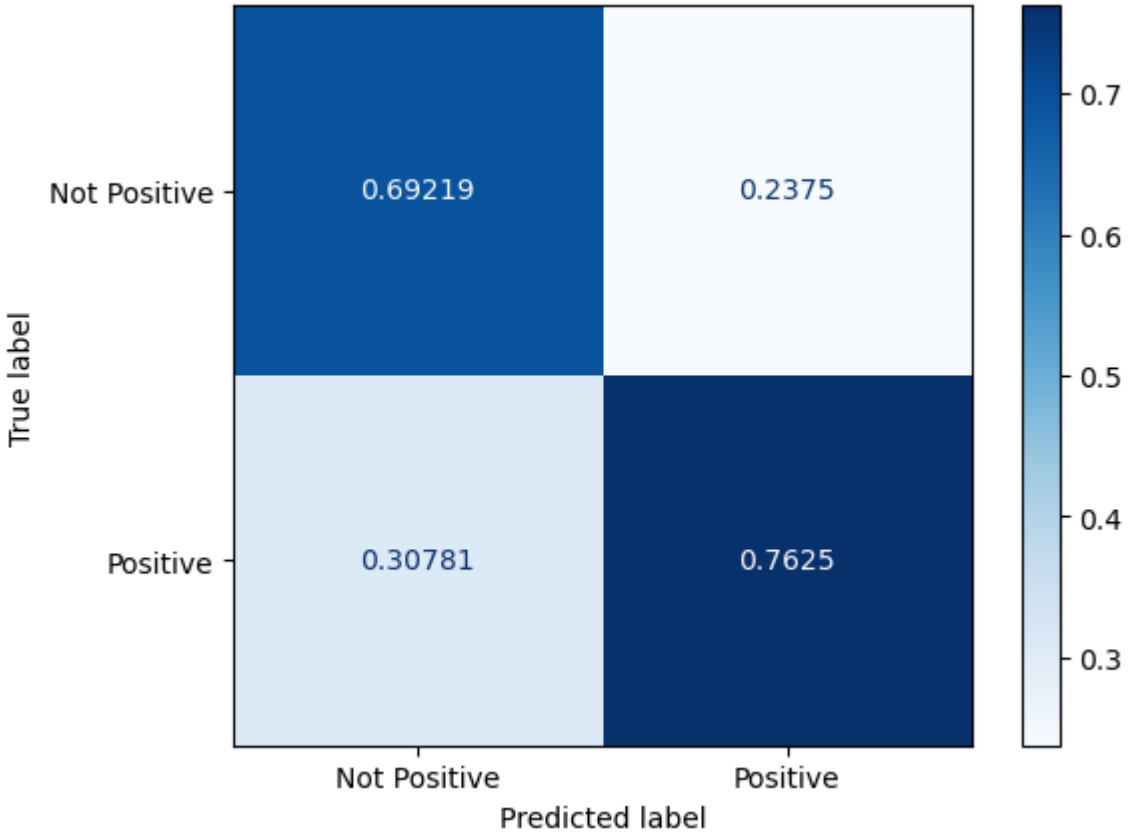


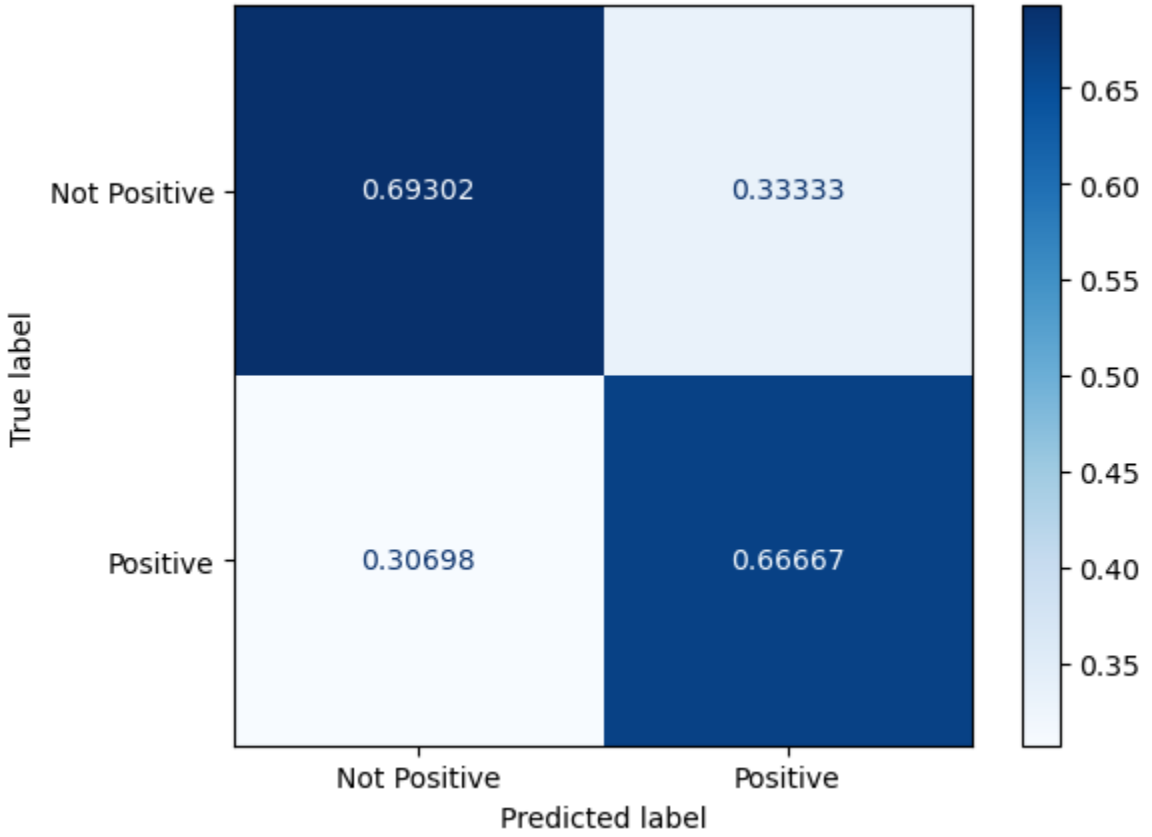
Based on the visualizations above, the best approach would be to make the Tweets my Independent variable, and the sentiments my target variable. I also made the sentiments binary – non-positive and positive sentiments



Data Modelling

1. Naïve Bayes Model






```
Mean train accuracy: 0.7232843137254902
Mean test accuracy: 0.7050653594771242
```

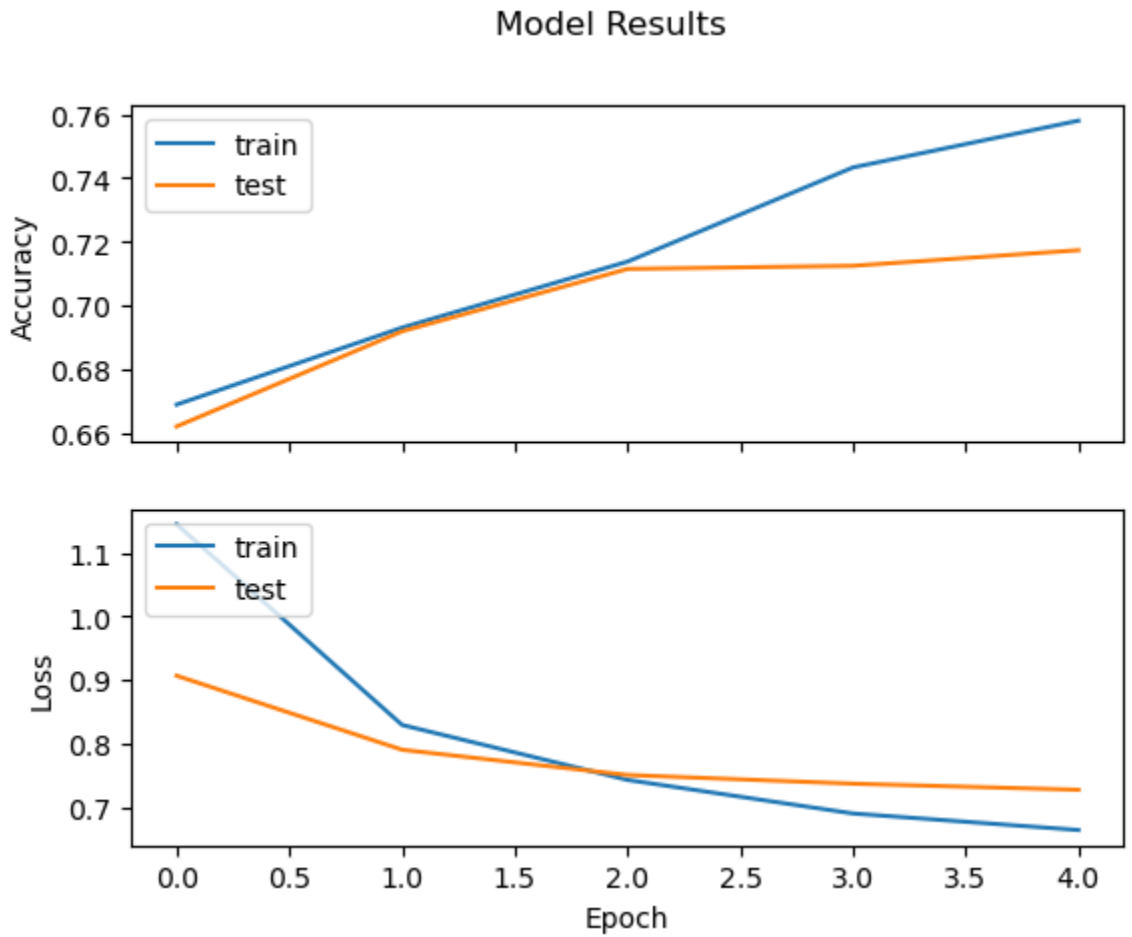
```
Training set accuracy: 0.7197712418300654
Validation set accuracy: 0.6903478686918177
```

The accuracy of the model is not better with training and validation data set accuracy , but there is a smaller margin of difference between the training and validation accuracy.

I decided to attempt using a neural network model next

Data Modelling

2. Neural Network Model



Final Training Loss: 0.663762867450714 1 Final
Training Accuracy: 0.7580065131187439 Final
Validation Loss: 0.727064549922943 1 Final
Validation Accuracy: 0.7172954678535461

Results Comparison

Naïve Bayes Results:
- Training set accuracy: 0.72 or 72%
- Validation set accuracy: 0.69 or 69%

Neural Network Results:
- Final Training Accuracy: 0.76 or 76%
- Final Validation Accuracy: 0.72 or 72%

- The Neural Network outperforms Naive Bayes, showing better performance on both training and validation sets.
- However, the validation accuracy is still lower than training accuracy, suggesting some overfitting.

Recommendations & Conclusion

The Neural Network Model would be the best for Twitter Sentiment Analysis

Neural Network performs better than Naïve Bayes in linking sentiment to brands (Apple & Google) but may need hyperparameter tuning for optimal results.

It will also benefit from more training data from Apple and Google

How Apple and Google can use the Neural Network Model

- **Brand Monitoring** – To track real-time sentiment on products for proactive issue resolution.
- **Customer Feedback Analysis** – To identify strengths and areas for improvement.
- **Marketing Strategy** – To leverage sentiment insights to refine campaigns.
- **Trend Analysis** – To detect shifts in perception to anticipate crises early.

Thank You

Brian Amani
https://github.com/papyrusleaf
https://www.linkedin.com/in/brian-amani-63a64987/