Think different

Deciphering the Public Opinion on Apple

Phase 4 Group 10 Members

- Nicholas Kirui
- Rose Kimondo
- Laura Kanda
- Charles Aloo
- Chemutai Annolyne
- Dave Omondi

Project Overview

As digital brand advisors, we have been hired by Apple with the goal of conducting an analysis that assesses public opinion about the company based on a Twitter sentiment in the South by Southwest (SXSW) conference, which celebrates the convergence of tech and other critical societal factors like education and culture.

One medium that companies can leverage to identify public opinion about their products and activities in social media platforms like Twitter. People from all walks of life share their varied opinions surrounding businesses, making Twitter a perfect source of data to obtain public sentiment.

For this project, Apple required us to conduct a sentiment analysis of tweets presented during the SXSW Conference

Project Objectives

Below are the objectives of our analysis:

1. Determine how Apple is perceived as a company based on the tweets presented during the SXSW Conference, in comparison to Google which is one of their main competitors.

2. Determine consumers' reactions to Apple's SXSW announcement (how are their new products perceived, and how do people react to new announcements?)

3. Seek out insights that the tweets can give into the challenges that Apple faces during these product announcements. Are there specific challenges that should be addressed?

4. Create a model that when given a tweet or series of tweets can determine the user's sentiment (positive, negative, or neutral). Apple can use these models to assess public opinion and stay ahead of its competition.

Data Understanding

We used a dataset from data world provided by CrowdFlower which has tweets about Google and Apple from a conference. The tweet labels were crowdsources and reflect the emotion they convey and what product/service/company the emotion is directed at based on the content.

Contributors evaluated tweets about multiple brands and products. The crowd was asked if the tweet expressed positive, negative, or no emotion towards a brand and/or product.

Data Limitation

The data was limited as explained below:

- Twitter is full of spelling errors, shortened words, hashtags, acronyms, tags, and very specific nouns and words. These are difficult for a model to learn.
- Emotions are also complicated and a tweet can have a mix of emotions or ambiguity.
- Emotions are also relative. For example, the phrase "IPhones are better than Android phones" could be negative or positive depending on the perception.
- A model also learns as well as its labels and our data was labelled by humans.
- These considerations make the problem tricky but the model was successful in classifying the tweets.
- Our dataset had highly imbalanced target values(emotions), which meant that we will had to address this during the modeling stage

Data Preparation

To prepare our data for modelling,we had to address various factors such as missing values and duplicate data.

For missing values,we had to drop some columns and fill other columns in order to ensure our model works as required.

Data Analysis

With our objectives in mind, we needed to isolate and analyze positive and negative tweets as a whole and on a company and product basis.

We utilized the TweetTokenizer functionality since its ideal for processing hashtags and handles correctly. We had to drop all handles from the tweets since our focus was on the content of the tweets.

We also had to utilize lemmatization to ensure that we captured any stop words that could be generated in our modelling process

World count with Positive tweets

We utilized the world cloud feature to generate tokens with positive sentiments as illustrated.

This showed us what the public was tweeting about in a positive way. We saw that a lot of people were excited about the iPad, iPhone, and iPad 2 launch.

It was also clear that people were tweeting about Apple and Google.

Negative sentiments

From our analysis,Apple, Google, iphone, ipad were common among the negative tweets. We visualized this with a word cloud as illustrated

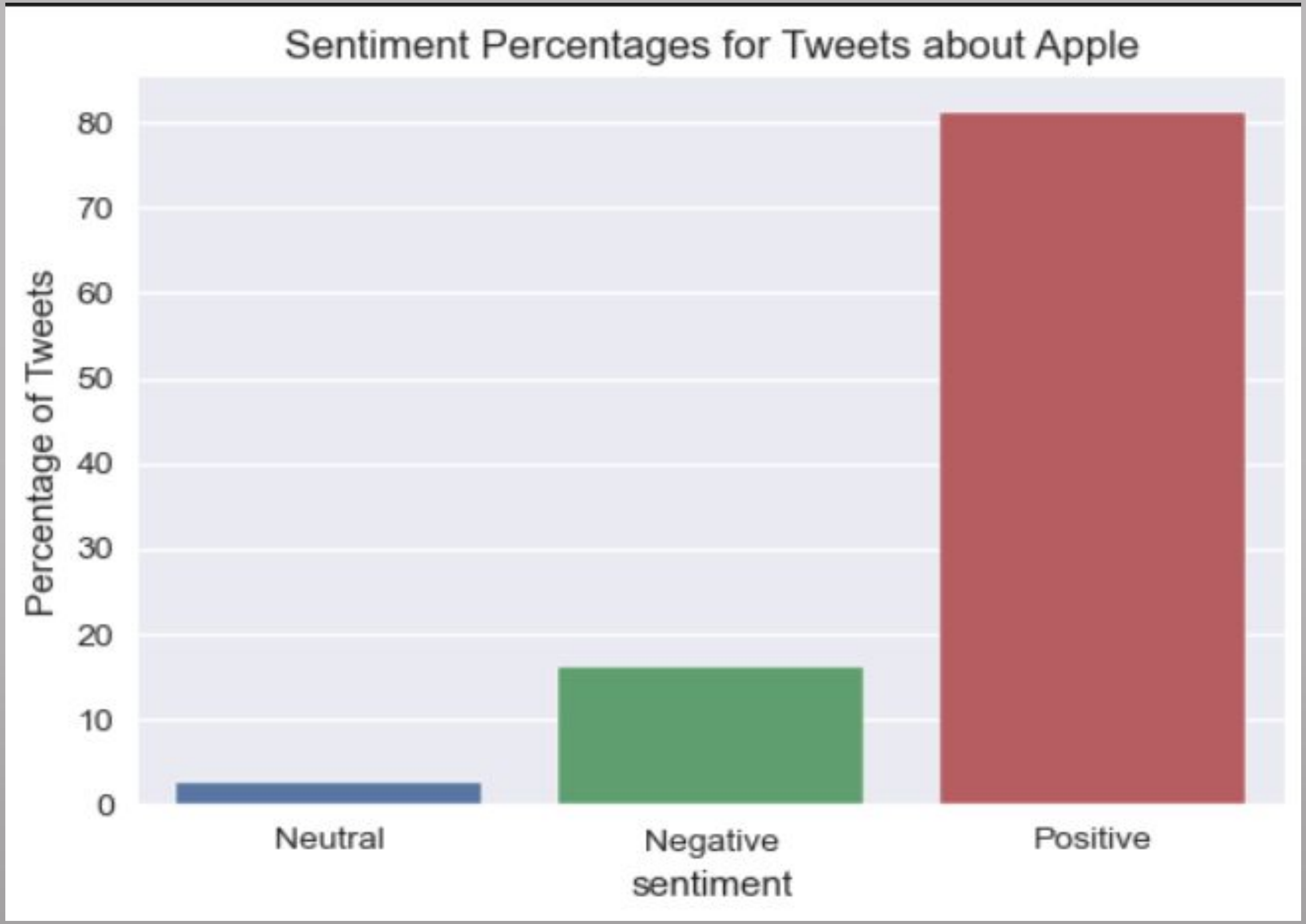
For further analysis we had to remove the company names to get additional information.

We further utilized to tweets related to specific brands.

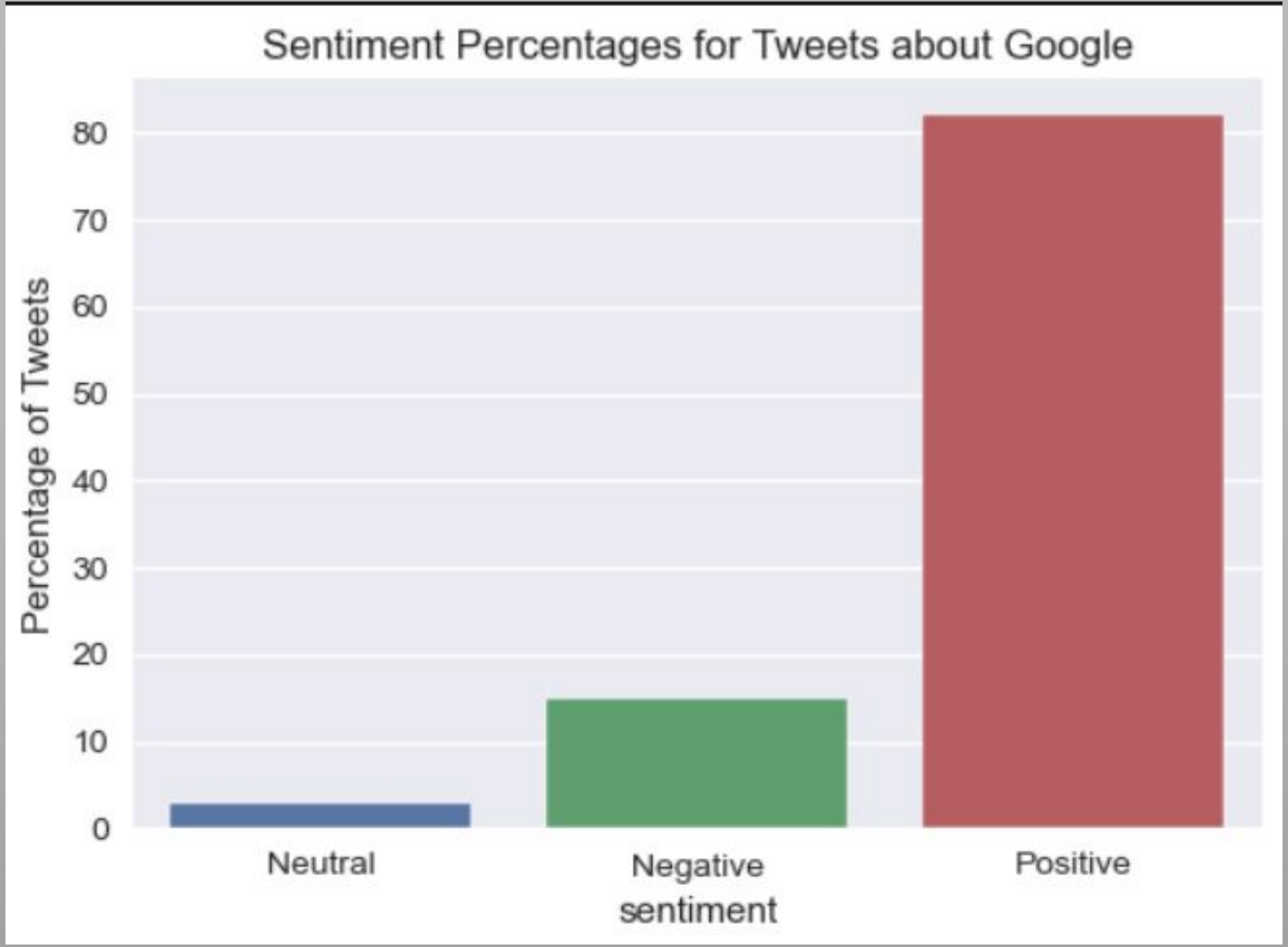
Google and apple sentiments visualisations

We visualized the percentage of positive,negative and neutral sentiments for apple and google as illustrated

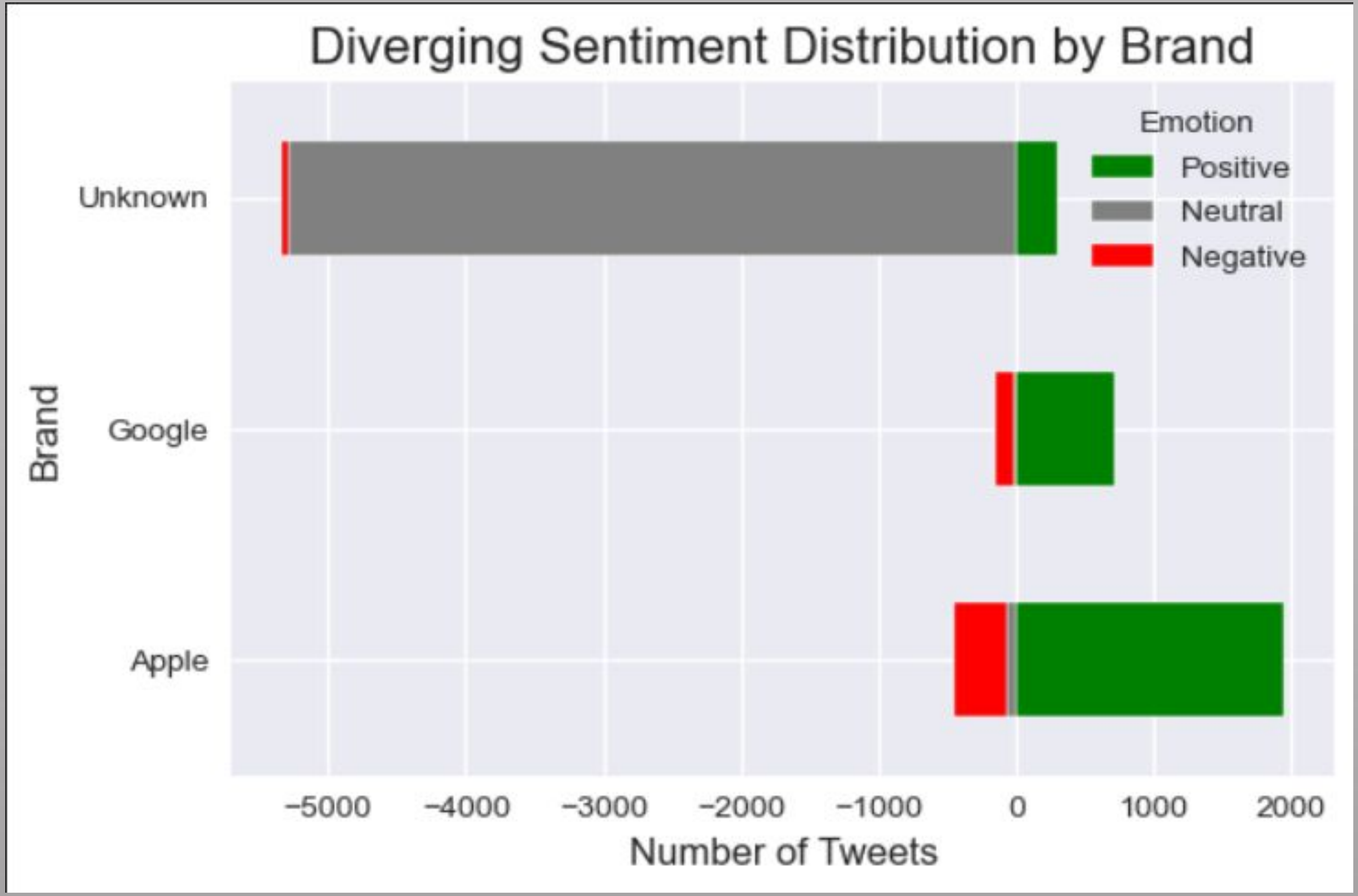
From the visualisations it was clear that both companies and their products and services are mostly perceived positively.



Google sentiments visualisations



Google and Apple sentiments visualisations



Apple - Positives:

- Positive tweets about Apple suggest that the temporary pop-up store in downtown Austin was generally received well.
- The iPad 2 was frequently discussed positively, with people excited about its launch.

Apple - Negatives:

- The iPhone's battery is frequently discussed in negative tweets.
- Design of the iPad was referred to as a "design headache."
- There are several references to Apple as a "fascist company."
- Several apps are referred to as "battery killer" and the design of the News app seems to have not been received positively..

Modelling

- To prepare our data for modelling we had to convert the text data into numerical data since most algorithms are designed to work with numerical data.
- In our case we used TF-IDF which weighs words based on their frequency in a document and their rarity across the corpus.
- To further explore the data and avoid data leakage, we split the data into train and test sets .
- We also had to address class imbalance using Random Oversampling since it preserves data distribution in a model.
- We worked with various models and later tuned them to ensure overall accuracy,generalization and overall effectiveness

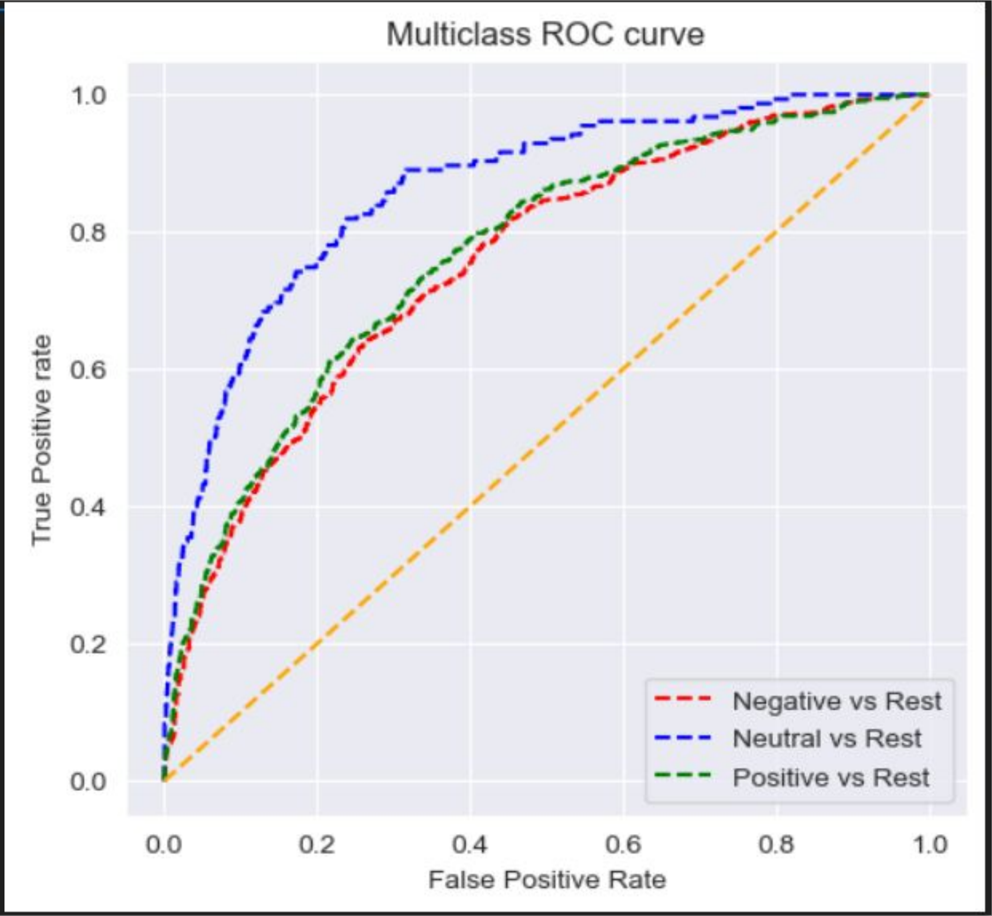
Visualisation of various model performance-LogisticRegression

```
Running... LogisticRegression Model
Cross Validation Score: 0.832
Train Accuracy Score: 0.9282
Test Accuracy Score: 0.6662

Classification Report:
              precision    recall  f1-score   support

           0       0.75      0.74      0.74      1335
           1       0.42      0.42      0.42       155
           2       0.57      0.59      0.58       739

    accuracy                           0.67      2229
   macro avg       0.58      0.58      0.58      2229
weighted avg       0.67      0.67      0.67      2229
```



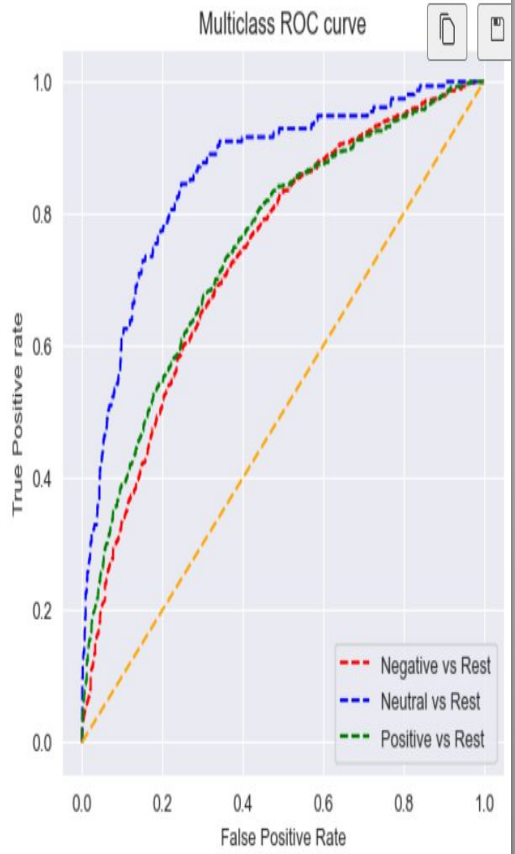
MultinomialNB Model performance

```
Running... MultinomialNB Model
Cross Validation Score: 0.7932
Train Accuracy Score: 0.9004
Test Accuracy Score: 0.6178

Classification Report:
              precision    recall  f1-score   support

           0       0.77      0.62      0.69      1335
           1       0.30      0.61      0.40       155
           2       0.55      0.61      0.58       739

    accuracy                           0.62      2229
   macro avg       0.54      0.61      0.55      2229
weighted avg       0.66      0.62      0.63      2229
```



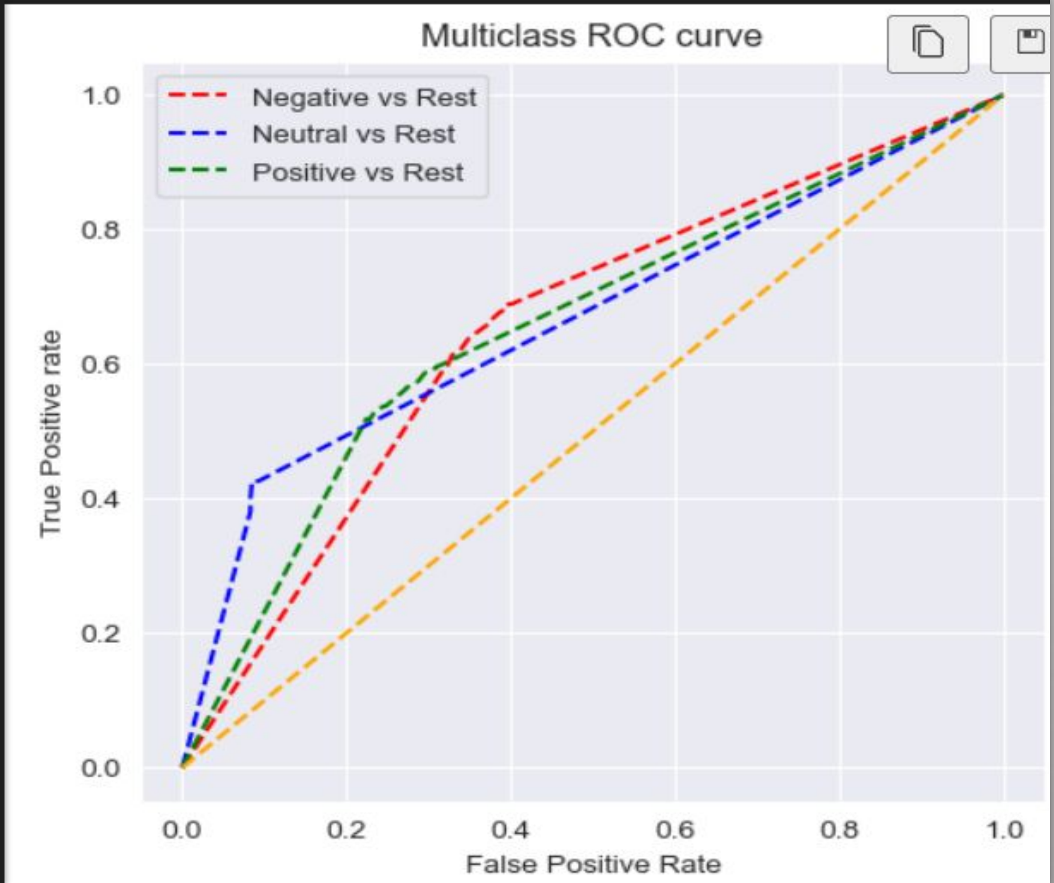
DecisionTree Model performance

```
Running... DecisionTree Model
Cross Validation Score: 0.8083
Train Accuracy Score: 0.958
Test Accuracy Score: 0.5989

Classification Report:
              precision    recall  f1-score   support

           0       0.73      0.64      0.68      1335
           1       0.27      0.43      0.33       155
           2       0.51      0.56      0.53       739

    accuracy                           0.60      2229
   macro avg       0.50      0.54      0.52      2229
weighted avg       0.63      0.60      0.61      2229
```



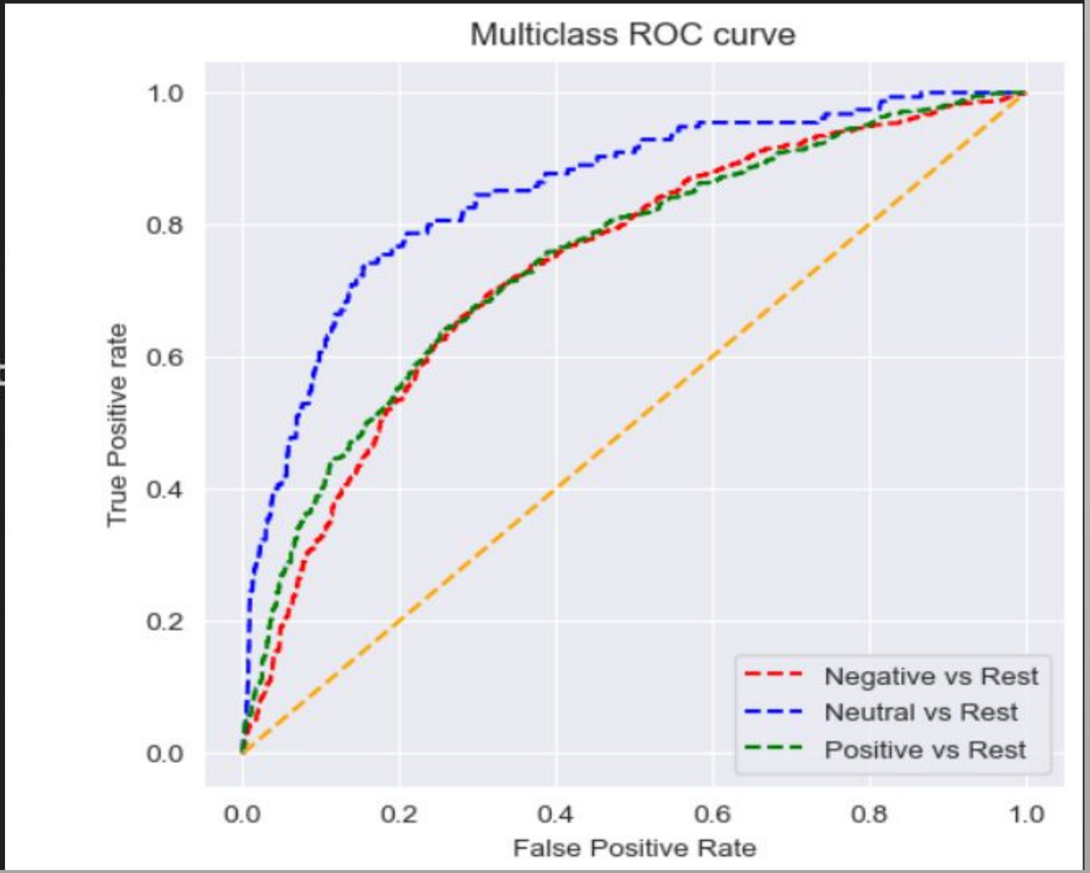
VectorClass Model performance

```
Running... VectorClass Model
Cross Validation Score: 0.8624
Train Accuracy Score: 0.9478
Test Accuracy Score: 0.6774

Classification Report:
              precision    recall  f1-score   support

           0       0.71      0.83      0.76      1335
           1       0.65      0.23      0.34       155
           2       0.61      0.50      0.55       739

    accuracy                           0.68      2229
   macro avg       0.66      0.52      0.55      2229
weighted avg       0.67      0.68      0.66      2229
```



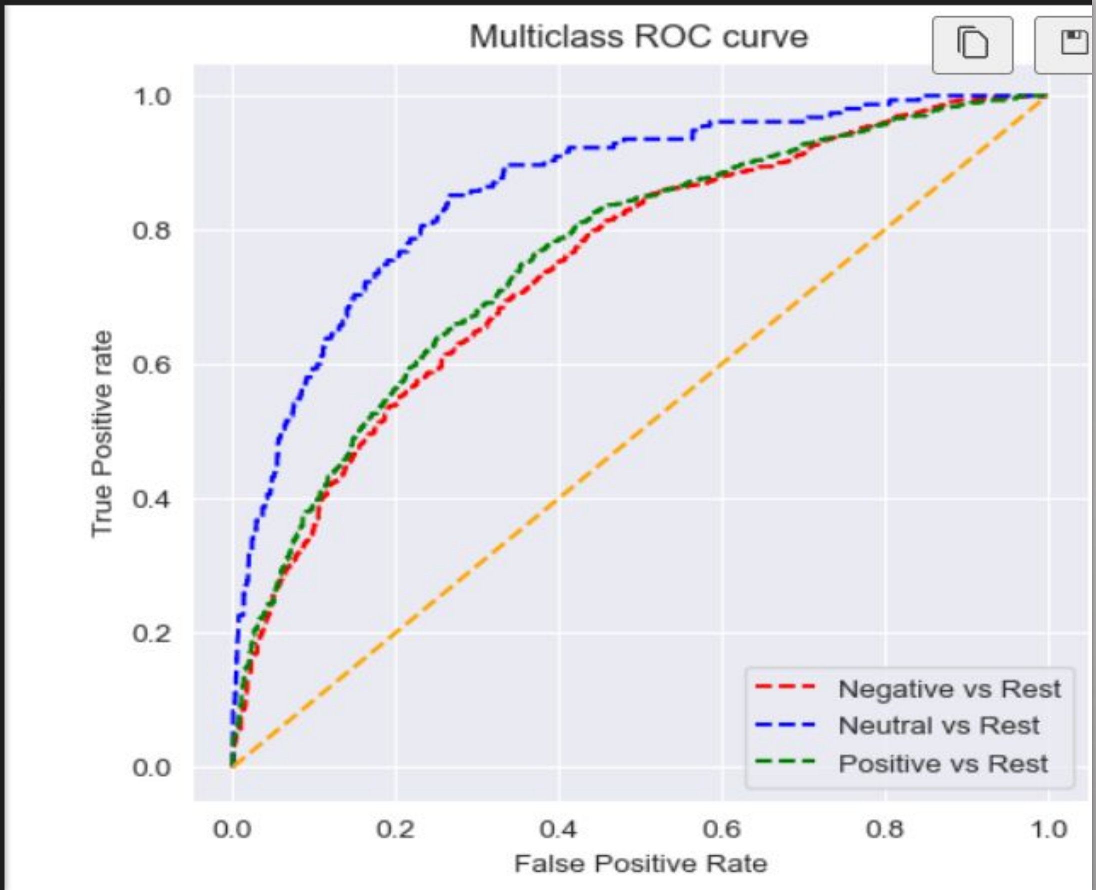
SDGClassifier Model performance

```
Running... SGDClassifier Model
Cross Validation Score: 0.8258
Train Accuracy Score: 0.9053
Test Accuracy Score: 0.6577

Classification Report:
              precision    recall  f1-score   support

           0       0.75      0.72      0.73      1335
           1       0.39      0.48      0.43       155
           2       0.57      0.58      0.58       739

    accuracy                           0.66      2229
   macro avg       0.57      0.59      0.58      2229
weighted avg       0.66      0.66      0.66      2229
```



Summary of model performance scores

	Model	CV Score	Train Accuracy	Test Accuracy	Type
0	LogisticRegression	0.8320	0.9282	0.6662	Default
1	MultinomialNB	0.7932	0.9004	0.6178	Default
2	DecisionTree	0.8083	0.9580	0.5989	Default
3	GradientBoost	0.6453	0.7225	0.6101	Default
4	VectorClass	0.8624	0.9478	0.6774	Default
5	SGDClassifier	0.8258	0.9053	0.6577	Default

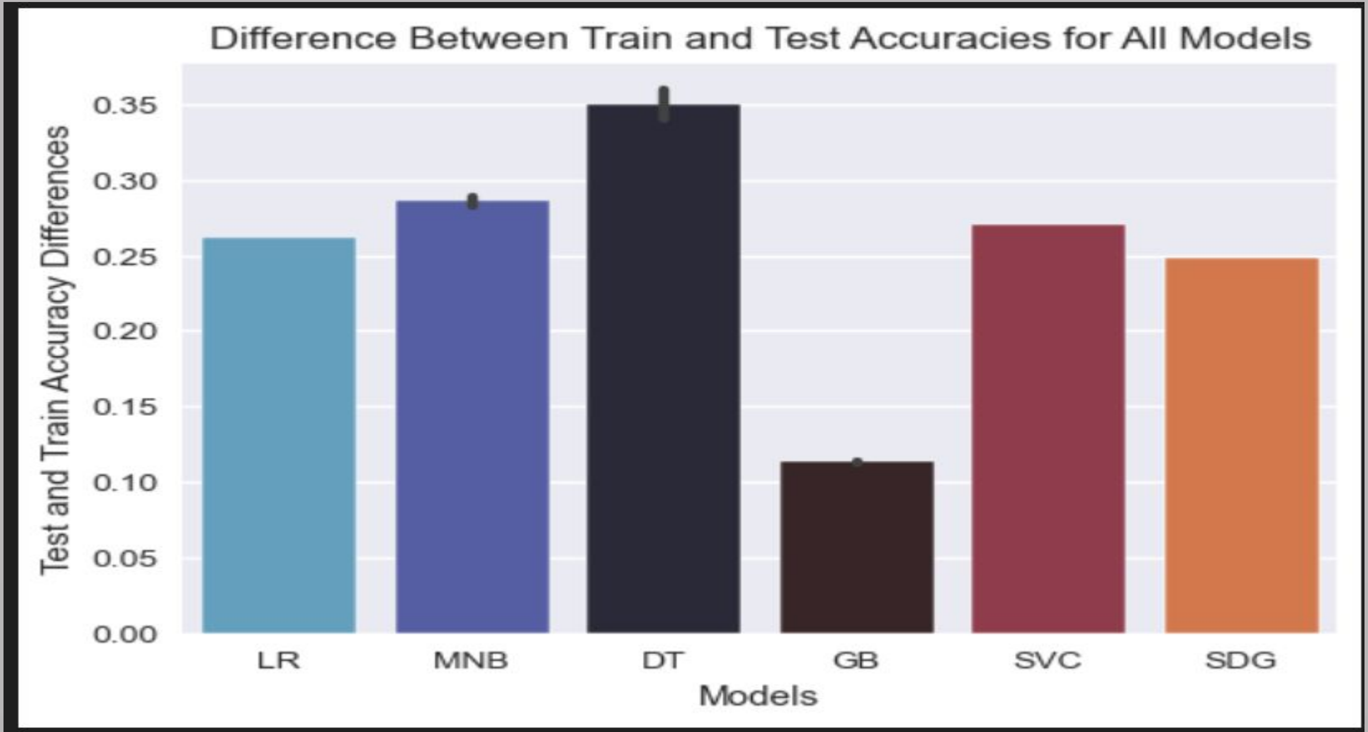
Model Tuning

Results after model tuning was done

	Model	CV Score	Train Accuracy	Test Accuracy	Type
0	LogisticRegression	0.8320	0.9282	0.6662	Tuned
1	MultinomialNB	0.8169	0.9264	0.6375	Tuned
2	DecisionTree	0.8167	0.9580	0.6182	Tuned
3	GradientBoost	0.6465	0.7233	0.6092	Tuned
4	VectorClass	0.8624	0.9478	0.6774	Tuned
5	SGDClassifier	0.8258	0.9053	0.6577	Tuned

Gradient Boost model

From the visualization above,the Gradient Boost model is the least overfit since it has the least difference between the train and test data



	Model	CV Score	Train Accuracy	Test Accuracy	Type	accuracy_differences
0	LogisticRegression	0.8320	0.9282	0.6662	Default	0.2620
1	MultinomialNB	0.7932	0.9004	0.6178	Default	0.2826
2	DecisionTree	0.8083	0.9580	0.5989	Default	0.3591
3	GradientBoost	0.6453	0.7225	0.6101	Default	0.1124
4	VectorClass	0.8624	0.9478	0.6774	Default	0.2704
5	SGDClassifier	0.8258	0.9053	0.6577	Default	0.2476
6	LogisticRegression	0.8320	0.9282	0.6662	Tuned	0.2620
7	MultinomialNB	0.8169	0.9264	0.6375	Tuned	0.2889
8	DecisionTree	0.8167	0.9580	0.6182	Tuned	0.3398
9	GradientBoost	0.6465	0.7233	0.6092	Tuned	0.1141
10	VectorClass	0.8624	0.9478	0.6774	Tuned	0.2704
11	SGDClassifier	0.8258	0.9053	0.6577	Tuned	0.2476

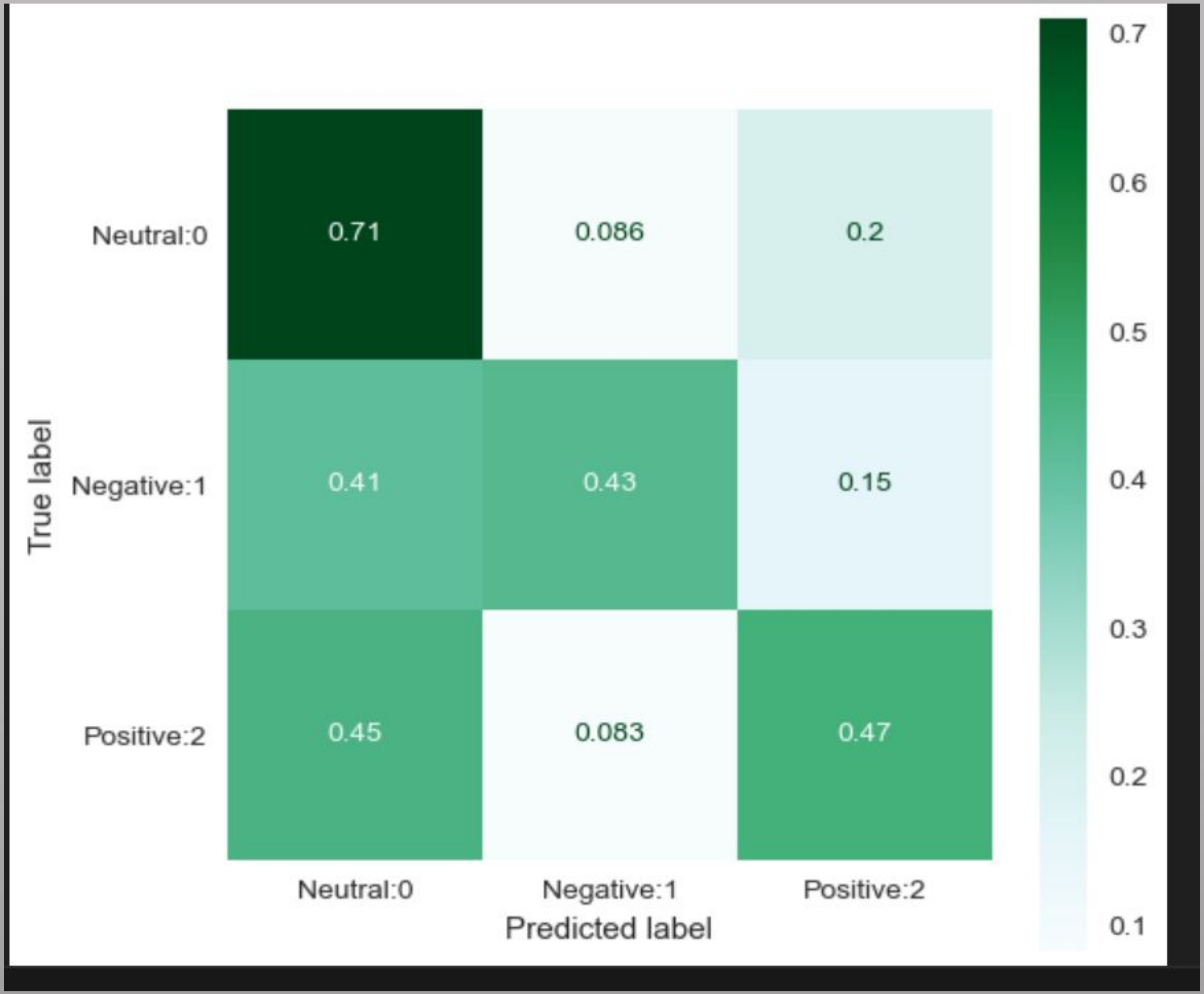
Model selection

- Even though the VectorClass model had the highest test accuracy score and the Multinomial Naive Bayes models had better true negative and true positive rates, they were extremely overfit and therefore, did not generalize well to real world data.
- It was important that our model performs well when presented with data that it has not yet seen. The true positive and negative rates of our final model were also somewhat high, which indicated good overall performance.
- We settled on the Gradient Boost model as our final model
- It had True Negatives value of 0.43 and a True Positives value of 0.47, Also, it had an accuracy score of 62% and only misclassified 15% of the negative sentiments as positive.

Confusion Matrix

Confusion matrix visualization for the final model

This model is recommended for use since it is fairly accurate in classifying tweet sentiments, especially compared to doing so by hand



Recommendation

GradientBoost model is recommended for use since it is fairly accurate in classifying tweet sentiments, especially compared to doing so by hand. As a result, Apple's product team can use this sentiment analyzer to target neutral consumers and convert them to buyers.

While the model improved from the first iteration, it still does not distinguish non-neutral tweets from neutral ones very well. Furthermore, the sentiment analysis only looked at tweets, while consumers post about Apple from other platforms as well.

Moving forward, it would be interesting to try and implement sentiment analysis from other platforms as well to gauge consumers.

Also, having more explicit sentiment labels could provide more insight into the customers' sentiments.

Further Research

For further research, the data should be updated with more recent tweets, and more advanced models should be explored, such as the recent transformers.