

# TWITTER SENTIMENT ANALYSIS



## **Authors:**

1. Aaron Onserio
2. Diana Mwaura
3. Joshua Rwanda
4. Samuel Kyalo
5. Stella Kitur
6. Stephanie Mbithe

GitHub repository: <https://github.com/R3TR0Quan/twitter-sentiment-analysis>

# Project Overview

This project seeks to build a Natural Language Processing (NLP) model, to analyze Twitter sentiments about Apple and Google products. The dataset comes from CrowdFlower to rate Tweets as positive, Negative or Neutral. It also analyses these tweets to pinpoint what about the product they enjoyed and what they disliked.

## Business Challenges

### *1. Brand reputation*

The brand reputation is crucial to the success of any company.

Therefore, Apple & Google would be interested in maintaining and/or improving their brand reputation, where they are lacking.

### *2. Product Development*

Not understanding the customer preferences when it comes to the development of new products.

### *3. Competition*

Not being able to monitor the sentiments that are related to their competitors.

## **Proposed solutions (based on the challenges)**

### *1. Brand reputation*

Ensure Quality Customer Service

Proactive Online Monitoring

### *2. Product Development*

Prioritize customer feedback in the Research & Development process

## ***Competitive analysis***

Monitor competitor sentiments and identify market gaps to differentiate the company's offerings.

# BUSINESS UNDERSTANDING

## Business Problem

In the era of social media, it is crucial for businesses to understand the sentiments expressed by customers towards their brands or products.

This sentiment analysis project aims to analyze Twitter data and extract valuable insights regarding the sentiments associated with Apple and Google products mentioned in tweets. By uncovering the public's opinions and emotions, businesses can make data-driven decisions to improve their market positioning and enhance customer satisfaction.

## Objectives

The main objective of this project is to create a model that when given a tweet or series of tweets and a product would determine how the user felt about that product. This is trivial for a human to accomplish; our model can do this for thousands or even millions of tweets in a short time.

1. To build a text classifier to accurately distinguish between positive, neutral, and negative sentiments.
2. Competitive Analysis – Compare the sentiment towards Apple and Google products to identify any significant differences in public perception.
3. Give insights as to where the company can increase customer satisfaction.

## Success metrics

The model performance of this project will be analyzed using the following performance metrics:

- Accuracy
- Precision and Recall
- F1 Score

For the multiclass classification, the recall macro average was used to compare the models. The recall macro average is useful in scenarios where you want to evaluate the overall performance of the classifier across all classes, regardless of the class distribution. It provides an aggregated measure of how well the model is able to identify positive instances for each class.

A weakness of this metric is that it fails to take into account the number of true negatives or false positives.

The evaluation therefore also considered weighted macro average, F1 scores and accuracy.

# DATA UNDERSTANDING

The dataset used was drawn from DataWorld provided by CrowdFlower which has tweets about Apple and Google from the South by Southwest (SXSW) conference. The tweet labels were crowdsourced and reflect which emotion they convey and what product/service/company this emotion is directed at based on the content.

The dataset comprises of Twitter data, including tweet texts, mentions of brands or products, and the associated sentiments or emotions.

It had 9093 rows and 3 columns.

Each row was a specific tweet entry, and the three columns were:

1. The tweet texts.
2. The product which the tweet is directed at.
3. A classification of the emotion of the tweet towards the product, i.e A “Positive emotion”, a “Negative emotion” and “I can’t tell.”

The various products that the tweets were directed to are:

- iPad
- Apple
- iPad or iPhone App
- Google
- iPhone
- Other Google product or service
- Android App
- Android
- Other Apple product or Service

All these products are from two distinct brands, i.e Apple and Google.

# DATA PREPARATION

## 1. Renaming of columns

The column names were renamed into shorter labels that can easily be referenced:

- “tweet\_text” was renamed to “Tweet”
- “emotion\_in\_tweet\_is\_directed\_at” was renamed to “Product/Brand”
- “is\_there\_an\_emotion\_directed\_at\_a\_brand\_or\_product” was renamed to “Emotion”

## 2. Dealing with missing values

- The dataset had 1 missing value in tweets and 5,802 missing values in the product category.
- To deal with these missing values:
  - a.) The row that had the missing tweet was **dropped**.
  - b.) The entries in the product column that had missing values were **filled** with “Unknown Product.”

## 3. Dealing with duplicate entries

- The dataset had 22 duplicate entries.
- The duplicate entries were **dropped**, while **keeping the first entry**.

## 4. Dropping of certain rows

- The “Emotion” row contained entries: “Positive”, “Negative”, “Neutral” and “unknown.”
- The rows with the 156 entries of “Unknown” emotion were dropped to be left with the three main classifications.
- This is because these entries only formed **1.72%** of the dataset, so this did not make a significant difference to the overall dataset.

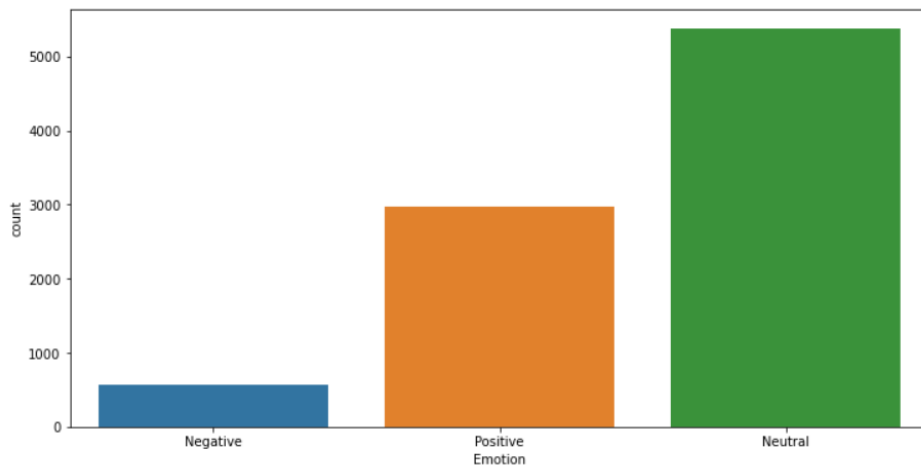
## 5. Feature engineering

- A 4<sup>th</sup> row was created named “**Brand.**”
- All the products were classified into two main brands, as either “Apple” and “Google”.

## 6. Exploratory data analysis

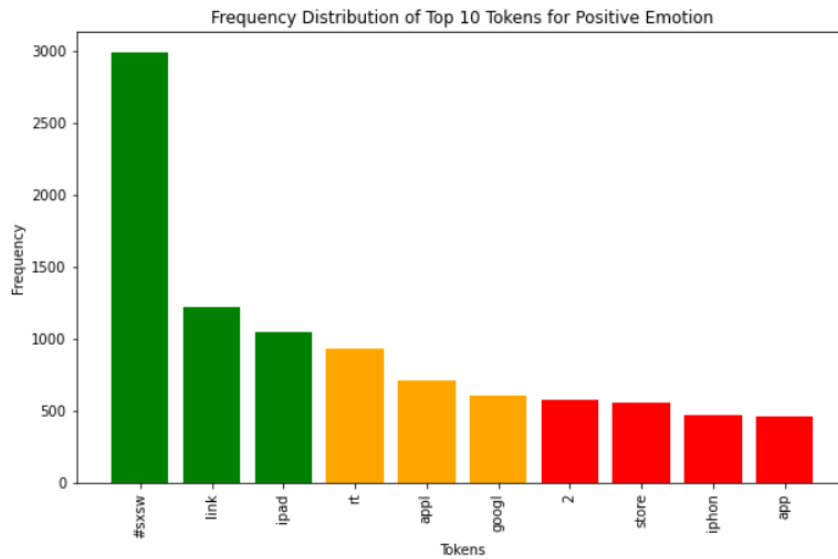
### 1. Distribution of tweets according to Emotion

The distribution of tweets according to emotion was done as follows:



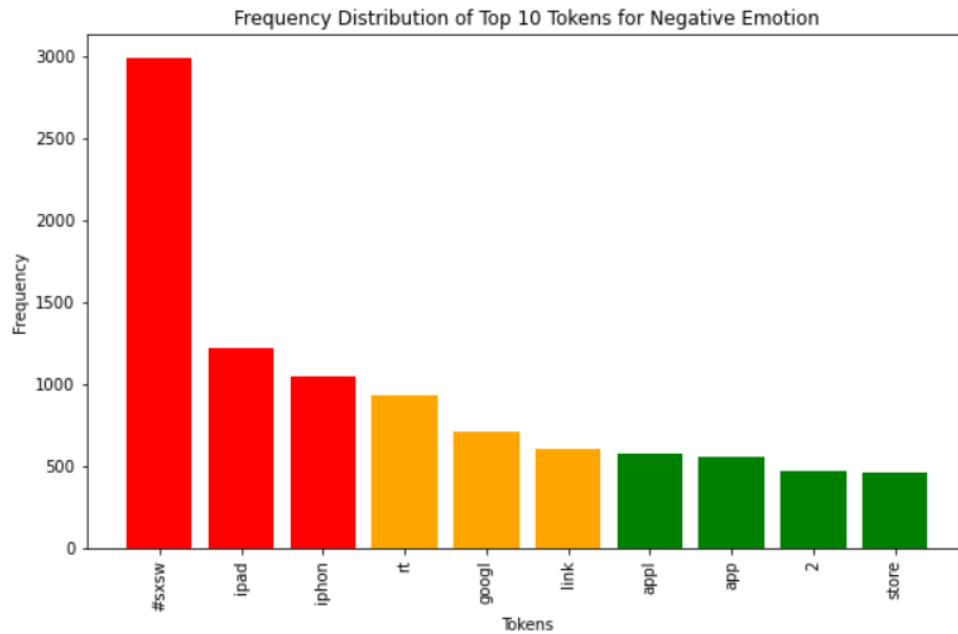
This shows that there is a class imbalance of the Positive, Negative and Neutral variables, with most of the tweets being classified as Neutral and a small percentage being Negative tweets.

## 2. Distribution of the top 10 words (tokens) used for positive emotion.



The most common words in the tweets classified as positive emotions are #sxsxw (South by Southwest), link, iPad.

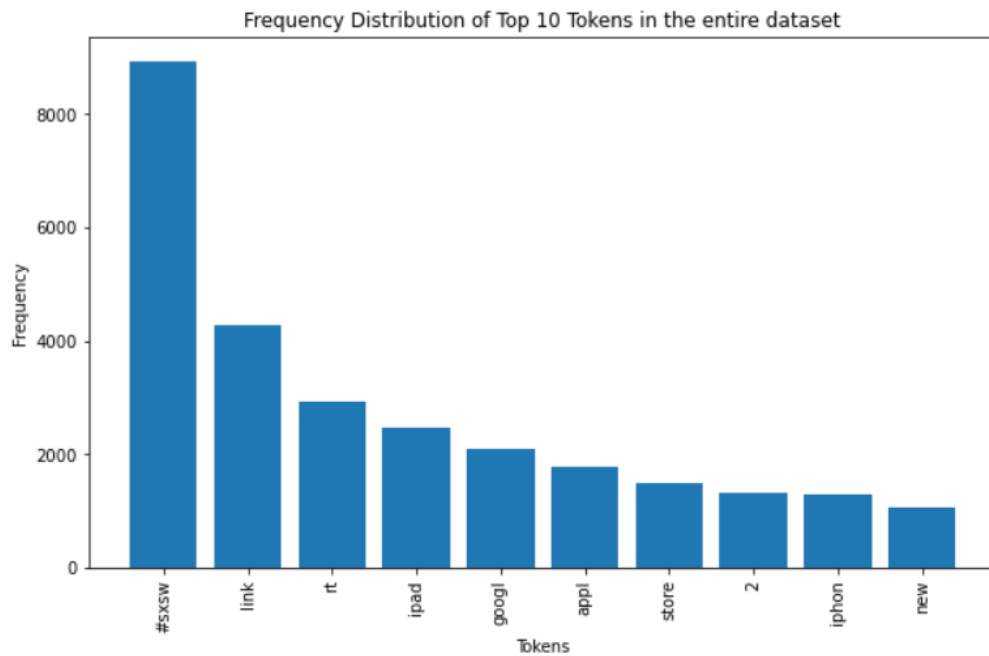
## 3. Distribution of the top 10 words (tokens) used for negative emotion.



The most common words in the tweets classified as negative emotions are #sxsxw (South by South West),ipad, iphone.



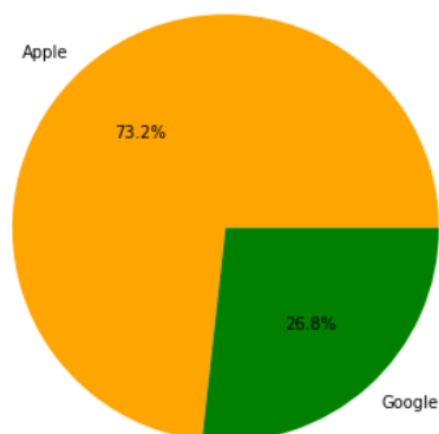
4. Frequency distribution of the top 10 words used in the entire dataset.



The most common words in all the tweets are #sxsxw (South by South West), link, rt.

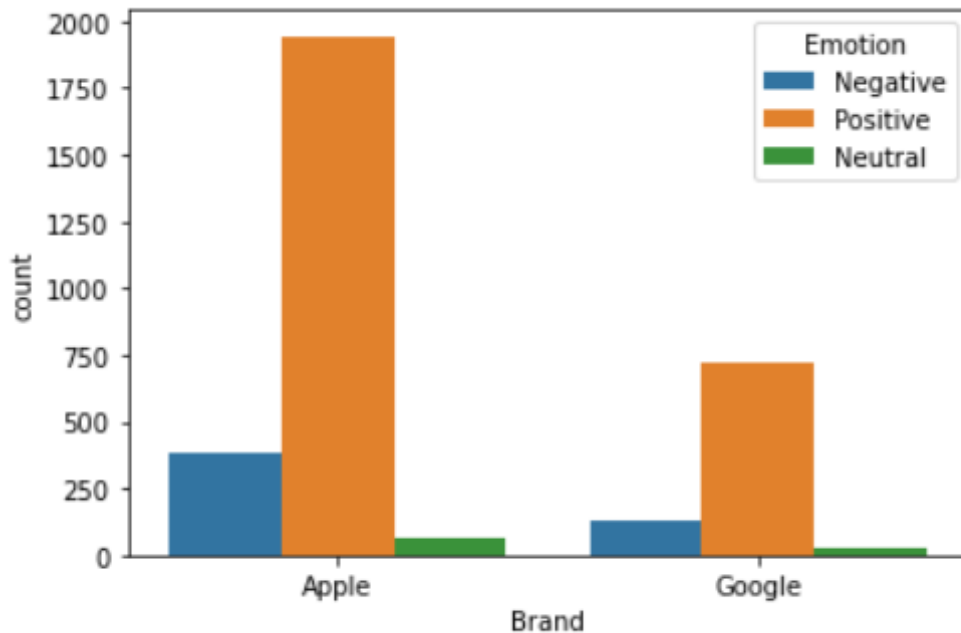
5. Distribution of tweets for apple and google products in the dataset.

Number of tweets for Apple and Google Brands



The Apple products are more by 73.2% compared to the Google brand at 26.8%

## 6. Distribution of emotions for each brand



For the Apple brand the Positive emotions were highest followed by the negative emotion then the neutral emotion. For the Google brand the positive emotions were highest followed by the negative emotion then the neutral being the least.

### Reviewing User sentiments on Specific Products and brands:

In this section, two functions were defined:

- i. A function that takes in the product/brand, and emotion and gives an output of a specified number of words about that brand, for example an input of Brand: “Apple”, Emotion: “Negative” and Number of tokens: “10” will give the top ten negative words used in tweets about apple products.
- ii. A function that takes in the “words” and produces a specified number of tweets where the word is used.

Analysis of this was performed and the following results were obtained:

a) **Praises** of Apple Products:

- The iPad 2 was launched, and customers seemed to enjoy the improvement in design made over the iPad.
- The new iPhone cases released during the SXSW Conference received a lot of attention on the tweets and the comments were generally very positive.
- There was a tweet thread of 'iPhone vs android' and iPhone was the more popular choice among users.
- The pop-up store in Austin created a lot of attention on the tweets and users in that area generally enjoyed the experience there.

b) **Complaints** of Apple Products:

- The battery life of the iPhone was a major source of complaint. Users complained that it does not last a whole day.
- The apple music app was described by some users as "one of the worst apps I have had to use in a long time".
- The increase in size of the new iPad 2 was disapproved by some users as 'too big'.

c) **Praises** of Google Products:

- The Packman app on Android Operating system was praised.
- Users were very interested in Google maps, and it proved to be the most popular Maps application with Marrison Mayer stating that, "Usage of google maps surpassed online use in the past couple of months" and that there were 150 million google map users.
- Google chrome browser had positive reviews when compared to Windows explorer with one user explaining that, "The switch is done immediately they buy a new laptop".
- A new Circle social media app the was rumoured to be launched created a lot of excitement among twitter users ahead of the launch event.

d) **Complaints** of Google Products:

- The Android operating system was the biggest pain point for most users who tweeted with bugs being the major source of concern, one user describing the experience as 'Painful'.
- Samsung products have been mentioned with the Android discussion and users stated that its implementation of android was better than that in google devices.
- Customer service at google was an issue with users complaining of a lack of refunds when returning faulty devices e.g. the Nexus smartphone.

# MODELING

In this section, the project was seeking to build both a binary and a multiclass classifier. The packages used in modelling were:

1. The NLTK (Natural Language Toolkit). This is a python library that provides resources for working with human language data. The features within the package that were used included:
  - i. Tokenization to split text into individual words. Here the TweetTokenizer was preferred as it is able to handle twitter handles adequately.
  - ii. Stemming to reduce the words to their base/root form. The PorterStemmer was used.
  - iii. Corpus to import a library of stop-words to be removed from the texts.
2. The sklearn (scikit-learn) package. This package was mostly used in the modelling process.
3. The imblearn (imbalanced-learn) package. This was used to address the challenges faced by an imbalanced dataset which could lead to bias. In this project, RandomOverSampler was used.

## 1. Binary Classification

A **baseline model** was used using the MultinomialNB (Multinomial Naive Bayes) class, from the package sklearn. It was used in this application of tweet sentiment analysis due to its ability to handle discrete features.

In this classification:

1. A **baseline model** was built by splitting the model into train and test and then after fitting the model. A **baseline train score** of **0.836495** was obtained.
2. Due to a **class imbalance**, **Random Over Sampler** was used to balance the model, and a second model was built with the oversampled data. A training accuracy score of 0.6079 was obtained and an F1 score of 0.57757.

## **2. Multiclass Classification**

Here, 4 models were built:

### **1. A dummy Model.**

For this model, a **DummyClassifier** from the **sklearn** package was used. This model was used as a benchmark to compare the performance of more advanced classifiers against the trivial random prediction strategy applied by the DummyClassifier.

For this model:

- i. A pipeline consisting of two steps: Vectorization and applying the dummy classifier was used.
- ii. A train test split was performed on the data
- iii. The model was evaluated using a function defined.

As this model was built using data with a class imbalance, The class imbalance was addressed using the Oversampling method and the model fit again.

### **2. A logistic regression Model.**

This model was built using the **LogisticRegressionCV** class from the **sklearn.linear\_model** module in scikit-learn. Unlike logistic regression, this model incorporates cross validation for automatic hyperparameter tuning.

For this model:

- i. A pipeline consisting of two steps: Vectorization and applying the LogisticRegressionCV model was used.
- ii. A train test split was performed on the data
- iii. The model was evaluated using a function defined.

Hyperparameter tuning was then performed on the model using the GridSearchCV class from sklearn. This was used to systematically search through a specified grid of hyperparameters to find the optimal combination that yields the best performance of a given model.

### 3. A Random Forest Classifier

This model was built using the **RandomForestClassifier** class from the **sklearn.ensemble** module in scikit-learn. It combines multiple decision trees to create an accurate classifier.

For this model:

- i. A pipeline consisting of two steps: Vectorization and applying the RandomForestClassifier model was used.
- ii. A train test split was performed on the data
- iii. The model was evaluated using a function defined.

Hyperparameter tuning was also applied to this model using Gridsearch CV

### 4. An XG Boost Classifier.

This model was built using the **XGBClassifier** class from the **xgboost** library. This model is based on the gradient boosting framework which sequentially combines weak prediction models like decision trees to create a powerful model.

For this model:

- i. A pipeline consisting of two steps: Vectorization and applying the XGBClassifier model was used.
- ii. A train test split was performed on the data.
- iii. The model was evaluated using a function defined.

Hyperparameter tuning was also applied to this model using Gridsearch CV.

# MODEL EVALUATION

## Evaluating the Binary models

In the binary classification section, we focused on optimizing our models for the F1 score since we would like our model to predict both negative and positive tweets correctly. For the binary classification problem, the best model was the tuned random oversampled Multinomial Naive Bayes model based on the score of 0.61 on the training dataset, and an F1 score of 0.58.

## Evaluating the Multiclass models

- In the multiclass classification the **best model** in this task was **the tuned oversampled logistic regression model** with an **accuracy score of 0.67**. The tuned random oversampled logistic regression model also had a score of 0.53 on the testing dataset and 0.54 on the training dataset; however, since it got the highest accuracy score compared to the rest, we are declaring the best model in this task.
- **XGBoost** model had the same score on the test dataset prediction of 0.65 and a test dataset prediction of 0.85 but with a lower accuracy score of 0.51 before and after GridSearchCV tuning.
- Since this is a multiclass model, it is more difficult to interpret how each word is affecting the prediction; however, it still provides insight into important words that **Apple and Google should keep an eye out for tweets which include words** like: “case”, “map”, “headaches”, “#fail”, “hate”, “battery” etc.



## CONCLUSION

In today's world, it is imperative for **businesses** to **listen to their customers attentively**. Actively paying heed to public opinion regarding their products and services is not only crucial for maintaining financial success but also opens doors for remaining competitive in the market.

By utilizing the models, we have developed, **Apple and Google** can **effectively monitor the sentiment** surrounding their events and products on various social media platforms. This approach would enable the company to stay well-informed about what people are expressing regarding their competitors, potentially granting them a competitive edge.

## RECOMMENDATIONS

1. From the analysis performed, users are not happy with iPhone's battery performance and therefore more research is needed in this area to **improve their product's battery life**.
2. Brands are advised to **monitor social media regularly** to stay informed about public opinions and respond promptly to customer feedback.
3. Brands are also advised to **extend sentiment analysis to include competitors' products and events** to understand customer perceptions and identify opportunities for differentiation and improvement.
4. **Improve Customer Service**, as there were complaints regarding customer service, ensuring quality service is provided would help improve the brand reputation.