billymwangi /
**nlp-project**

<> **Code**    ⊙ Issues    ⑂ Pull requests    ▷ Actions    ⊞ Projects    📖 Wiki    ⊘ Security    ∿ Insights    ⚙ Settings

👁    ⑂    ☆

☆ **0** stars    ⑂ **0** forks    👁 **1** watching    ∿ Activity

🌐 Public repository

⑂ main ▾

⑂ Branches    🏷 Tags

👤 **billymwangi**    ...                                        1 hour ago  🕑

View code

≡  **README.md**                                                                ✏

# Twitter sentiment natural language processing analysis for Google Products.



## Stakeholder:

A marketing team at a technology company.

# Business Overview

Let's understand a Sentiment Analysis problem from a business standpoint. The good news is: with the power of the internet, businesses today get a huge number of customer feedback through their business website, social media page, business listings, etc. However, the bad news is: a majority of businesses do not even know how to use this information to improve themselves.

# Business Problem Statement

The marketing team wants to understand consumer sentiment toward Apple and Google products. They wish to explore emotions, and sentiments expressed by users of Apple and Google products. They need insights that can inform marketing strategies, brand perception, and customer satisfaction initiatives for their target audience.

Currently, the marketing team collects sentiments manually by having a team classify each tweet as either positive, negative, or neutral. This is both expensive and takes a lot of time. It also means the team doesn't get real-time information to make quick and strategic course-corrections

The answer to the question of manual sentiment analysis is automation. Automating the process of sentiment analysis will also ensure that the team is able to get real-time sentiment insights -e.g. on a biweekly basis:

Knowing the number and actual positive sentiment tweets can enable the team to reinforce these positive attributes in advertising via consumer testimonials. This will serve to boost brand credibility and boost sales among new customers

Knowing the number and actual negative sentiment tweets will enable the team to know what product improvements to make, and to reach out to consumers with the issues to have them sorted out. This will reduce customer churn

## The main business objective is:

How do we use available data to create a model that will be able to classify future tweets with the correct sentiments? This will save the team both time and money by moving away from the manual process This will also help grow loyalty and sales, as articulated above in the problem statement

The specific business objectives to be answered include:

## What is the overall sentiment towards Apple and Google products on Twitter?

Are there any recurring themes or topics associated with positive or negative sentiments towards these brands? Are there any notable differences in sentiment between Apple and Google products?

### Why is Twitter Sentiment Analysis Important?

Understanding Customer Feedback: By analyzing the sentiment of customer feedback, companies can identify areas where they need to improve their products or services.

Reputation Management: Sentiment analysis can help companies monitor their brand reputation online and quickly respond to negative comments or reviews.

Political Analysis: Sentiment analysis can help political campaigns understand public opinion and tailor their messaging accordingly.

Crisis Management: In the event of a crisis, sentiment analysis can help organizations monitor social media and news outlets for negative sentiment and respond appropriately.

Marketing Research: Sentiment analysis can help marketers understand consumer behavior and preferences, and develop targeted advertising campaigns

## Data Understanding

We work with the Twitter sentiment dataset that involves customers' emotions and tweets towards different brands and products. The dataset contains 9000 columns and 3 rows With the columns being : tweet: Has the tweet text and message from customers is_there_an_emotion_directed_at_a_brand_or_product: has either positive, negative,no emotion, can't tell brand/product: that included iPhone,iPad, google

## Dataset Description

The data is in CSV format. In computing, a comma-separated values (CSV) file stores tabular data (numbers and text) in plain text. Each line of the file is a data record. Each record consists of one or more fields, separated by commas.

Attribute Information

tweet text: The tweets in the given dataset

emotion in a tweet is directed at The tweets collected from various sources and the types of devices associated with('Negative emotion or positive emotion)

emotion directed at a brand or product: having either positive or negative sentiments associated with Google and Apple products.

## Project Overview

The main aim of this project is to analyze the tweet-product-dataset, obtained from Twitter.

Through this analysis, we are to achieve the following:

we plotted the most frequent word distribution across different sentiments
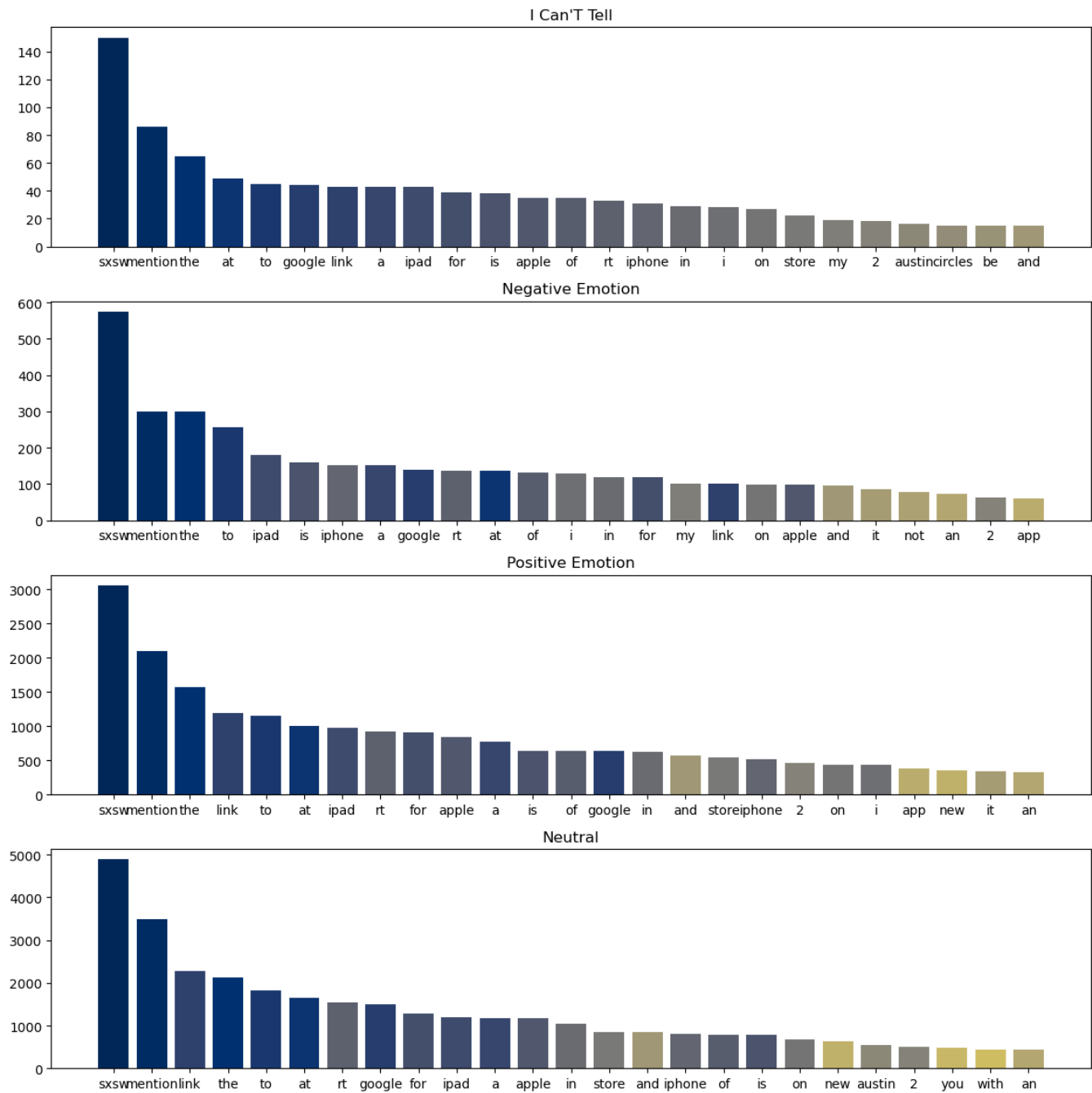


Identify some of the sentiments associated with the tweets of the various products.

Process the data and analyze it to determine its key characteristics.

Come up with a model that accurately evaluates and predicts the sentiments towards given product lines.
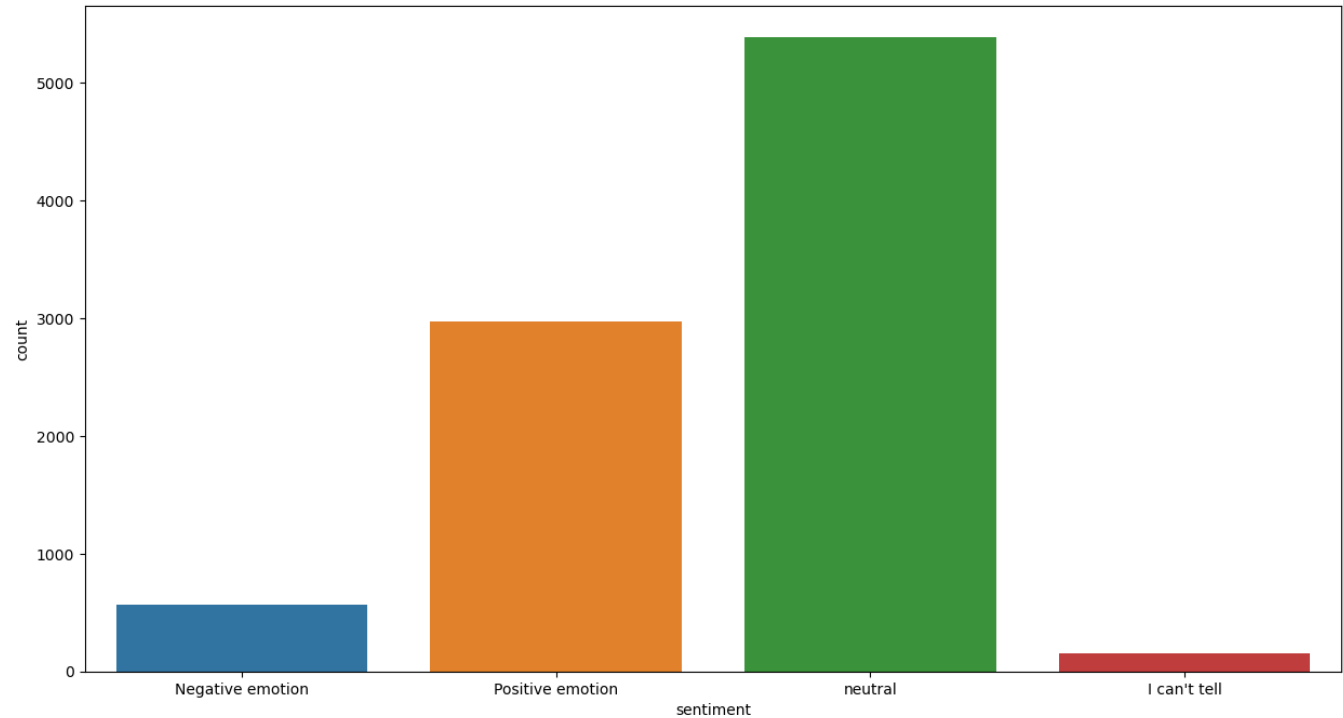
## The project involves the following steps:

Data preparation:

The dataset is split into training and testing sets. The training set is used to train the model, and the testing set is used to evaluate the model's performance.

Feature extraction:

Various techniques are used to convert text reviews into numerical features that can be used by machine learning algorithms. These techniques include bag-of-words, TF-IDF, and word embeddings.

Model selection:

Different machine learning algorithms are experimented with to find the best-performing model. These algorithms include logistic regression, support vector machines, and neural networks.

Model training and evaluation:

The selected model is trained on the training data and evaluated on the testing data. The model's performance is assessed using evaluation metrics such as accuracy, precision, recall, and F1 score.

The goal of this project is to provide businesses with valuable insights into customer sentiment. This information can be used to make data-driven decisions, improve customer satisfaction, and enhance brand reputation.

In addition to the steps mentioned above, the following data preprocessing steps were performed:

Special characters were removed.

Stopwords were removed.

Tokenization was performed.

Lemmatization was used to reduce words to their common root form

Vectorization Techniques

Term Frequency-Inverse Document Frequency (TF-IDF Vectorizer)

Classification Models

we plotted word clouds for different sentiments

## Success Metrics

The team decided to use accuracy as the performance metric for the models. This is the case as we give equal importance to correctly identifying each sentiment. This is because although it seems, from a business perspective, correctly identifying neutral and negative tweets is more important than correctly identifying positive sentiment tweets, these categories are codependent on each other as there are three sentiment categories. Our success metric for our project is 60% model accuracy.

# Evaluation

To evaluate our NLP sentiment analysis model, we used a variety of evaluation metrics to conduct thorough testing and analysis. The results of these evaluations provide insights into the effectiveness of our approach. Here is a breakdown of the key terms in the original text: Evaluate: To evaluate the performance of our Twitter sentiment analysis project Performance: Evaluate the accuracy of our model to correctly classify tweets. NLP sentiment analysis model: A model that can automatically customer reviews as positive, negative, or neutral. Evaluation metrics: Measures that are used to assess the performance of a model. Insights: Understandings or realizations that are gained from our data.

# Conclusion

The main objective of this project was to create a model that can classify Twitter product reviews/ tweets into the appropriate sentiments-positive, negative, and neutral. Through this project, we have met this objective, and also obtained the following results:

- Most of the customers have neutral emotions towards the products, followed by those with positive emotions. There are customers with negative emotions as well.

- The most reviewed/tweeted about products are Apple followed by Google products which are expected in the US market demographic.

- The accuracy of predicting the product user sentiment is at around `61%`, following the best modeling outcome, after optimizing through tuning. This meets our success metric of `60%` accuracy.

# Recommendations

Based on the above conclusions, we recommend the following:

- The stakeholders should target majorly the users whose sentiments towards their products are negative so as to retain and further add to their customer base.

- Our model will help the team to access faster and real-time insights on product reviews since they won't do this manually anymore.

- Knowing the number and actual positive sentiment tweets will enable the team to reinforce these positive attributes in advertising via consumer testimonials. This will serve to boost brand credibility and boost sales among new customers.

- Knowing the number and actual negative sentiment tweets will enable the team to know what product improvements to make, and to reach out to consumers with the issues to have them sorted out. This will reduce customer churn.

-

- We also expect that the team will have better customer engagement since they will be responding with strategies based on the voice of the customer. This will make customers feel heard and improve loyalty and engagement with the brand.

# Next Steps

Try to use neural networks and transfer learning to improve the accuracy of the model.

- Deploy the model to enable the marketing team to have an interface to work with.

- Develop dashboards to enable the real-time display of insights gleaned from customers' product reviews/tweets.

**Note on our Next steps**

- We found some time to also run another model(a neural network model) -RoBERT(Robustly Optimized BERT approach) is a state-of-the-art natural language processing (NLP) model- on our data this model achieved an accuracy of  66%
- code for that model is included here- ROBERTA-MODEL

# Collaborators

The project was a collaborative effort, and the following people contributed their expertise:

```
Billy Mwangi

Fiona Njuguna

Kinoti Mwenda

Mary Wairimu

Patrick Okore

Nyokabi Waiganjo

Brain Njau
```

**Releases**

No releases published
Create a new release

**Packages**

No packages published
Publish your first package

**Contributors**  7

**Languages**

- **Jupyter Notebook** 100.0%