Understanding Twitter Sentiment on Apple and Google Products.

Exploring user sentiment towards tech giants' products through social media analysis.

Content

Introduction

- Public opinion shapes business strategies in today's interconnected world, where social media influences daily decisions. Consumers constantly share opinions about the products they use, creating a rich pool of data.
- For companies like Google and Apple, analyzing this sentiment is crucial to understanding customer perceptions, staying relevant, and adapting to trends. By leveraging sentiment analysis, businesses can align their strategies with public opinion, enhancing their competitiveness and ensuring long-term success.

Business Overview

Sentiment Flow Analysis

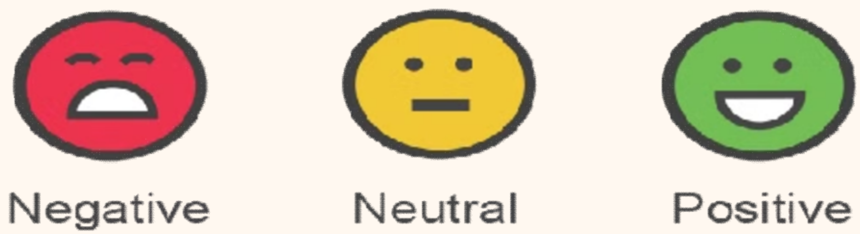
Aims to tackle the real world problem related to understanding public sentiment towards Apple and Google products on Twitter .

Key Stakeholders

Apple, Google, and their dedicated teams (marketing& Customer support).

Business Problem

This project analyzes Twitter sentiments to understand customer satisfaction with Apple and Google products. We classify tweets as positive, negative, or neutral, providing insights into customer feedback. This data can help identify emerging issues, guide product improvements, and inform marketing and customer support strategies.understand customer satisfaction, identify potential issues and tailor their responses accordingly.



Objectives

Key Objective

Develop an NLP multiclass classification model for sentiment analysis to categorize sentiments as Positive, Negative, or Neutral, with a target accuracy and recall above 80%.

Specific Objectives

1. To Identify the most common words used in the dataset using Word Cloud .
2. To confirm the most common words that are positively and negatively tagged
3. To recognize the most opined products by users
4. To spot the distribution of sentiments

Data Understanding

1. Source of Data: The dataset originates from CrowdFlower via data.worldInitial Dataset

2. Products being focused on: Apple & Google Products

3. The data : Twitter sentiments

   Data Description: Over 9000 rows and 3 columns

4. The sentiments : Positive, Negative & Neutral.

5. The columns :

   - Tweet text (Tweet)
   - Emotion in tweet is directed at (Product)
   - Is there an emotion directed at a brand (Emotion)

Data Preparation

STEP

1. Data Cleaning

TASKS

- Removing corrupted records
- Neutral sentiment adjustment
- Dropped irrelevant records
- Dropped rows with missing values
- Removing duplicates

OUTPUT

- Dataset with **8,439 rows,** no missing values or duplicates, and columns with consistent naming and content

2. Text Preprocessing

TASKS

- Lemmatization
- Stop word removal
- Tokenization
- Part-of-speech tagging

OUTPUT

- Cleaned tweets stored as lemmatized tokens in a new column, with the final cleaned text saved in the 'clean_tweet' column.

3. Visualizations

TASKS

- Visualization using frequency distributions and bigrams, highlighting product-related terms
- Word cloud visualization

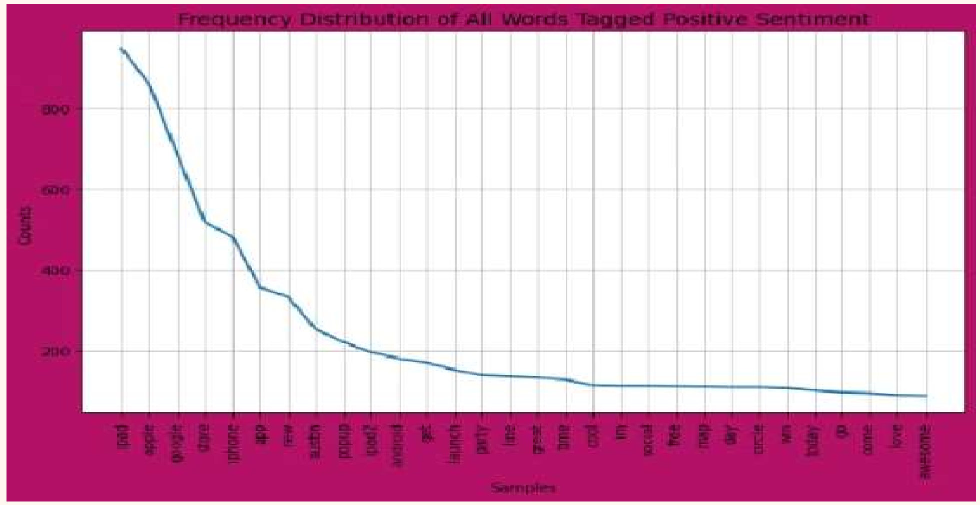
OUTPUT

- Bigram
- Word cloud

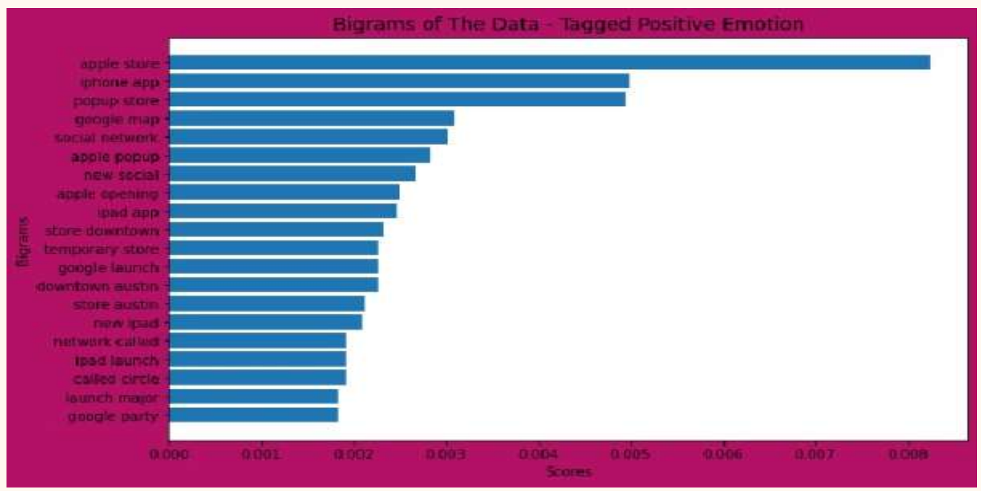
Data Visualization.

Positive Emotion Visualization

The most frequent words used were 'ipad', 'apple', 'google'. The count is between 600-870

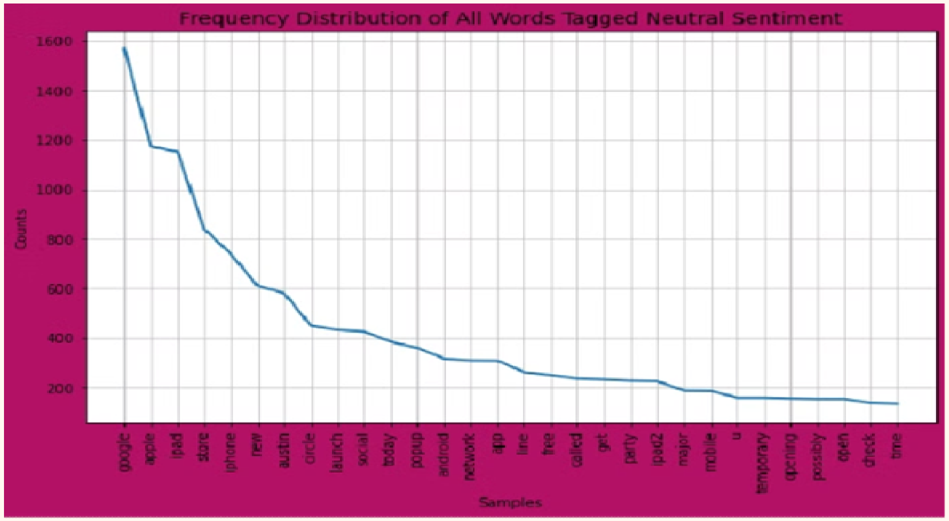


'Apple store', 'iphone app' and 'popup store' were the most frequent word combinations used.

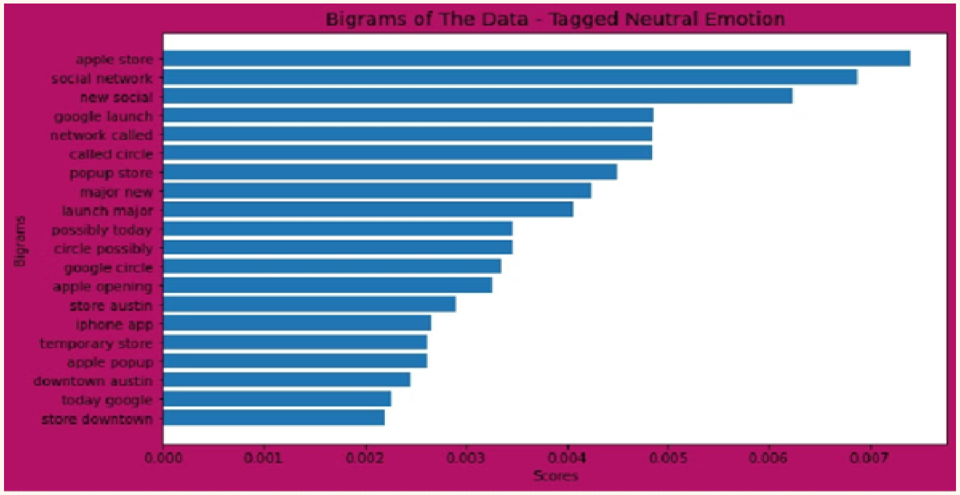


Neutral Emotion Visualization

The words 'google', 'apple' and 'ipad' appeared more frequently than all other words. The count being over 1000 thus proving class imbalance

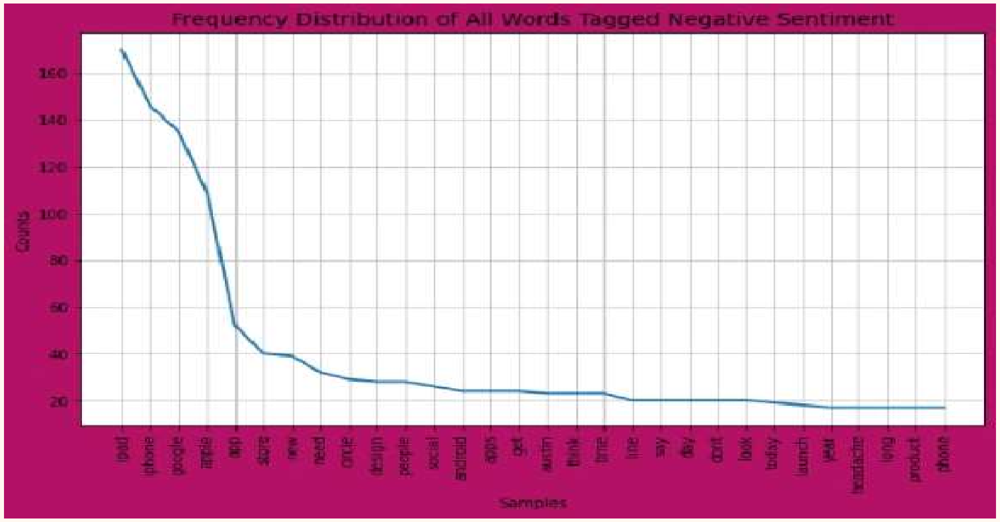


'Apple store', 'social network' and new social' had the highest scores. Note the scores are slightly lower than Positive Emotion.

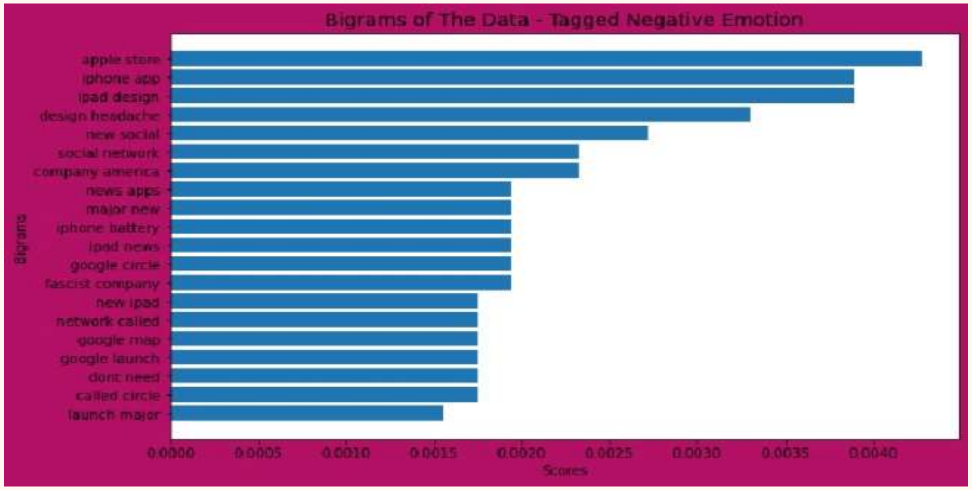


Negative Emotion Visualization

The words 'ipad', 'iphone', 'google' and 'apple' appeared more frequently than all other words. however the count is less than neutral and positive sentiments

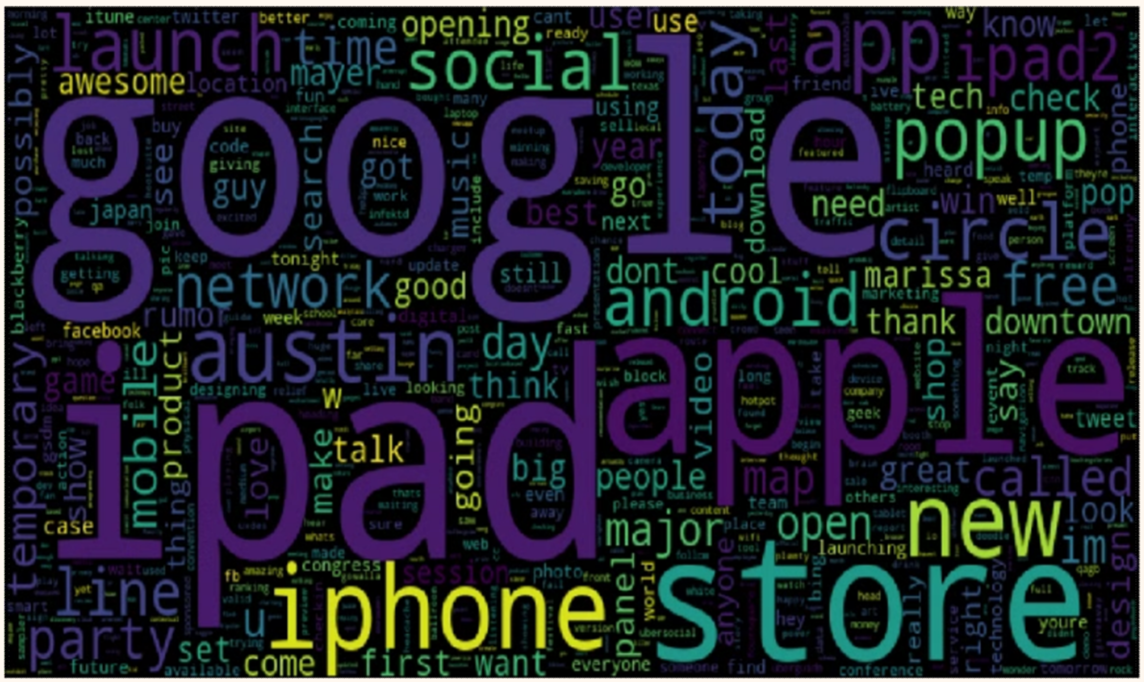


"Apple store', 'iphone app', 'ipad design' had the highest scores.



Word Cloud

The words google, ipad, apple, store and iphone are the most visible words used in the dataset



Modelling

Text Preparation for Modeling

1. **Label Encoding:**

   Converting emotion labels into numerical values.

2. **Vectorization:**

   Using TF-IDF and Count Vectorizer to transform text data into numerical vectors.

3. **SMOTE:**

   Applying SMOTE (Synthetic Minority Over-sampling Technique) to handle class imbalance.

4. **Train-Test Split:**

   Split the dataset into training and testing sets for model evaluation.

Model Performance

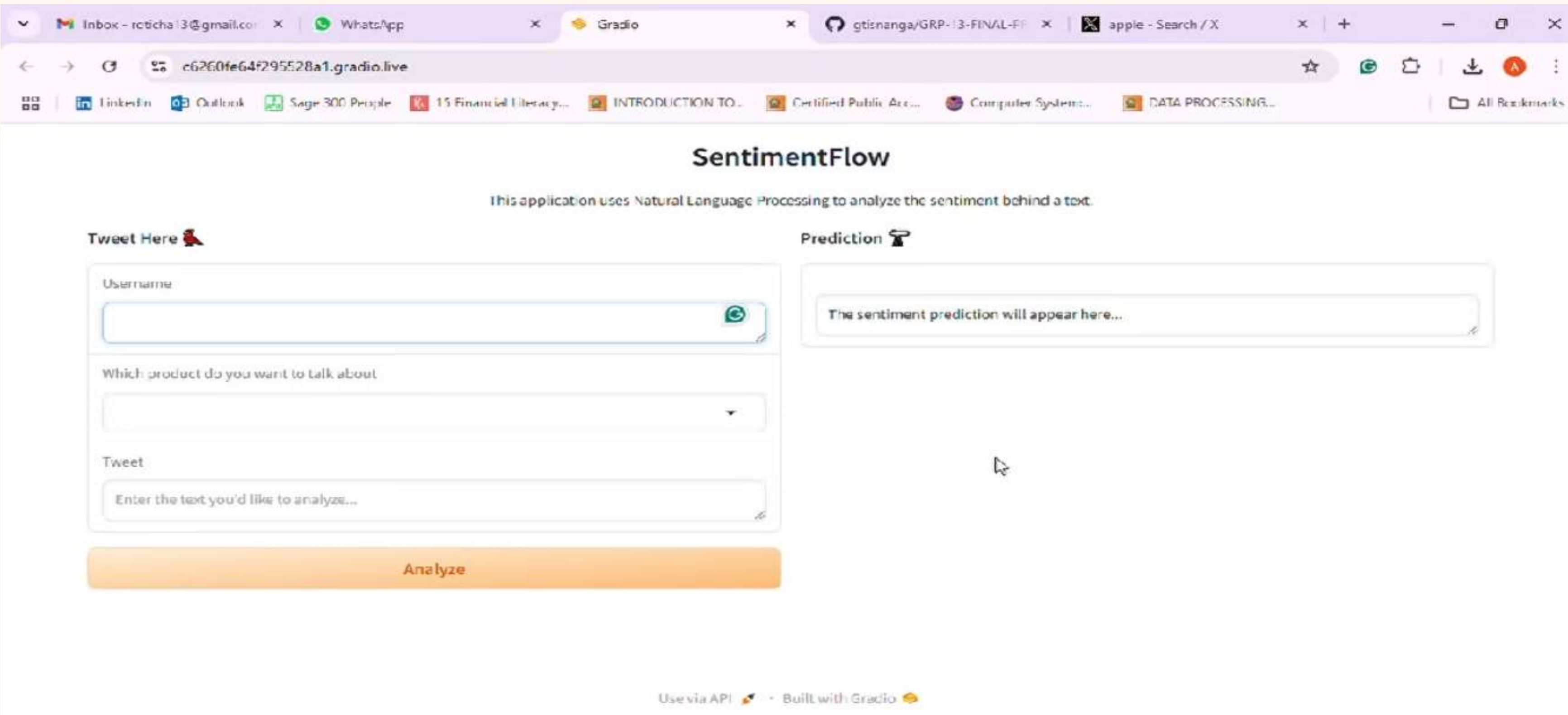
The Best Model was the Random Forest Model.

Logistic regression was quite close but due to this project being about solving a classification problem, Random Forest classifier was the best option for deployment.

Models were produced after hyper parameter tuning and TF-IDF Vectorization .

Model	Accuracy Score	Recall Score
Tuned Random Forest	83.7%	83.6%
Tuned Logistic Regression	82.7%	82.6%
Tuned Naïve Bayes	80%	79.9%
Tuned Decision Trees	76.5%	76.4%

Deployment


SentimentFlow
This application uses Natural Language Processing to analyze the sentiment behind a text
Tweet Here
Username
Which product do you want to talk about
Tweet
Enter the text you'd like to analyze...
Analyze
Prediction
The sentiment prediction will appear here...
Use via API
Built with Gradio

Conclusion

1

Project Summary

We successfully employed machine learning model to classifying tweets into emotion categories for Apple and Google Products.

2

Project Impact

The project's results offer valuable insights into effective sentiment analysis for social media data. The models developed provide robust tools for classifying tweets and understanding public sentiment, contributing to advancements in sentiment analysis and machine learning applications.

3

Room for Improvement

The project highlights the importance of effective data preprocessing, model evaluation, and hyper parameter tuning in achieving high-performance sentiment analysis. The findings emphasize the potential of machine learning models in practical applications and pave the way for future research and enhancements in the field.

Recommendations

Real-Time Monitoring

Implement alert systems to flag negative sentiments for prompt customer issue resolution.

Scalability:

Optimize models to handle large-scale data efficiently for high-volume tweet processing.

Real-Time Processing:

Enable real-time sentiment analysis to support timely decision-making and trend response.

Continuous Monitoring:

Regularly track model performance, retraining with new data to ensure accuracy and relevance.

Platform Integration:

Connect models to social media APIs for seamless data collection and continuous insights.

Team

- Amos Kipng'etich
- Alex Miningwa
- Angela Maina
- Charles Ndegwa
- Gloria Tisnanga
- Sandra Koech
- Sylvia Manono

THANKYOU Questions & Clarification.