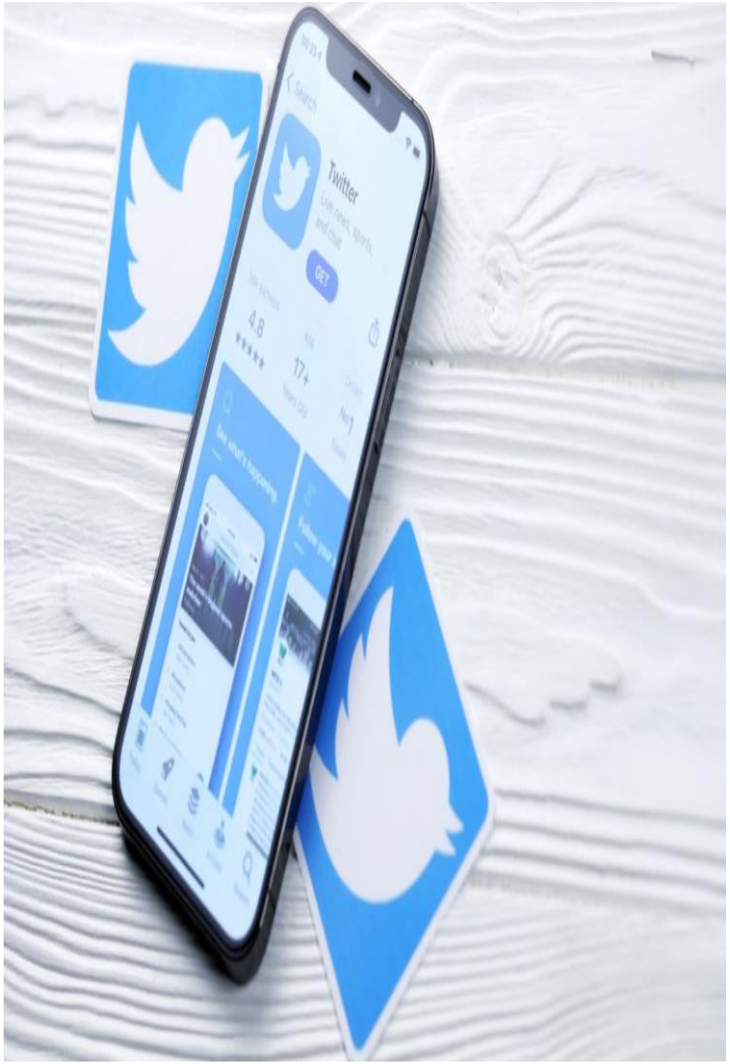
Customer Sentiment Analysis

Outline

- Background
- Business Problem
- Data
- Methods
- Results
- Conclusions



Background

- Apple and Google has released products.
- Consumers are tweeting their opinions on twitter.
- Using Natural Language Processing (NLP), we can determine what are the most common words being used and what tone.
- Identify the best model for future predictions.

Business Problem

- Identify what customers are saying about the new phones and company.

Data

- Using the tweet product dataset. Dataset has over 9000 tweets, and 3 columns.
- Contains data including
  - Tweets
  - Focus of the tweet
  - Sentiment

Methods

- Removal of special characters (@,#, etc.)
- Applying NLP Techniques such as Stemming, Lemmatization, and Tokenization.
- Removal of small words (less than 3 characters in length)
- Removing stop words
- WordClouds – Visualizations of most important words

Methods

Build machine learning models with Word Vectorizers to predict future sentiments (using CountVectorizer and Term Frequency - Inverse Document Frequency (TFIDF)

Developed several models to evaluate performance

- Precision – correct predictions vs total predictions
- Recall – Correct predictions vs actual positive predictions
- F1 score – harmonic mean of Precision and Recall
- Accuracy

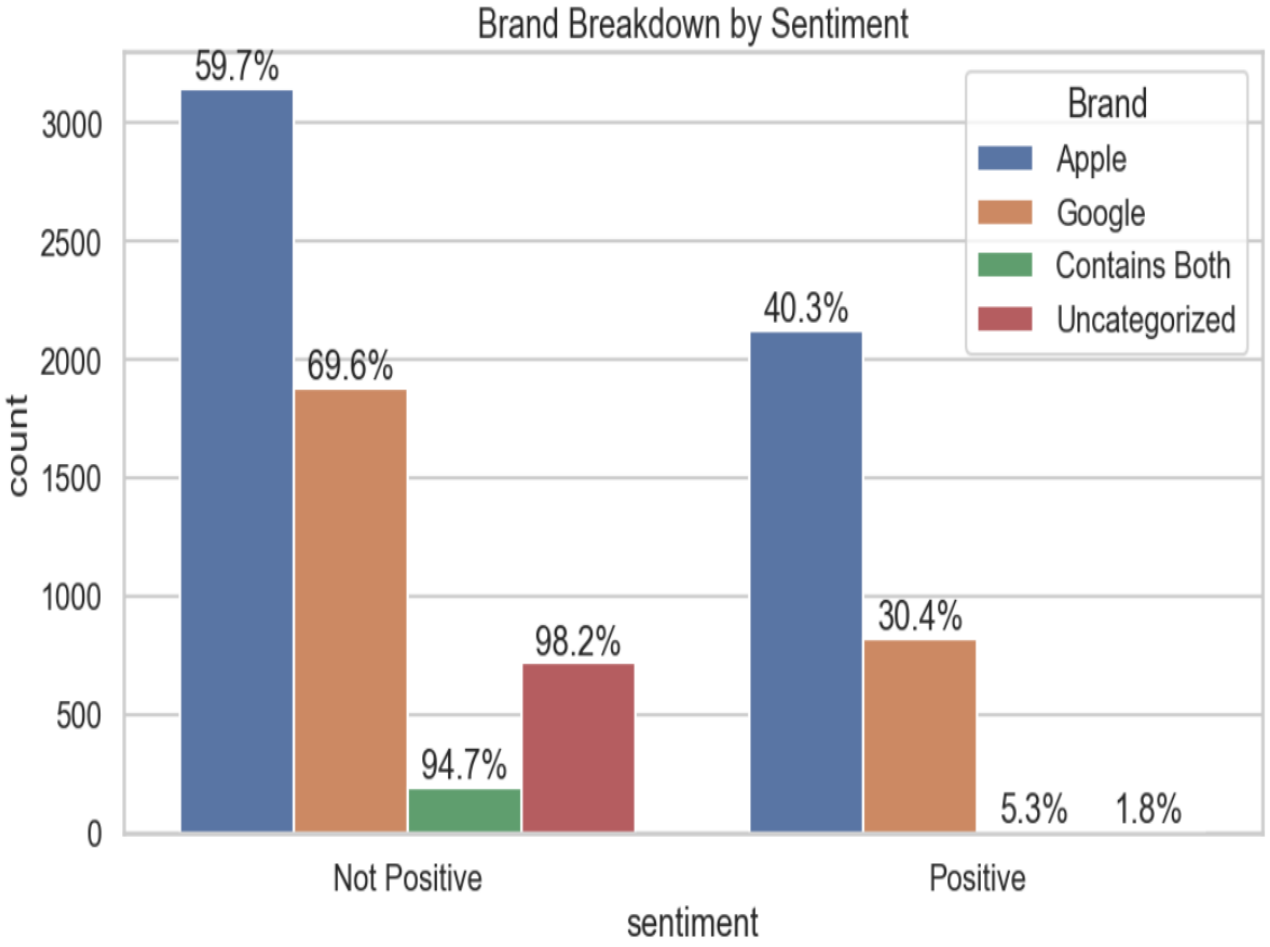
Methods

A Little bit about the Vectorizer models…

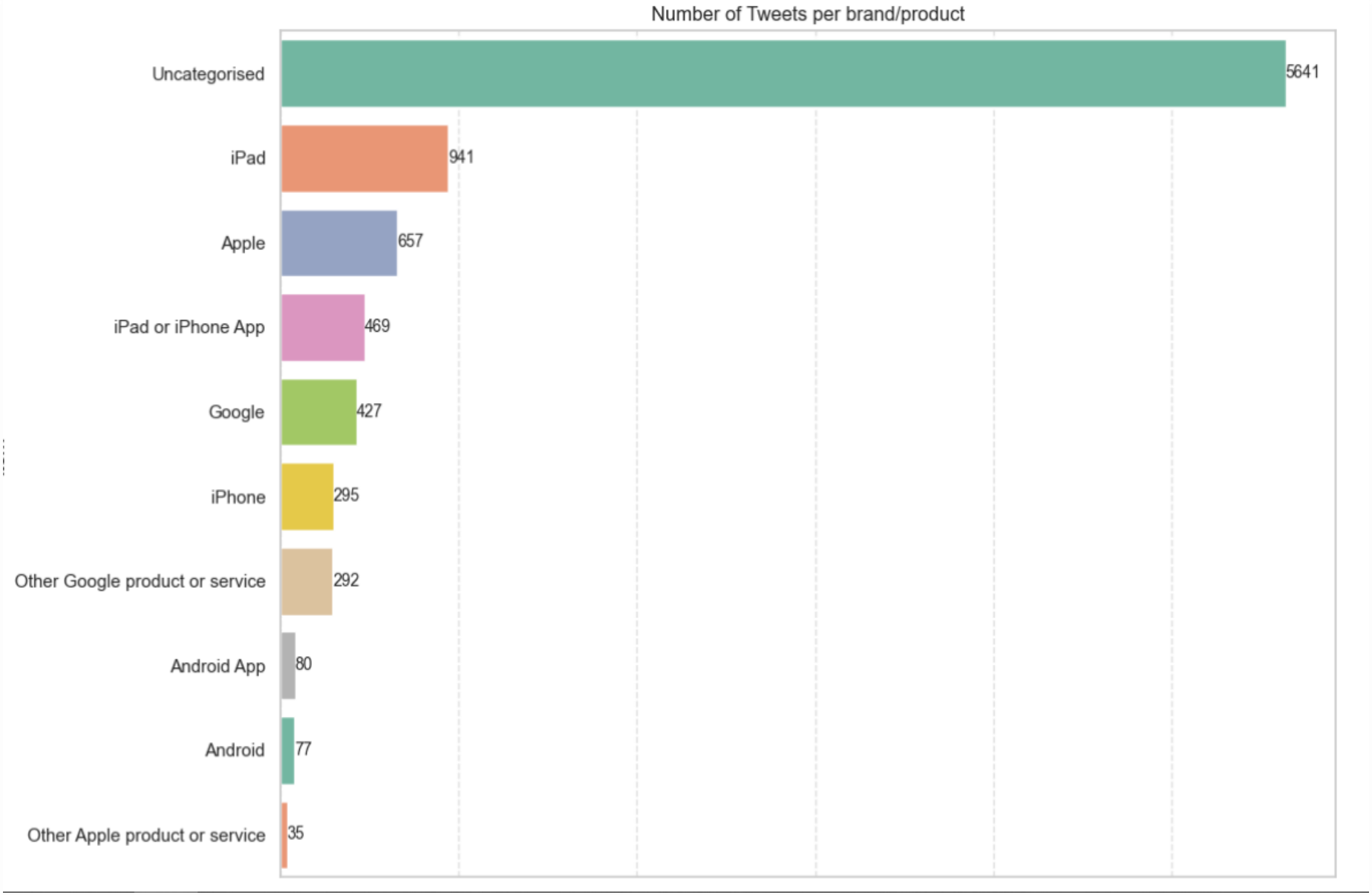
CountVectorizer counts number of times a word appears.

TFIDF (Term Frequency – Inverse Document Frequency) – counts the number of words but considers overall document weightage

Brand Breakdown Sentiment



Number of Tweets per Product



Results

dinst

WordCloud for positive Emotion Sentiment (Apple)

Results

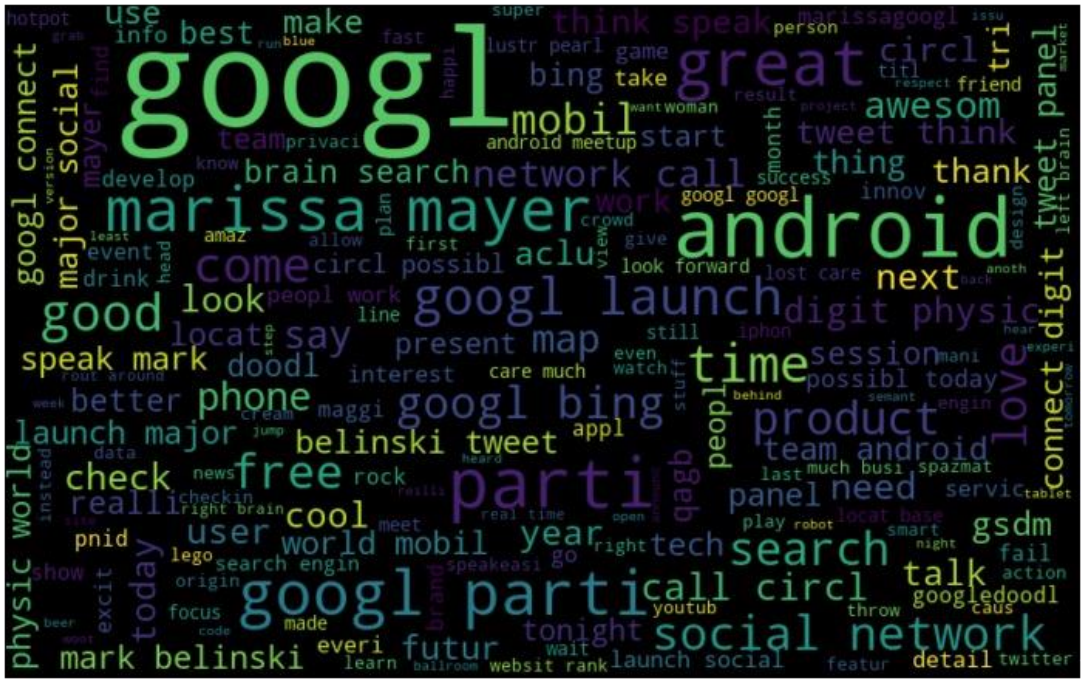
WordCloud for positive Emotion Sentiment (Google / Android)

Results

WordCloud for all Apple Sentiments

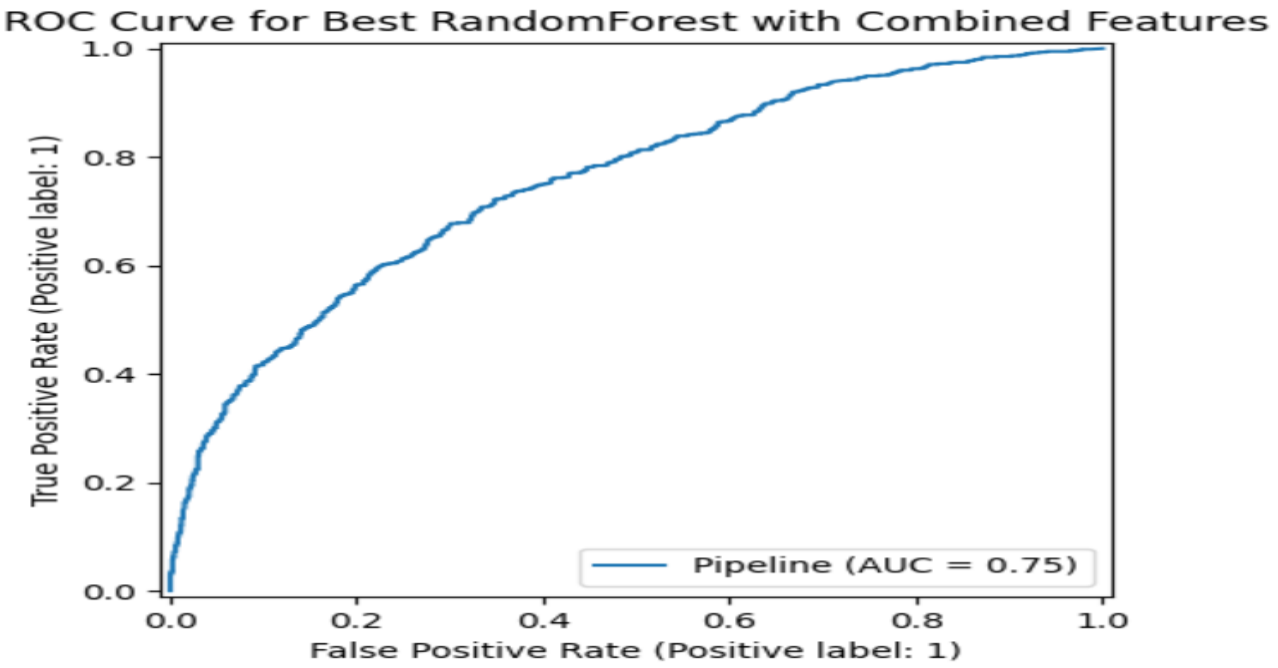
Results

WordCloud for all Google Sentiments

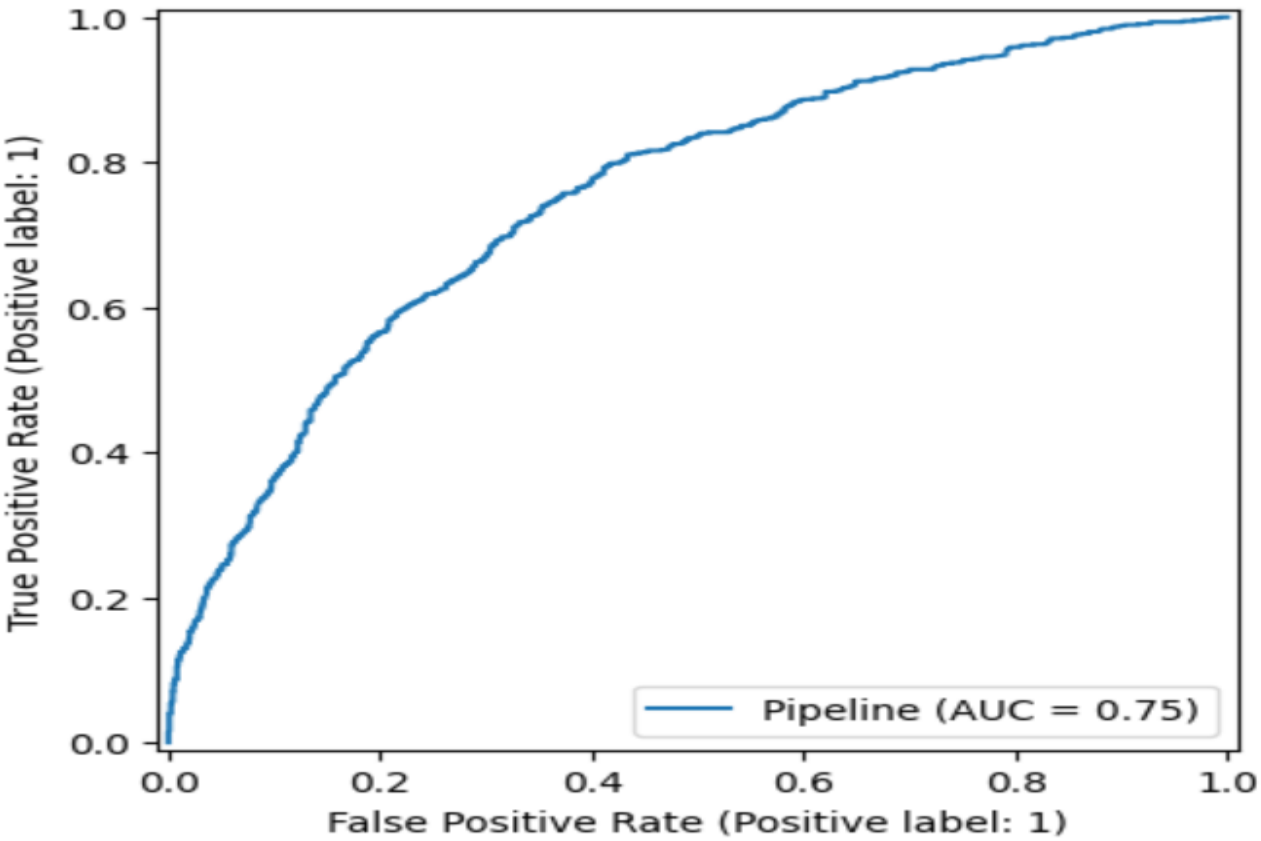


Results

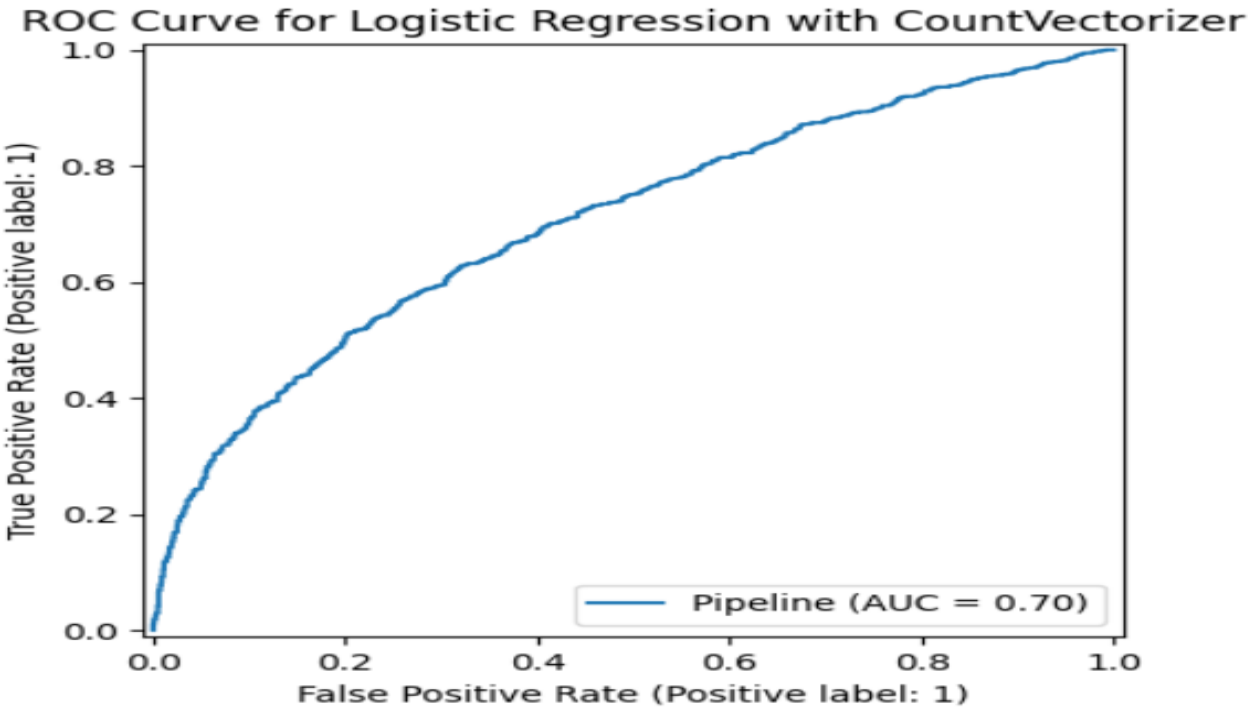
overall accuracy of 71%, correctly classifying 71% of the tweets

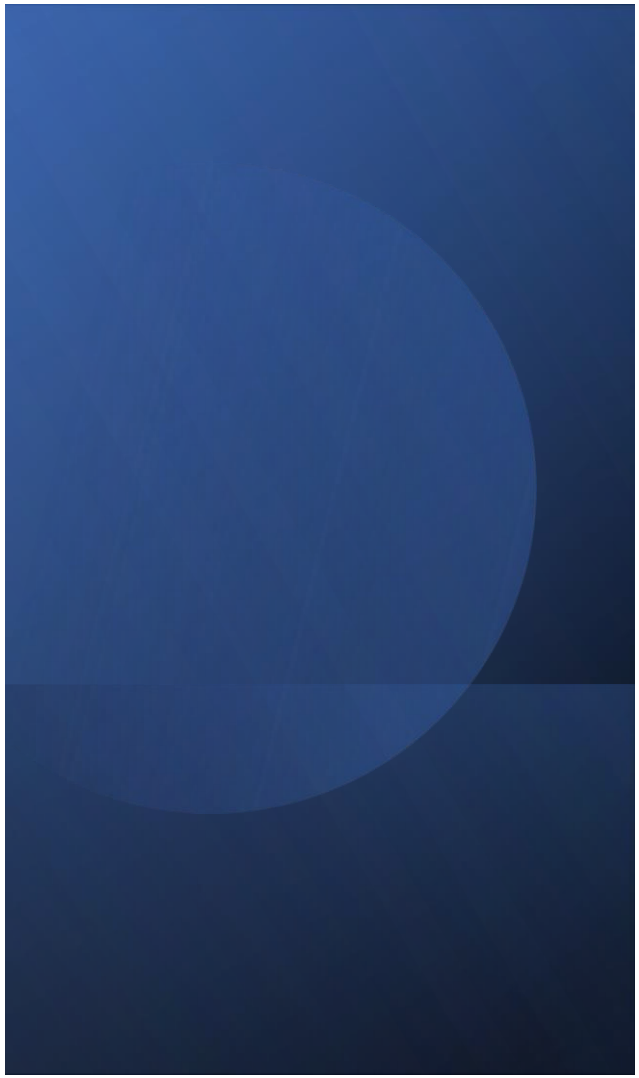


Roc curve on grid search



Logistic model with TfidfVectorizer



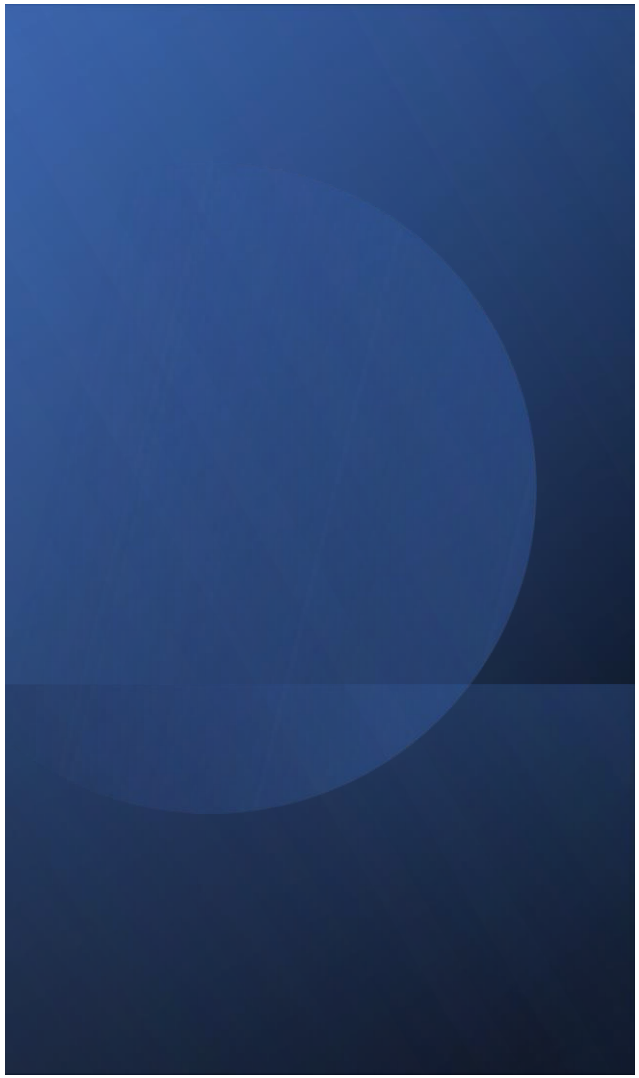


Result

TfidfVectorizer is the better choice in this scenario as it demonstrates higher accuracy, better precision, recall, and F1-score for both classes.

Summary of Findings / Conclusion

- Most positive words were associated with events for Google and Apple
- Most negative words associated with the product or negative words (such as battery) for Apple.
- Most negative words were associated with the event for Google.
- The TFIDF Vectorizer is the best for predicting future sentiments.
    - TFIDF Vectorizer outperformed Precision, Recall, F1 score against CountVectorizer



Recommendations

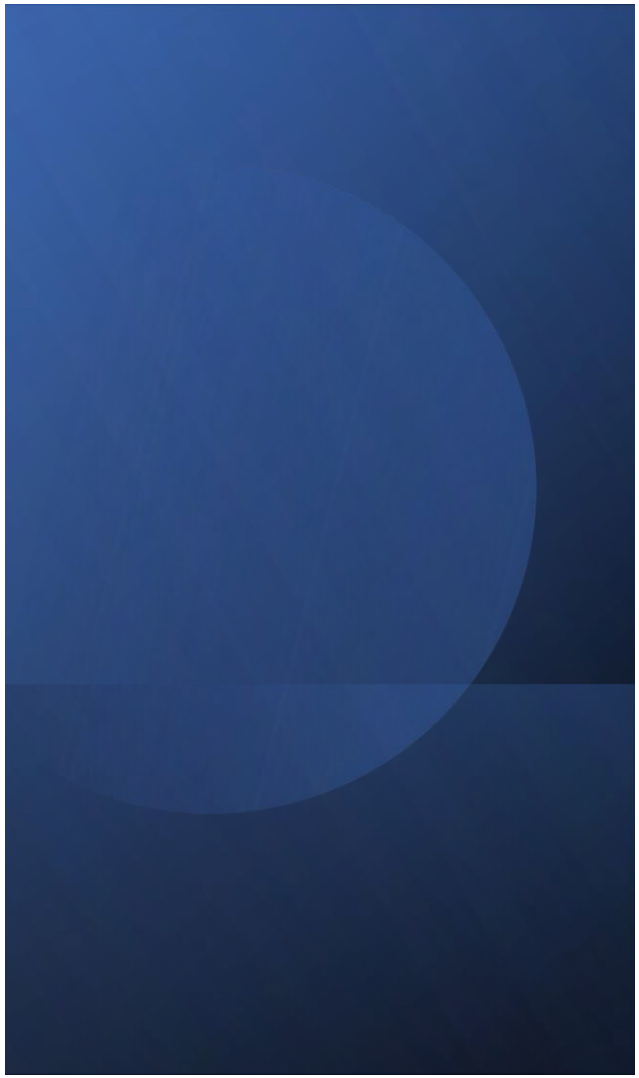
For Apple:

1.Address Battery Life Issues: Prioritize enhancements in battery technology and customer education on battery management.

2.Improve Store Experience: Reduce wait times and enhance customer service.

3.Promote Positive Experiences: Continue with promotional giveaways and leverage positive feedback in marketing campaigns.

4.Respond to Criticisms: Be proactive in addressing criticisms about business practices and product design.



For Google:

Enhance Product Functionality: Focus on improving the mapping product and resolving any issues with new product launches.

Leverage Positive Feedback: Use successful events and launches to strengthen marketing and customer engagement strategies.

Address Negative Sentiment: Actively respond to criticisms and controversies to maintain customer trust and satisfaction.

Thank you!