

Sentiment Analysis of Apple and Google Product Tweets

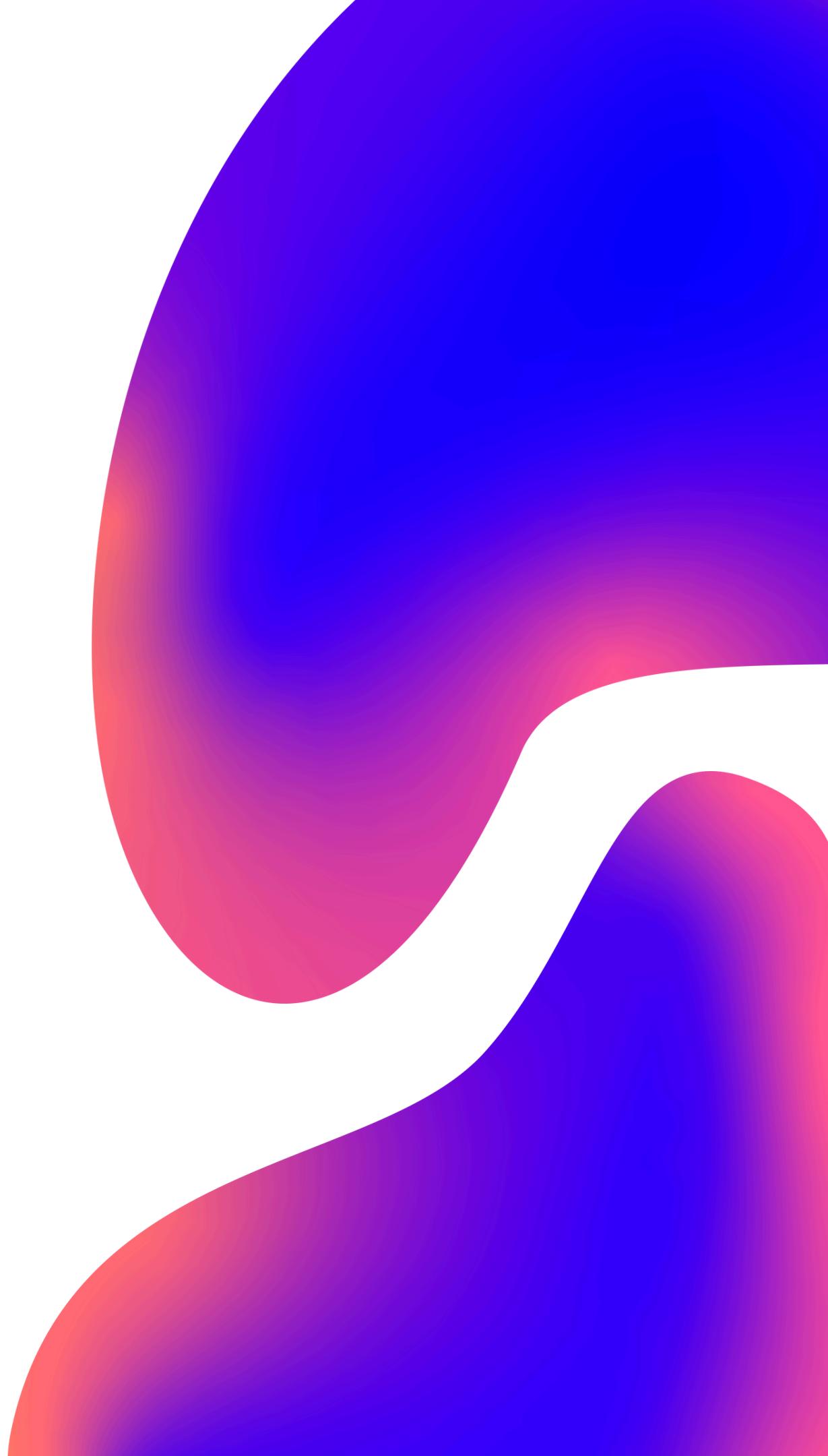
Introduction

- This project focuses on building a Natural Language Processing (NLP) model to classify the sentiment of Twitter posts related to Apple and Google products.
- Using a labeled dataset of tweets, the model predicts whether a tweet expresses positive, negative, or neutral sentiment.
- The project demonstrates an end-to-end NLP workflow, from data understanding and preparation to feature engineering and modeling, with the goal of producing a practical proof of concept for sentiment analysis.

Problem Statement

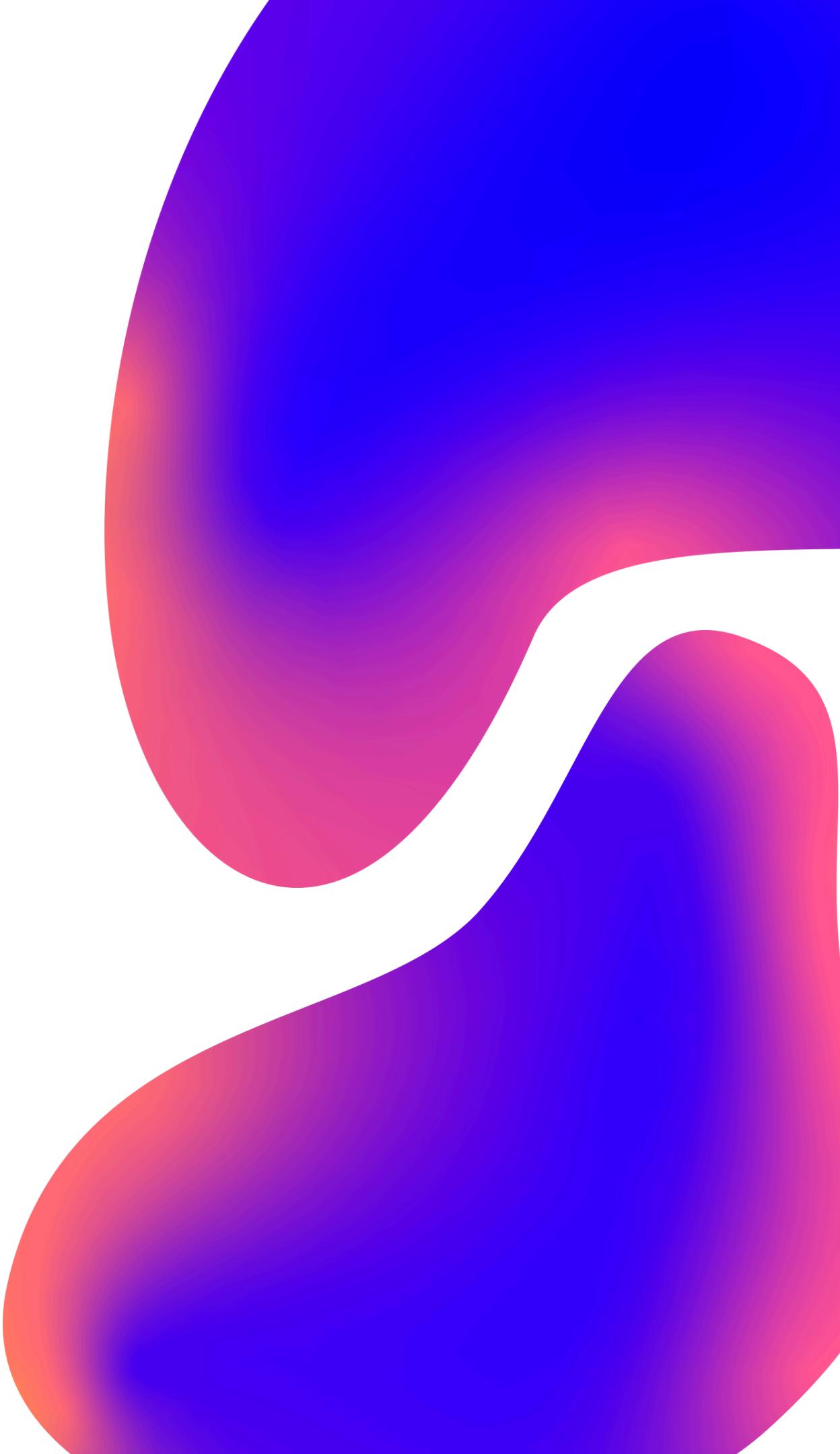
Technology companies receive thousands of tweets daily from users expressing opinions about their products.

Manually analyzing this feedback is inefficient, subjective, and not scalable. As a result, organizations struggle to quickly understand public sentiment and respond to customer concerns in a timely manner. This project addresses the challenge of automatically classifying tweet sentiment to help stakeholders efficiently analyze large volumes of social media data.



Business Understanding

- 1.** This project addresses the challenge of monitoring customer sentiment at scale on social media platforms such as Twitter. It supports stakeholders including marketing teams, product managers, and social media analysts at Apple and Google by automatically classifying tweets as positive, negative, or neutral.
- 2.** By interpreting consumer sentiment, businesses can gain actionable insights, address issues promptly, and refine marketing strategies.



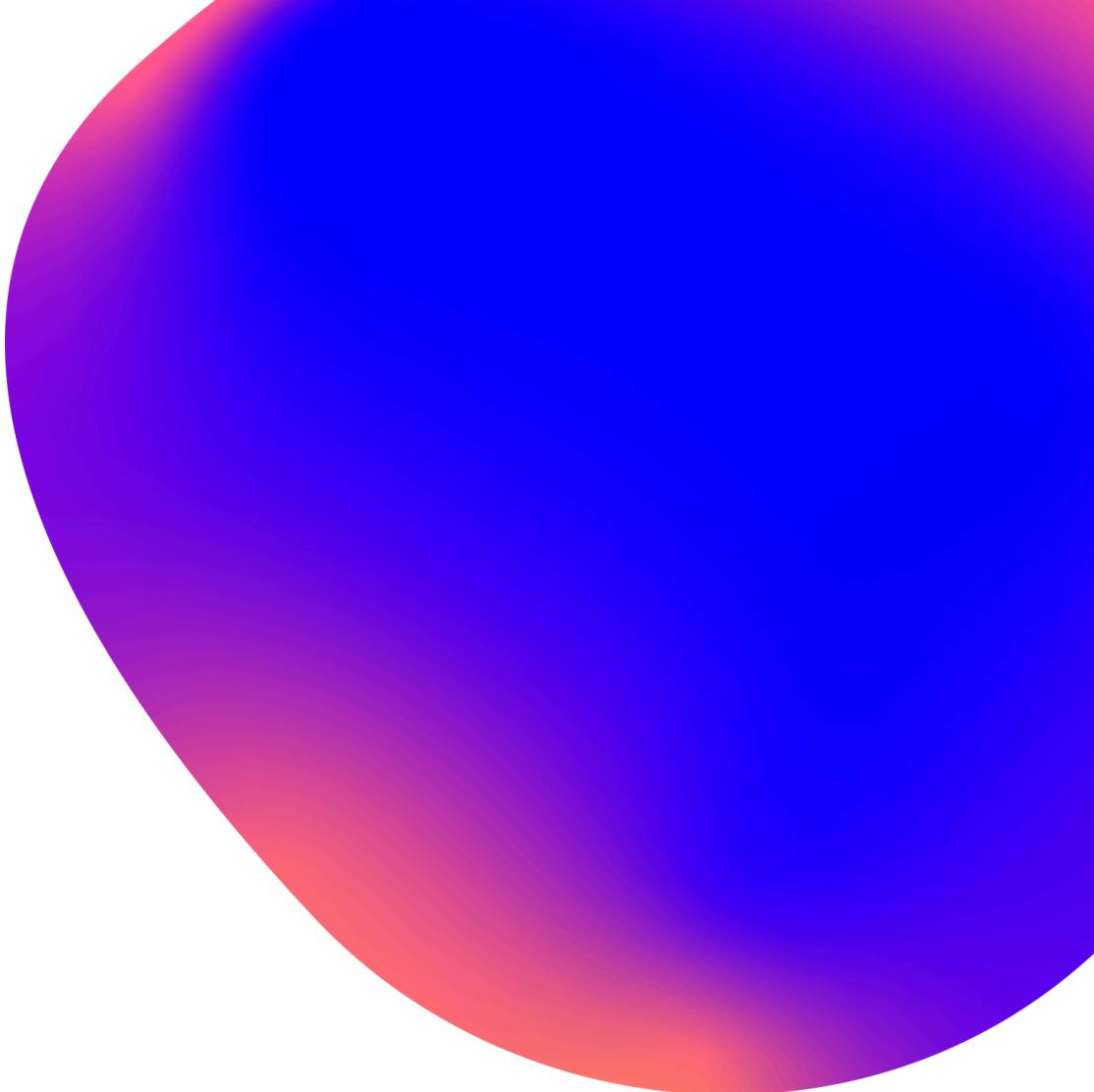
The Objectives

Build an accurate sentiment analysis model

Classify tweets as positive, negative, or neutral

Extract meaningful insights from social media data

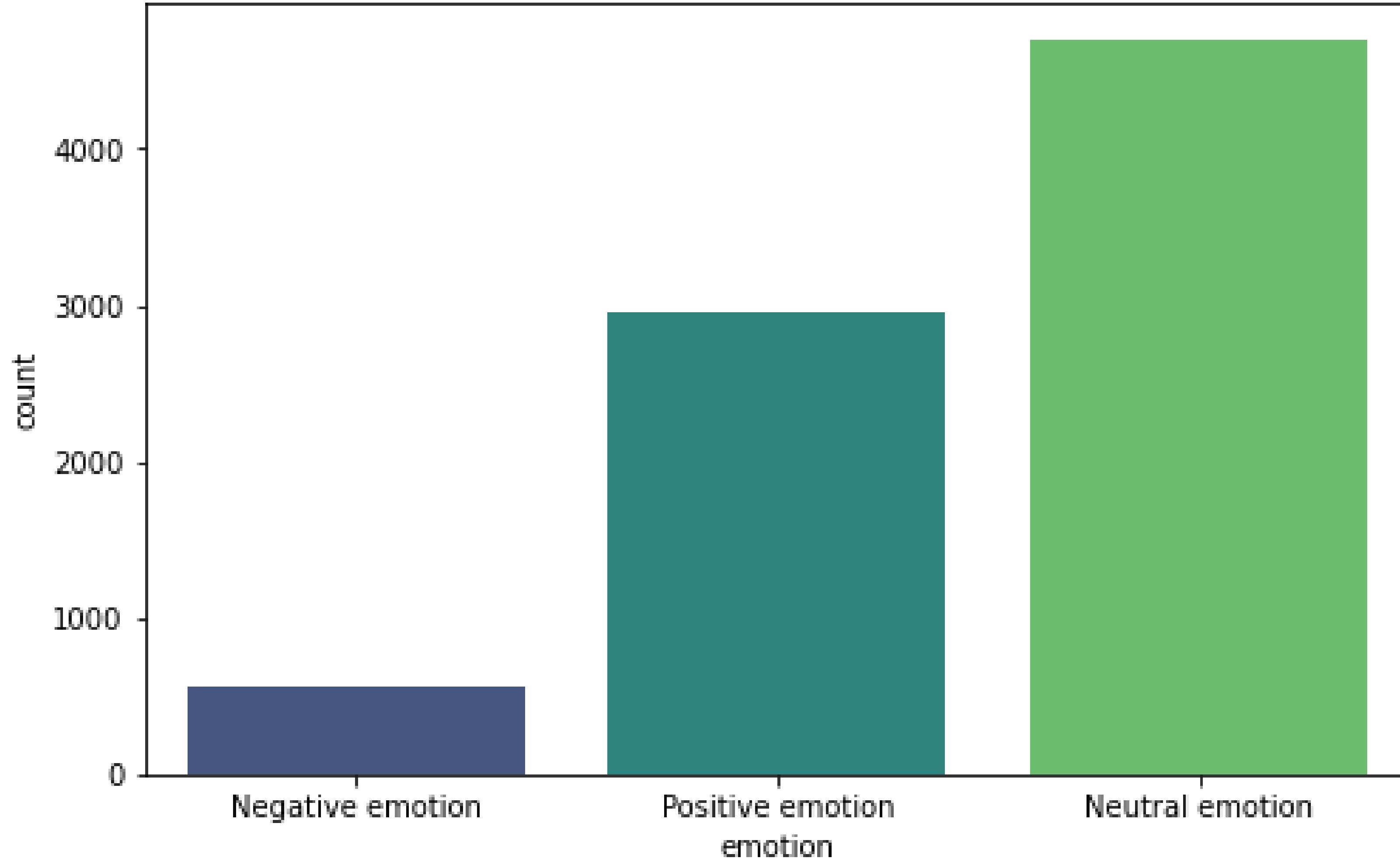
Support improved brand perception and customer satisfaction



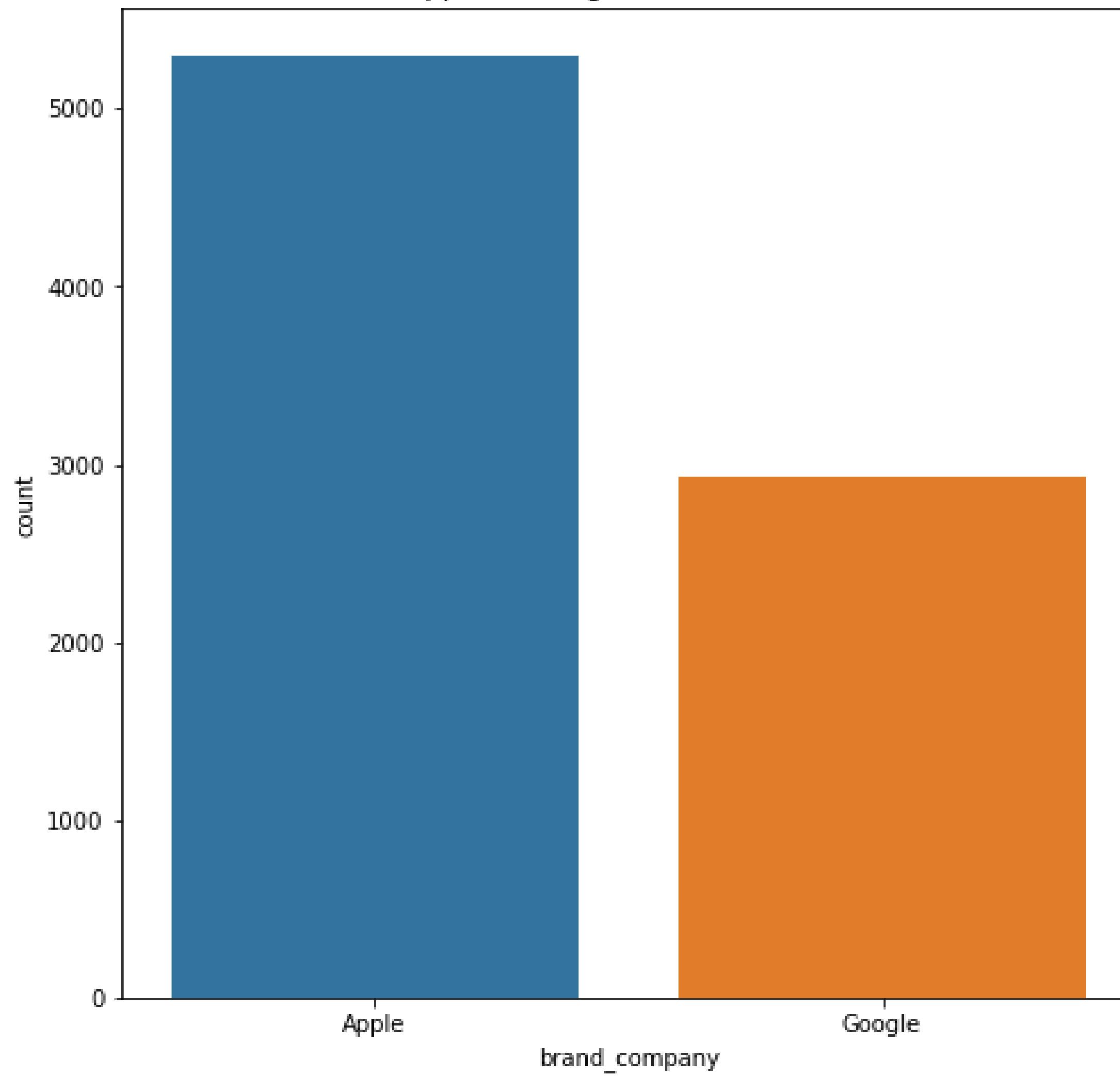
Data Understanding

The dataset used in this project comes from CrowdFlower (via data.world) and contains over 9,000 tweets related to Apple and Google products, each labeled by human annotators as positive, negative, or neutral, making it suitable for supervised learning. The tweets are short, informal, and typical of social media, with the sentiment label as the target variable. This data directly represents real customer opinions, making it highly relevant for analyzing public sentiment toward technology brands. However, the dataset has limitations: tweets may lack context, sentiment labels can be subjective, and there may be class imbalances across sentiment categories.

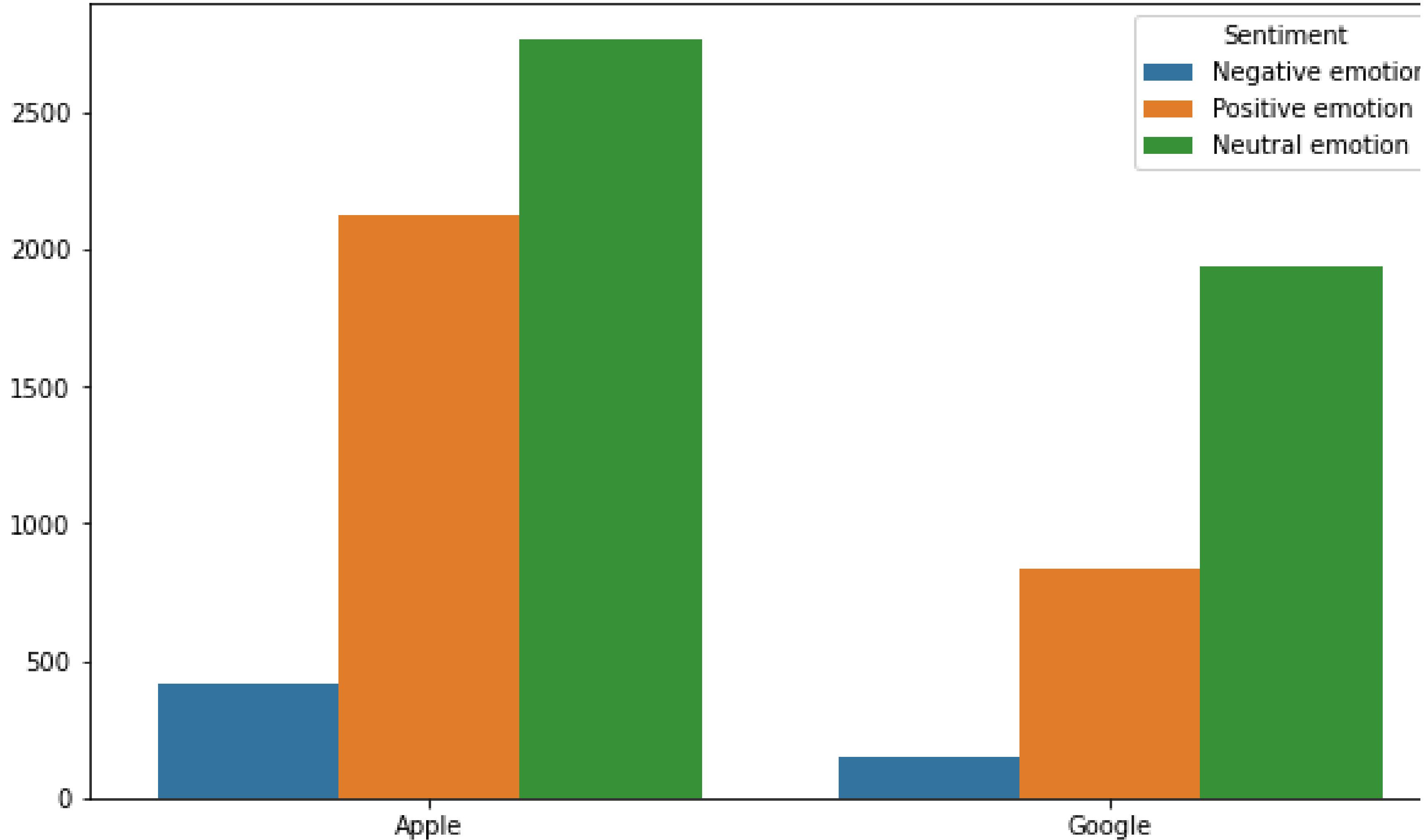
Distribution of Sentiments



Apple vs Google Tweets Emotion



Sentiment Distribution: Apple vs. Google



Conclusion

The Support Vector Machine Model slightly outperforms Complement Naive Bayes and the Random forest model with a macro F1-score of 57%

Recommendations

- SVM is recommended as the primary sentiment classification model due to its balanced performance across sentiment classes.
- Adopt a hybrid modeling approach, combining Naive Bayes for risk screening and SVM for accurate sentiment confirmation and reporting.
- Align model choice with business objectives, prioritizing recall for risk monitoring and balanced performance for overall sentiment analysis.

Thank You