sentiment analysis

Sentiment Analysis of Twitter Conversations About Google and Apple Products

prepared by: group 2

Introduction

- This project performs sentiment analysis on tweets directed to Apple and Google products using a labeled dataset.
- By classifying tweets as positive, negative, or neutral, we aim to understand public perception and identify trends in consumer sentiment.

problem statement

- Analyzing textual data especially from social media platforms to support informed decisions is more challenging compared to quantiative data.
- This project leverages Natural Language Processing concepts and Machine Learning models to build, tune, evaluate, and deploy a robust, generalizable framework for predicting underlying sentiment and the product the emotion is directed at based on tweet context.
- The goal is to support data informed decisions for Apple.

data set overview

“tweet_product_company.csv”

This dataset captures real-world tweet data mentioning Apple and Google products, offering a rich source of public sentiment expressed through social media

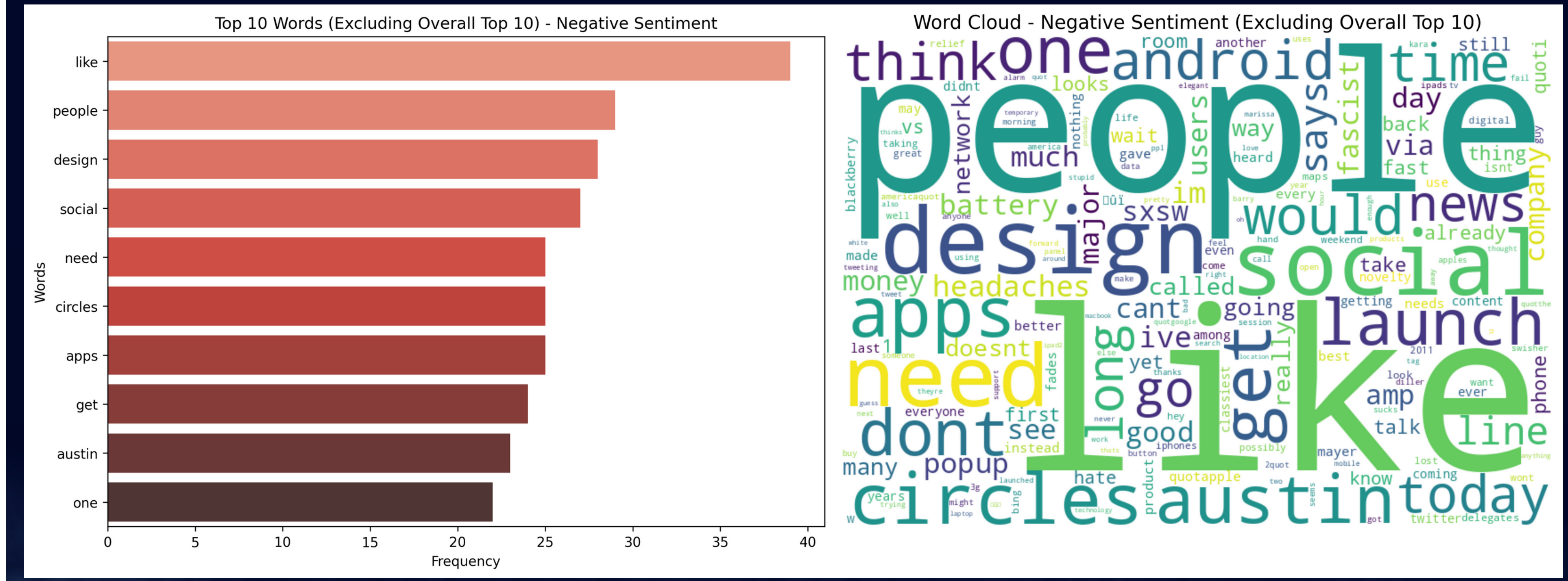
Data Source

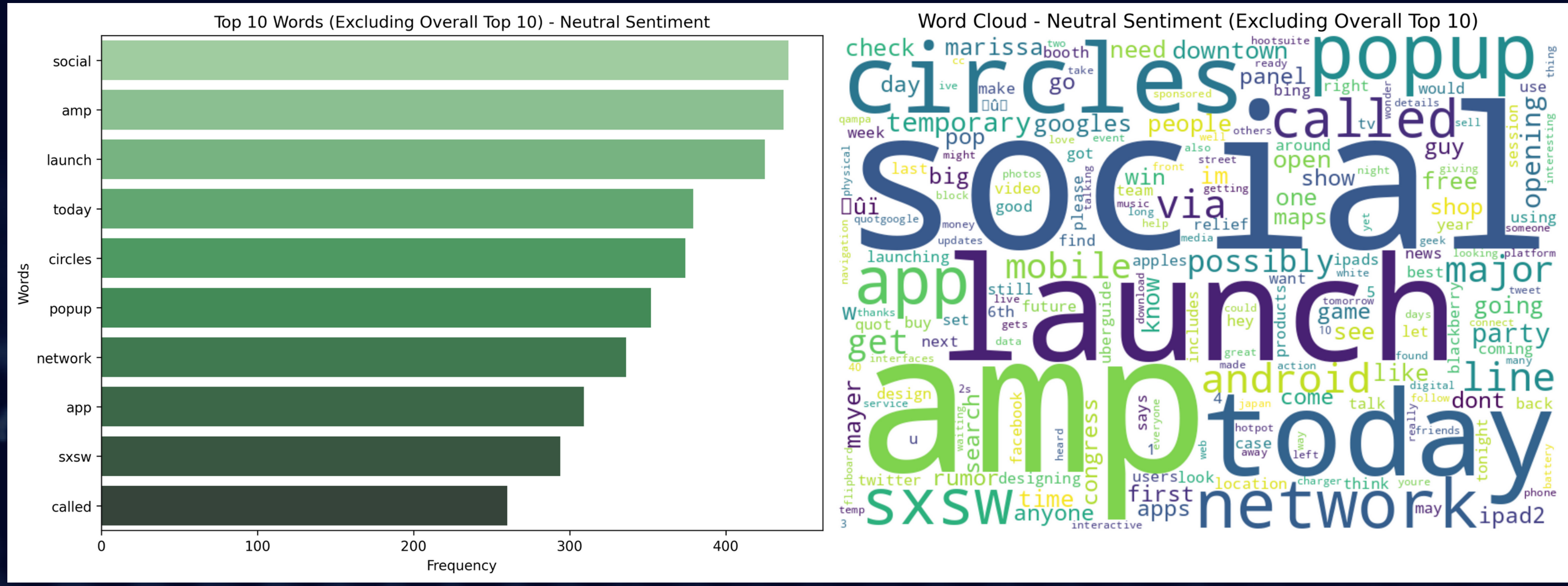
- sourced from CrowdFlower via data.world https://data.world/crowdflower/brands-and-product-emotions and consists of over 9,000 human-rated tweets.
- It reflects organic user opinions and consumer reactions across various Apple and Google product releases, updates, and experiences.

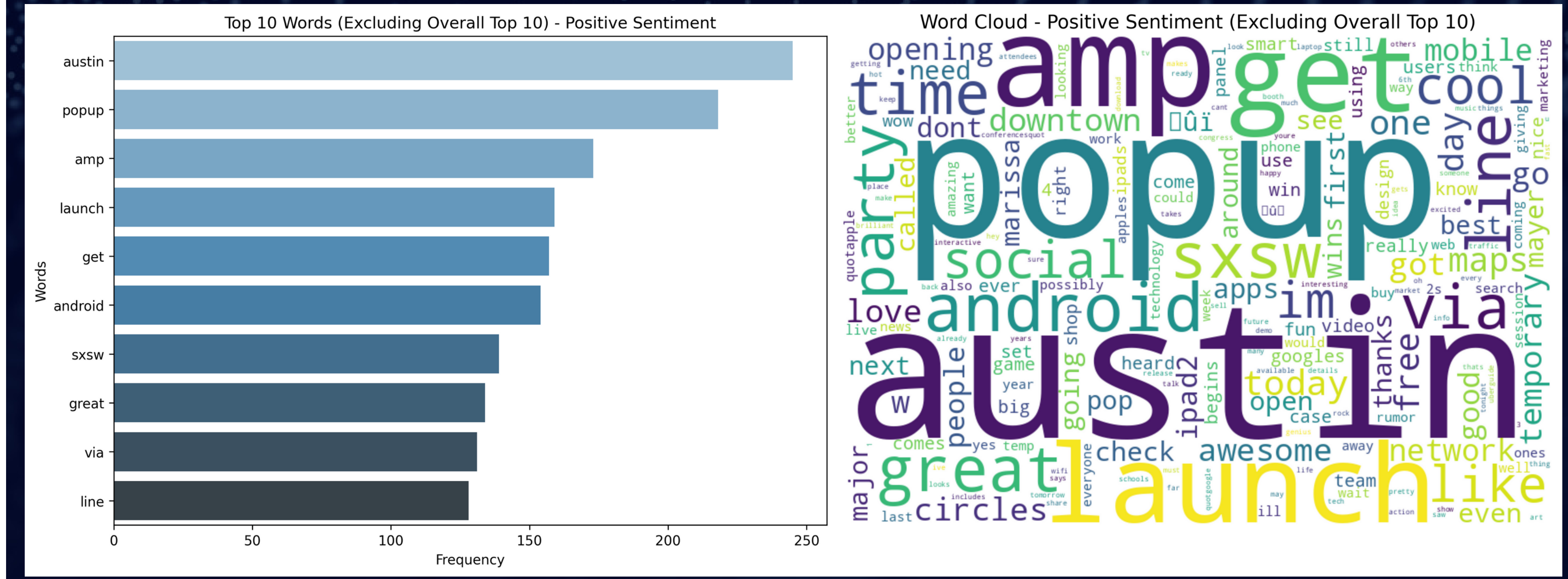
Why this dataset

- Tweets are short, noisy, and opinion-driven—ideal for testing robust NLP techniques.
- It supports both binary classification (positive vs negative) and multiclass sentiment prediction.

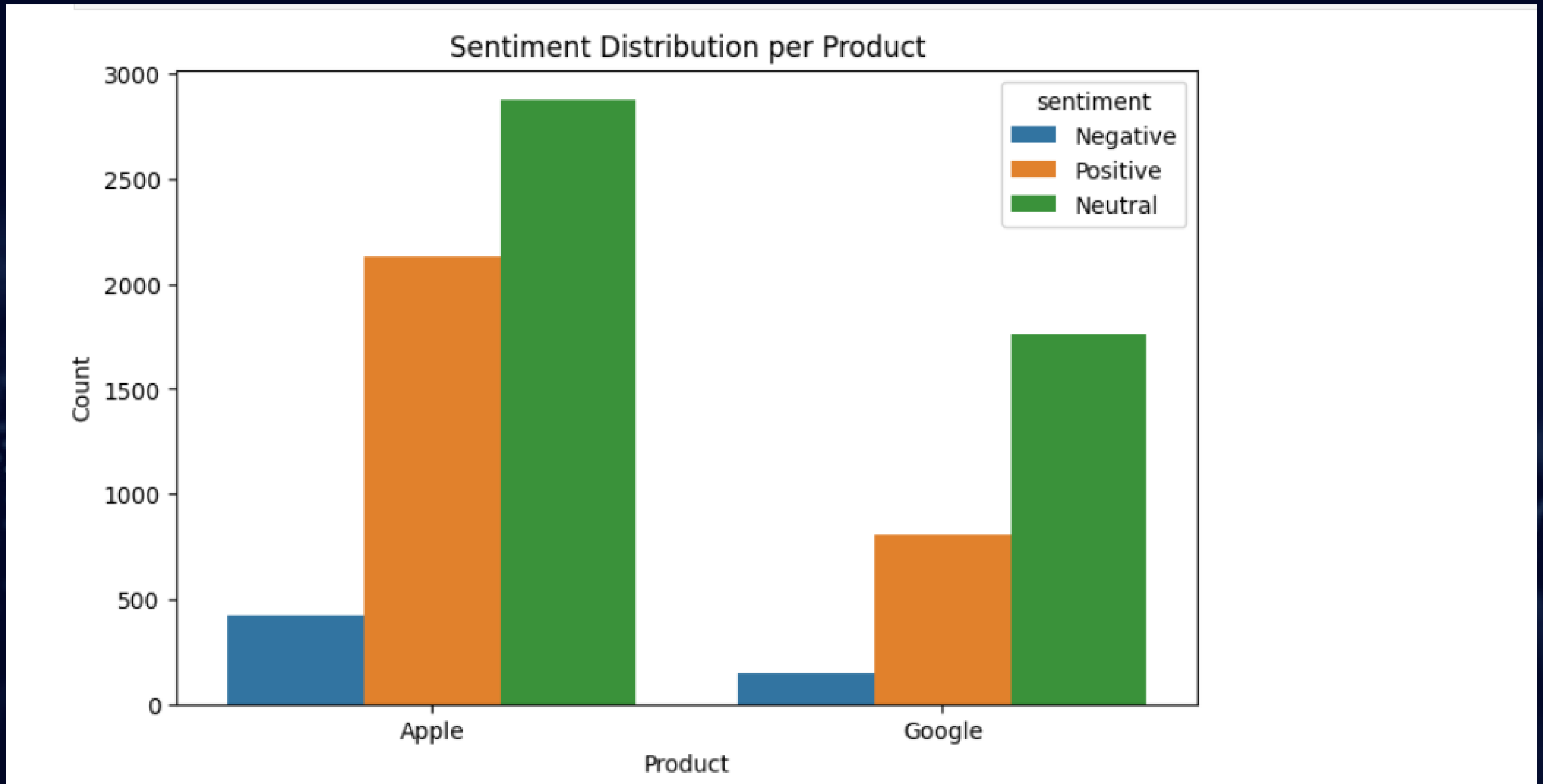
Top 10 Words Based on Sentiment







Sentiment Distribuction per Product



- **Apple Stood Out with More Tweets and Stronger Sentiment**

Apple-related tweets were not only more frequent than Google's—they also leaned more positive. This trend points to higher engagement and brand loyalty among Apple users, making their voices louder and more enthusiastic on Twitter.

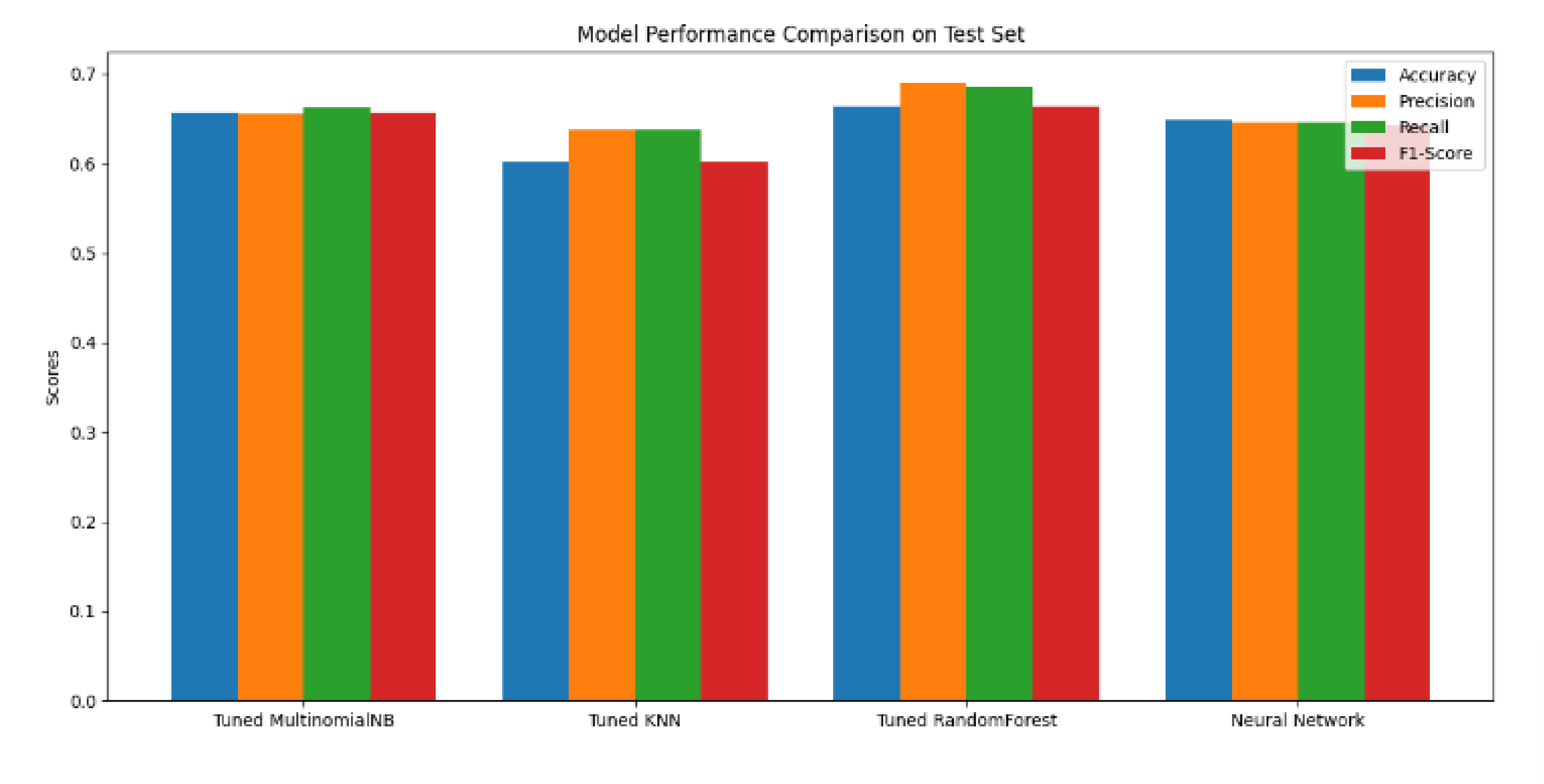
Modelling

- Logistic Regression – Predicting product emotion is directed at.
- Multinomial Naive Bayes – Baseline model for sentiment prediction.
- Random Forest Classifier – Ensemble model for sentiment prediction.
- K-Nearest Neighbors – Instance-based algorithm for sentiment prediction
- Sequential Neural Network – Deep learning model For sentiment prediction.

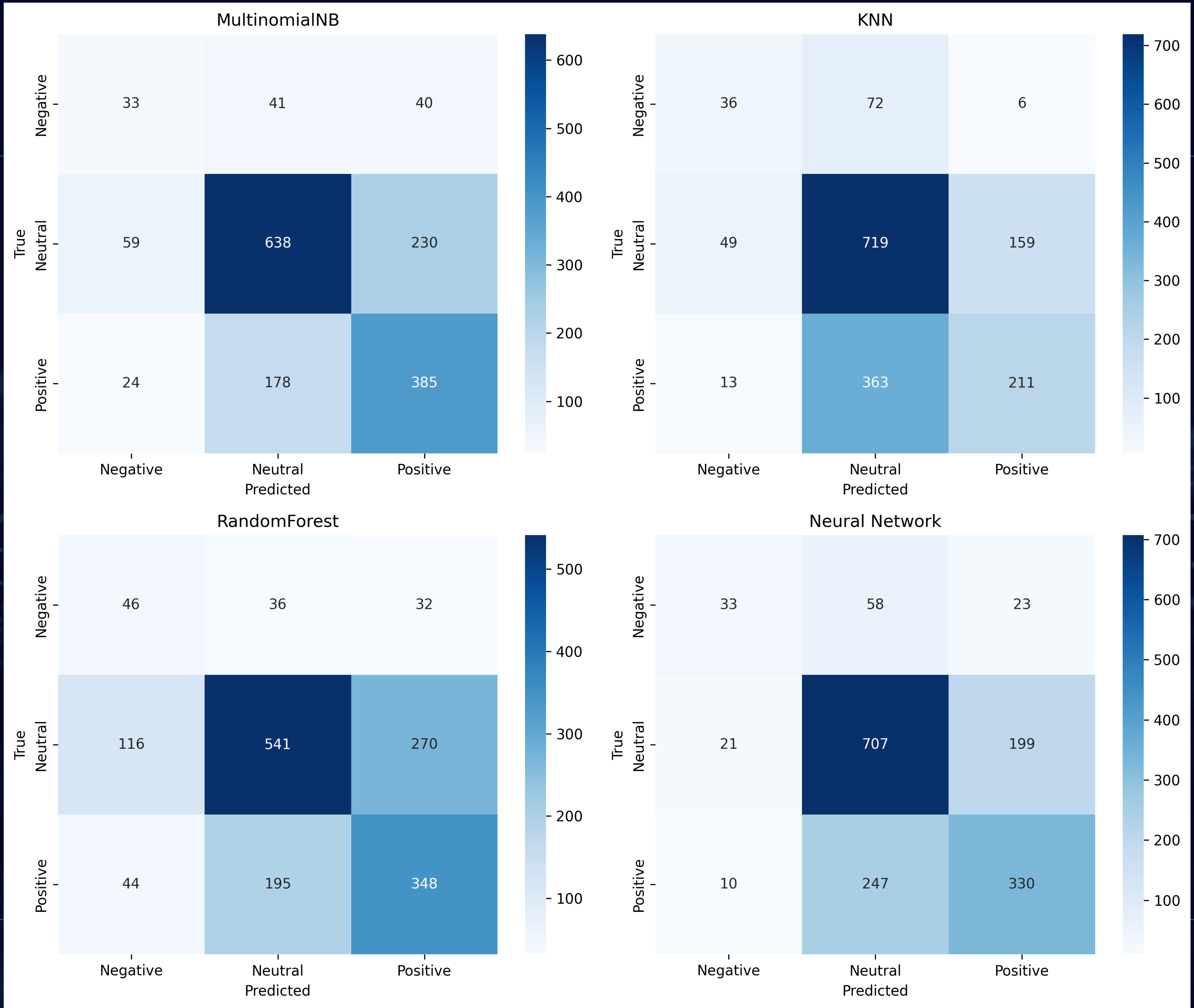
0.985	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.980	0.985	**0.975**	0.980
Multinomial NB	0.6509	0.6558	0.6486	0.6509
KNN Classifier	0.5869	0.6133	0.6192	0.5869
Random Forest	0.659	0.667	0.6726	0.6591
Neural Network	0.654	0.6464	0.6493	0.6434

Model Evaluation

- Tuned Random Forest delivered the best overall performance across all metrics (accuracy, precision, recall, F1-score).
- Multinomial Naive Bayes performed well and is a lightweight, efficient alternative.
- Neural Network showed promising results, but did not outperform Random Forest in this setup.
- K-Nearest Neighbors (KNN) had the lowest performance, indicating it's less effective for this task.



Model Evaluation



Conclusion

Applied NLP techniques to analyze tweets related to Apple and Google products

Focused on two main objectives:

- Classify tweet as pertaining to Apple or Google products using a Logistic Regression model
- Predict tweet sentiment (Positive, Neutral, Negative) using:
  - Multinomial Naive Bayes
  - K-Nearest Neighbors
  - Random Forest
  - Neural Network (Sequential model)

Selected deployment by compared models' performance based on:

- Accuracy, Precision, Recall, F1-score
- Confusion Matrices for visual validation

Tuned Random Forest achieved the best performance and was selected for deployment

Deployment through Streamlit:

- streamlit_app_rf.py: Logistic Regression + Random Forest

Recommendations

- Larger Training Dataset: Incorporate a larger and more diverse dataset to reduce potential overfitting, optimize real-world robustness, and enhance generalizability with noisy social media text data.
- Product-Specific Sentiment: Further analysis could delve into sentiment towards specific products (e.g., "iPhone 15" vs. "Google Pixel 8") rather than just the company, providing more granular insights.
- Temporal Analysis: Incorporate time-series analysis to track sentiment trends over time, especially around product launches or major company announcements, to identify immediate public reactions.

Next Steps

- Deploy the Streamlit deployment app to a Cloud Service.
- Build a Tableau Dashboard to visualize sentiment and product trends in realtime to support data informed decisions.
- Retrain the deployed models with latest tweets data to promote progressive accuracy improvement/ advancements.

Thank You