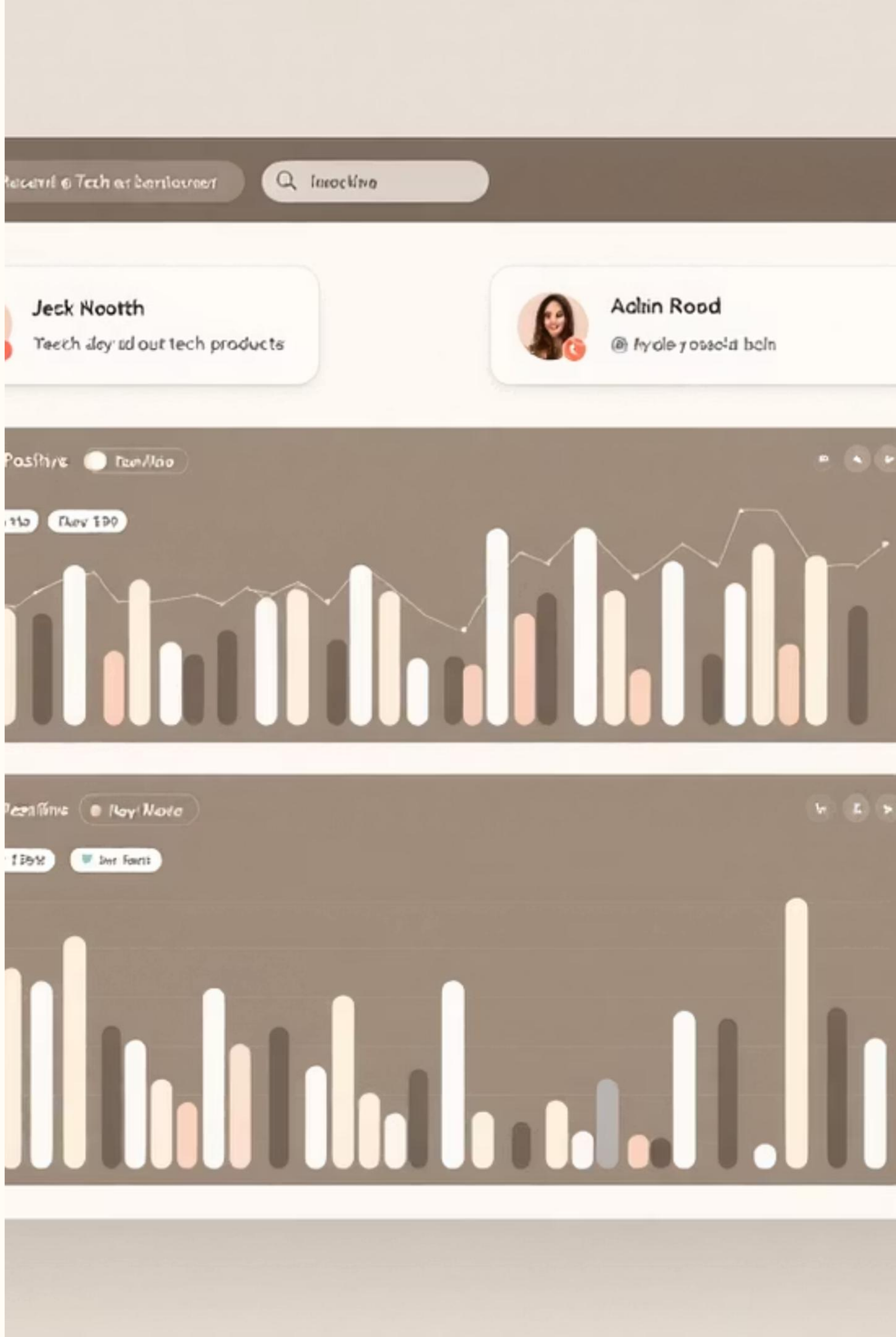
Product Brand Sentiment Analysis

Presented by:

1. Patrice Okoiti
2. Kevin Oguda
3. Dennis Osebe

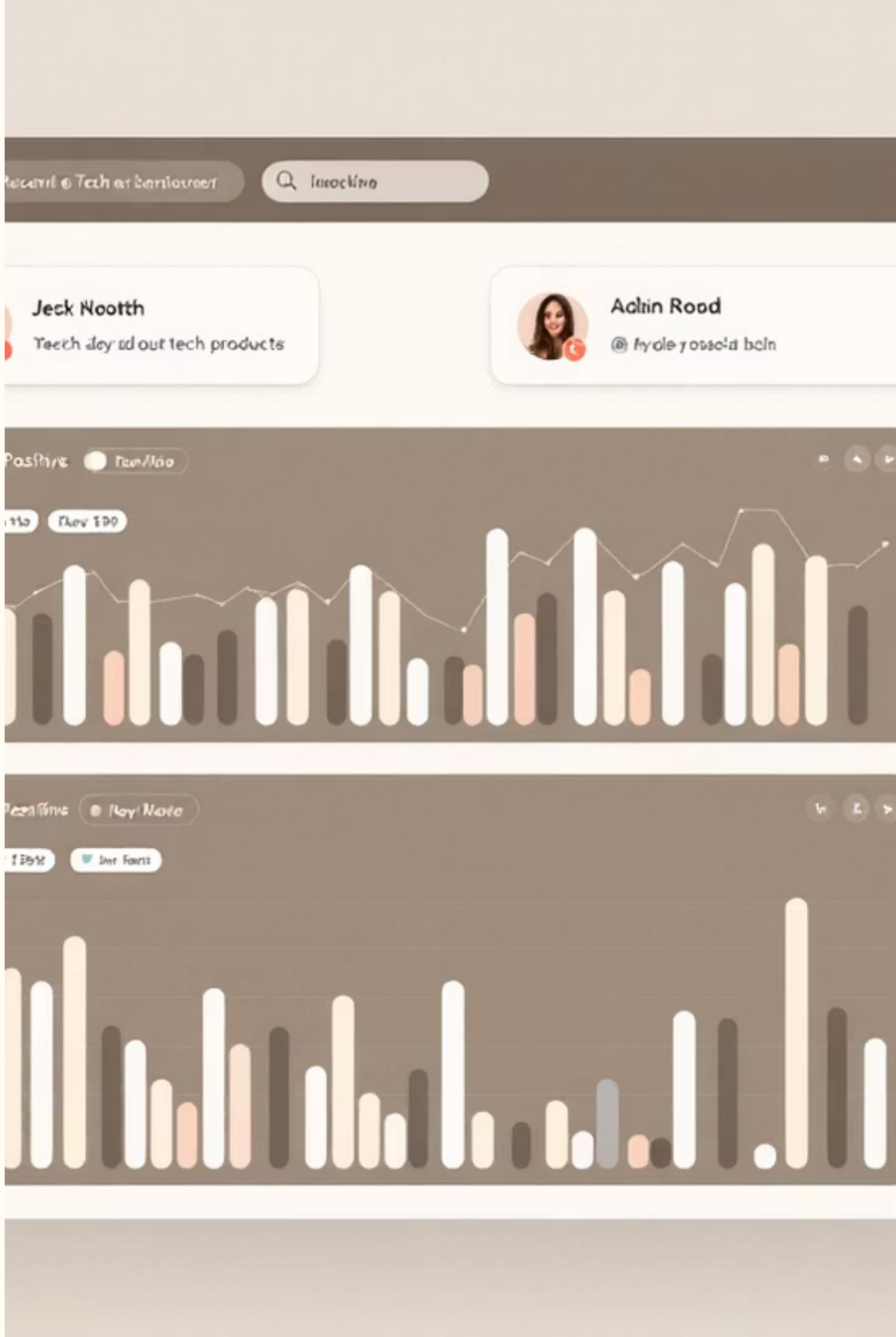


Overview

In today's digital ecosystem, social media, especially Twitter, serves as a public platform where customers freely share their sentiments about products and brands.

This project analyzes over 9,000 tweets mentioning tech products during the SXSW conference to uncover sentiment patterns using NLP and machine learning.

Purpose: Provide real-time, data-driven insights into brand perception that help tech companies improve strategy and product development.



Problem Statement

Current Limitation

Tech brands struggle to manually track customer sentiment due to the high volume and informal nature of social media content.

Challenges Identified:

- Difficulty in interpreting unstructured and noisy text
- Sarcasm and ambiguity in tweets
- Class imbalance in sentiment labels

Goal

Build a scalable, automated sentiment analysis model capable of accurately classifying brand-related tweets.

Objectives

Explore and visualize sentiment distribution across tech brands.

Analyze sentiment by product category and brand.

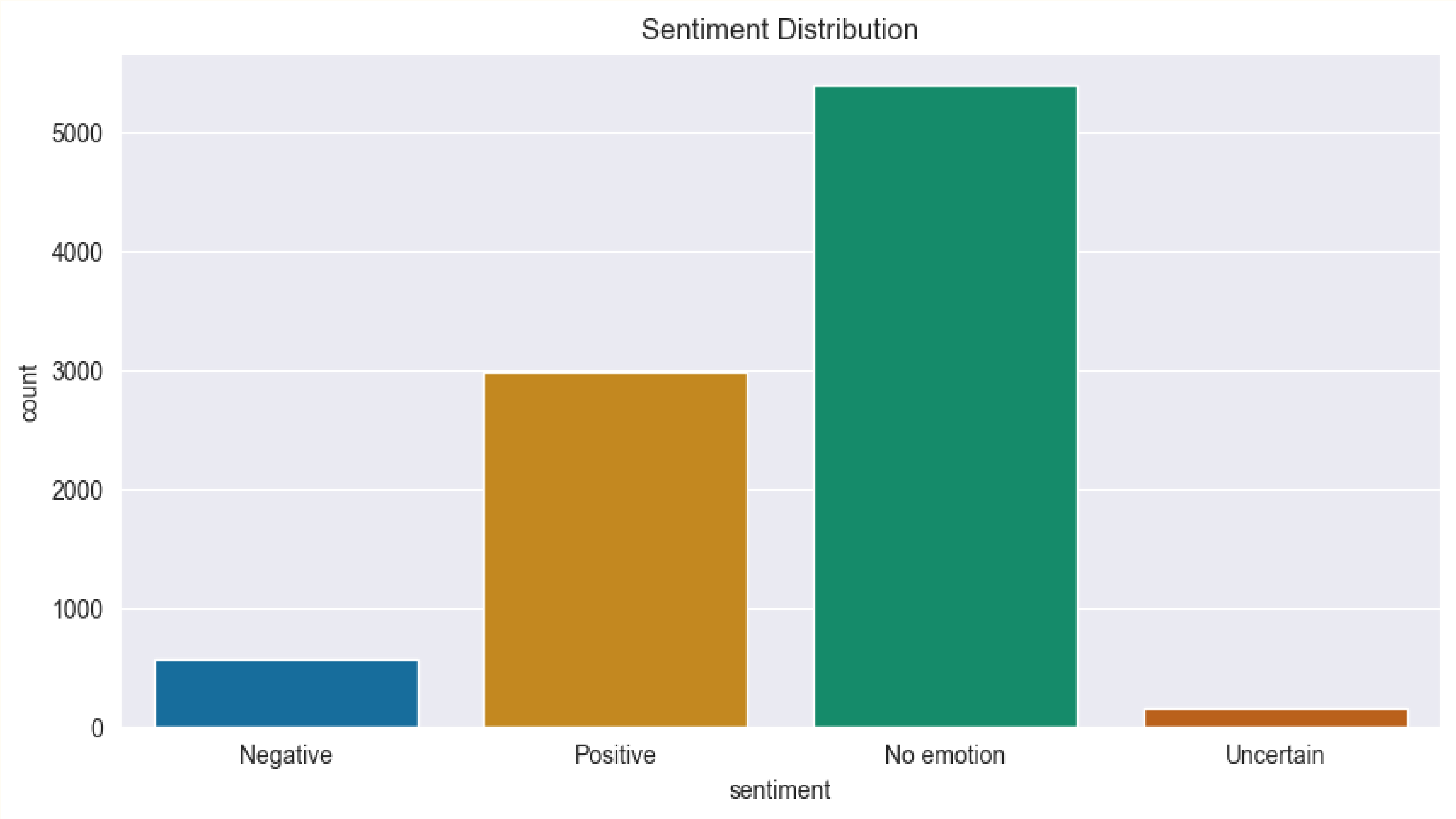
Identify the most positively/negatively perceived brands.

Develop and evaluate multiple machine learning models to accurately classify sentiments

Deploy the best model via an API for real-time use.

Findings / Analysis

Sentiment Distribution:

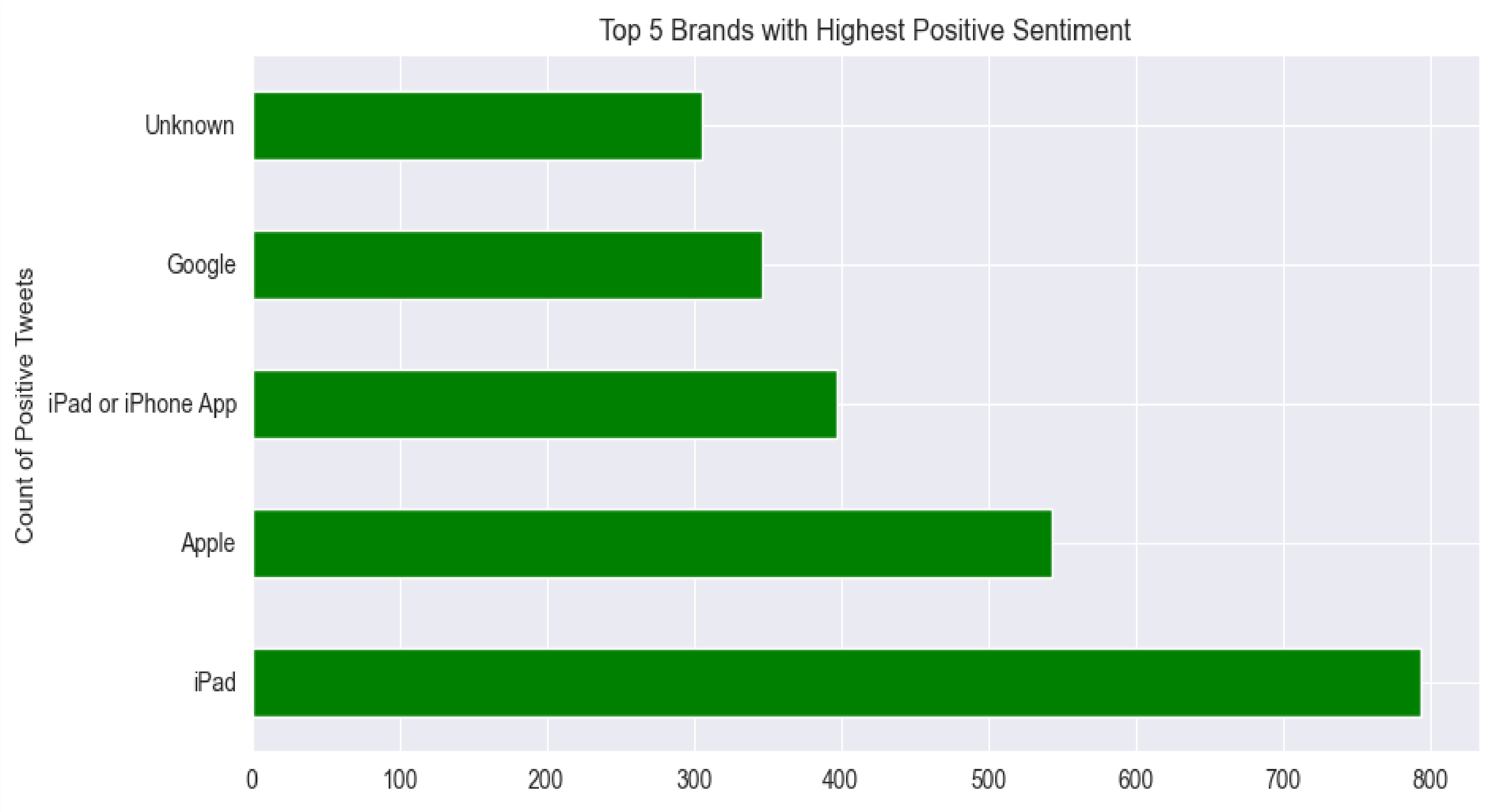


Brand Trends:

Imbalanced Sentiment Distribution

Findings / Analysis

Positive Sentiment Distribution:



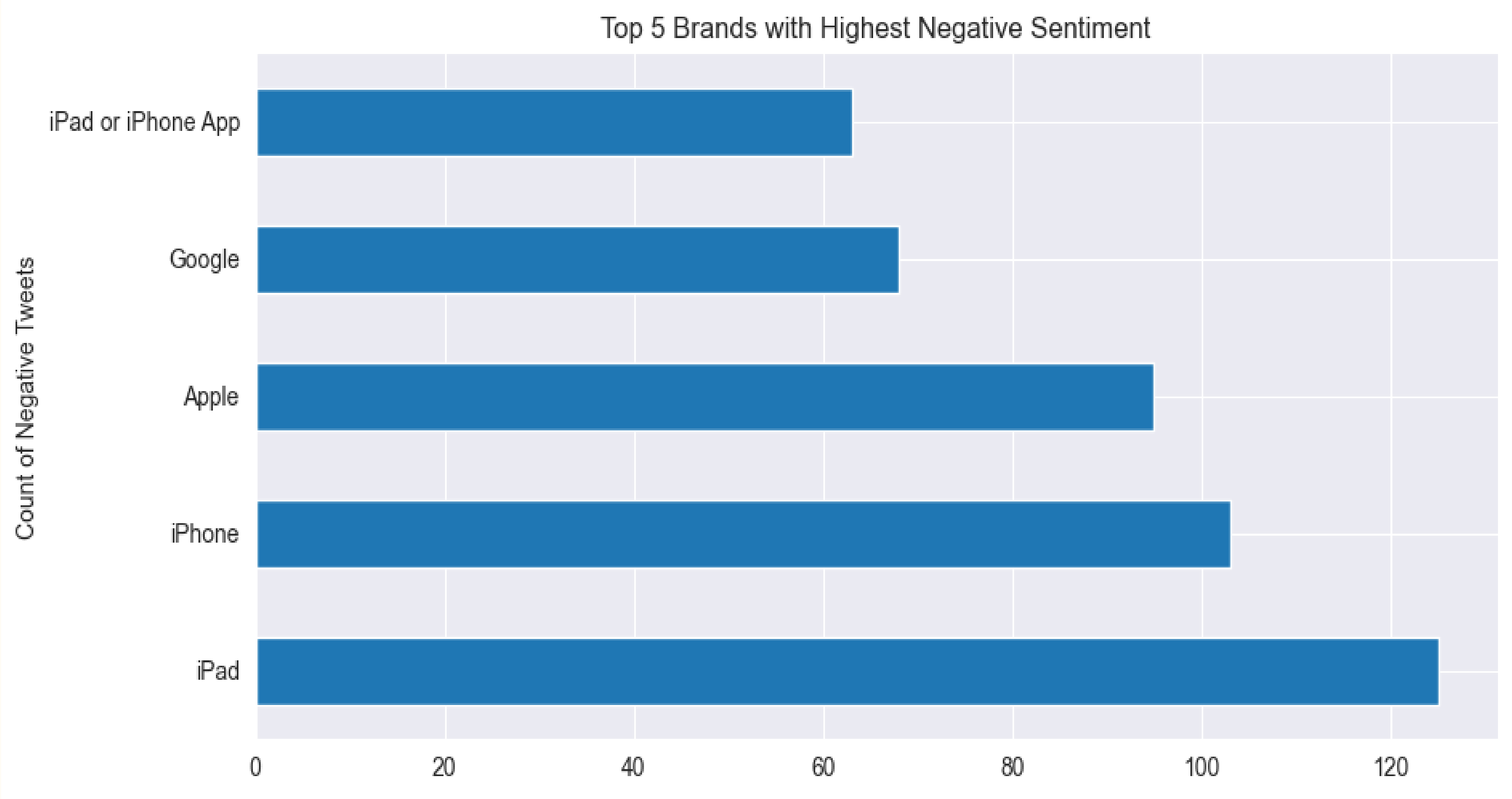
Brand Trends:

Most positive sentiment:

iPad, **Apple**, **Google**

Findings / Analysis

Negative Sentiment Distribution:



Brand Trends:

Most negative sentiment:

iPad, iPhone, Apple

Modeling

Pipeline Stages:

Text Preparation

Text cleaning, tokenization, lemmatization

Feature Engineering

TF-IDF + feature engineering (char/word/sentence counts)

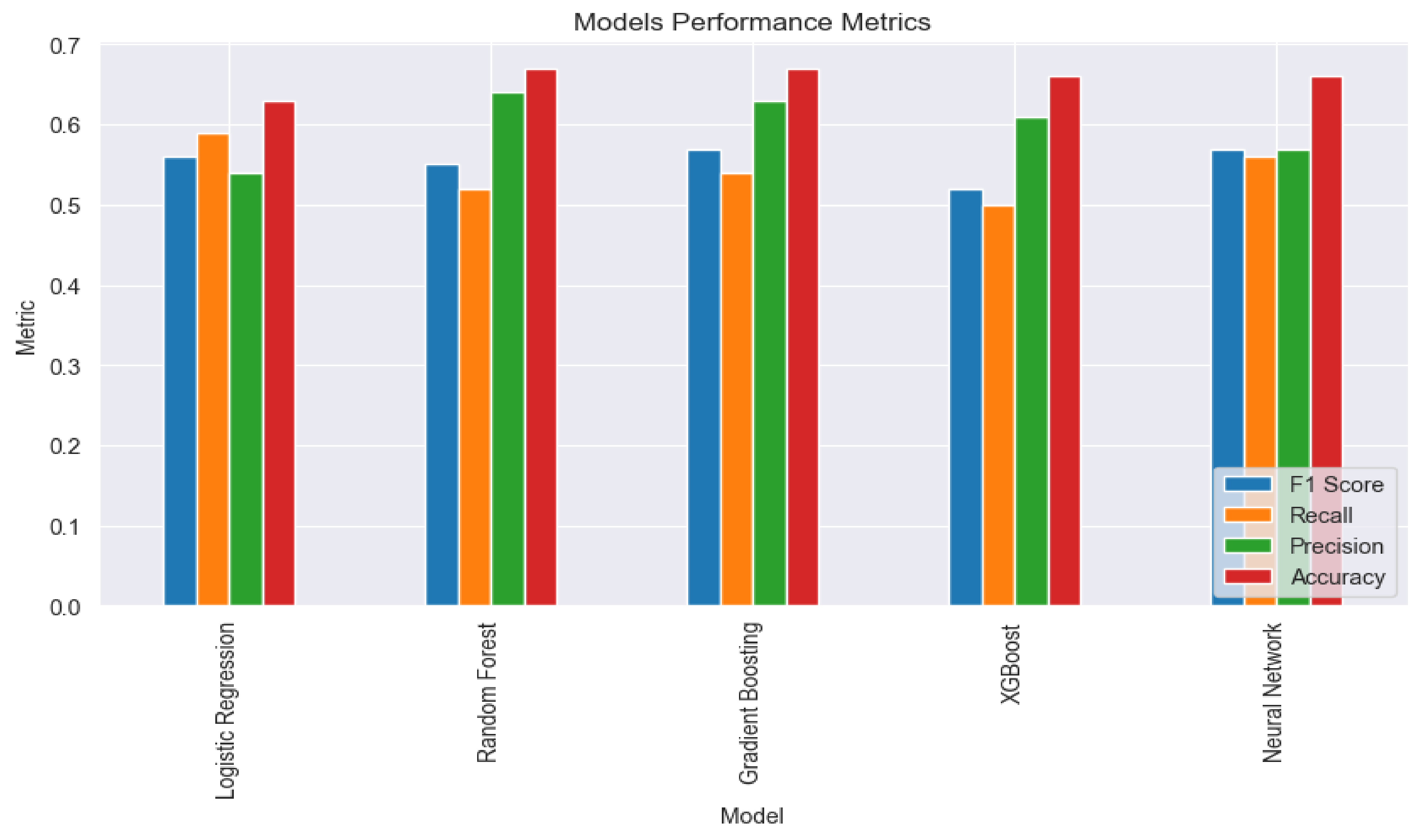
Class Balancing

SMOTE for class imbalance

Models Evaluated:

- Logistic Regression
- Random Forest
- Gradient Boosting
- XGBoost
- Neural Network

Modeling



Best Model:

Gradient Boosting

- F1 score: **0.56**
- Accuracy: **0.67**
- Precision: **0.62**
- Recall**: 0.54**

Most balanced performer across all evaluation metrics

Conclusion

Sentiment analysis of tweets offers timely and actionable insights for tech brands.

Gradient Boosting was the most balanced model in terms of performance.

While models perform moderately well, improvements can be made through:

- Advanced NLP (e.g., transformers)
- Improved data quality
- Better handling of class imbalance

Recommendation

Short-term:

Deploy current Gradient Boosting model for real-time sentiment monitoring.

Scrape for better quality data and find alternative ways of handling imbalance.

Long-term:

Incorporate Transformer-based models (e.g., BERT, SetFit).

Continuously retrain on fresh data to improve accuracy.

Integrate sentiment feedback into marketing and product cycles.