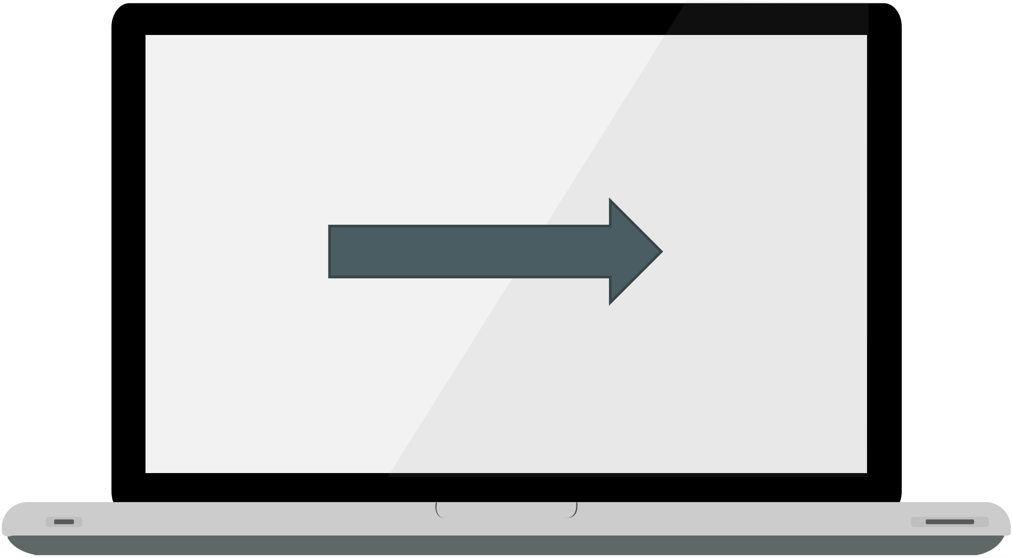
SentimentFlow Analysis Presentation

A sentiment analysis project using Natural Language Processing

Business Overview

SentimentFlow aims to address a real-world problem related to understanding public sentiment towards Apple and Google products on Twitter. The stakeholders include companies, marketing teams, and decision-makers who want to gauge public opinion and make informed strategic decisions based on social media sentiment.

Business Problem



The problem is to accurately classify the sentiment of tweets related to Apple and Google products. We want to determine whether a tweet expresses a positive, negative, or neutral sentiment. This classification can help companies understand customer satisfaction, identify potential issues, and tailor their responses accordingly

Objectives

Specific Objectives

Objective 1

To idenitfy the most common words used in the dataset using Word cloud.

Objective 2

To confirm the most common words that are positively and negatively tagged.

Objective 3

To recognize the products most opined by the users.

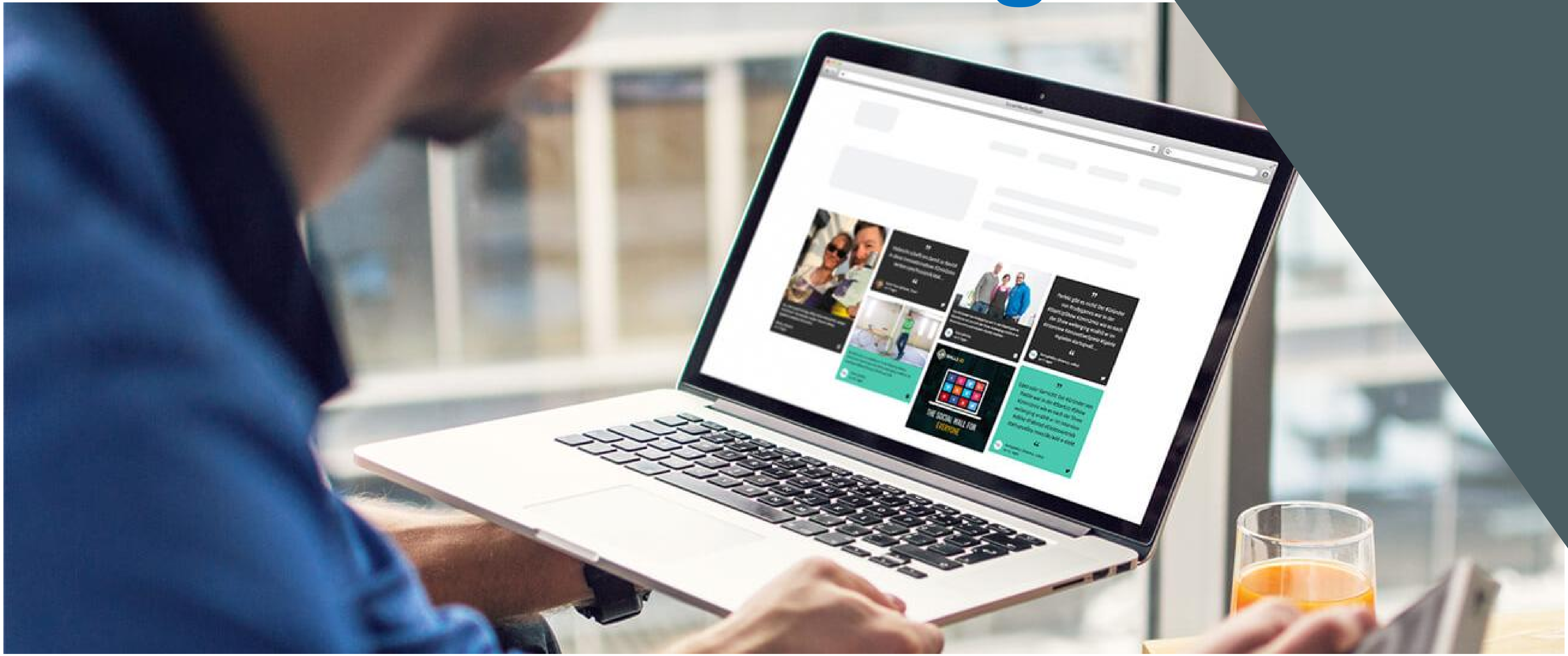
Objective 4

To spot the distribution of the sentiments.

Main Objective

To develop a NLP (Natural Language Processing) multiclass classification model for sentiment analysis, aim to achieve a recall score of 80% and an accuracy of 80%. The model should categorize sentiments into three classes: Positive, Negative, and Neutral.

Data Understanding



Data Understanding

1

Dataset Description

9000 rows and three columns

2

Source Of Data

The dataset originates from CrowdFlower via data.world

3

The columns

- tweet text
- emotion in tweet is directed at
- is there an emotion directed at a brand

4

Products

The dataset focused on Google and Apple products

The data

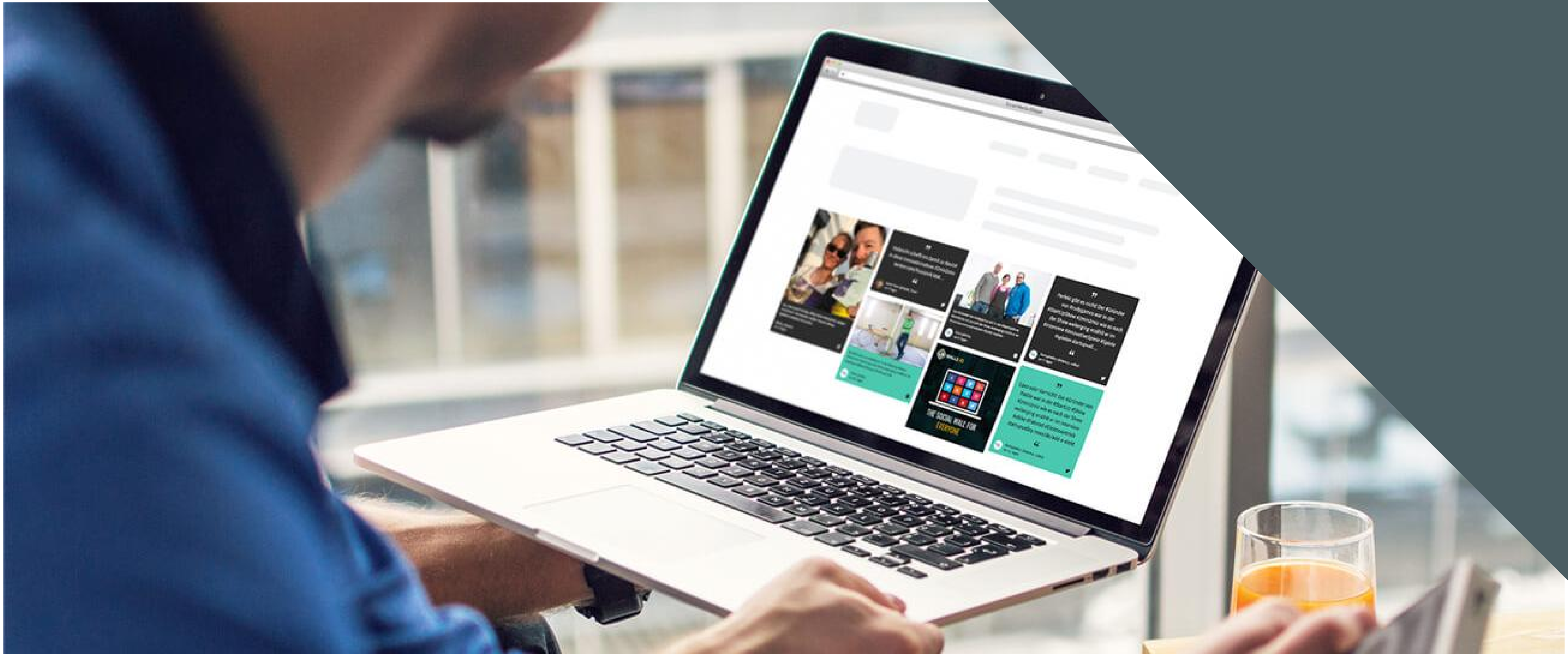
Twitter Sentiments

5

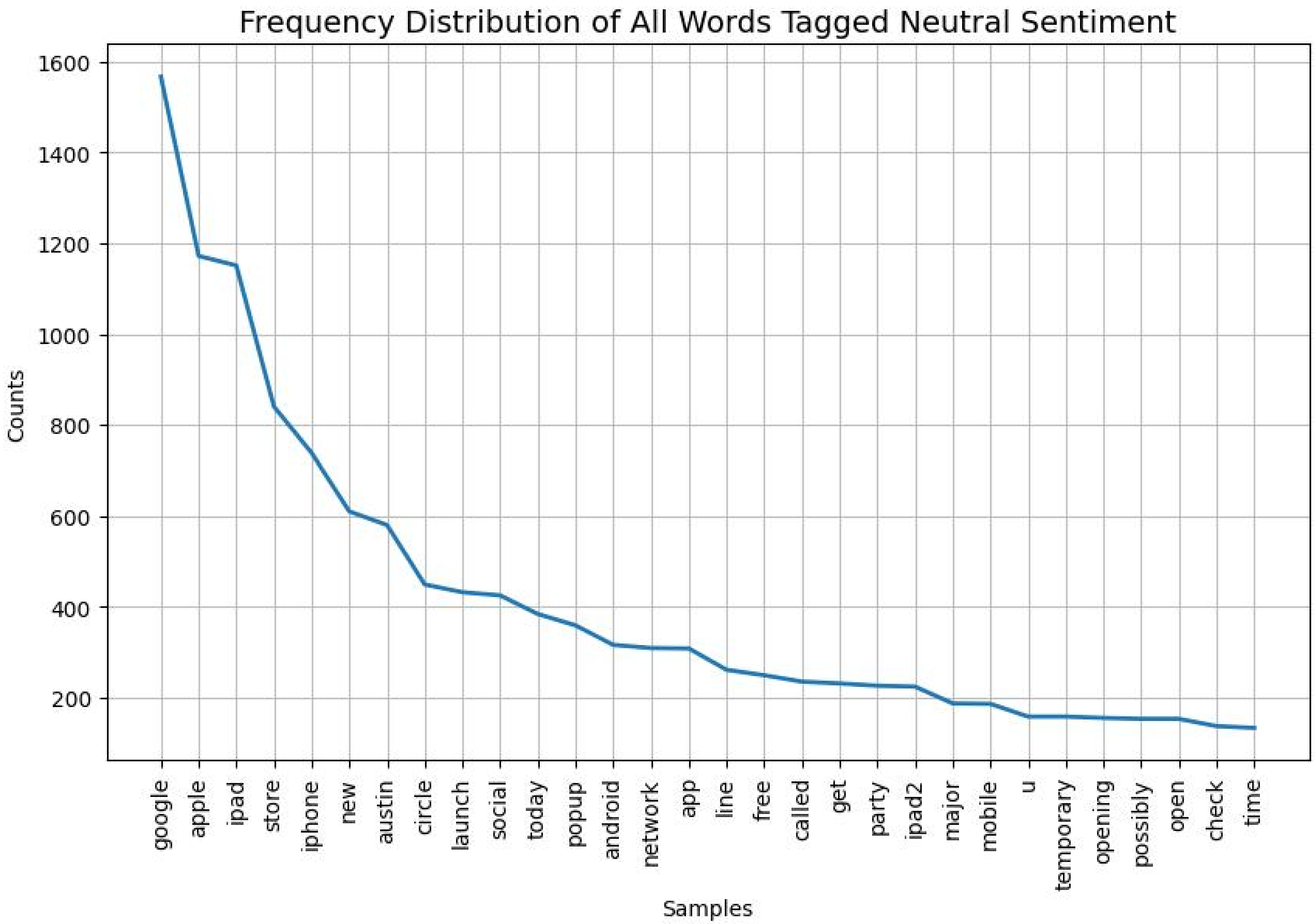
The Sentiments

- Positive
- Negative
- Neutral

Data Visualization

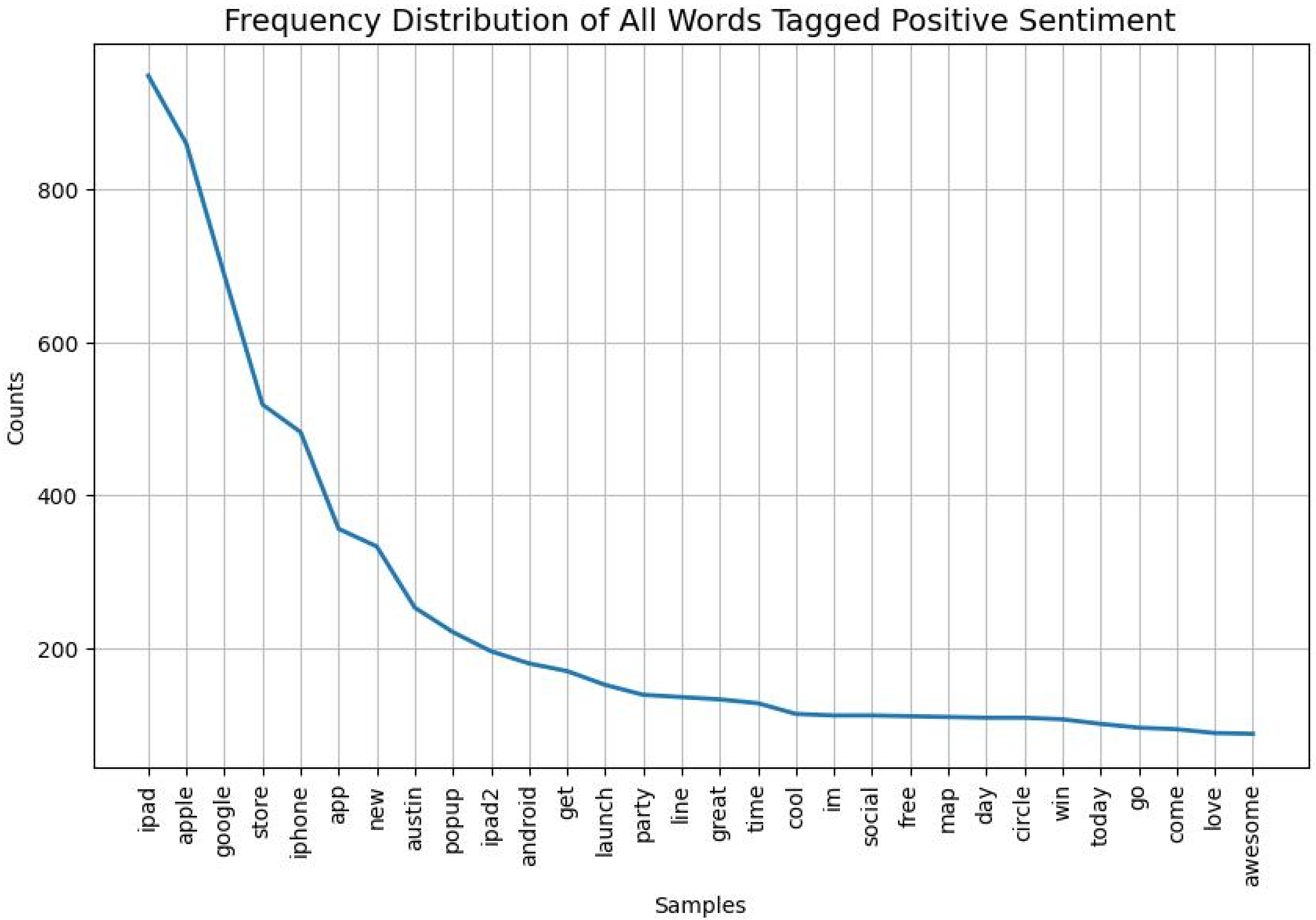


Neutral Sentiments



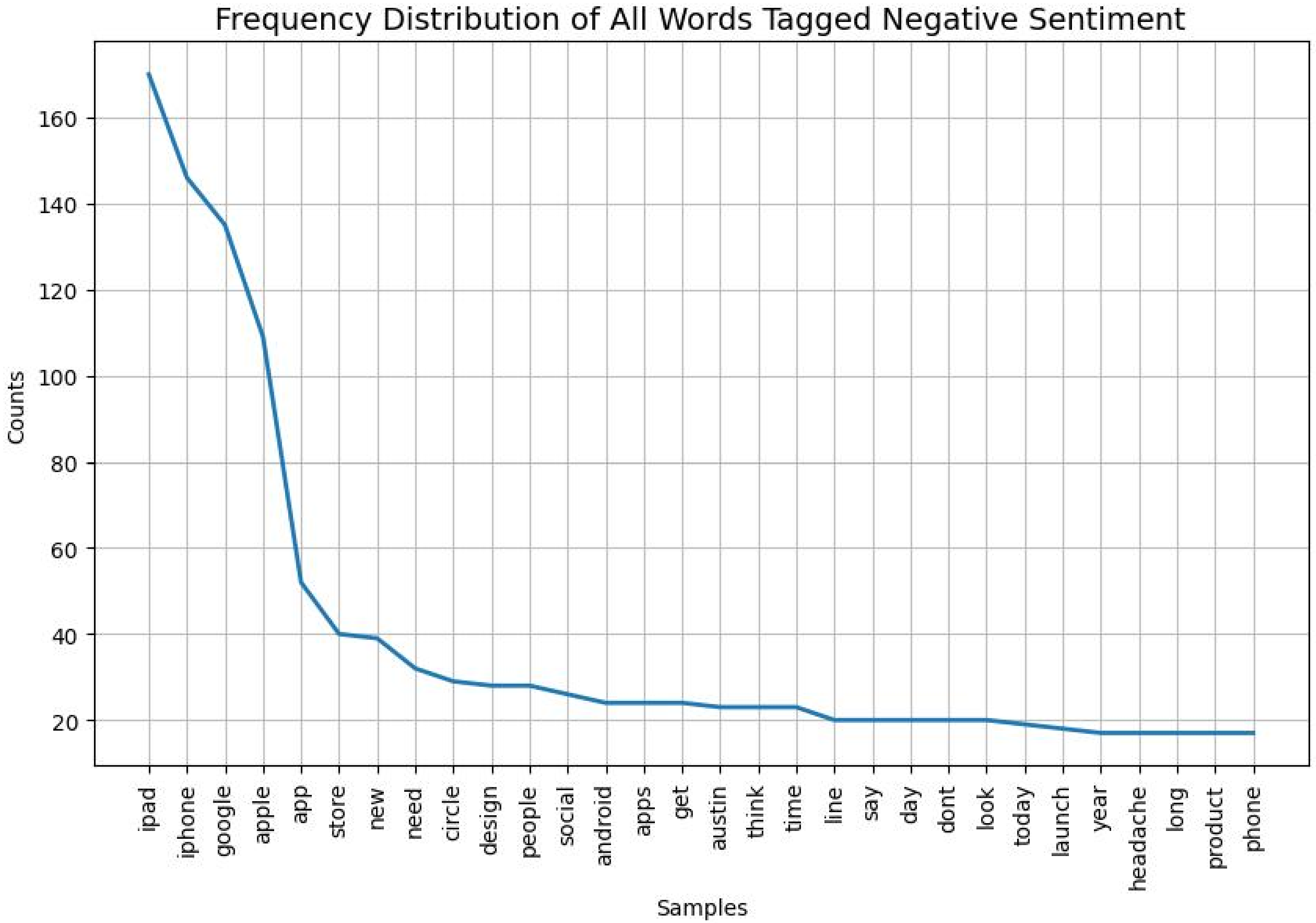
With respect to all the data categorised as 'neutral', the words 'google', 'apple', 'ipad' and 'store' appeared more frequently than all other words.

Positive sentiment



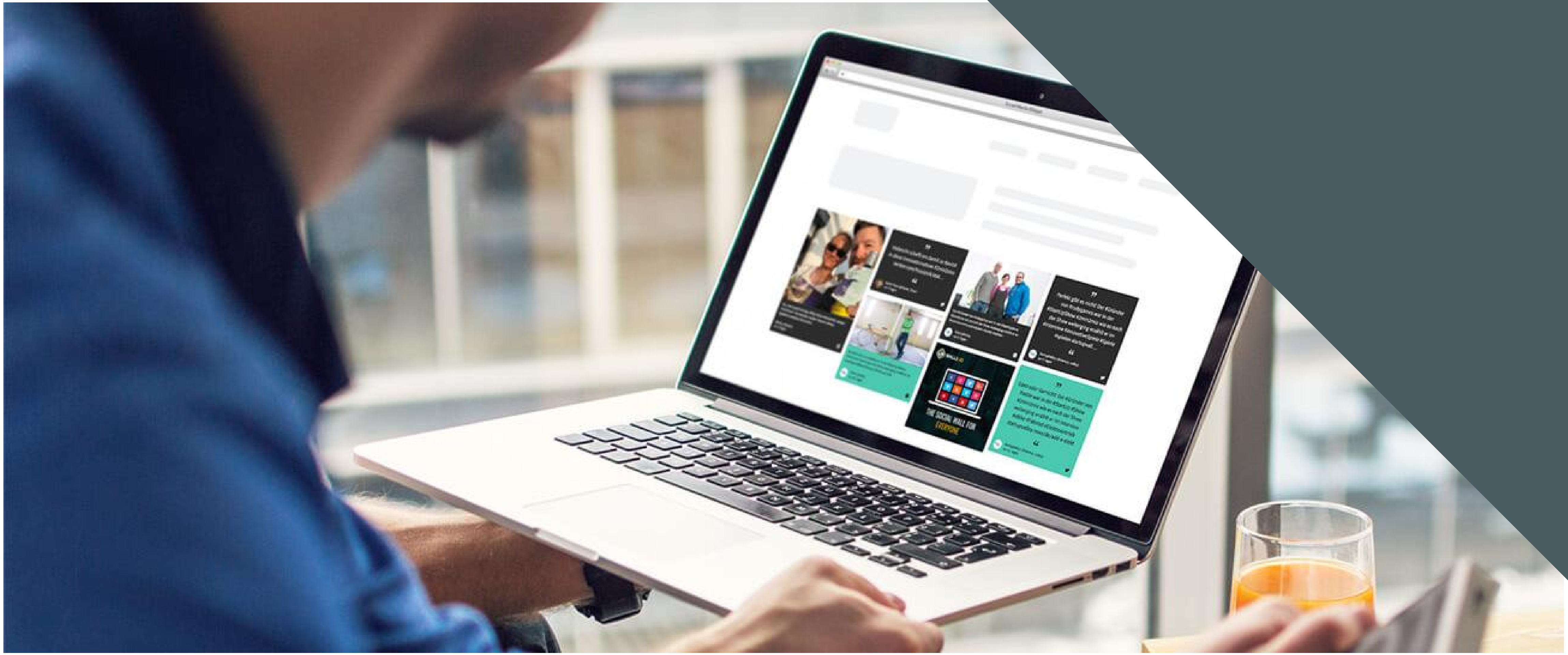
With respect to all the data categorised as positive, the words **ipad**, **apple**, **google** and **store** appeared more frequently than all other words.Other key positive words introduced in this section include **awesome**, **love**, **win**, c**ool**, **great**, **party**

Negative Sentiments



With respect to all the data categorised as 'negative', the words **ipad**, **iphone**, **google** and **apple** appeared more frequently than all other words. But were less than the counts recorded in the Neutral Frequency Distributions.

Modelling



Modelling Results

The Best Model

We found the best model to be the Random Forest Model and the Logistic Regression - both with the highest accuracy scores of 83.7%.

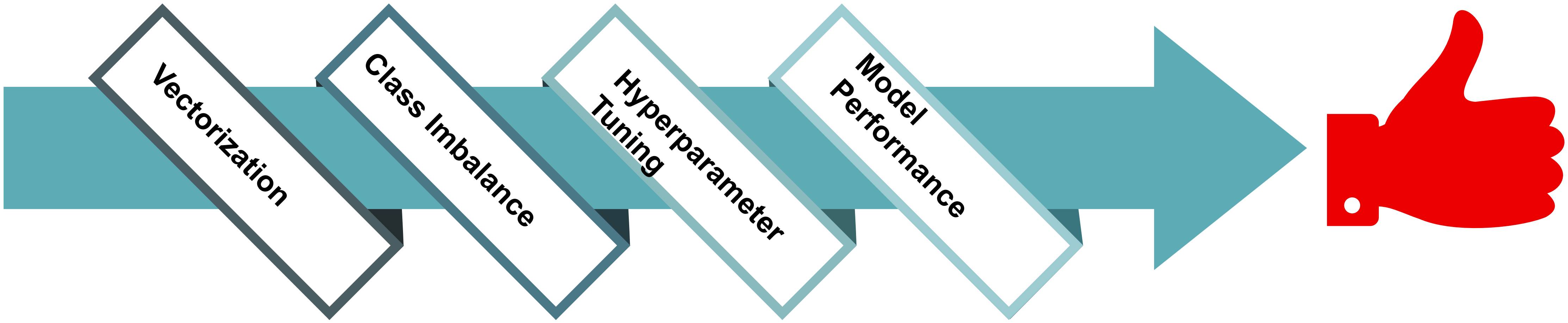
Tuned Logistic Regression- 87.3%

Tuned Random Forest - 87.3%

Logistic Regression - 80.8%

Tuned MultinomialNB - 80%

Conclusion



Vectorization

01

TF-IDF Vectorization consistently outperformed CountVectorizer in all models. It has demonstrated its superior capability in feature representation for sentiment analysis.

02

Class Imbalance

Applying SMOTE was effective in handling class imbalance, ensuring that the models did not bias towards the majority class and provided balanced performance across all emotion categories.

03

Hyperparameter Tuning

Hyperparameter tuning significantly improved model performance, as seen in the Random Forest and Logistic Regression models where accuracy and recall improved by more than 10% in some cases.

04

Model Performance

Tuned Random Forest and Tuned Logistic Regression models achieved the highest accuracy and recall scores with TF-IDF vectorization, both scoring approximately 83.7% in accuracy and 83.6% in recall.

Recommendations

Scalability

Optimize the models for performance and efficiency to ensure they can process a high volume of tweets quickly and accurately.

Monitoring Negative Sentiments

This allows for prompt interventions and resolution of consumer issues.

Continuous Model Monitoring

Implement continuous monitoring of the deployed models to detect any performance degradation over time.

Real-Time Processing

Explore real-time processing capabilities to provide up-to-date sentiment analysis, which is crucial for timely decision-making and responding to emerging trends.

Social Media Platform Intergration

Integrate the sentiment analysis models with social media platforms' APIs for seamless data collection and analysis, enabling continuous monitoring and real-time insights.

Next Steps

Step 1

Collect tweets from different geographical areas to understand regional sentiment variations

Step 2

Use more advanced models such as BERT, Deep Neural Network and LSTM models.

Step 3

Implement continuous monitoring of the deployed models

Incorporate user feedback to iteratively improve the model's performance

Step 4

Developing solutions for real-time sentiment analysis to provide timely insights and responses.

Thank You