Sentiment Analysis of Tweets Related to Apple & Google Products

Business Understanding

Our Stakeholders

- Marketing and Product Development teams at Apple and Google

Business Problem

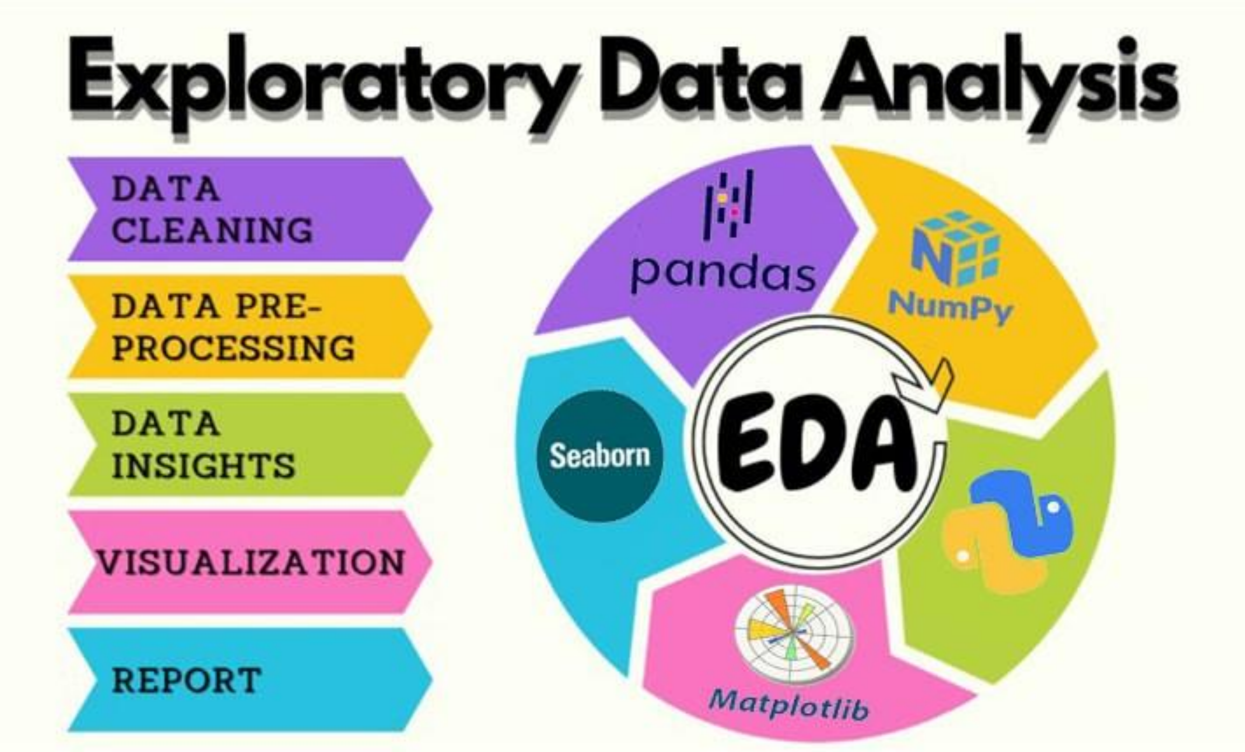
- Apple and Google aim to monitor and analyze general consumer sentiments on social media platforms to gauge public perception and identify trends on how their products are received.

Data Understanding

- Dataset: CrowdFlower dataset of over 9,000 Tweets about Apple and Google products (sourced from CrowdFlower) labeled as positive, negative, or neutral by human raters

- Suitability: Provides a broad sample of real-time consumer opinions, facilitating both binary and multiclass classification tasks.

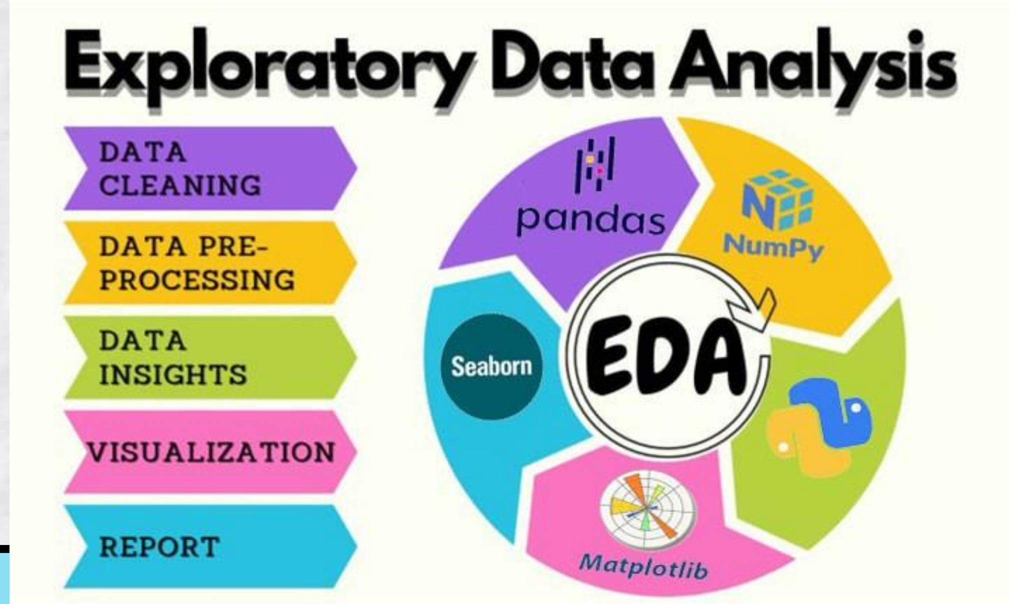
DATA PREPARATION AND EXPLORATORY DATA ANALYSIS



Pre-processing Steps:

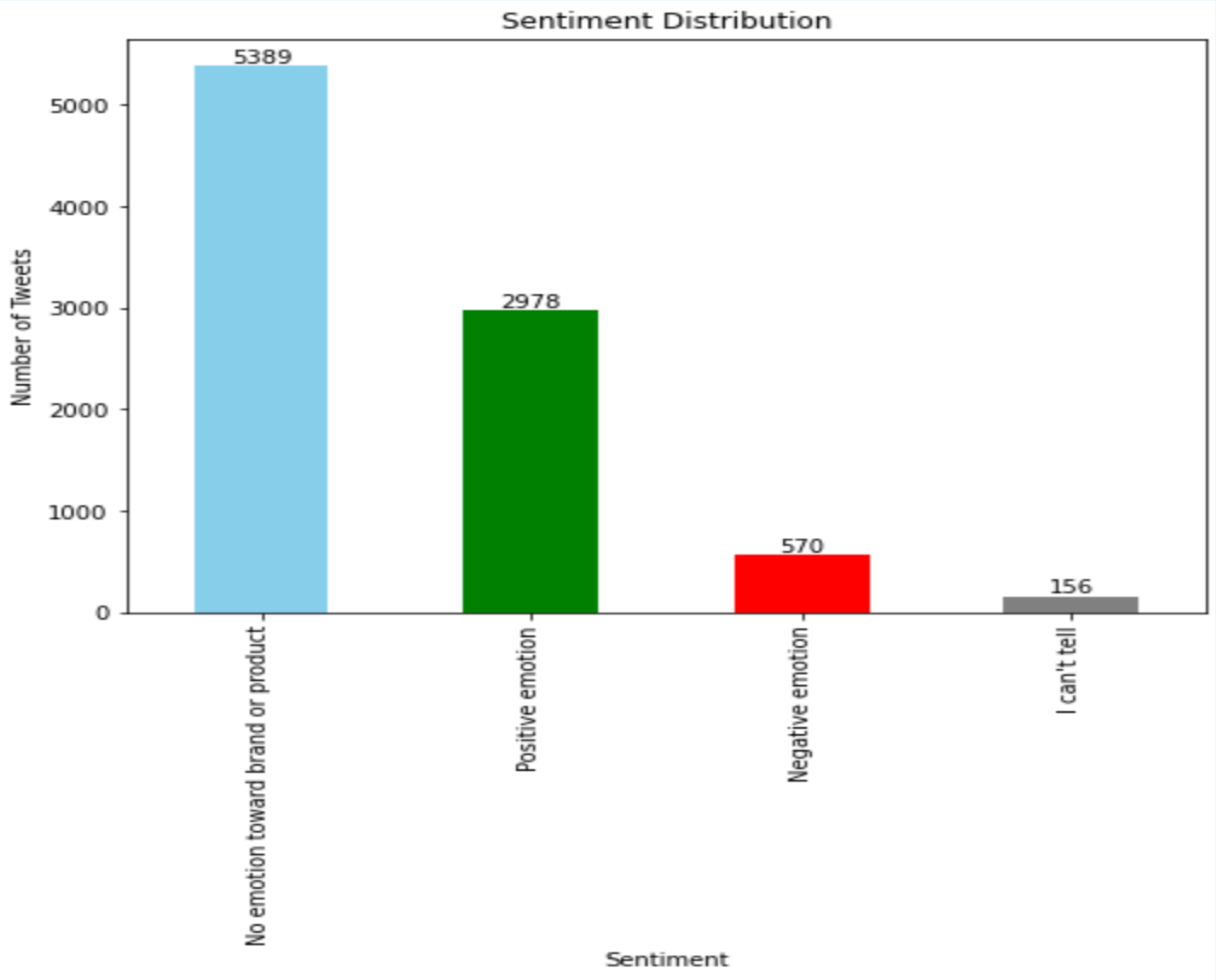
- Converting text to lowercase
- Text Tokenization
- Stopword and punctuation removal
- Missing Data: Dropped emotion_in_tweet_is_directed_at column due to high missing values.
- transformation using TF-IDF vectorization
- Class Balancing: Used SMOTE to balance imbalanced sentiment classes.

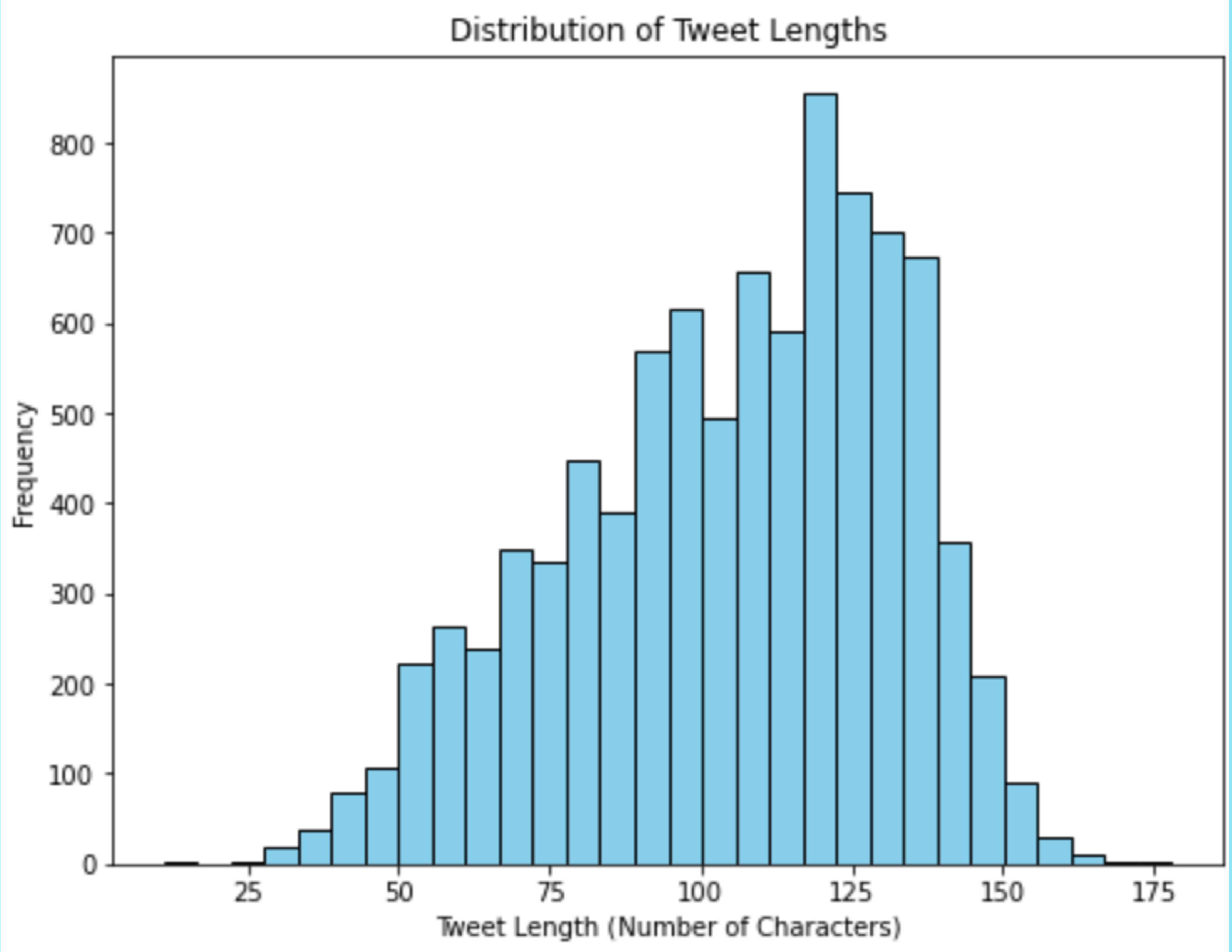
EXPLORATORY DATA ANALYSIS



- *The word clouds show that event-related terms like "SXSW", "mention", and "link" dominate, with limited sentiment-specific words, suggesting that removing non-informative terms and words during pre-processing could improve sentiment analysis.*

SENTIMENT AND TWEET LENGTH DISTRIBUTION



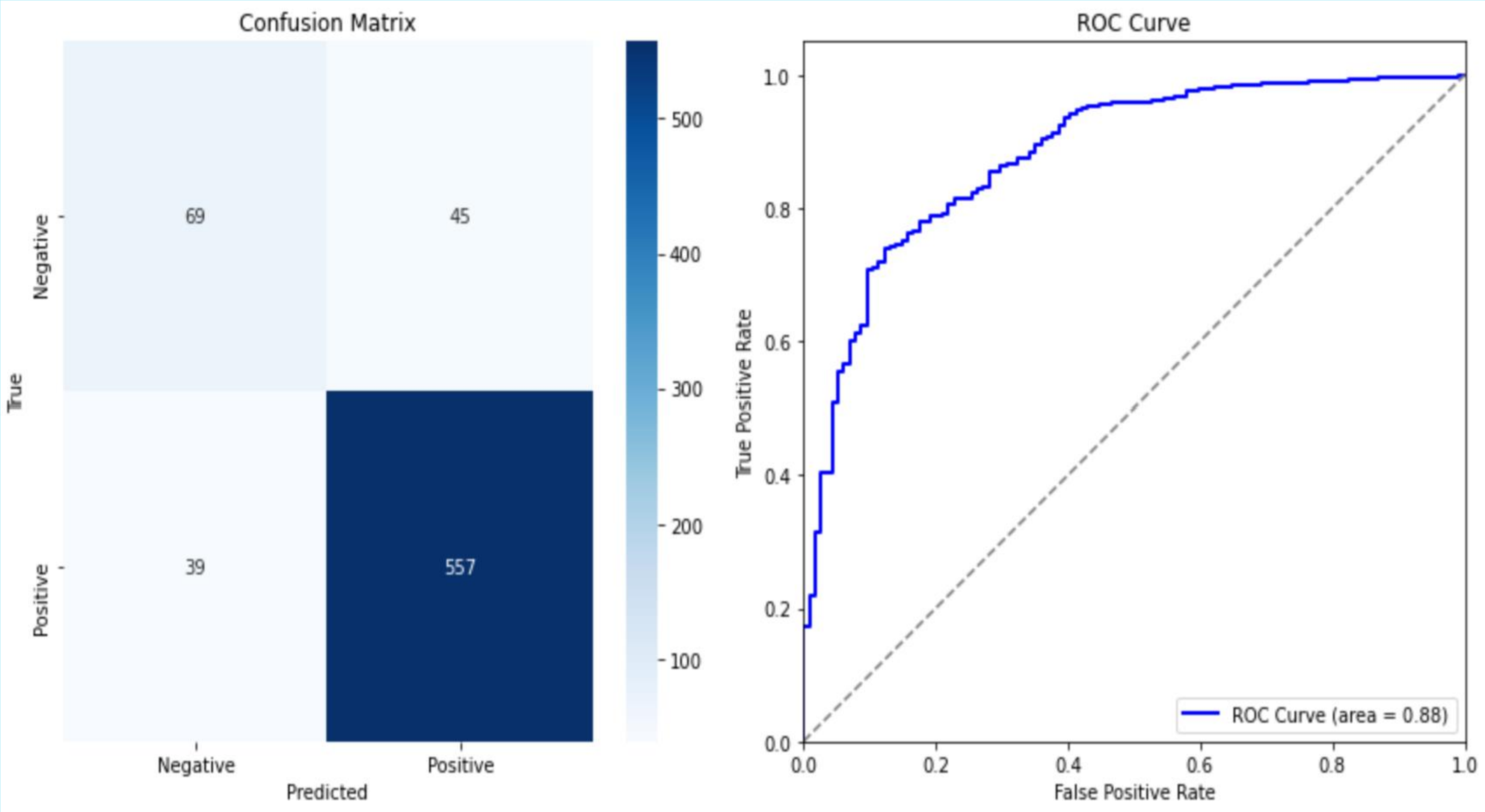


The sentiment distribution is imbalanced, with most tweets having no emotion or positive sentiment as per the chart on the left, while tweet lengths mostly fall within a moderate range as per chart on the right, negating the need for truncation during tokenization.

Modeling

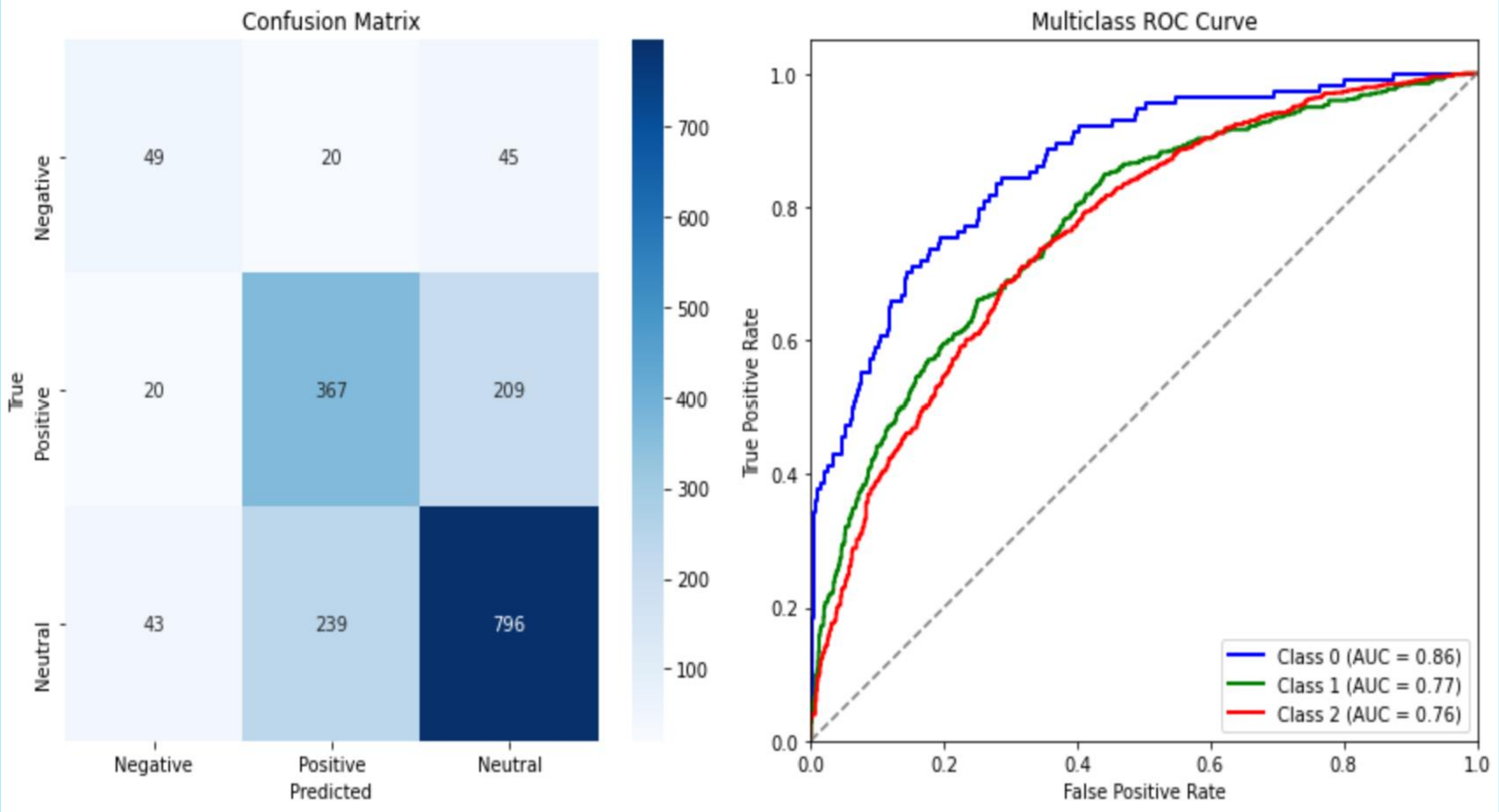
1. Binary Classification Models

- Logistic Regression, Random Forest, and SVM were tested.
- Logistic Regression performed best with **88% accuracy**, excelling in detecting negative sentiments (F1-score: **0.62**) and has AUC-ROC score of **0.88**.



2. Multiclass Classification Models

- Multinomial Naive Bayes and an **ensemble model** (Naive Bayes + Random Forest) were evaluated.
- The ensemble model outperformed with **68% accuracy**, offering balanced performance across all sentiment classes and better detection of negative (F1-score: **0.43**) and neutral sentiments (F1-score: **0.75**), making it the preferred multiclass solution. AUC scores for each class are also depicted in below chart.



Evaluation

<u>KEY METRICS</u>:

Accuracy: Proportion of correctly classified instances.
Precision: Correct positive predictions out of all positive predictions.
Recall: Model's ability to identify all instances of a class.
F1-Score: Balances precision and recall, important for class imbalance.

KEY FINDINGS

- Logistic regression performed best as a binary classifier with an accuracy of 88% , precision

- Ensemble model performed best as a
- multiclass classifier with an accuracy of 68%

CONCLUSION

The project developed effective NLP sentiment classification models for analyzing consumer sentiment on Twitter. Logistic Regression excelled in binary classification with 88% accuracy, particularly in detecting negative sentiment, while the ensemble model (Naive Bayes + Random Forest) outperformed in multiclass classification with 68% accuracy. These models will help provide actionable insights for Apple and Google, supporting informed decisions on product strategies and customer engagement.

RECOMMENDATION

- ❖ Enhance Negative Sentiment Detection: Explore advanced methods like phrase extraction or models like BERT to capture nuanced negative expressions.
- ❖ Address Class Imbalance: Collect more diverse data to improve representation of underrepresented sentiment classes, especially negative sentiment.
- ❖ Refine Pre-processing: Implement more sophisticated techniques, such as handling negations and domain-specific language.
- ❖ Ongoing Monitoring and Retraining: Regularly update the model to ensure accuracy as consumer sentiment evolves.

Any Questions?

Phase_4, Group_6:

- ✓ ***Mary Musyoka***
- ✓ ***Julian Kilyungi***
- ✓ ***Tabitha Kariuki***
- ✓ ***John Kul***
- ✓ ***Norah Oluoch***
- ✓ ***Joseph Ngige***

THANK YOU