TWITTER SENTIMENT ANALYSIS – GOOGLE AND APPLE PRODUCTS

UNVEILING CONSUMER SENTIMENTS THROUGH MACHINE LEARNING

TEAM MEMBERS

- GEORGE CHIRA
- MAYA BORORIO
- GEORGE MACHOKA
- LEILA NYAMBURA
- CATHERINE WANJAU
- PUPITY MWENDWA
- ALEX IRUNGU

TABLE OF CONTENT

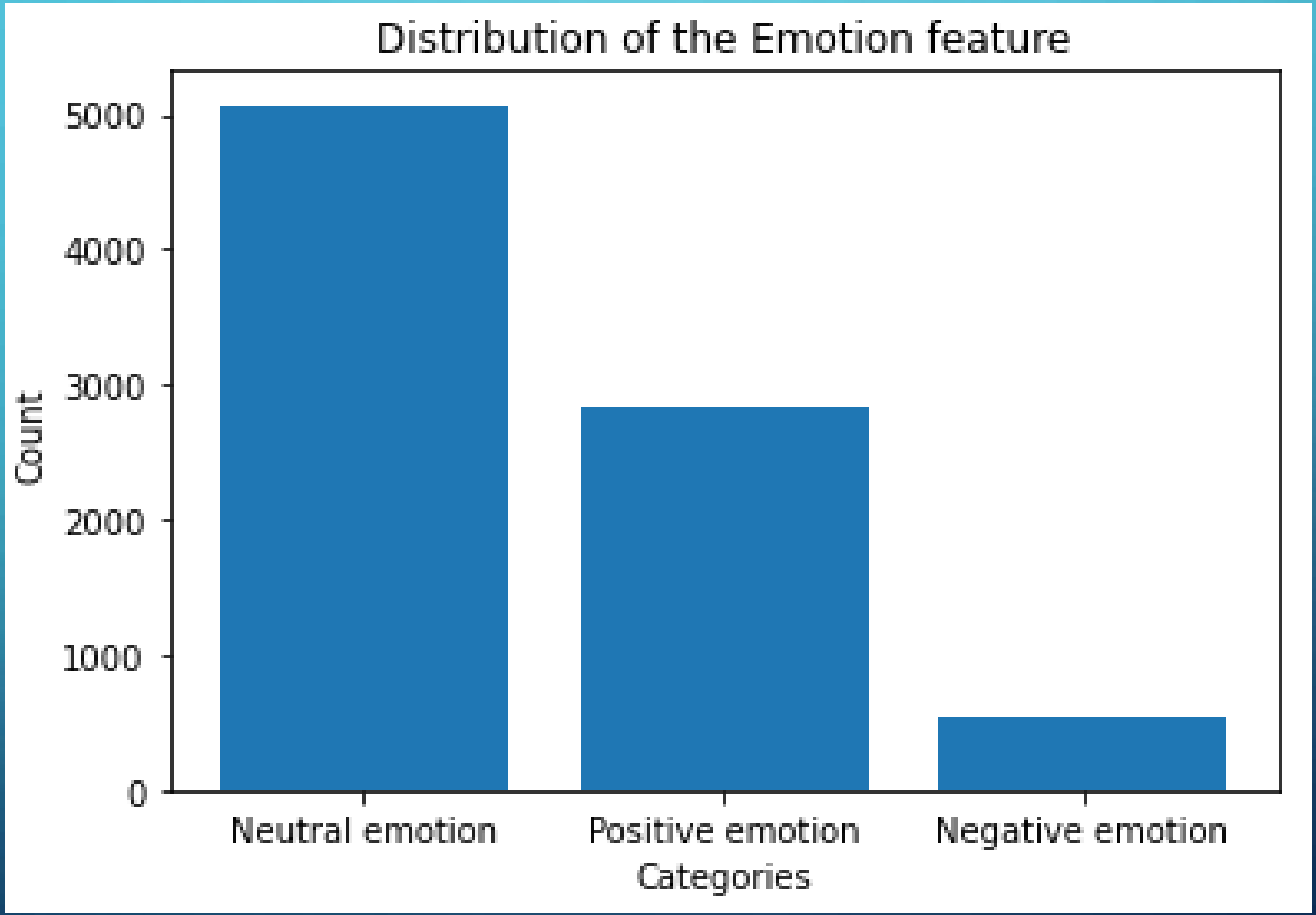
OVERVIEW

- This project aimed to analyze Twitter sentiment about Apple and Google products. The sentiment analysis helps us understand if customers are expressing positive, negative, or neutral feelings toward these companies' products on social media.
- **OBJECTIVE:** By understanding sentiments, Apple and Google can gain insight into customer satisfaction, identify product strengths and weaknesses, and make informed decisions to improve customer experience.

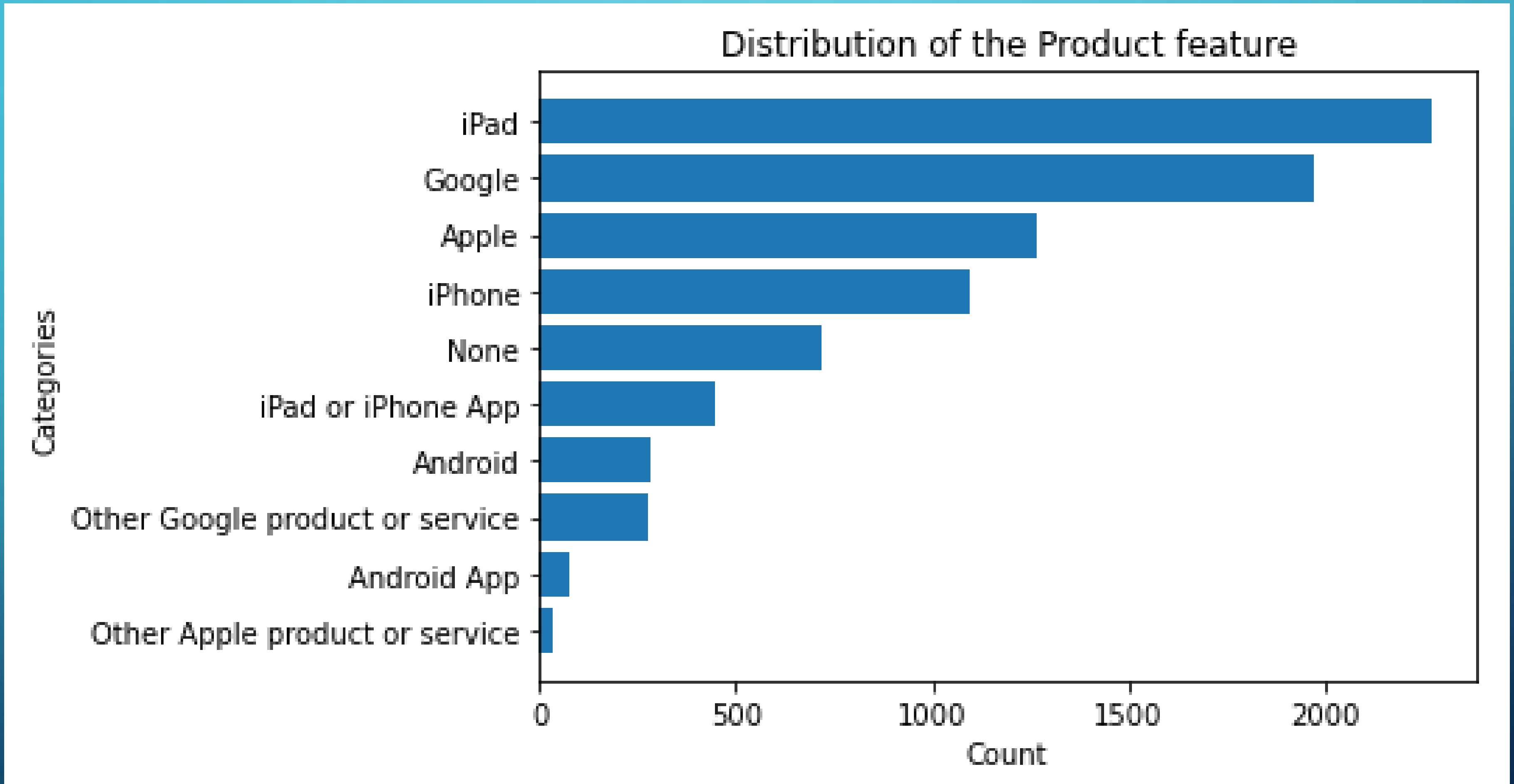
BUSINESS AND DATA UNDERSTANDING

- **Business Context:** Both Apple and Google strive to maintain positive customer relations and improve their products. With increasing engagement on social media, Twitter sentiment offers real-time feedback on how customers feel about new releases, updates, and the general perception of each brand.
- **Data Source:** We used **secondary data** from CrowdFlower with a dataset of over 9,000 tweets mentioning Apple and Google products. Each tweet was manually rated as positive, negative, or neutral by human reviewers, providing us with labeled data that accurately reflects customer sentiment.

EMOTION DISTRIBUTION



PRODUCT DISTRIBUTION



DATA PREPARATION

- **Data Cleaning:** The initial data contained various non-essential elements like special characters, hashtags, and URLs. We removed these to focus on the core text content of each tweet. Cleaning ensures that the data is consistent and relevant for analysis.
- **Text Transformation:** Text data isn’t inherently understandable to machine learning models. We transformed each tweet into a numerical format through methods like tokenization, which breaks down text into smaller pieces (words or phrases). This conversion allows the model to detect patterns and associations within the text.

MODELING

- <u>**Model Choices:**</u>
  - <u>**Binary Classification Model:**</u> First, we built a model to categorize tweets as either positive or negative. This is useful for understanding general customer sentiment and is relatively straightforward to implement.
  - <u>**Multiclass Classification Model:**</u> We then expanded our analysis by including neutral sentiments, creating a three-category classification model. This provides more detailed insights, distinguishing between tweets that are strongly opinionated and those that are more neutral.
- <u>**Model Functionality:**</u> These models work by identifying patterns in the words and phrases people use. The binary model focused on simpler patterns, while the multiclass model required deeper analysis to separate neutral comments from more strongly positive or negative sentiments.

EVALUATION

- **Accuracy:** This metric tells us how often the model's predictions align with the actual sentiment in the tweets. High accuracy means the model is effective at recognizing and classifying sentiments while low accuracy means its not as effective. For instance, if we achieved 85% accuracy, it means the model correctly classified 85 out of every 100 tweets.
- **Confusion Matrix:** To assess how the model performed across all categories (positive, negative, and neutral), we used a confusion matrix. This tool helps us see where the model made mistakes, such as mixing up neutral tweets with positive ones. It provides insight into areas where the model can be improved or where sentiment is harder to distinguish.

KEY FINDINGS

- **Tuned Random Forest model** and **Tuned Logistic Regression model** achieved the highest accuracy and recall scores, both scoring approximately **83.7%** in accuracy and **83.6%** in recall. Making these our models of choice.
- This high performance was achieved after model hyperparameter tuning which significantly improved model performance, where the Random Forest and Logistic Regression models' accuracy and recall improved by more than **10%** in some cases.

NOTE: Hyperparameter tuning is the process of selecting the optimal values for a machine learning model's configuration variables.

RECOMMENDATIONS

- For Apple and Google:
  - **Product Improvements:** For both companies, we suggest focusing on areas that received frequent negative feedback. For example, if users complained about specific features, these could be reviewed and improved in future updates or versions.
  - **Marketing Strategies:** Highlight aspects that customers praise, such as design, functionality, or unique features. Understanding what drives positive sentiment allows Apple and Google to reinforce these aspects in their marketing efforts.

NEXT STEPS

- **Regular Monitoring:** We suggest implementing ongoing sentiment analysis to keep track of customer feedback.
- **Broader Analysis:** Consider analyzing sentiment across different platforms, like Facebook or Reddit, for a more comprehensive understanding.
- **Customer Support Integration:** Use this analysis to proactively address customer complaints and improve support.

THANK YOU