# NATURAL LANGUAGE PROCESSING PROJECT

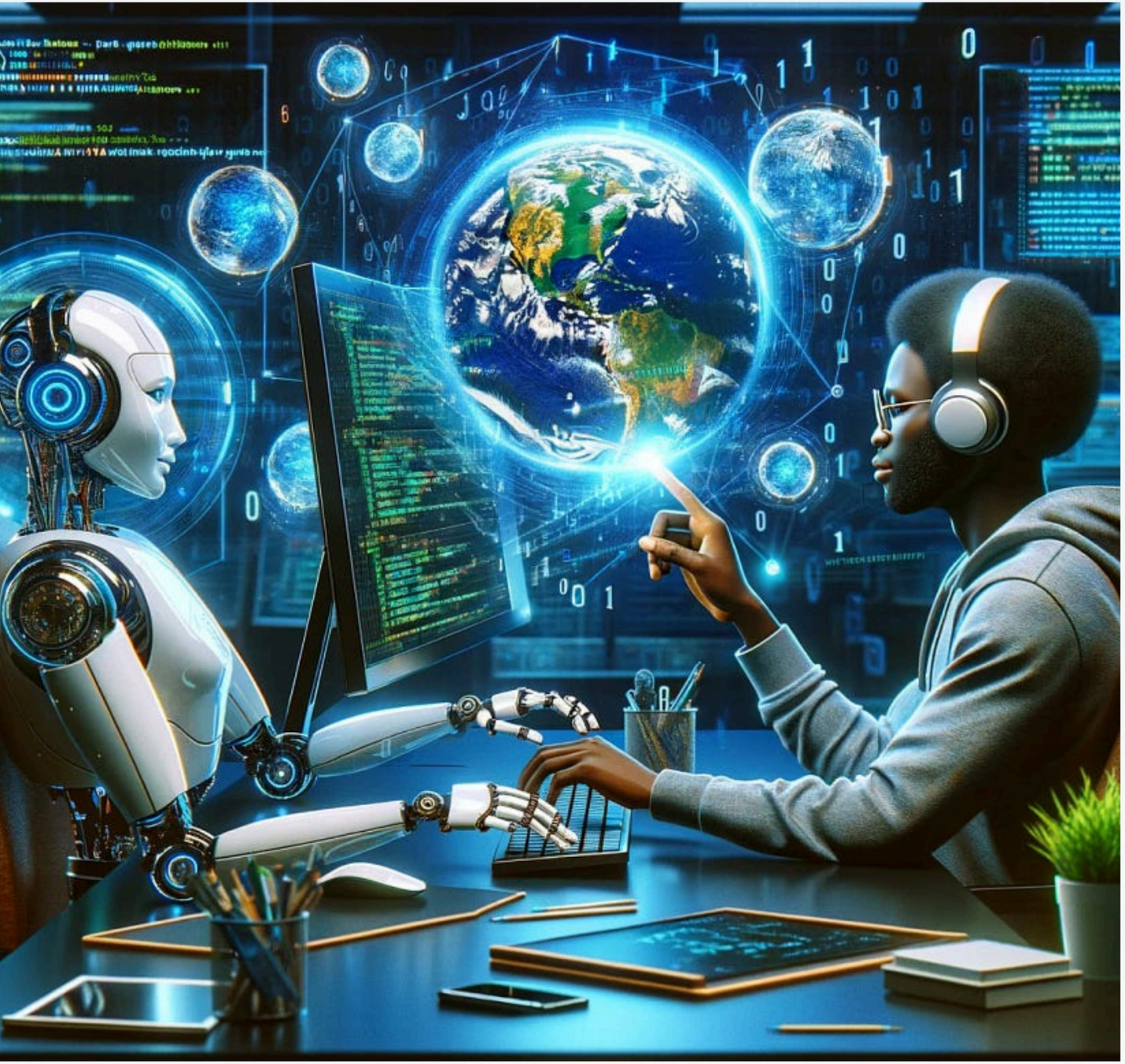
BY PETER MBUTHIA
MARION MACHARIA
MERCY JEPKOSGEI
MITCH MATHIU

# BUSINESS CONTEXT.

- Sentiment Analysis of Apple and Google Products on Twitter

- The primary aim is to analyze public sentiment towards Apple and Google products through Twitter data using Natural Language Processing (NLP).

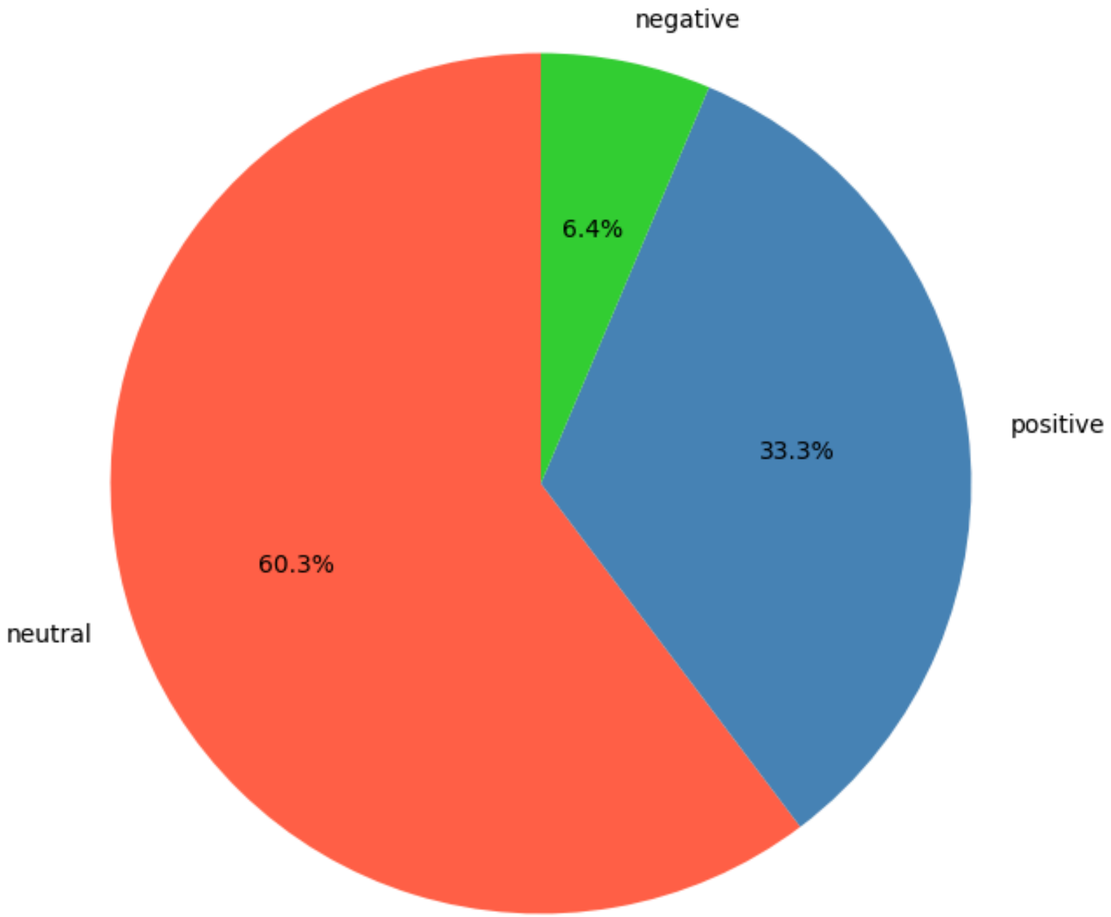
- By understanding how users feel about these products, businesses can make informed decisions regarding marketing strategies, product development, and customer engagement.



# OBJECTIVE

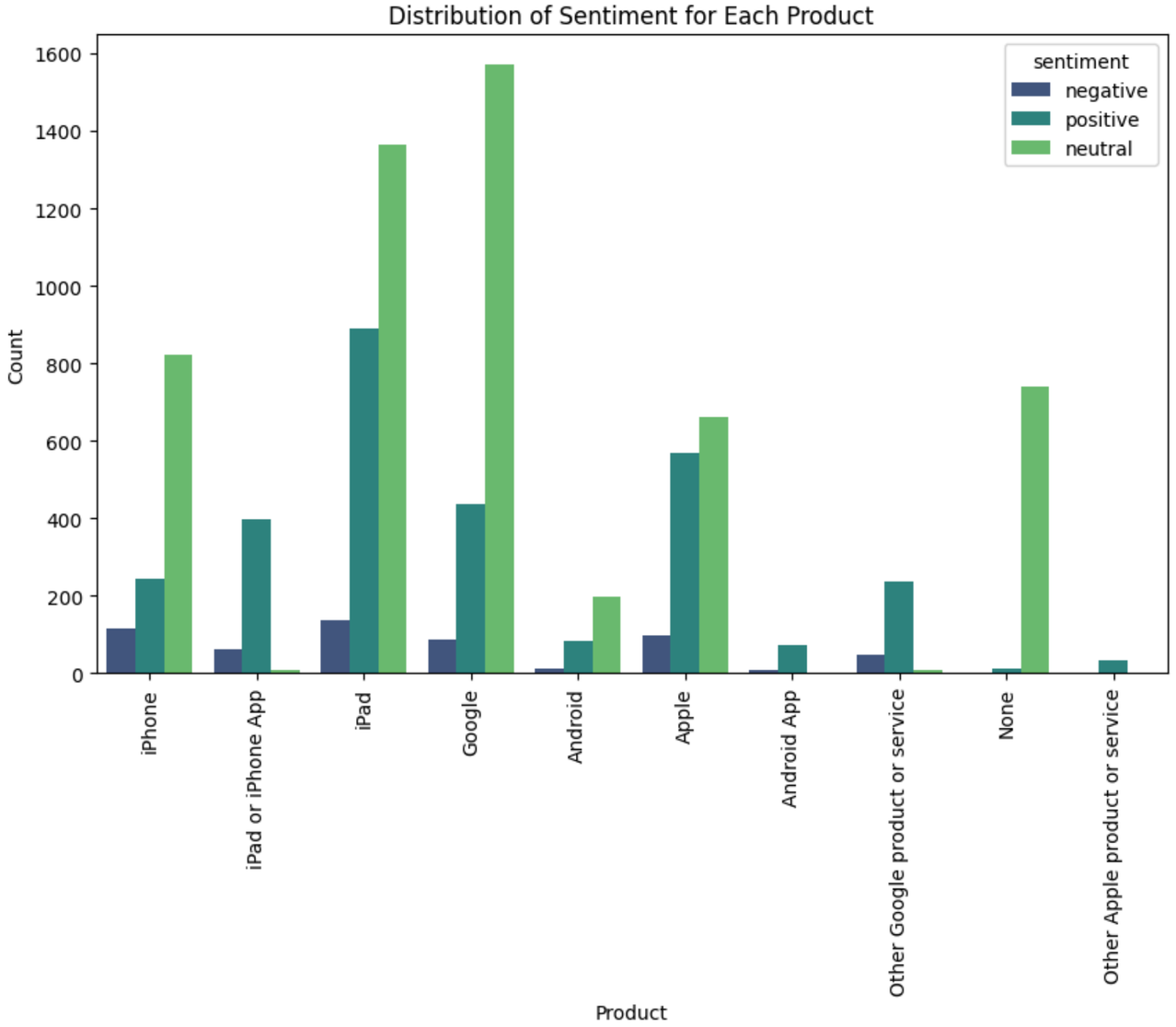
**To develop a Natural Language Processing (NLP) model that accurately classifies the sentiment of tweets related to Apple and Google products as positive, negative, or neutral.**

# SENTIMENT DISTRIBUTION




**The Pie Chart above displays that the percentage of neural sentiments is higher (60.3%), followed by positive (33.3%) then negative at (6.4%). With such a collection, there is likelihood of a class imbalance challenge when analyzing the sentiments.**

# SENTIMENT ACROSS TOP 10 PRODUCTS

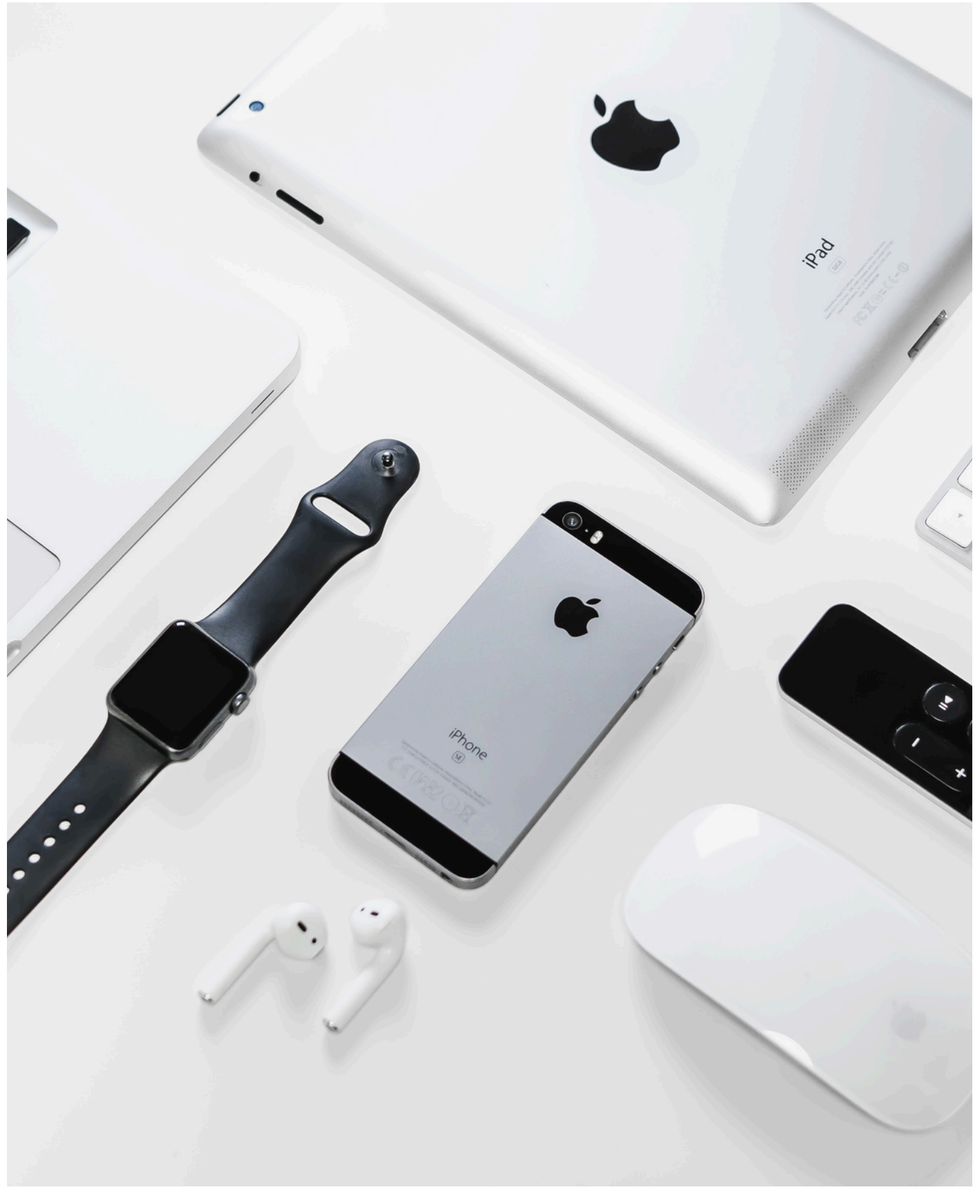



**The plot demonstrates that the distribution of neutral sentiments across the product is high. This may be as a result of a class imbalance, whereby the number of neutral sentiments are more than those of either positive or negative. According to the plot, the ipad product has the highest positive sentiment of 900 whilst the highest negative sentiments are observed to affect the iphone product.**

# MODEL PERFORMANCE

- After testing various models for sentiment analysis, both simple (logistic regression) and advanced models (XGBOOST, SVM, Random Forest) show strong binary classification accuracy between 84% and 89% after tuning.
- However, multiclass trials consistently stagnate at 67%, with class imbalance, particularly the dominance of neutral sentiments, posing a challenge across all models.

# RECOMMENDATION




- To get better insights, companies like Google and Apple may decide to focus on high impact users such as relevant product influencers and tech communities whom can provide more valuable insights about their products. This is usually informed by the high product engagement these target segmentation has.

- Seek to interogate the neutral comments further using aspect-based sentiment analysis with the aim of contextualizing the comment. This may help identify valuable information/feedback about a product's performance or design.

# RECOMMENDATION


- For the sentiment analysis model to perform better, introducing sentiment weighting might aid in distinguishing feedback from high-engagement users to reduce the influence of low-impact neutral responses influencing the overall insights.

- Lastly, enhancing the sentiment classification models from the conventional(positive, negative and neutral) responses, where it can be able to interpret more nuanced emotions within text.