

NLP

FROM TWEETS TO INSIGHTS

*SENTIMENT ANALYSIS OF APPLE & GOOGLE
PRODUCTS*



PROBLEM STATEMENT

"Every tweet is a review. But with thousands of voices speaking at once, how do we filter the noise and measure how people really feel about Apple and Google products?"

PROJECT OBJECTIVE

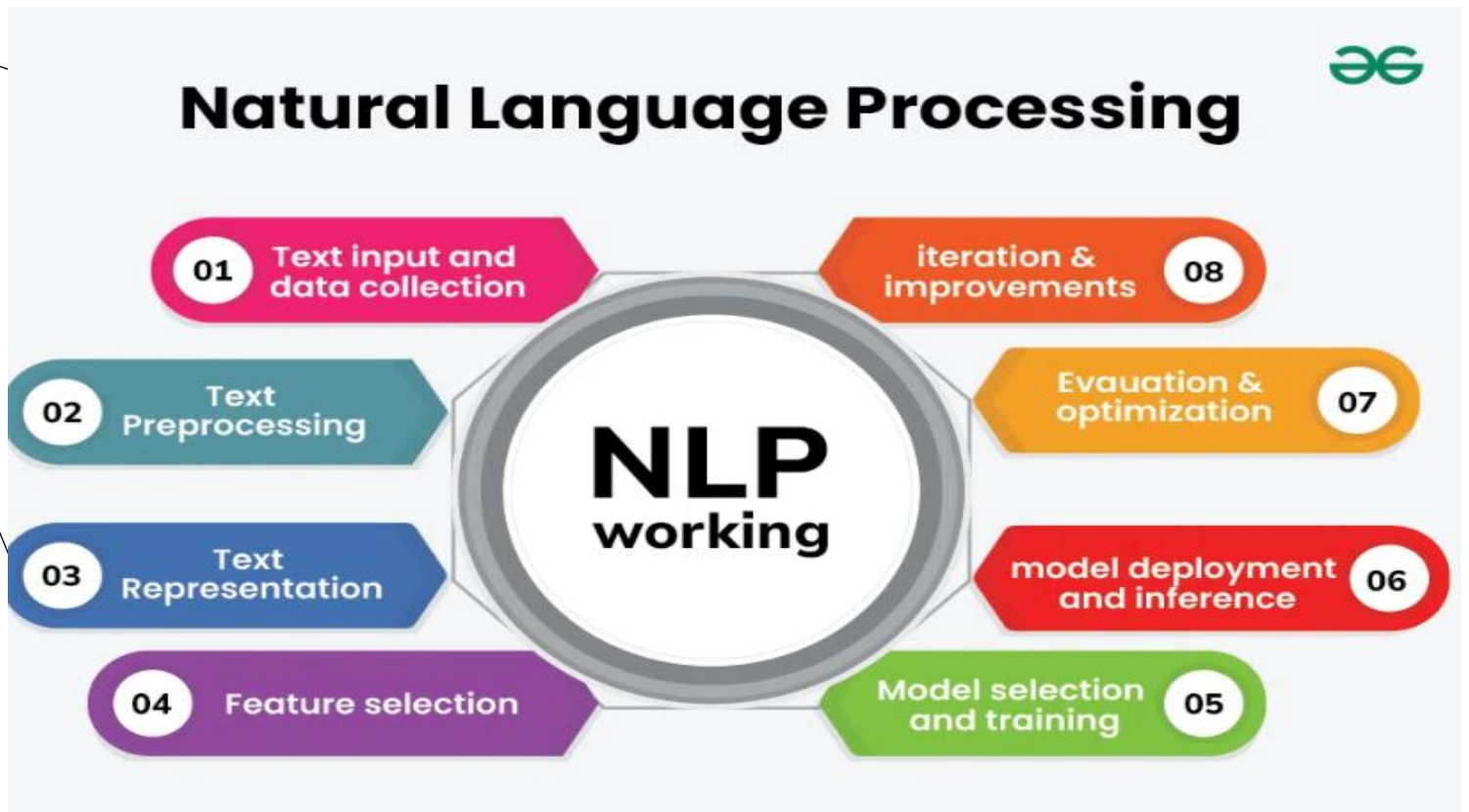
The objective of this project is to develop a Natural Language Processing (NLP) model capable of classifying tweets about Apple and Google products into positive, negative, or neutral sentiment categories. By leveraging advanced text processing and machine learning techniques, the model aims to provide accurate, automated insights into consumer opinions expressed on Twitter.



DATASET OVERVIEW

- Source: *CrowdFlower (Data.world)* - 9,000+ tweets about Apple & Google products.
- Size: ~9,100 tweets, 3 columns.
- Features
 - `tweet_text` → raw tweet
 - `emotion_in_tweet_is_directed_at` → product/brand mentioned (iPhone, iPad, Google, etc.).
 - `is_there_an_emotion_directed_at_a_brand_or_product` → sentiment (Positive, Negative, Neutral).

METHODOLOGY

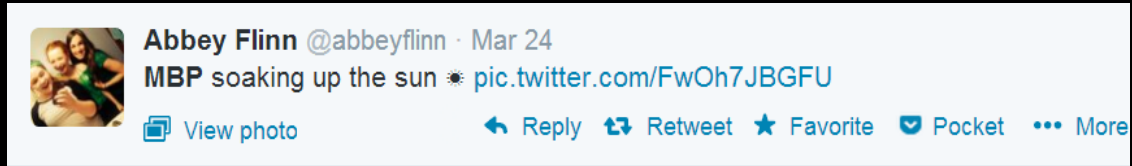


TOOLS USED

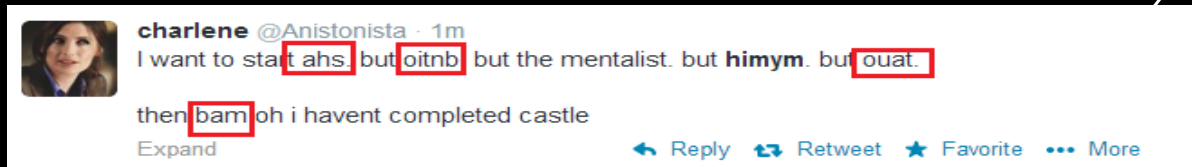


CHALLENGES

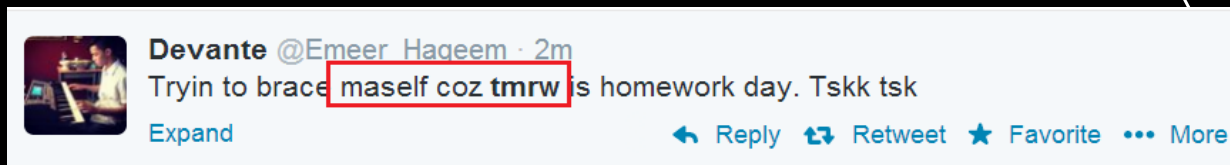
- Tweets are highly unstructured and also non-grammatical



- Out of Vocabulary Words



- Lexical Variation



- Extensive usage of acronyms like *asap*, *lol*, *afaik*, *noisy* (hashtags, emojis, URLs) texts, *sarcasm*, *imbalanced data*.

EXPLORATORY DATA ANALYSIS

SENTIMENT DISTRIBUTION



WORD CLOUD FOR POSITIVE & NEGATIVE EMOTIONS

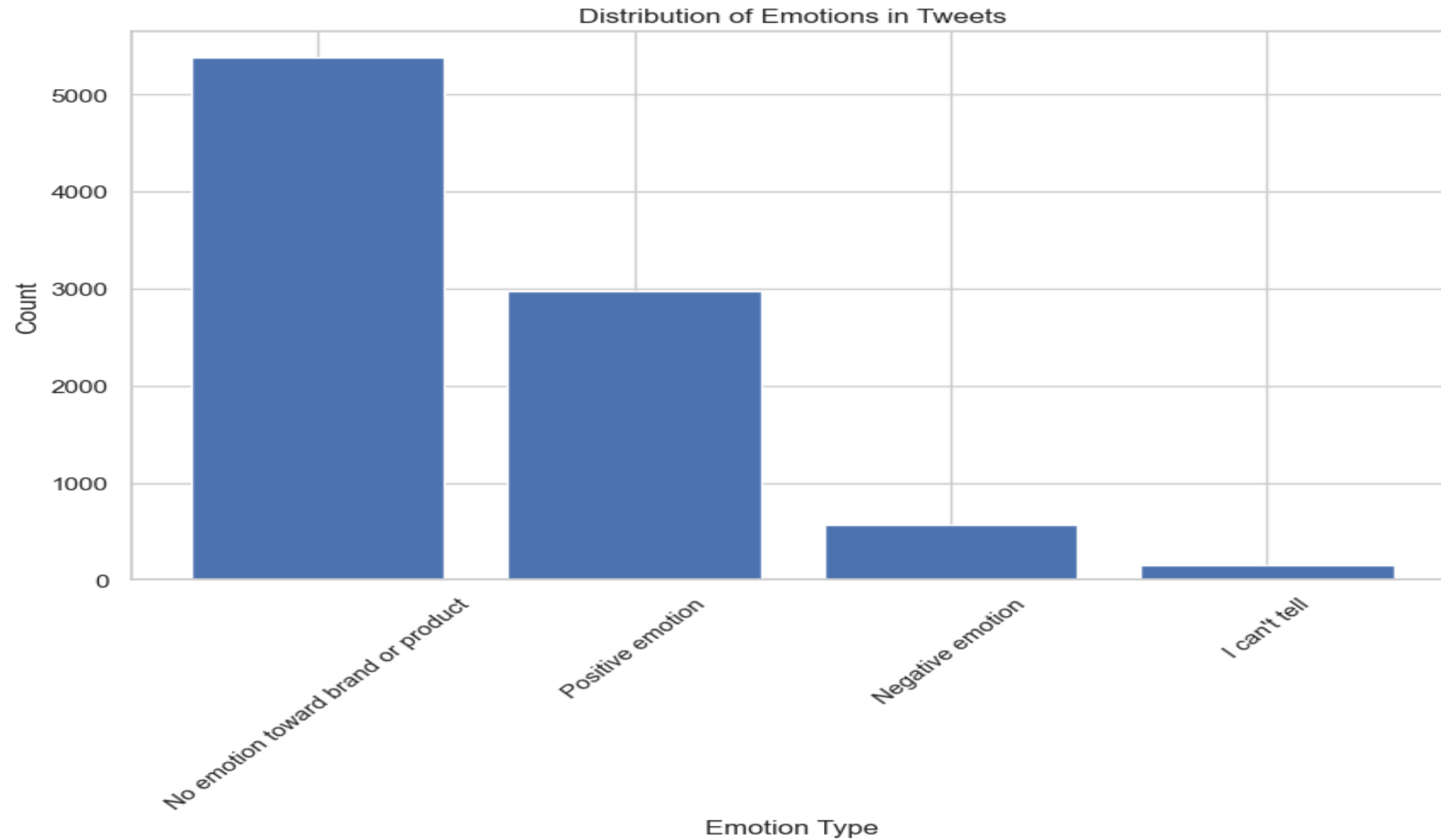
Word Cloud for Positive Emotions



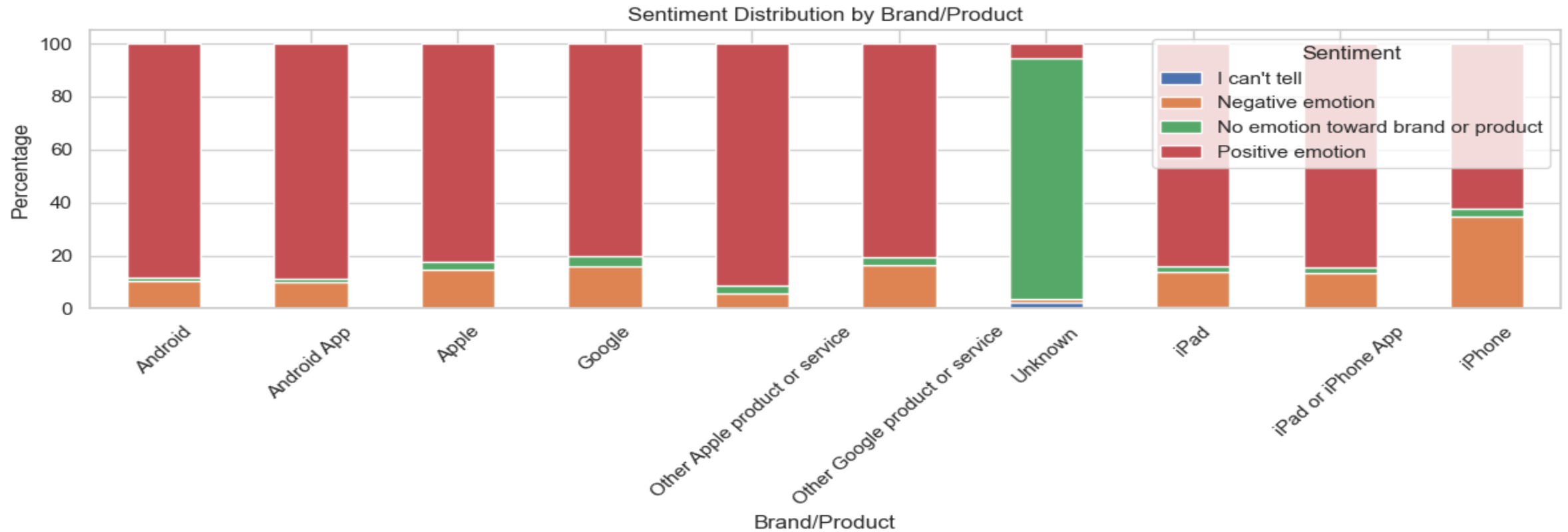
Word Cloud for Negative Emotions



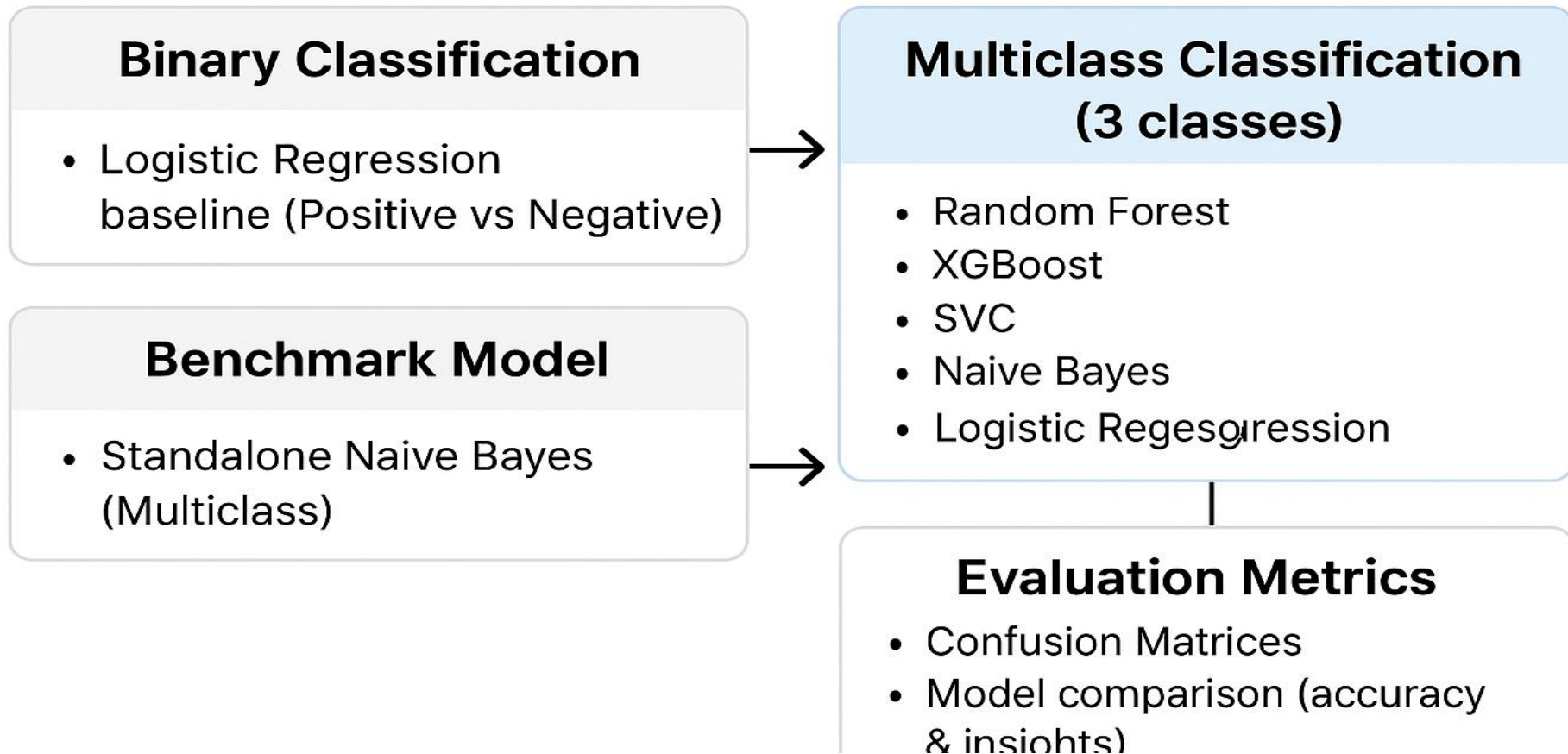
EMOTION DISTRIBUTION IN THE TWEETS



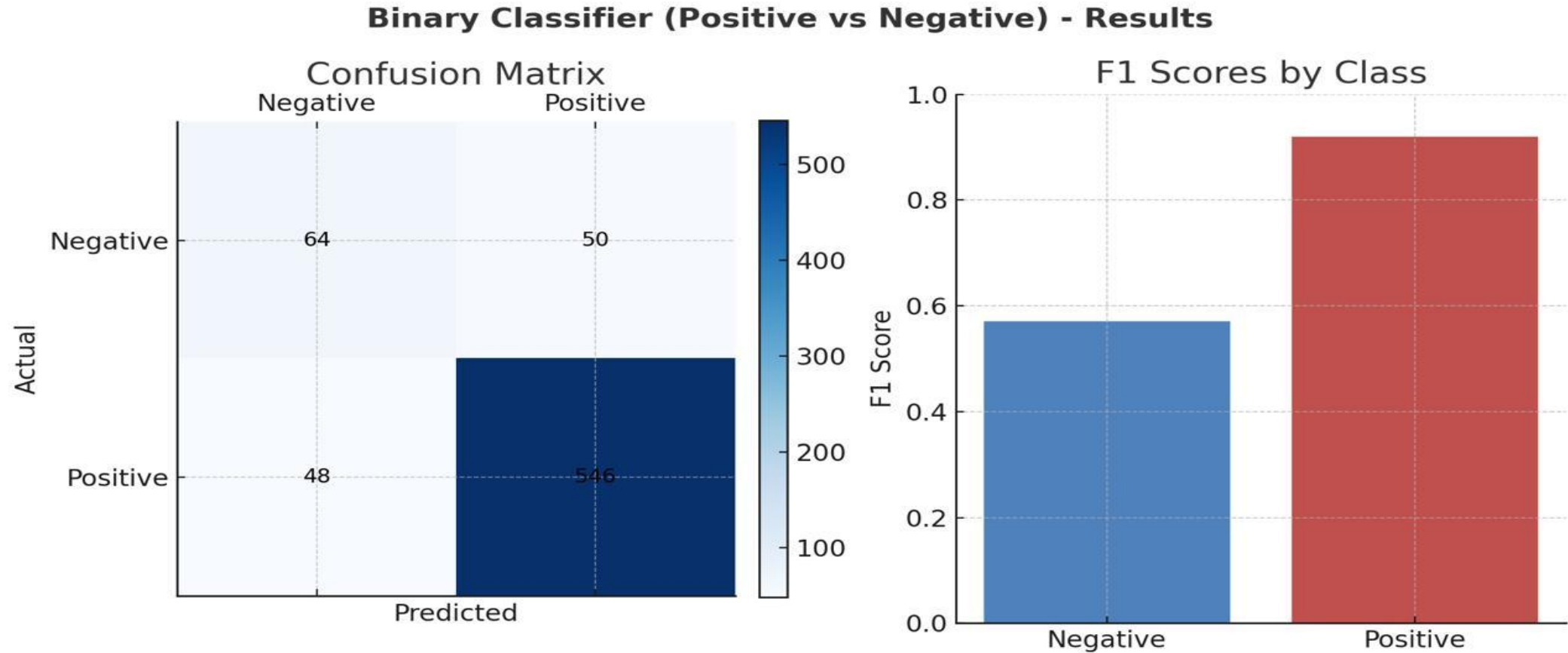
SENTIMENT DISTRIBUTION BY BRAND/PRODUCT



Modeling & Evaluation

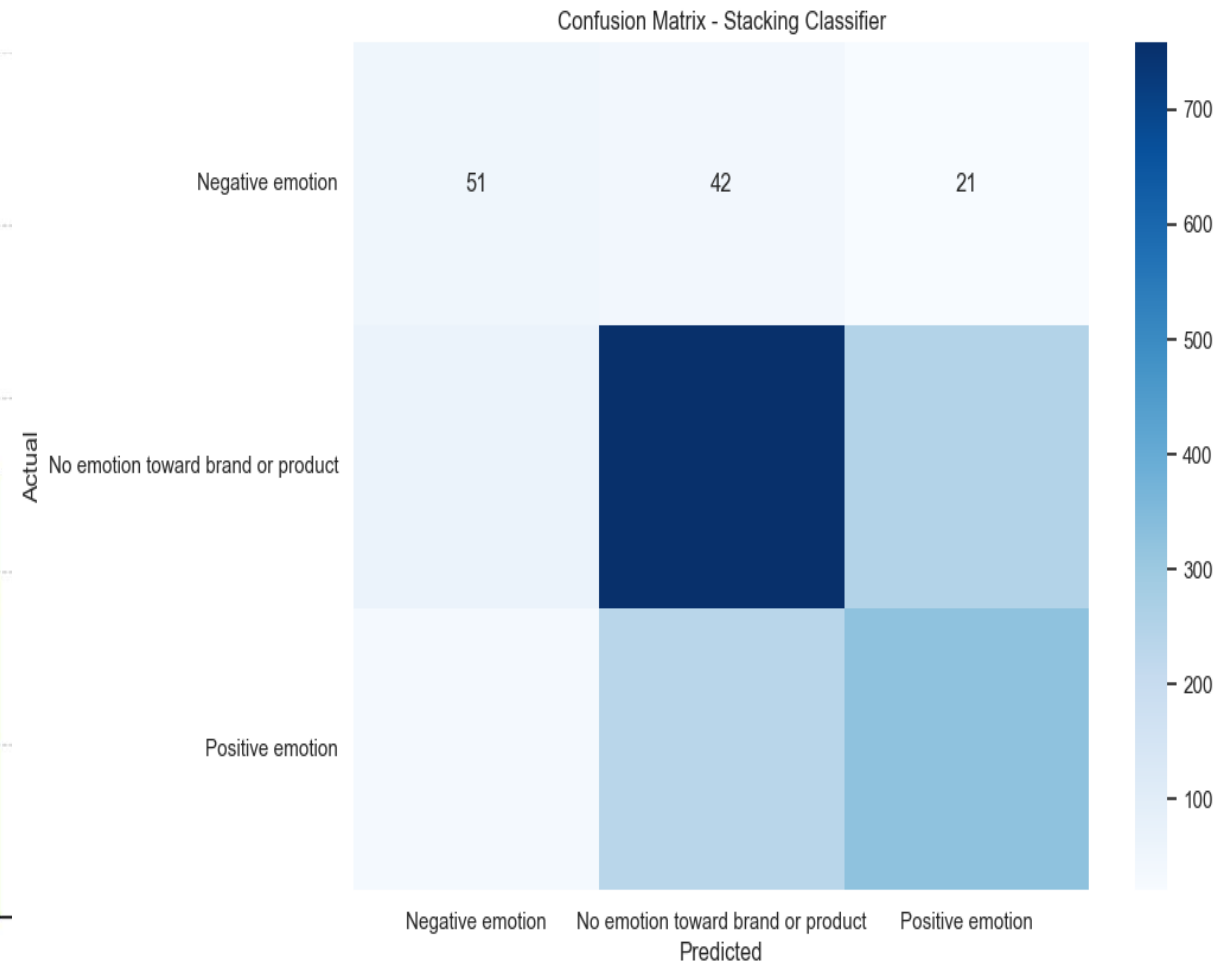
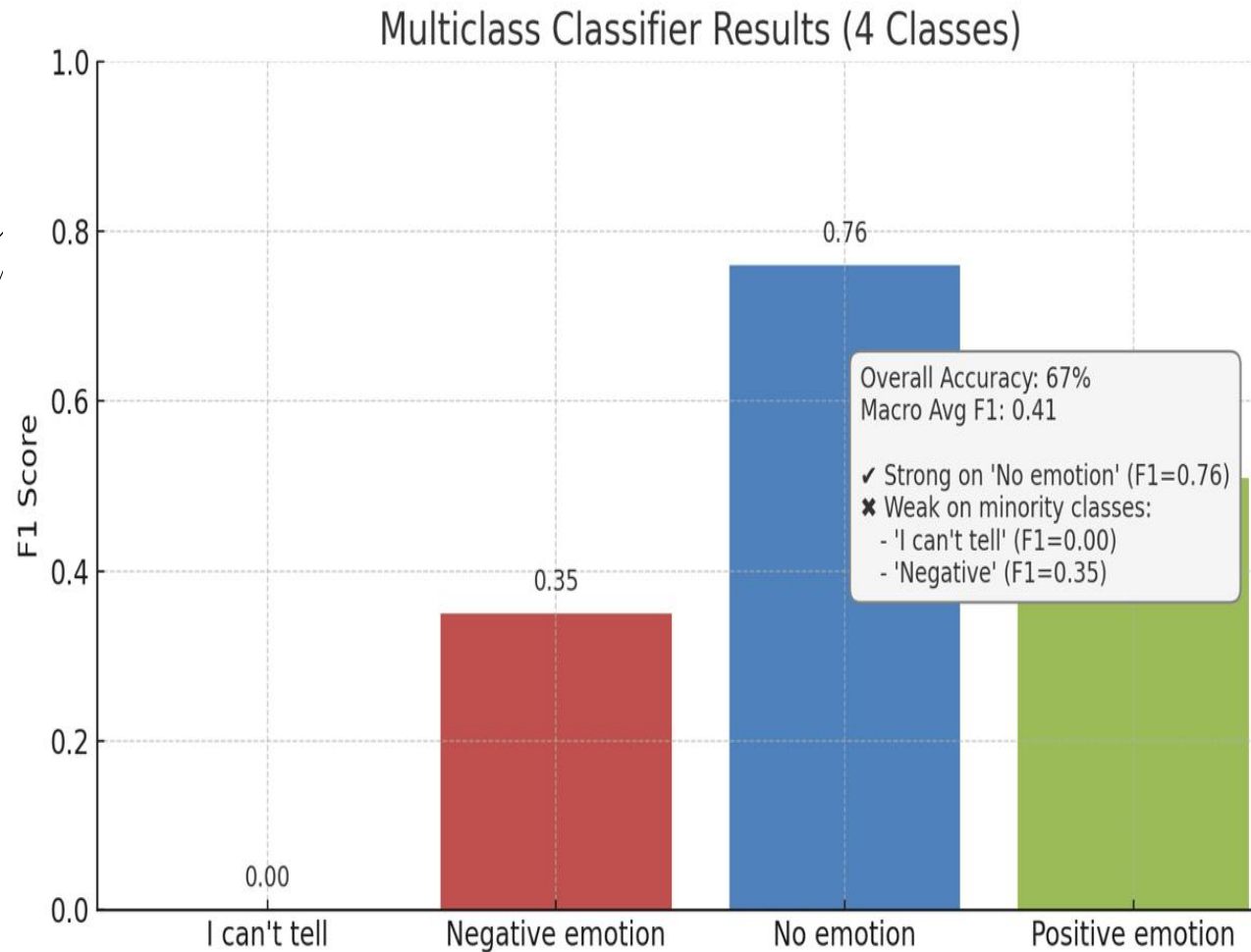


RESULTS-BINARY CLASSIFIER



"Model achieves high accuracy (86%) with strong detection of positive tweets, but struggles with negative sentiment due to class imbalance."

RESULTS-MULTICLASS CLASSIFIER



"Model achieves 67% accuracy, performing well on 'No emotion' but failing to capture minority classes, especially 'I can't tell'."

SUMMARY

Conclusion & Recommendations

Conclusion



Sentiment analysis = insights into Google & Apple perception



Preprocessing improved accuracy



Final model (pipeline_multi2) strong on pos/neg



Real-time monitoring helps business act fast

Recommendations



Deployment:
Cloud API (Flask/FastAPI)



Monitoring:
Dashboards (Streamuits /Power BI/τα)



Continuous Learning:
Retrain with new tweets



Business Use:
Marketing, support, risk detection

THANK YOU

