PROJECT TITLE:

MEASURING PUBLIC PERCEPTION OF APPLE AND GOOGLE ON TWITTER USING MACHINE LEARNING.

PRESENTED BY:

Faith Kanyuki, Mary bennah Kuloba, Sharon Thiga, Sharon Maina

BUSINESS UNDERSTANDING

OBJECTIVE

BUILD A NATURAL LANGUAGE PROCESSING (NLP) MODEL TO CLASSIFY TWEETS MENTIONING APPLE OR GOOGLE INTO POSITIVE**, **NEGATIVE**, OR **NEUTRAL** SENTIMENTS. THIS WILL HELP UNDERSTAND PUBLIC PERCEPTION, MONITOR BRAND REPUTATION, AND INFORM MARKETING AND PRODUCT STRATEGIES.

KEY QUESTIONS

- HOW DO PEOPLE FEEL ABOUT APPLE VS GOOGLE PRODUCTS ON SOCIAL MEDIA?

- WHICH WORDS OR PHRASES ARE MOST INFLUENTIAL IN EXPRESSING SENTIMENT?

- DO SPECIFIC EVENTS OR PRODUCT LAUNCHES TRIGGER CHANGES IN SENTIMENT?

BUSINESS IMPACT

- MARKETING STRATEGY: ADJUST CAMPAIGNS BASED ON TRENDS IN SENTIMENT.

- CUSTOMER SUPPORT: PRIORITIZE RESPONSES WHEN NEGATIVE SENTIMENT SPIKES.

- PRODUCT DEVELOPMENT: IDENTIFY COMMON ISSUES OR REQUESTS FROM USER FEEDBACK.

- CRISIS MANAGEMENT: DETECT NEGATIVE TRENDS EARLY TO MITIGATE PR RISKS.

DATA UNDERSTANDING

Dataset Overview

- Source: CrowdFlower via data. World

- Total tweets:9,000 (Apple and Google mentions)

- Labels: Positive, Negative, Neutral (human-annotated)

Key Columns

Column	Description	Example
tweet_text	The content of the tweet	"I love my new iPhone!"
emotion_in_tweet_is_directed_at	Brand or product mentioned	"iPhone"
is_there_an_emotion_directed_at_a_brand_or_product	Sentiment label	Positive, Negative, Neutral

Dataset Suitability

- Real user opinions, unstructured text

- Labeled for sentiment → suitable for supervised ML

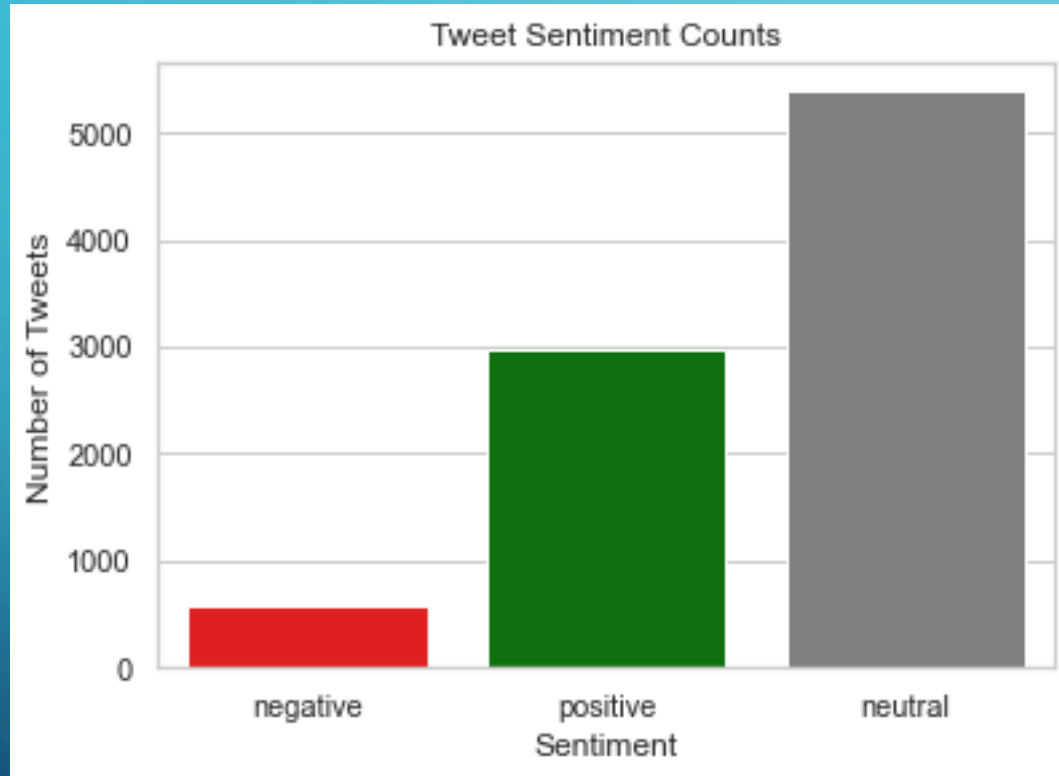- Allows tracking trends over time and across products

DATA PREPARATION

- Data Cleaning

This included;

1. Dropping unwanted columns

2. Handling missing values

3. Renaming columns and sentiment categories

4. Removing unwanted categories

5. Cleaning text data

6. Text vectorization

DATA ANALYZATION

1.Interpretation of Tweet Sentiment Counts (Bar Chart)

Neutral (5,500 tweets):
The majority of tweets are neutral, indicating that most users are either sharing information or not expressing strong opinions about Apple or Google products.
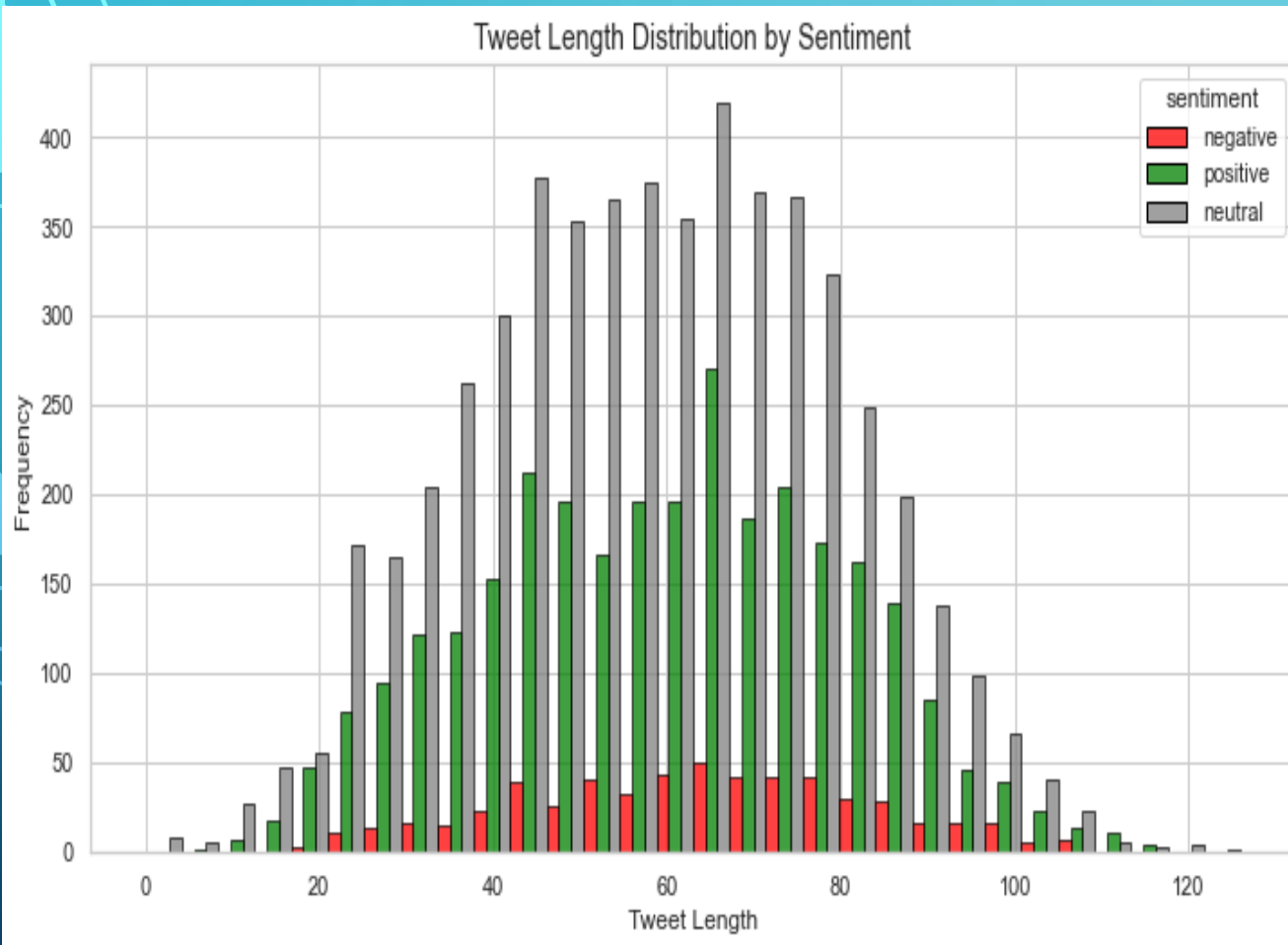
- **Positive (3,000 tweets):**
A significant portion of tweets show favorable sentiment, reflecting user satisfaction or approval.

- **Negative (600 tweets):**
Only a small fraction of tweets are negative, suggesting limited dissatisfaction expressed on Twitter.

2. ANALYZING TWEET LENGTHS BY SENTIMENT

Tweet Length Distribution by Sentiment

Interpretation of Tweet Length Distribution by Sentiment

Positive Tweets:

These tweets tend to have moderate to slightly longer lengths, mostly between 50 and 100 characters.

This suggests that users expressing positive emotions often write a bit more, possibly sharing detailed feedback, praise, or experiences.

Negative Tweets:

Negative tweets are generally shorter, clustering around 20 to 80 characters.

This indicates that users expressing dissatisfaction often post brief, direct messages, likely reflecting frustration or immediate reactions.

Neutral Tweets:

Neutral tweets show a broader range of lengths, including some of the longest tweets in the dataset.

These are likely factual, informational, or descriptive messages that require more words to convey context without conveying emotion.

Overall Patterns:

- Positive and neutral tweets are generally longer than negative ones.

- Negative tweets are concise, often signaling quick reactions

MODEL TRAINING AND EVALUATION

We used the following models;

- ✓ Model 1 Logistic Regression (Baseline) Text +Numerical Features
- ✓ MODEL 2 Random Forest Classifier (TF-IDF + Numeric Features)
- ✓ MODEL 3 SVM Classifier (TF-IDF + Numeric Features)

- ✓ MODEL 4 XGBoost Classifier (TF-IDF + Numeric Features)

MODELING INTERPRÉTATION (MULTI CLASS SENTIMENT CLASSIFICATION)

Overall Comparison of Models

Model	Test Accuracy	Macro F1	Observation
Logistic Regression	0.6544	0.5723	Baseline model. Performs fairly balanced across classes.
Random Forest	0.6734	0.5204	Accuracy slightly higher, but macro F1 lower → struggles on minority classes (Negative).
SVM	0.6437	0.5663	Similar to Logistic Regression; better recall on Negative.
XGBoost	0.6695	0.4687	Accuracy okay, but macro F1 low → very poor on Negative and Positive classes.

Observation:
- Accuracy can be misleading due to class imbalance (Neutral dominates).
- **Macro F1** is a better metric as it gives equal weight to all classe

BUSINESS INTERPRETATION

Overall Interpretation

- Best models for minority classes (Negative, Positive): Logistic Regression & SVM
- Best for majority class (Neutral): XGBoost & Random Forest
- Best overall balance (Macro F1):** Logistic Regression (0.5723)
- Key challenge: Class imbalance causes models to favor Neutral, hurting minority classes.

Recommendations Based on Models

1. Use Logistic Regression or SVM if balanced performance across all sentiments is desired.
2. Consider class weighting or resampling(e.g., SMOTE) for tree-based models to improve Negative/Positive detection.
3. Focus on improving Negative recall → more labeled data, oversampling, or threshold tuning.
4. Always report Macro F1 in addition to Accuracy, as it is more informative for imbalanced multiclass sentiment data.

CONCLUSION

Model Performance Conclusions

- Logistic Regression is the most suitable model for practical business use, even though it is not the highest in accuracy.
  - It achieved a macro F1 score of 0.5723, meaning it balances performance across negative, neutral, and positive classes better than the other models.
  - This balance is essential for understanding the full sentiment landscape around Apple and Google products.

- Accuracy alone is misleading in imbalanced datasets like this one.
  - For example, XGBoost achieved 66.95% accuracy which seems high,
    but it only had 10% recall for negative tweets— meaning it failed to detect 9 out of 10 negative tweets.
  - In real business scenarios such as crisis management, customer support, or PR monitoring,
    missing negative tweets is a critical failure, making XGBoost unsuitable despite its high accuracy.

Overall, the results show that balanced performance (macro F1) is more important than accuracy when analyzing public sentiment, especially when negative feedback is rare but highly important.

BUSINESS INSIGHTS

Sentiment Distribution Patterns

- Neutral tweets dominate (60.3%), showing that most discussions around Apple and Google are: Informational
- Positive sentiment (33.3%) is much higher than negative, suggesting:
  - A generally **favorable public perception of both brands

- Negative sentiment is very low (6.4%), but should be monitored, since shifts here can indicate early reputational or product issues.

Model Limitations Impact

- Detecting negative sentiment remains challenging — even the best model (SVM) reached only 54% recall for negative tweets.

Implication:

For Apple and Google, failing to catch negative sentiment early can lead to delayed crisis response, damaged brand reputation and lost customer trust.

BUSINESS RECOMMENDATIONS

- Use Logistic Regression or SVM as the primary model due to their balanced performance across all sentiment classes.
- Set up sentiment alerts to detect spikes in negative sentiment early, compensating for the 51–54% recall rate on negative tweets.
- Collect more negative sentiment data to fix severe class imbalance (only 6.4% negative tweets).
- Apply class balancing techniques such as SMOTE or class weighting to improve minority class detection.
- Enhance feature engineering with emojis, punctuation, capitalization, and better embeddings to reduce confusion between Negative and Neutral tweets.
- Compare Apple vs Google sentiment directly to identify brand-specific perception differences.
- Track sentiment during product launches to measure campaign success or detect rising issues.
- Extract high-impact keywords to guide marketing language, ad copy, and content creation.
- Prioritize negative tweets using a sentiment-based routing system for faster response.
- Use a two-stage system: automatic model screening + human review for borderline cases to improve accuracy.
- Analyze sentiment by product features to pinpoint what users love or dislike.
- - Identify recurring negative themes(bugs, complaints, frustrations) for roadmap and quality improvements.
- Create brand-specific sentiment models for Apple and Google to capture unique vocabulary and patterns.
- Build real-time dashboards showing sentiment trends, alerts, and activity for executives and marketing teams.

ANY QUESTIONS?