DATA REPORT

Social Media Sentiment Analysis for Apple and Google Products

Executive Summary

This report details the process and findings of a sentiment analysis project focused on understanding public perception of Apple and Google products based on Twitter data. The primary challenge was the significant imbalance in sentiment classes, particularly the scarcity of negative tweets. Through iterative modeling, starting with baseline machine learning models and progressing to advanced deep learning techniques, the project successfully developed a robust classifier. The final champion model, a fine-tuned **RoBERTa**, demonstrated strong performance, notably achieving **60% recall** for the challenging 'Negative emotion' class, thereby addressing the core business need to identify critical feedback.

Problem Statement

Social media users frequently express opinions and emotions about tech products on platforms like Twitter. With thousands of daily tweets, manual sentiment assessment is impractical. This project aimed to automate sentiment classification to determine whether a tweet conveys positive, negative, or neutral feelings toward Apple and Google products.

Business Objectives

1. Discover sentiment patterns associated with specific products/brands.
2. Develop an automated system to classify public sentiment whether Positive, Negative or Neutral.
3. Identify potential brand crises or opportunities from sentiment trends.

4. Build and iterate on models, starting with a baseline such as Logistic Regression, and moving to advanced NLP such as RoBERTa.

Data Overview

The dataset contains about 9,000 labeled tweets with three main columns:

1. `*tweet_text*` – raw content of the tweet
2. `*emotion_in_tweet_is_directed_at*` – the product or brand referenced.
3. `*is_there_an_emotion_directed_at_a_brand_or_product*` – sentiment label i.e. positive, negative, neutral.

The dataset initially contained missing values especially in `*emotion_in_tweet_is_directed_at*`, duplicates, and significant class imbalance.

The data was cleaned in the following ways:

1. The duplicates were removed.
2. The irrelevant rows were dropped.
3. The missing values were handled.
4. The text was standardized by removing white spaces and lower casing them.
5. Noise was removed, such as hashtags, URLs, mentions, emojis and punctuations.
6. Standardizing the data i.e. grouping specific names into the simpler parent categories.

Tweets were processed and the preprocessing steps included tokenization, lemmatization, stopword removal, and standardization of the `tweet_text` column.

Methodology

The analysis followed a structured, iterative approach:

1. **Data Preparation:** Cleaned raw tweets (removed URLs, mentions, hashtags, punctuation), standardized text (lowercase, whitespace removal), handled null values, grouped specific product names into broader brand categories (Apple, Google), and renamed columns. Created a *`brand_classification`* feature based on keywords.

2. **Exploratory Data Analysis (EDA):** Analyzed data distributions (univariate analysis), relationships between variables (bivariate analysis), and interactions among multiple variables (multivariate analysis) using counts, plots such as bar charts, heatmaps, boxplots, as well as word clouds.

3. **Hypothesis Testing:** Conducted Chi-Square, t-tests, and ANOVA to statistically validate observed differences in sentiment distributions between brands and tweet lengths across sentiments.

4. **Text Preprocessing:** Tokenized, lemmatized, and removed stopwords while retaining sentiment-specific words from the standardized tweets, to create the *`processed_tweet`* feature for modeling.

5. **Feature Engineering:** Used TF-IDF vectorization with unigrams and bigrams on the `processed_tweet` data to create numerical features for

machine learning models. Later, GloVe word embeddings were also tested.

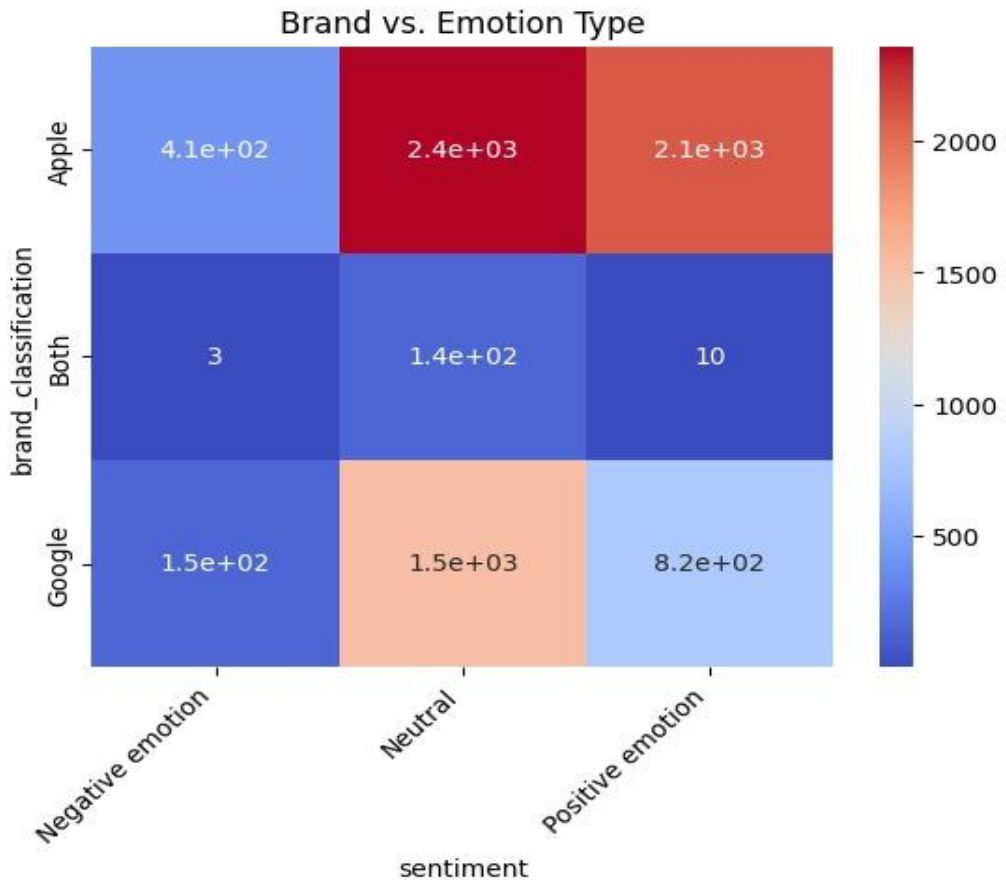
6. **Modeling and Evaluation:**

- **Baseline (Multiclass):** Trained classic ML models; Logistic Regression, SVM, XGBoost, Random Forest and Naive Bayes on the 3-class problem i.e. Positive, Negative or Neutral, using class weighting to handle imbalance. Evaluated using classification reports and Macro F1-score.
- **Binary Classification:** Simplified the problem to Positive vs. Negative (removing Neutral) to establish a better baseline and identify a champion architecture. Models were re-trained and evaluated.
- ·**Hyperparameter Tuning:** Used GridSearchCV on the top 3 binary models i.e. Logistic Regression, SVMand Random Forest to optimize performance.
- **Deep Learning (Multiclass):** Fine-tuned pre-trained Transformer models i.e. DistilBERT and RoBERTa on the original 3-class problem, incorporating weighted loss for imbalance.

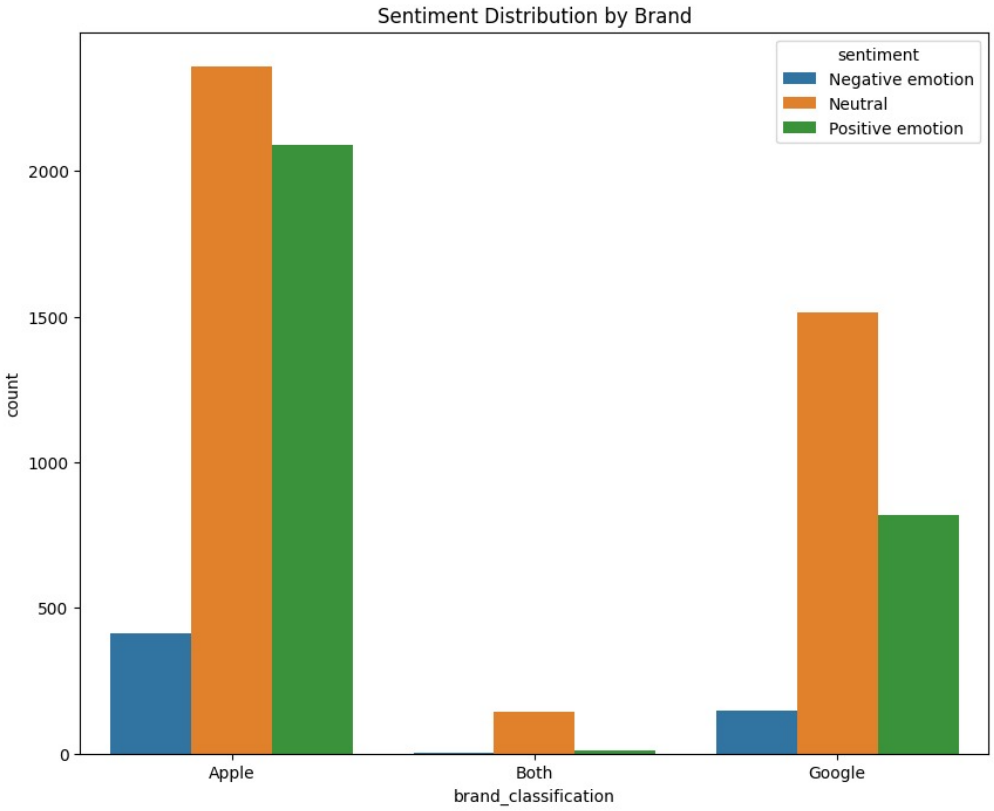
7. **Model Explainability:** Used LIME (Local Interpretable Model-agnostic Explanations) to understand the word contributions for the final champion model's predictions on specific examples.

Exploratory Data Analysis (EDA) Highlights

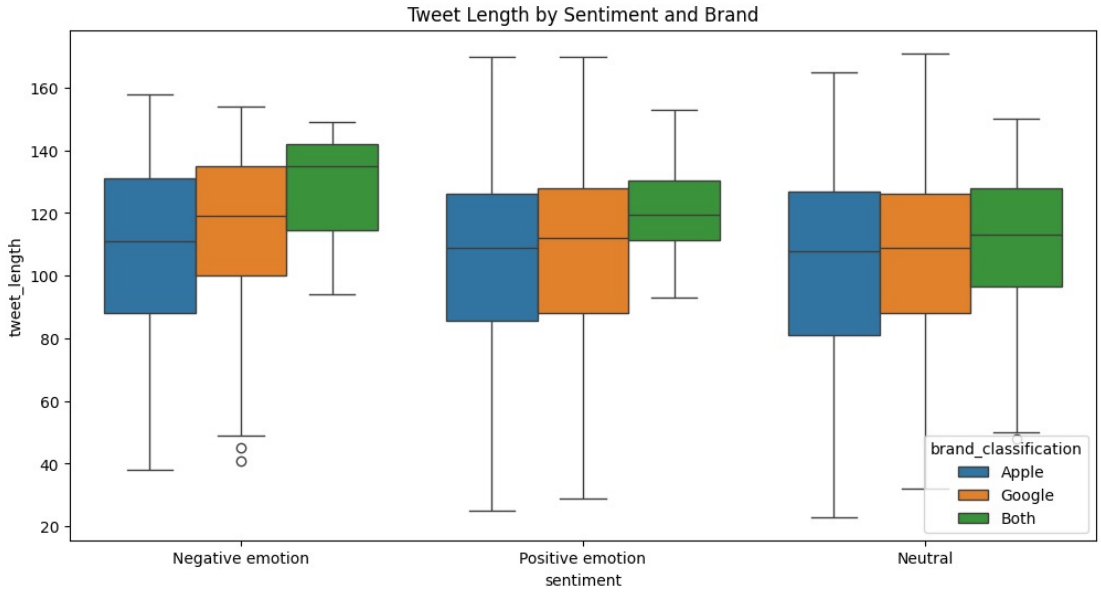
- **Sentiment Imbalance:** "Neutral" sentiment heavily dominated the dataset, followed by "Positive", with "Negative" being the least frequent class.
- **Brand Imbalance:** Apple was mentioned roughly twice as often as Google.



- **Brand Sentiment Profile:** Apple generated more emotional responses, both positive and negative, compared to Google, whose mentions were predominantly neutral.



- **Tweet Length:** Tweets mentioning *both* brands tended to be longer, especially negative ones. Negative tweets, in general, showed slightly longer median lengths than positive or neutral ones.



- **Word Clouds:** Revealed key terms associated with sentiments such as 'great', 'good', 'like' for positive; 'bad', 'hate' for negative, and intensity words such as 'too', 'very', 'really'.

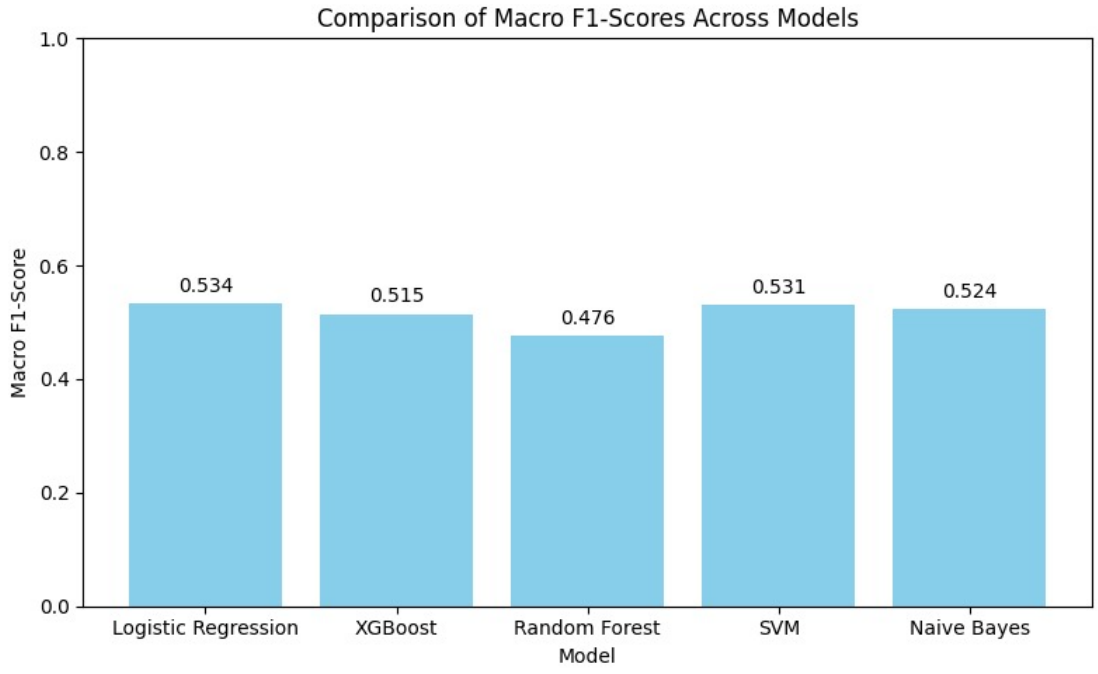
Hypothesis Testing Results

- **Brand vs. Sentiment (Chi-Square):** Rejected the null hypothesis, confirming a statistically significant difference in sentiment distribution between Apple and Google tweets.
- **Tweet Length vs. Sentiment (t-test & ANOVA):** Rejected the null hypothesis, indicating a significant difference in mean tweet length between positive and negative tweets, and across positive, negative, and neutral groups.

Modeling Results

1. **Baseline Multiclass Models:**

- Suffered significantly due to class imbalance, particularly showing poor recall and F1-scores for the "Negative emotion" class.
- **Best Baseline:** Logistic Regression achieved a Macro F1-Score of approximately **0.53**.



2. **Binary Classification Models (Positive vs. Negative):**

- Performance dramatically improved after removing the neutral class.
- **Champion Architecture:** Tuned **Logistic Regression** achieved the best Macro F1-Score of **0.72**. Tuned SVM also performed well (0.70).

3. **Deep Learning Models (Multiclass):**

- **DistilBERT:** Showed improvement over baselines, reaching a Negative recall of **0.53**.
- **RoBERTa (Final Champion):** As a model pre-trained specifically on Twitter data, it achieved the best performance on the challenging 3-class problem, notably reaching **0.60 recall** for "Negative emotion" and a strong overall **Macro F1-Score of 0.70**.

Model Explainability using (LIME)

- LIME was applied to the champion **RoBERTa** model.
- Analysis on a sample negative tweet correctly showed that the model focused on the sentiment-bearing word **"headache"** as the primary reason for its negative prediction, confirming its ability to identify relevant terms.

Conclusions

1. The project successfully created an effective sentiment classifier for tech brand tweets, even with an uneven distribution of sentiments in the data.
2. Traditional machine learning models found it difficult to identify the rare negative tweets accurately.
3. Simplifying the task to just positive vs. negative classification confirmed the project's approach was viable.
4. Fine-tuning a specialized Twitter-RoBERTa model produced the best results for the original three-category classification i.e. positive, negative, neutral.

   The model achieved 60% recall for negative sentiment, meaning it correctly identified 60% of all negative tweets. Business Need Addressed: This significantly helps the business identify and respond to critical customer feedback on social media.

Recommendations

1. **Deploy RoBERTa Model:** Implement the fine-tuned Twitter-RoBERTa model via an API called Hugging Face for real-time sentiment monitoring. Prioritize alerts and monitoring for **Apple-related tweets** due to their higher emotional engagement volume.
2. **Human-in-the-Loop Workflow:** Integrate the model's output, especially negative predictions, into a dashboard for human review e.g., customer support, to verify sentiment and enable timely customer engagement.
3. **Continuous Improvement Pipeline:** Establish a system for periodic model re-training on new Twitter data to prevent performance degradation (model drift). Consider incorporating tweet length as an additional feature in future iterations.