NLP SENTIMENT ANALYSIS OF APPLE VS GOOGLE TWEETS

TEAM MEMBERS

- Victor Onyango
- Godfrey Onsundwa
- Ray Onsongo
- Hellen Kavaya
- Elizabeth Kiilu

EXECUTIVE SUMMARY

This project analyzed 9,093 tweets about Apple and Google to understand how people feel about the two brands.

A sentiment classification system was built to label tweets as positive, negative, or neutral.

The goal is to help Apple's Marketing & Product teams understand real-time customer emotions.

The most accurate model was DistilBERT, which performed best in detecting subtle language and context.

BUSINESS PROBLEM

Apple needs to monitor customer sentiment in real time because:

- Negative reactions spread quickly online
- Public perception impacts brand reputation and product adoption
- Comparing Apple vs Google sentiment helps understand competitive positioning
- Early detection of issues helps prevent larger problems

PROJECT OBJECTIVES

This project aims to build an NLP-based sentiment analysis model capable of automatically classifying Apple- and Google-related tweets as positive, negative, or neutral. The solution will help Apple quantify customer emotions, detect shifts in public perception, and benchmark Apple's reputation relative to Google in fast-evolving online conversations.

STAKEHOLDERS

Primary Stakeholders:

- Apple Marketing Team
- Product & Engineering Teams

Secondary Stakeholders:

- Customer Support Teams
- PR & Communications
- Market Research & Competitive Intelligence

DATA PREPARATION

DATA PROCESSING

Loading data

Cleaning and processing which included

- Lowering
- Removing URLs
- Removing Hashtags and @ emotions
- Removing Punctuations
- Removing Stop words
- Tokenization
- Lemmatization

NLP FEATURE ENGINEERING

- Tweet length
- Number of Hashtags
- Emoji Count
- VADER Sentiment Scores

MODELING PPROACH

ADVANCED MODELS FOR MULTICLAS NLP

- Multinomial Naive Bayes
- Linear SVM
- Improved Logistic Regression
- Apply SMOTE to handle class imbalance
- Richer Features
- Strong Model Upgrade :Mini-Bert

CLASSICAL MACHINE LEARNING ODELS

- Logistic Regression
- Class-Balanced
- Linear SVM

Logistic Regression

Accuracy: 0.6891133557800224
F1-score (weighted): 0.6654851743632886

Classification Report:

	precision	recall	f1-score	support
0.0	0.54	0.06	0.11	114
1.0	0.63	0.54	0.58	594
2.0	0.71	0.84	0.77	1074
accuracy			0.69	1782
macro avg	0.63	0.48	0.49	1782
weighted avg	0.67	0.69	0.67	1782

The model achieved an overall accuracy of appr 69% and a weighted F1-score of 0.66, performing best on the majority (neutral) class. It struggles with the minority negative class, which has very low recall, indicating many negative tweets are misclassified. Positive tweets are moderately well-predicted, but class imbalance likely limits overall performance.

Class-Balanced Logistic Regression

Accuracy: 0.6447811447811448

Classification Report:

	precision	recall	f1-score	support
0.0	0.33	0.59	0.42	114
1.0	0.57	0.66	0.61	594
2.0	0.78	0.64	0.70	1074
accuracy			0.64	1782
macro avg	0.56	0.63	0.58	1782
weighted avg	0.68	0.64	0.65	1782

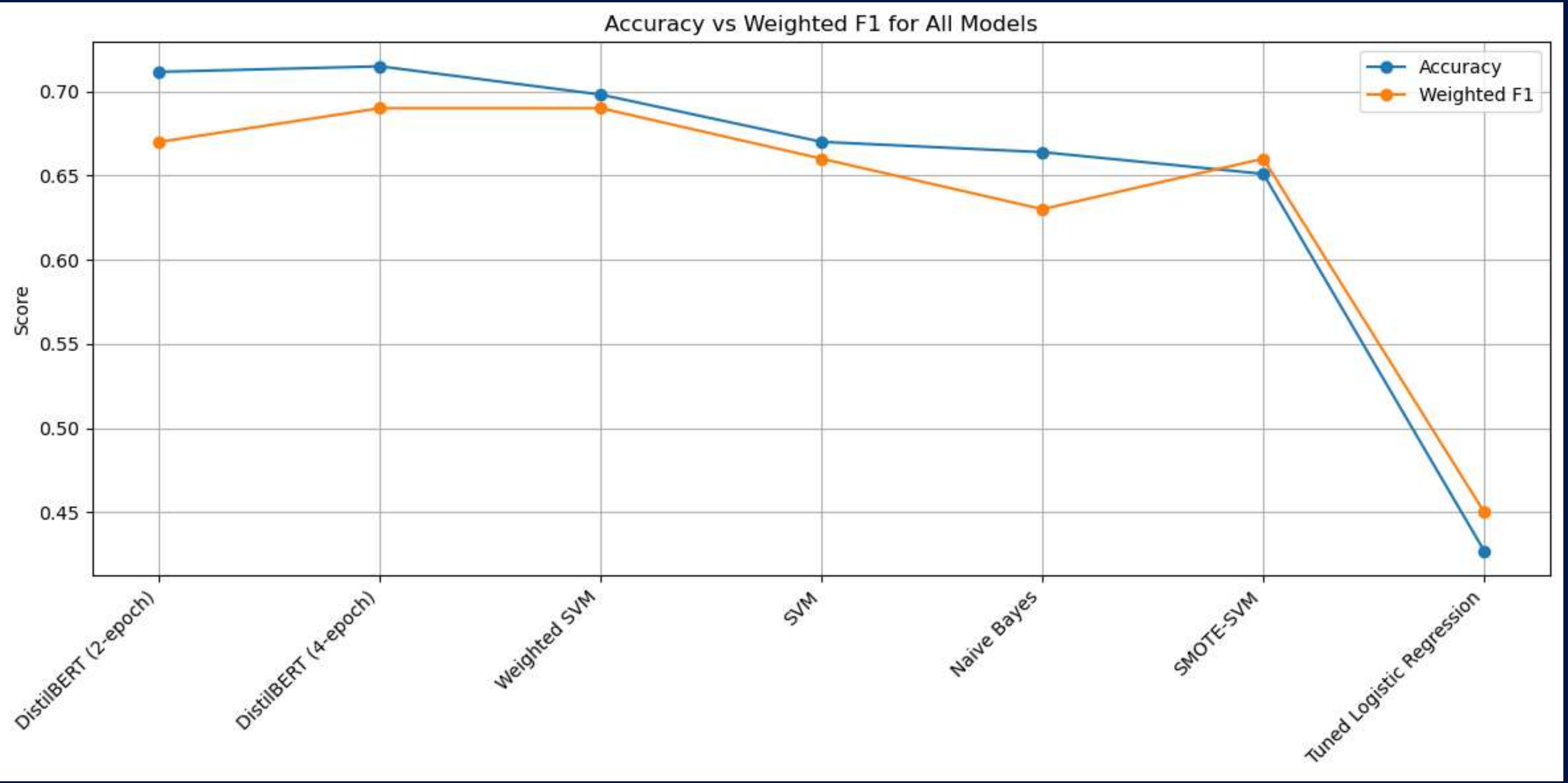
1. Accuracy & F1 Accuracy appr 64%: dropped slightly compared to the previous baseline (appr 69%).
2. Weighted F1 0.65: similar to before.Minority Class Performance (Negative, 0.0)

Precision: 0.30, Recall: 0.58, F1: 0.40

Recall improved a lot (was 0.07 before), meaning the model is now detecting more negative tweets.

Precision is lower, meaning some predicted negatives are actually positives or neutral.

MODEL EVALUATION



Across all experiments, transformer-based models (DistilBERT) consistently outperformed traditional machine-learning approaches. The 4-epoch DistilBERT model achieved the highest performance, with:

- Test Accuracy: 0.7149
- Strong loss reduction over epochs (0.78 to 0.34)
- Excellent stability and generalization on unseen data

This confirms the advantage of pretrained contextual embeddings for sentiment tasks, especially when dealing with the informal, context-rich nature of Twitter language

RECOMENDATIONS TO THE STAKEHOLDERS

- Deploy the BERT Model to monitor real time cutomer Sentiment
- Prioritize Investigation of Negative Sentiment Mentions
- Break down Sentiment by product Category
- Use Sentiment Insights to Enhance Customer Engagement Campaigns
- Continuously Retrain the model with New Tweet Data

LIMITATIONS

Key challenges of the sentiment analysis model:

- Neutral tweets are unclear → hard to classify
- Short, informal language → slang & emojis reduce meaning
- Sarcasm is hard to detect
- Model needs frequent updates to stay current
- Few negative tweets → weak accuracy for negative sentiment

FUTURE WORK

To improve and extend the system, future work will focus on:

- Enhancing the model by further fine-tuning transformer models for better handling of neutral and sarcastic tweets.
- Expanding the dataset with newer Apple and Google tweets, especially after product launches.
- Tracking sentiment over time to detect trends before and after major Apple events.
- Improving Apple-specific language understanding using a domain-focused vocabulary.
- Building a real-time dashboard to visualize live positive, neutral, and negative sentiment for faster decision-making.

THANK YOU