NLP Sentiment Analysis of Apple vs Google Tweets

GROUP 4 PHASE 4 PROJECT

Executive Summary

SUMMARY

- This project builds a sentiment classification system to analyze public emotions in Apple-related tweets. Using a dataset of 9,093 tweets labeled as positive, negative, or neutral, the goal is to help Apple's Marketing & Product teams better understand customer perception and identify emerging issues in real time. Tweets are short, informal, and often noisy, making NLP-based machine learning an appropriate solution.

Stakeholder

Primary Stakeholders

- Apple Marketing Team Uses real-time sentiment insights to understand public perception, monitor brand reputation, evaluate campaign performance, and respond quickly to public feedback after product launches or events.
- Apple Product Management & Engineering Teams Gain visibility into customer frustrations, feature requests, recurring complaints, and performance issues (e.g., crashes, battery problems). This helps prioritize product improvements and software patches.

Secondary Stakeholders

- Apple Customer Support & Care Teams Benefit from early detection of negative sentiment spikes related to device malfunctions or software bugs. They can proactively prepare responses, FAQs, or issue statements.
- Apple PR & Corporate Communications Use sentiment monitoring to manage crises, track public reaction during controversies, and craft timely communication strategies compared to competitors like Google.
- Competitive Intelligence & Market Research Teams Analyze Apple vs Google sentiment trends to understand market positioning, customer loyalty drivers, and areas where Apple is outperforming or lagging behind.

Business problem

Apple's Marketing & Product Intelligence teams require a system that can continuously monitor and compare public sentiment toward Apple products against competing brands—particularly Google. With social media platforms like Twitter influencing customer perception in real time, Apple needs early visibility into negative trends, emerging complaints, and user frustrations before they scale. Likewise, understanding which Apple products generate the strongest positive engagement—compared to similar Google offerings—provides insights for campaign optimization, product refinement, and competitive positioning. 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.

Data Preparation

Data Preprocessing

- **Loading data**
- **Cleaning & preprocessing which included;**
- Lowercasing
- Removing URLs
- Removing hashtags and @mentions
- Removing punctuation
- Removing stopwords
- Tokenization
- Lemmatization

NLP Feature Engineering

- Tweet length
- Number of hashtags
- Emoji count
- VADER sentiment scores

Modeling Approach

- **Classical Machine Learning Models:**
- Logistic Regression
- Class-Balanced Logistic Regression
- Linear SVM
- **Advanced Classical Models for Multiclass NLP**
- Multinomial Naive Bayes
- Linear SVM
- Improved Logistic Regression (Tuned)
- Apply SMOTE to handle class imbalance
- Richer Features
- Strong Model Upgrade: Mini-BERT

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

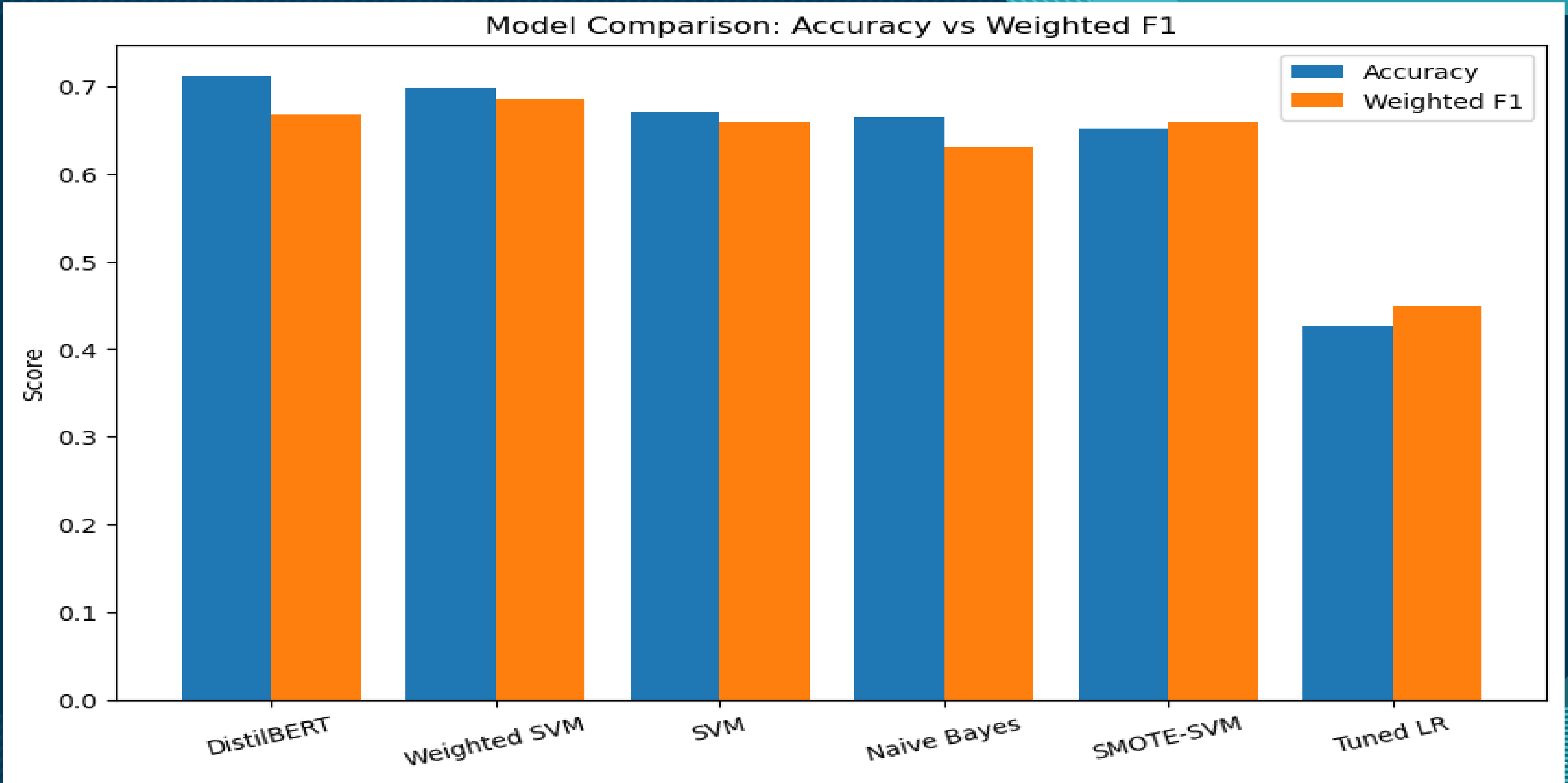
Class-Balanced Logistic Regression

- Accuracy: 0.6447811447811448

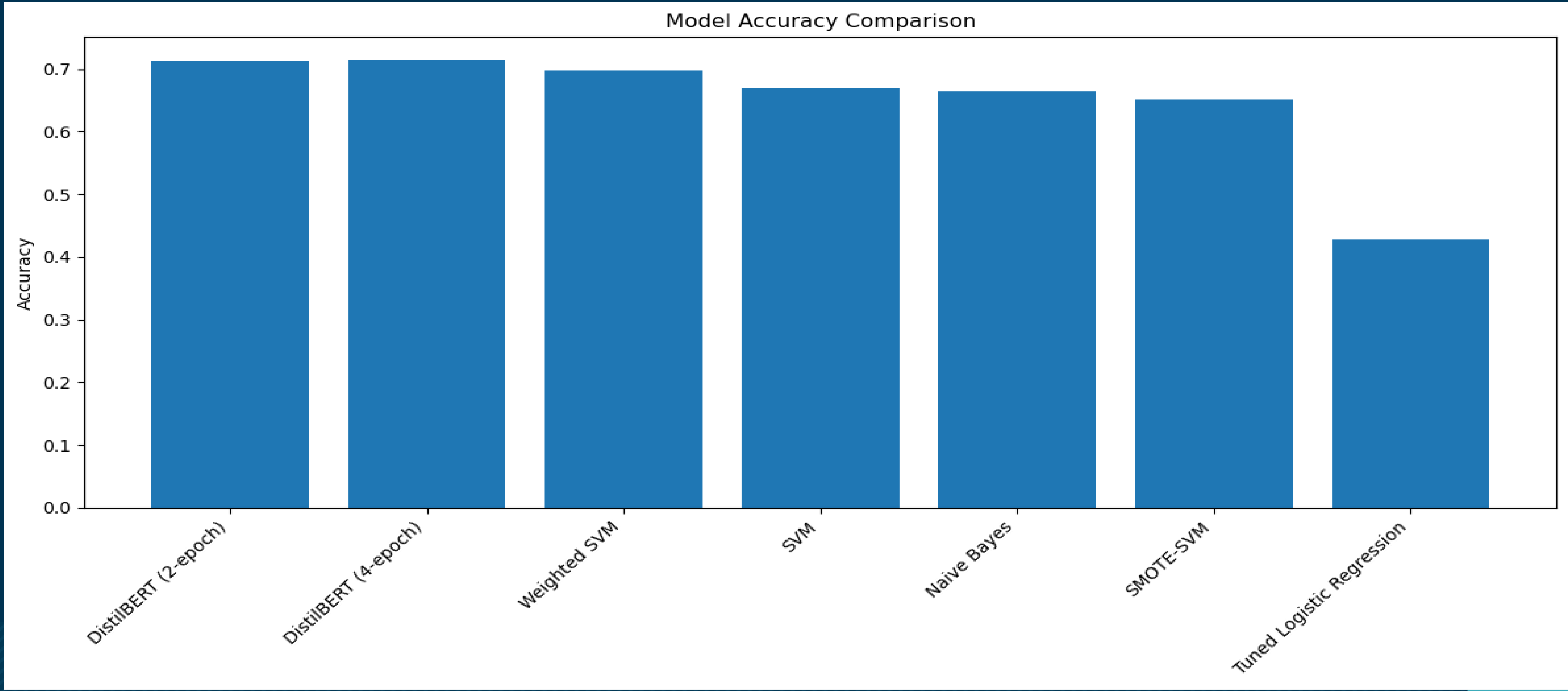
Classification Report:

	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

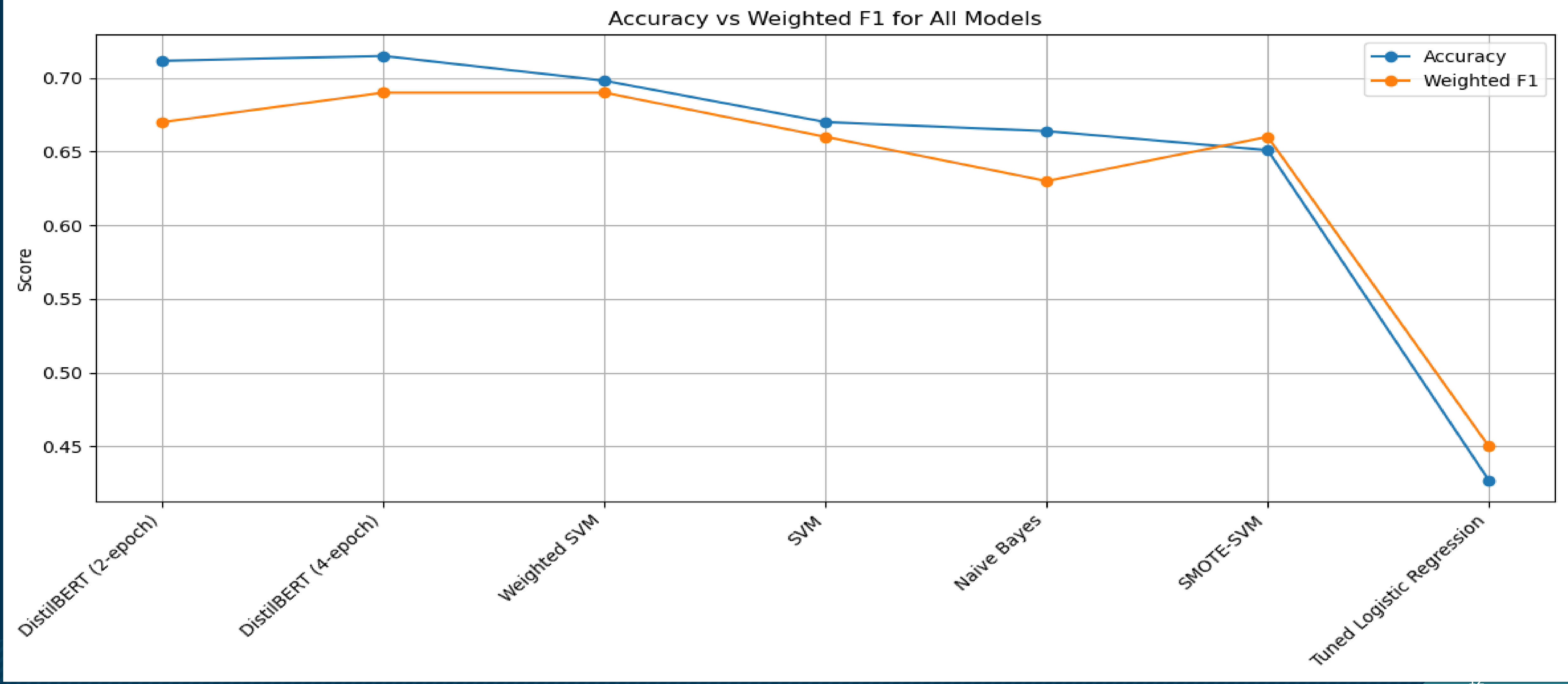
Model Comparison: Accuracy vs Weighted F1



Model Evaluation



Model Evaluation



Overall Model Performance

- 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.
- Following DistilBERT, the Weighted SVM model emerged as the strongest classical approach, achieving:
- Accuracy: 0.6981
- Competitive Weighted F1: 0.69
- Improved recall on the minority class without overly sacrificing the majority class
- This demonstrates that class-weighted optimization effectively compensates for label imbalance, a core challenge of real-world sentiment datasets.

Comparative Strengths and Weaknesses

- Models such as Naive Bayes and SVM provided respectable baselines, but their inability to capture semantic nuance limited performance. Logistic
- Regression significantly underperformed due to linear decision boundaries that cannot adequately model complex emotional expressions.
- The SMOTE-SVM model successfully improved minority-class recall but at the cost of overall accuracy, showing that synthetic oversampling must be applied carefully to avoid feature-space distortion.
- The transformer-based DistilBERT models, however, demonstrated:
- Superior semantic understanding
- Strong contextual sensitivity
- Better performance on ambiguous or indirect sentiment expressions
- Minimal preprocessing requirements. These advantages were especially evident in class 2 (positive sentiment), where BERT-based models achieved industry-standard recall levels.

Recommendations to Stakeholder

- Deploy the BERT Model to Monitor Real-Time Customer 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

Future Work

- Fine-tune a transformer model further. Increasing training epochs, expanding the dataset, and adjusting learning rates can significantly improve BERT-based performance—especially for neutral and sarcastic tweets.

- Collect more and fresher Apple-specific tweets. Expanding the dataset with recent product launches (iPhone, Mac, iPad, Vision Pro) will help the model stay current with evolving customer language and sentiment trends.

- Implement time-series sentiment tracking. Monitoring sentiment over weeks, months, or around product events (e.g., WWDC, product launches) would allow Apple to detect shifts earlier and respond quickly.

- Domain-specific lexicon enhancement. Incorporating Apple-related terms (e.g., “FaceID,” “AirDrop,” “iOS update,” “battery health”) can improve contextual understanding and reduce misclassifications.

- Build a real-time dashboard. A live system visualizing positive, neutral, and negative sentiment would give the marketing team actionable insights during major announcements or crises.

Thank you