TWITTER SENTIMENT ANALYSIS

Presentation by the Mini-Avengers(Group 6)

BUSINESS UNDERSTANDING

Dataset:	Tweets mentioning Apple and Google products
Goal:	Understand public sentiment (positive, negative, neutral) toward these brands
Problem Statement:	Volume of tweets too large for manual analysis, need automated sentiment classification
Business objectives:	Develop an automated and effective sentiment analysis system to identify sentiment patterns, detect brand opportunities or crises, and improve brand communication through advanced NLP techniques
Metrics of success:	Achieve actionable and measurable insights into public sentiment that enhance targeted marketing, early risk detection, campaign evaluation, and competitive brand benchmarking.
	T

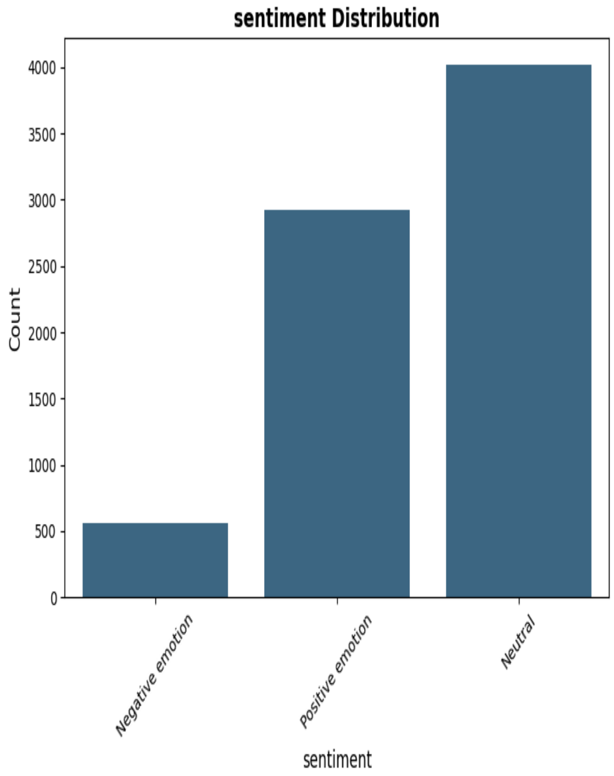
TABLE OF CONTENTS

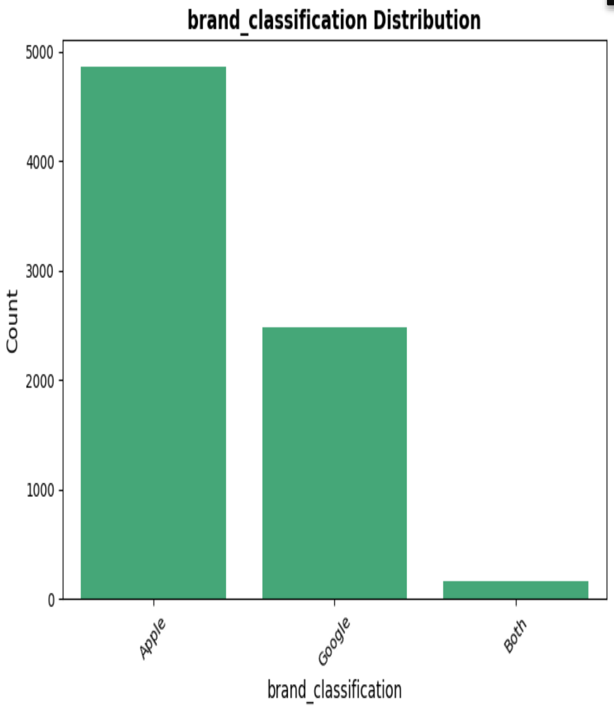
Stacy Mogeni : Team leader
Jeff Mogaka : Member
Ann Felicity : Member
Kitts Kikumu : Member
Fridah Njung'e : Member

EXPLORATORY DATA ANALYSIS

UNIVARIATE ANALYSIS

BRAND AND EMOTION DISTRIBUTION SUMMARY



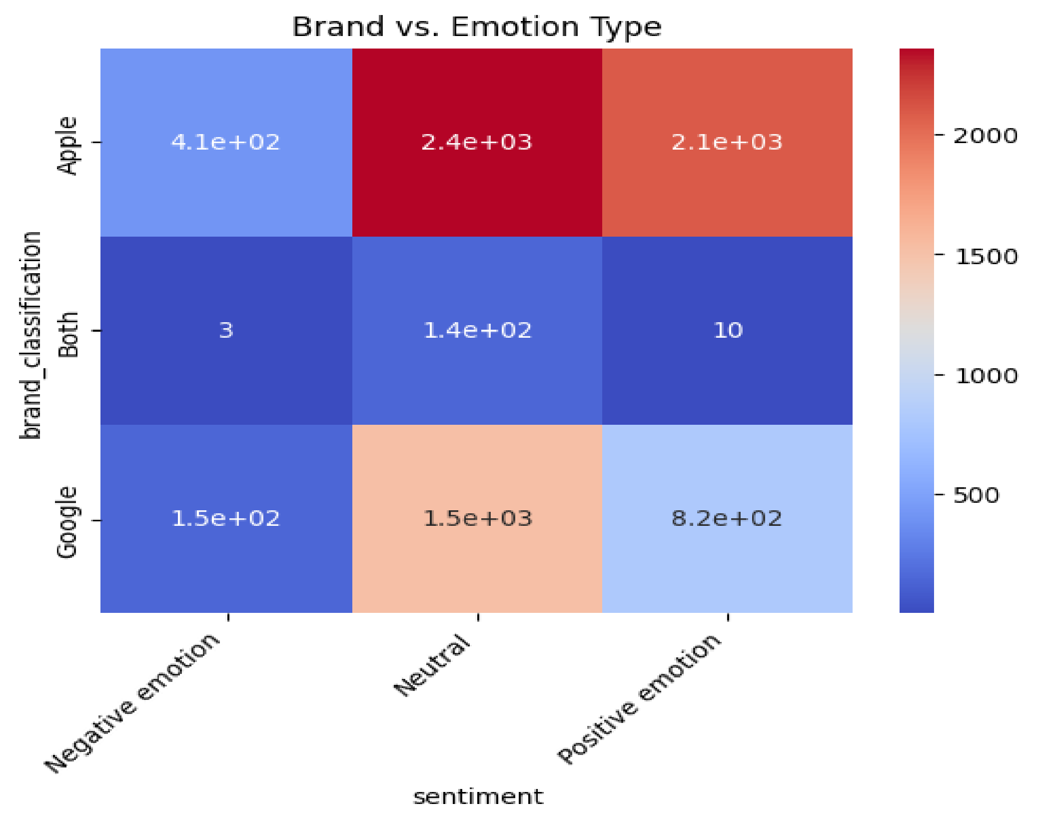


OBSERVATIONS

- The sentiment in tweets is largely neutral, with positive emotion being significantly stronger than negative emotion.
- Apple dominates the conversation, appearing about twice as often as Google in the brand mentions.
- The analysis shows focused brand attention, with Apple and Google accounting for the vast majority of emotional expressions detected.

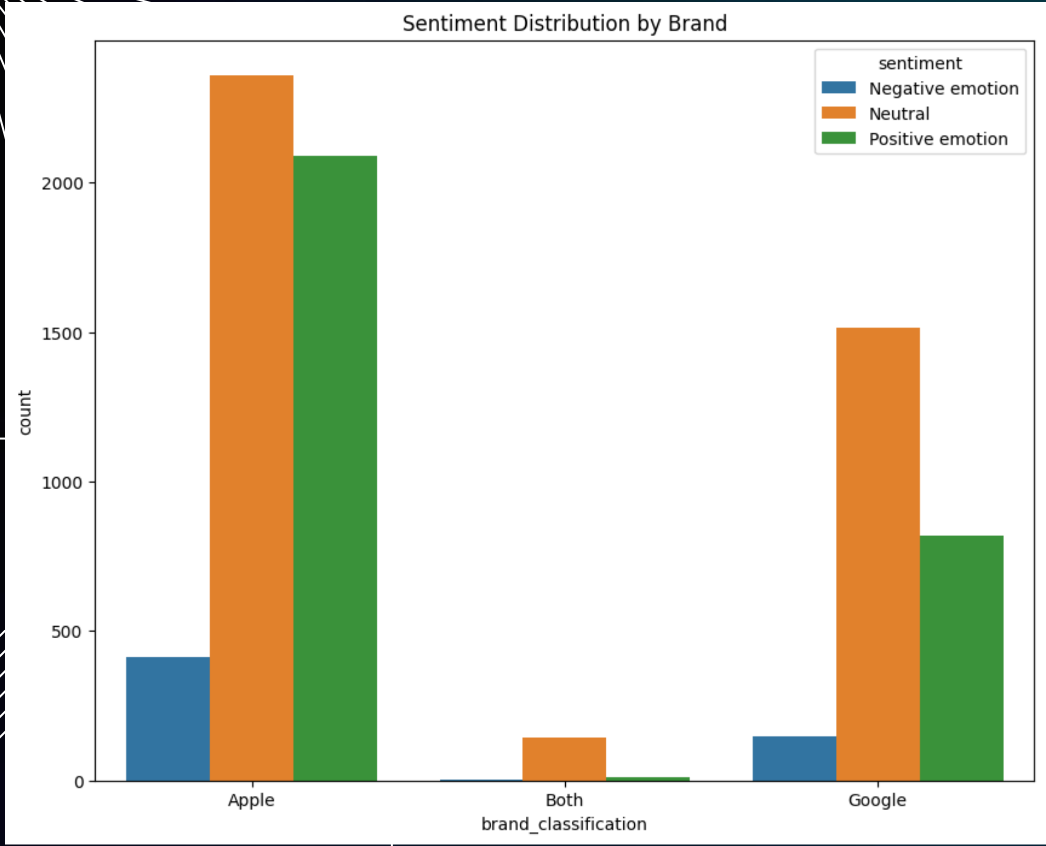
BIVARIATE ANALYSIS

BRAND-SPECIFIC SENTIMENT ANALYSIS



- Neutral sentiment is the most common for both Apple and Google, with Apple receiving more mentions overall, especially in neutral and positive categories.
- Positive sentiment for both brands far exceeds negative sentiment, with positive tweets outnumbering negatives by approximately five to one.
- Mentions of both brands together are rare and predominantly neutral, reflecting minimal direct comparisons within the dataset.

BRAND-SPECIFIC SENTIMENT ANALYSIS



Sentiment distribution by brand Summary

- ✓ Apple receives about twice as many mentions as Google, with both brands dominated by neutral sentiment; Apple's share of positive emotion is notably high, while negative sentiment is rare for both brands.
- ✓ Apple discussions are more emotionally charged and generate higher engagement, whereas Google conversations tend to be more neutral and fact-based.
- ✓ Mentions of both brands together are minimal, indicating few direct brand-to-brand comparisons in the dataset.

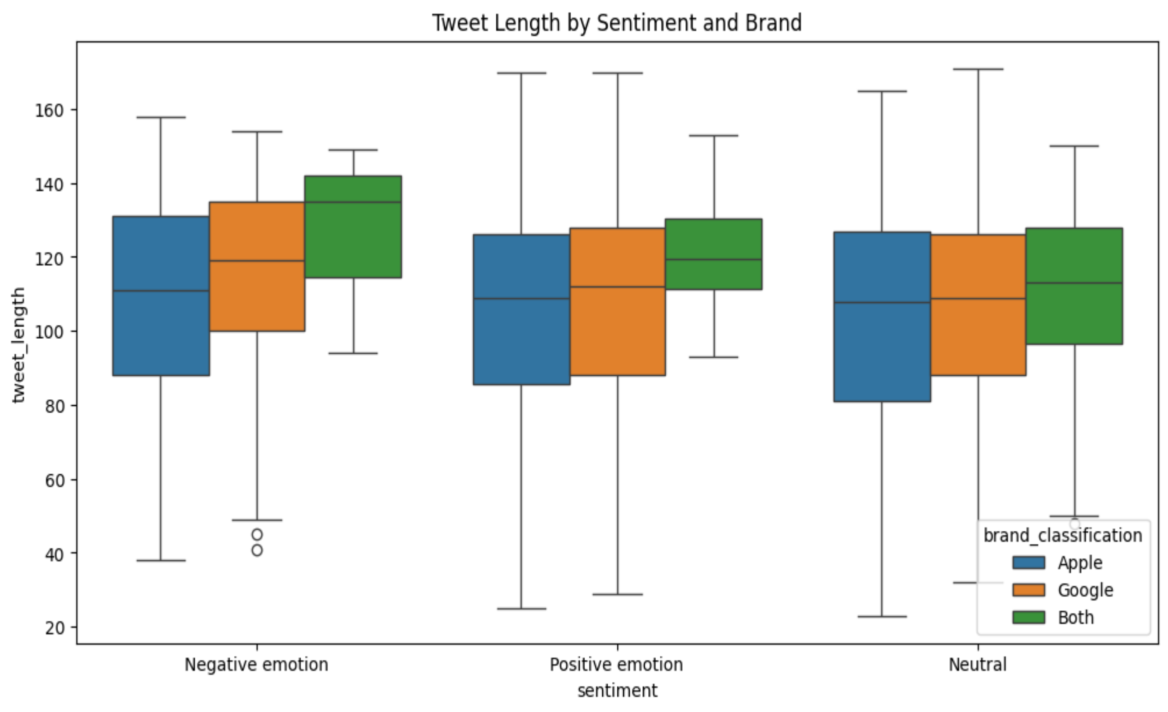
MULTIVARIATE ANALYSIS

TWEET LENGTH ANALYSIS

- Median tweet lengths are similar for Apple and Google, but tweets mentioning both brands are consistently longer, particularly for negative sentiment.

- Negative tweets about both brands tend to be the longest, suggesting that complaints or comparisons require more explanation.

- Overall, tweet length does not distinguish between Apple and Google, except when both are mentioned together, which prompts more detailed commentary.



HYPOTHESIS TESTING

Hypothesis Test Description	P-value	Conclusion
Difference in sentiment distribution between Apple and Google tweets	< 0.05 (CHI-square)	Significant difference; reject null hypothesis
Difference in mean tweet length between positive and negative tweets	0.009 (T-Test)	Significant difference; reject null hypothesis
Difference in mean tweet length among sentiment groups (positive, negative, neutral)	~0.005 (ANOVA test)	Significant difference; reject null hypothesis

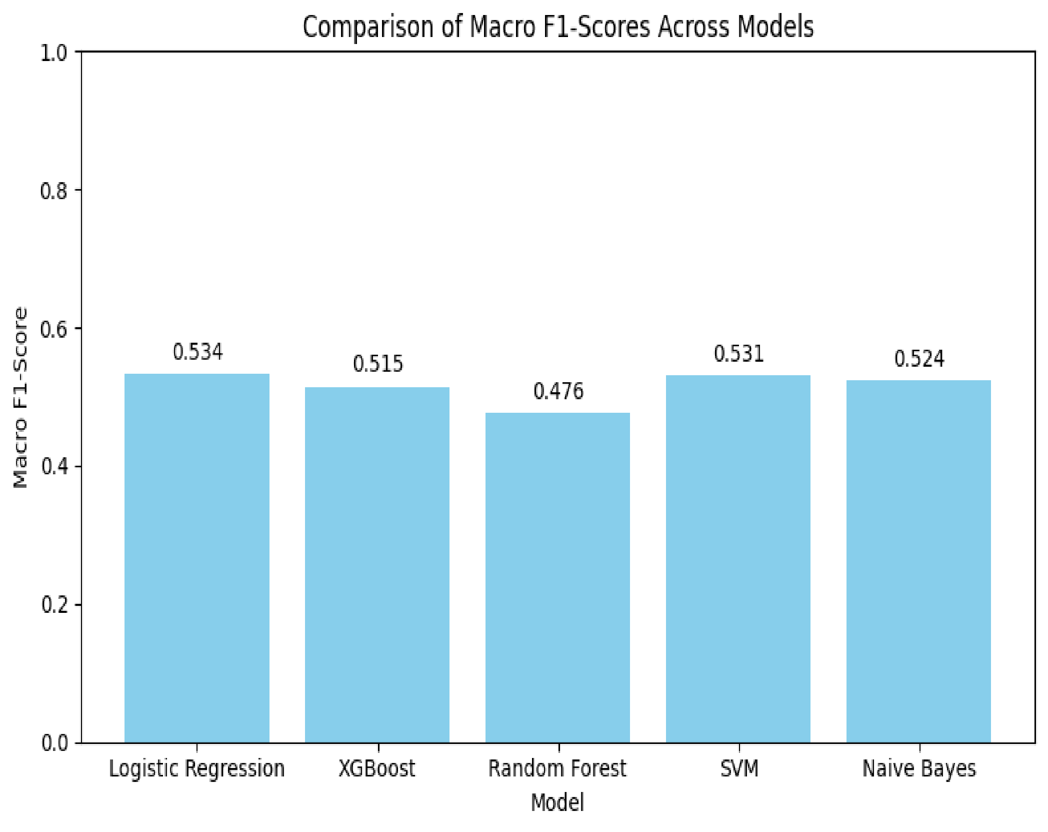
MODELING

MULTICLASS MODELLING

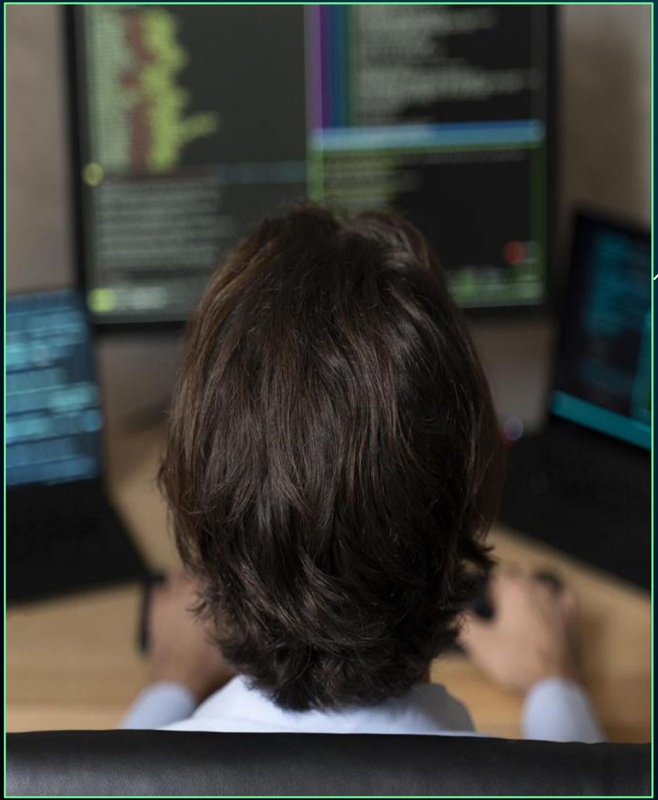
MACRO F1-SCORE COMPARISON ACROSS MODELS

- The baseline sentiment classification models struggled due to severe class imbalance, achieving only about 52% accuracy and a Macro F1-Score of 0.53 at best.

- All models performed very poorly at detecting negative sentiment, confirming that distinguishing less frequent classes in this three-way classification problem remains a significant challenge.



BINARY CLASSIFICATION



Model	Accuracy	Precision (Neg)	Recall (Neg)	F1-Score (Neg)	F1-Score (Pos)	Macro Avg F1-Score
SVM (Linear)	0.83	0.50	0.58	0.54	0.90	0.72
Logistic Regression	0.84	0.52	0.53	0.53	0.90	0.71
XGBoost	0.84	0.52	0.53	0.52	0.90	0.71
Random Forest	0.79	0.40	0.56	0.47	0.87	0.67
Naive Bayes (CNB)	0.85	0.60	0.26	0.36	0.91	0.64

- The SVM and Logistic Regression models show the most balanced performance, achieving the highest macro average F1-scores around 0.71–0.72 with strong recall on negative sentiment.
- While Naive Bayes has high precision for negative class, its low recall reduces overall effectiveness.
- Random Forest lags behind across most metrics, and XGBoost performs similarly to Logistic Regression but with slightly lower negative F1, indicating a need to improve recall for minority negative classes across all models.

SUMMARY OF TOP 3 BEST TUNED MODELS

Model	Accuracy	Recall (Negative)	F1-Score (Negative)	Macro Avg F1-Score
Tuned Logistic Regression	0.82	0.66	0.56	0.72
Tuned SVM (Linear)	0.82	0.55	0.51	0.70
Tuned Random Forest	0.78	0.67	0.51	0.68

. 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

DEPLOYMENT

1. The deployed SENTINEL AI app uses a fully trained Twitter-RoBERTa model to analyze tweet sentiment instantly.

2. Users input any tweet text, which the model processes by understanding Twitter-specific language nuances, including slang and context.

3. The model outputs a sentiment prediction—Positive, Negative, or Neutral—along with confidence scores indicating certainty, providing fast, expert-level analysis without the overhead of training during use.

Important Note:

While our model provides a significant advantage in analyzing sentiment, it doesn't classify every tweet perfectly.

Be aware that some results may be incorrect; an example is slide 24.

Further improvements in accuracy can be achieved through regular retraining of the model with more diverse data.

The next four slides contain some of the results of the deployed model:

localhost:8501

Deploy

SENTINEL AI

A Twitter Sentiment Analyzer powered by RoBERTa

Enter a tweet to analyze:

Is the MacBook Air M2 keyboard clicky enough, or is it still mushy? I need tactile feedback.

Press Ctrl+Enter to apply

Analyze Sentiment

Analysis Result

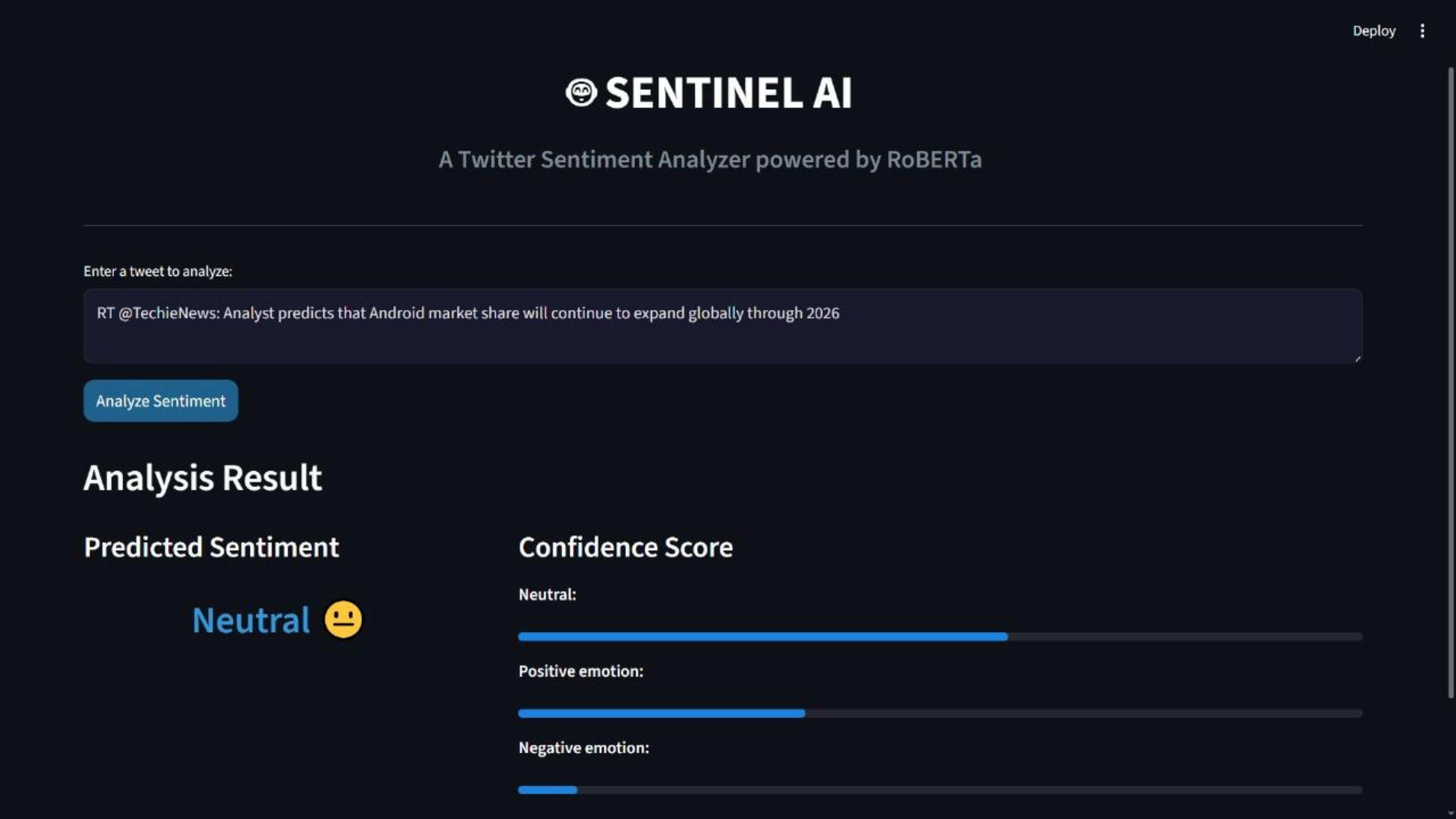
Predicted Sentiment

Positive emotion 👍

Confidence Score

Positive emotion:

Neutral:


Deploy
SENTINEL AI
A Twitter Sentiment Analyzer powered by RoBERTa
Enter a tweet to analyze:
RT @TechieNews: Analyst predicts that Android market share will continue to expand globally through 2026
Analyze Sentiment
Analysis Result
Predicted Sentiment
Neutral
Confidence Score
Neutral:
Positive emotion:
Negative emotion:

Deploy

SENTINEL AI

A Twitter Sentiment Analyzer powered by RoBERTa

Enter a tweet to analyze:

Just saw a huge drop in GPU prices. Might finally upgrade my PC graphics card!

Analyze Sentiment

Analysis Result

Predicted Sentiment

Positive emotion 👍

Confidence Score

Positive emotion:

Neutral:

Negative emotion:

Deploy

😊 SENTINEL AI

A Twitter Sentiment Analyzer powered by RoBERTa

Enter a tweet to analyze:

Is the MacBook Air M2 keyboard clicky enough, or is it still mushy? I need tactile feedback.

Analyze Sentiment

Analysis Result

Predicted Sentiment

Negative emotion 👎

Confidence Score

Negative emotion:

Neutral:

Positive emotion:

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.
5. The model achieved 60% recall for negative sentiment, meaning it correctly identified 60% of all negative tweets.
6. 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 TwitterRoBERTa model via an API called Hugging Face for real-time sentiment monitoring. Prioritize alerts and monitoring for Applerelated 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


THANK
YOU


QUESTIONS
WE HAVE THE ANSWER