ASPECT-BASED SENTIMENT ANALYSIS OF TWEETS USING NLP REPORT.

Title: ASPECT-BASED SENTIMENT ANALYSIS OF TWEETS USING NLP DATA REPORT

Authors: *Gabriel Tenesi, Sharon Wathiri, Celine Sitina, Wesley Kipsang.*

Date: October 18, 2025.

1.Introduction

This report explores how customers feel about Apple and Google based on their posts on social media. These opinions reflect real experiences with products and services, helping companies understand customer satisfaction and brand image.

Using Natural Language Processing (NLP) and machine learning, the analysis classifies tweets about Apple and Google as positive, negative, or neutral. The goal is to find the most effective model for understanding public sentiment and supporting better business decisions.

1.1 Problem Statement

Apple and Google continuously monitor customer satisfaction to stay ahead in the technology market. However, given the massive volume and speed of data generated on Twitter, manual tracking of sentiment is impractical. Without automated systems, valuable insights into customer satisfaction, emerging issues, and product perception may be overlooked.

This project aims to address this challenge by developing a machine learning model capable of classifying tweets related to Apple and Google as positive, negative, or neutral. The outcome will support organizations in understanding real-time consumer opinions, measuring brand perception, and identifying areas for improvement based on public feedback

1.2 Business objectives

1.2.1 Main objective:

To develop an NLP-based sentiment analysis model that automatically classifies tweets about Apple and Google into positive, negative, or neutral categories.

1.2.2 Specific objectives:

1. To explore and clean the tweet dataset, handling missing values, duplicates, and irrelevant characters.

2. To preprocess textual data through tokenization, stopword removal, and lemmatization.

3. To convert cleaned text into numerical features using appropriate vectorization techniques such as TF-IDF.

4. To train and evaluate multiple classification algorithms (e.g., Logistic Regression, Naive Bayes, SVM) to identify the best-performing model.

5. To interpret and visualize model predictions, identifying which features most influence positive and negative sentiment.

6. To provide actionable insights that can guide Apple and Google in improving customer experience and brand perception.

1.2.3 Research question

1. How can the dataset be explored and cleaned to ensure data quality and reliability for sentiment analysis?

2. What preprocessing techniques are most effective for preparing Twitter text data for modeling?

3. Which text vectorization method (e.g., TF-IDF) produces better numerical representations for tweet classification?

4. Which classification algorithms yield the highest accuracy and robustness in predicting tweet sentiment?

5. Which textual features (words, phrases, or hashtags) most strongly influence model predictions of sentiment?

6. How can the resulting sentiment insights be applied by Apple and Google to improve customer satisfaction and brand reputation?

1.3 Success Criteria

Model Performance: Achieve at least 85% classification accuracy and a macro F1-score ≥ 0.80 across all sentiment classes (positive, negative, neutral).

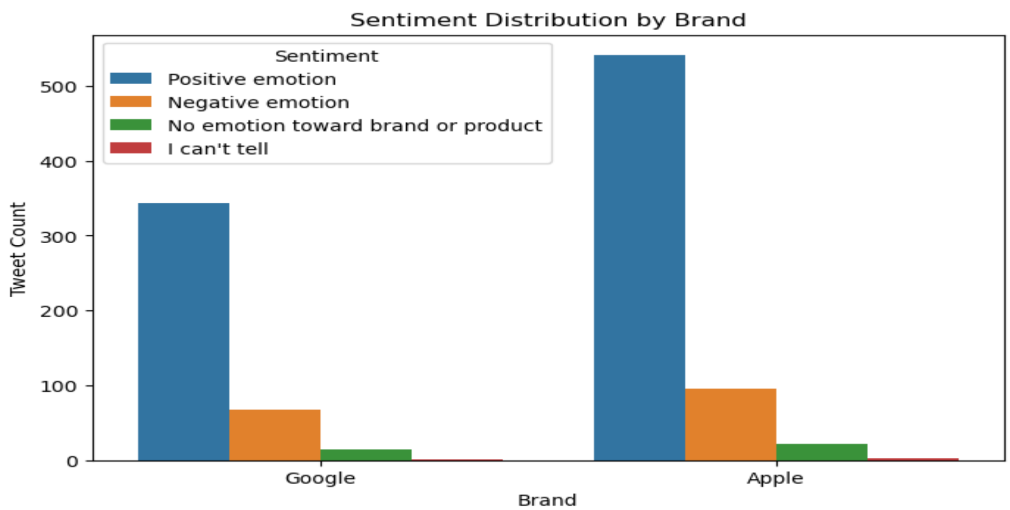
Model Interpretability: Clearly explain which features (words, hashtags, expressions) most affect sentiment predictions .

Business Value: Provide insights that help Apple and Google understand customer sentiment, identify common issues, and track brand reputation effectively

2.Findings & Analysis

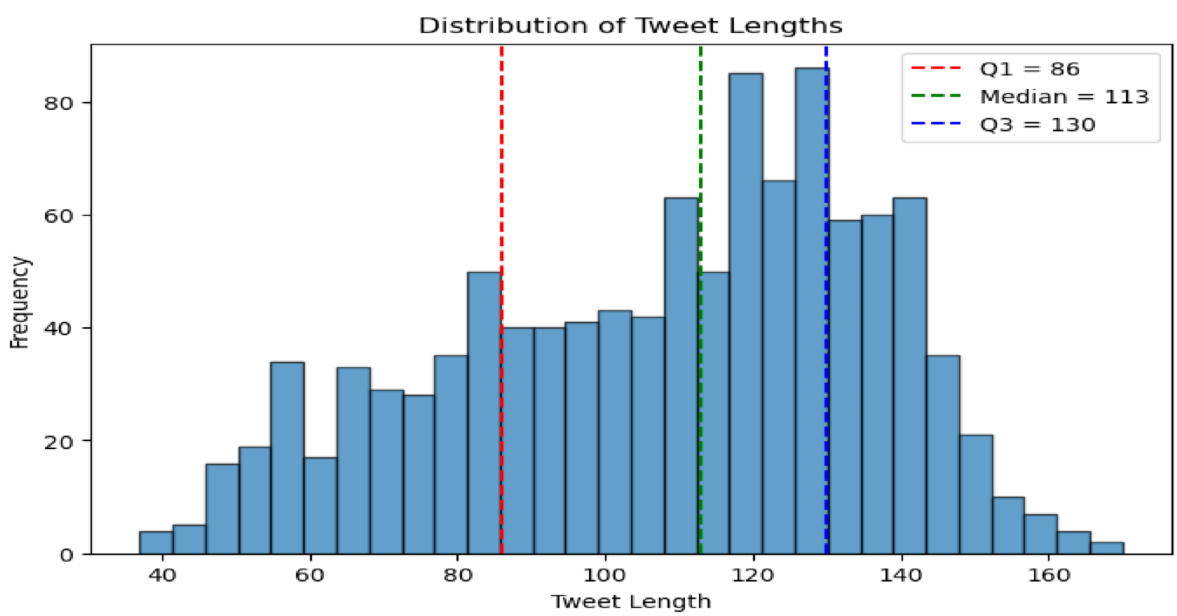
2.1 Sentiment Distribution by Brand

From the distribution of sentiments both brands get a mix of sentiments with the I can't tell being the least, the positive sentiments dominate in both brands, but the Apple brand receives more positive sentiments compared to the Google brand.



2.2 Tweet length Distribution

The distribution of tweet lengths is concentrated, with most tweets ranging between approximately 80 and 130 characters. This range corresponds to the interquartile range (Q1–Q3), indicating that most tweets are of moderate length.

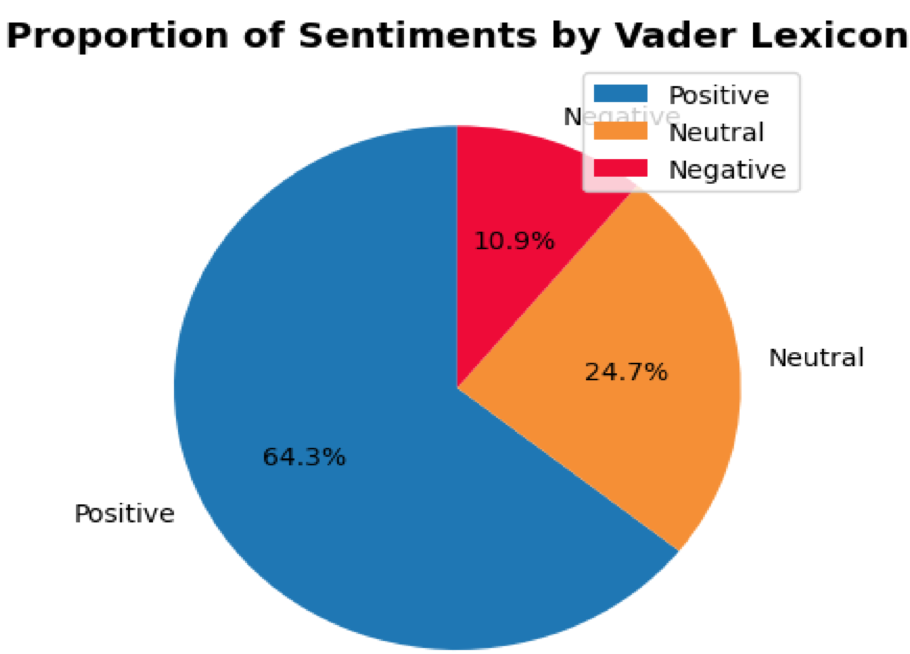


2.3 World Cloud of Tweets

Words like google and SXW are among the top words appearing in tweets.

2.4 Proportion of sentiments by Vader.

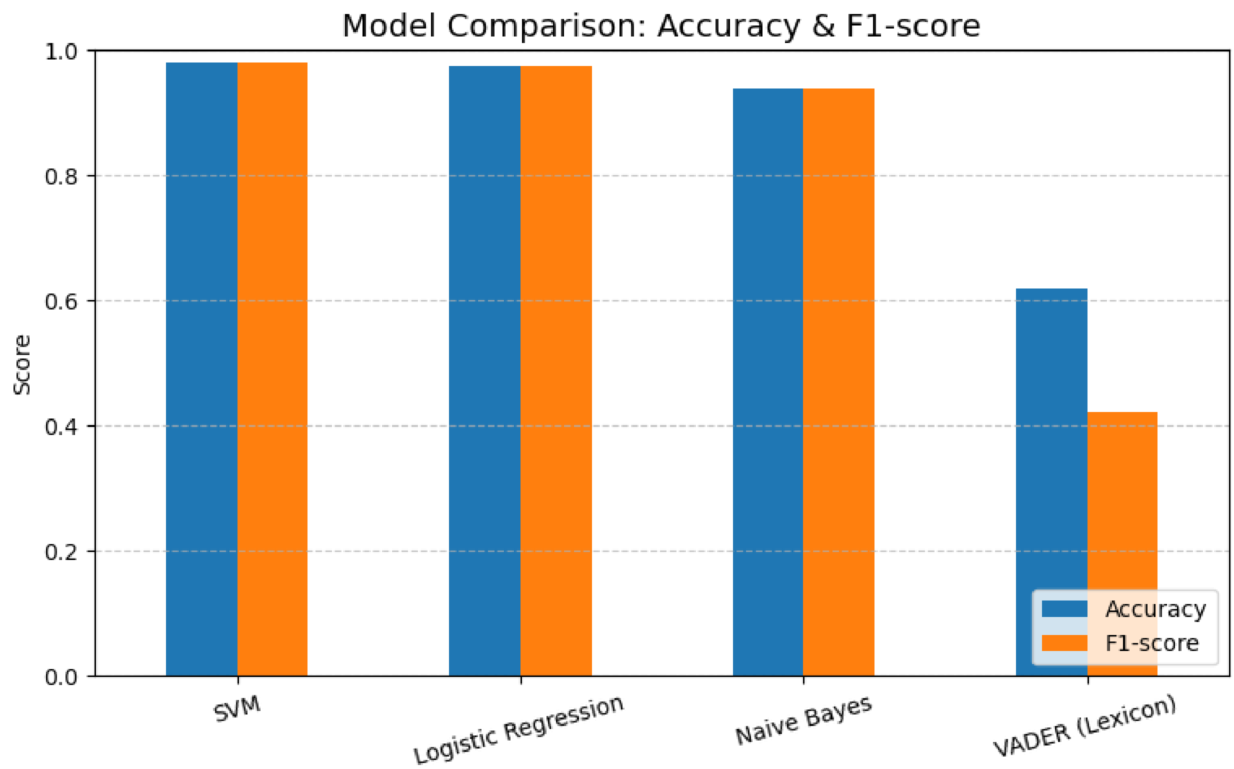
Distribution of sentiments on a pie chart showing the positive class dominate with 64.3%, followed by the Neutral with 24.7% and Negative being the least with 10.9%.



3.Model Selection

From our comparison of all models tested, SVM achieved the highest accuracy of 98.12%, slightly outperforming Logistic Regression of 97.55% and Naive Bayes of 93.97%.

This shows that SVM handled the high-dimensional TF-IDF features most effectively, making it the best overall model for this multiclass sentiment classification task.



4. Modelling & Evaluation

To identify the most effective approach for sentiment analysis, the following comparison ML models were developed and compared:

1. Logistic regression
2. Naïve Bayes
3. Support Vector Machine

<u>The following was how the models performed based on accuracy and F1 Score.</u>

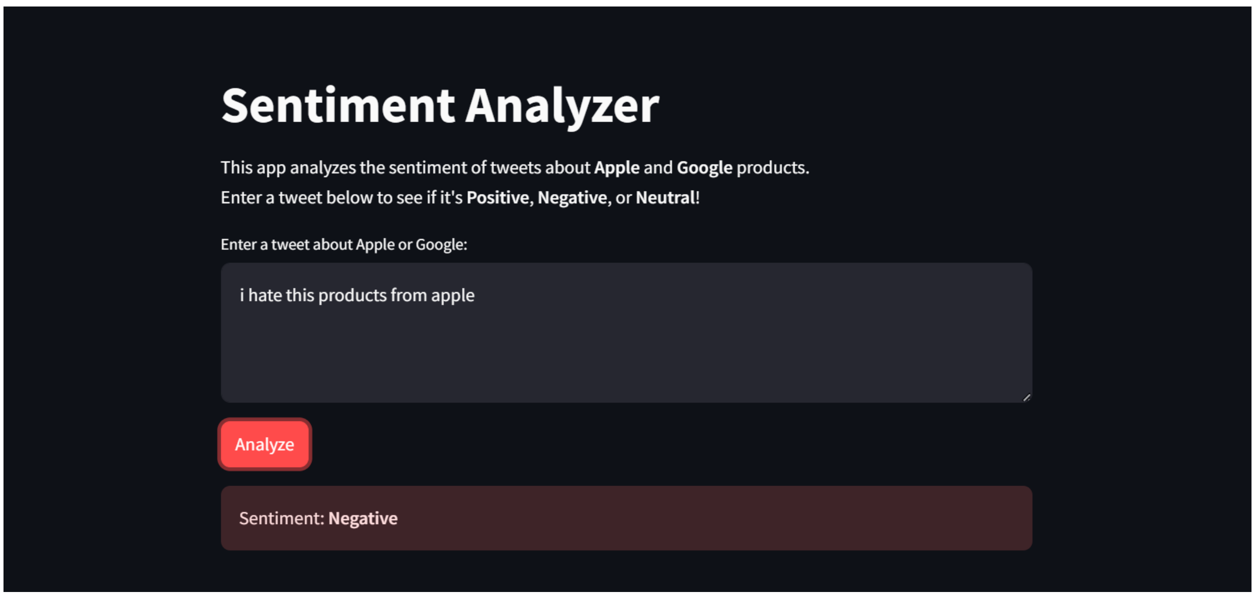
	VADER	Logistic regression	NAIVE-BAYES	SVM
Accuracy	0.620055	0.975518	0.939736	0.981168
F1 Score	0.420786	0.975460	0.939685	0.981151

5.Model Deployment

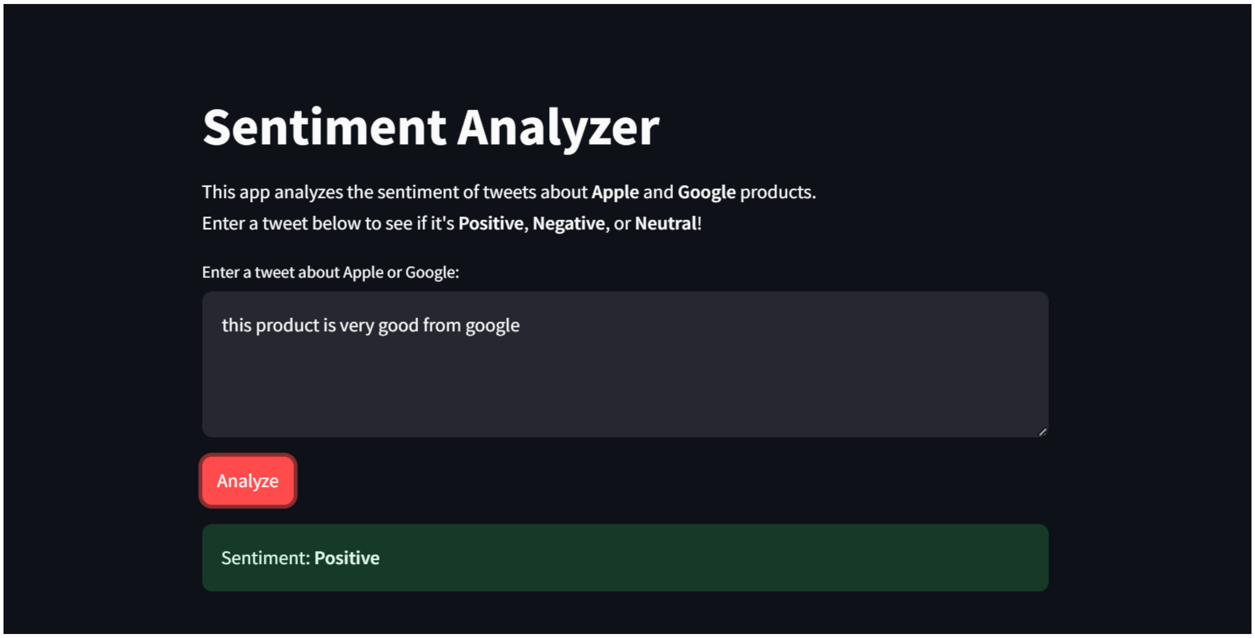
The sentiment analysis model was deployed as a web app using Streamlit and hosted on Render. Users can enter tweets about Apple or Google products to get real-time sentiment predictions (Positive, Negative, or Neutral).

Deployment Link**:** https://sentiment-analyzer-3zs1.onrender.com/

5.1 Negative sentiment fig



5.2 Positive sentiment fig



The figures above show the deployed model, which analyzes tweets about Apple and Google. The interface allows users to input text and instantly receives sentiment results.

6.Recommendations

6.1 Conclusion

- This project explored different models for sentiment analysis, including VADER, Naive Bayes, Logistic Regression, and SVM. The SVM model performed best, achieving an accuracy of 0.98%, showing strong ability to understand patterns in the text.

- Overall, machine learning models, especially those using TF-IDF features, performed much better than rule-based approaches like VADER.

6.2 Recommendations

I. Use SVM

- SVM achieved the highest performance with an **Accuracy and F1-score of 0.98**, making it the most reliable for classifying tweets about Apple and Google.
- **Business impact:** Ideal for real-time sentiment tracking and decision-making due to its high precision and consistency.

II. Use Logistic regression when interpretability matters

- Logistic Regression performed strongly with Accuracy **of 0.98 and F1 of 0.97**.
- Recommendation: Use alongside SVM when transparency and stakeholder understanding are priorities.

III. *Use Naïve Bayes as a supporting model*

- Naive Bayes achieved moderate results **Accuracy of 0.94**
- Recommendation: Suitable lightweight analysis tools, but not for production sentiment monitoring.

IV. Keep *improving* the model with new data to maintain accuracy.