Sentiment Analysis of Tweets Directed at Apple & Google Products

Harnessing customer insights from social media

By : Group 3 DS-PT11

Members:

1. Irene Kibengo
2. Erastus K Njuguna
3. Benson Mwihia
4. Daniel Akwabi
5. Luciana Ndanu
6. Sydney Were

Contents

Executive Summary

- Built a sentiment analysis model for Apple & Google tweets.
- We tested different models on Binary and Multiclass analysis and Regression delivered the best performance (~84% accuracy).
- Strong on Positive detection, weaker on Negative due to imbalance.
- Provides a foundation for real-time brand monitoring and insights.
- Next: improve Negative detection,deploy as dashboard/API, expand scope.

Business Context & Objectives

Problem Statement

- Customers constantly share opinions on Twitter, but the volume and noise make manual monitoring impossible. Without automation, businesses risk missing early warnings of negative sentiment, overlooking positive advocacy, and losing actionable insights. A sentiment analysis model enables real-time classification of tweets, empowering proactive reputation management and data-driven decisions.

Project Objectives

Automate Sentiment Detection – Classify tweets as positive, negative, or neutral in real time.

Enhance Brand Monitoring – Track how customers perceive Apple & Google products across Twitter.

Identify Actionable Insights – Highlight negative feedback for rapid response and improvement.

Support Strategic Decisions – Provide data-driven insights to inform marketing and product strategy.

Data Understanding and Preparation

Methodology

- Preprocessing: clean tweets (remove URLs, mentions, stopwords).
- Feature Extraction.
- Models we tested(5): Logistic Regression, Naïve Bayes, Random Forest, Gradient Boosting, AdaBoost and XGBoost.
- Class balancing: Oversampling of minority classes.

Data Set Overview

- 9,000+ labeled tweets from CrowdFlower dataset.
- Sentiment classes: Positive, Negative, Neutral/Unclear.
- Imbalance: Neutral dominates (~60%).
- Tweets are short, noisy, and contain slang/hashtags.

Data Cleaning

✓ Dropped 1 null tweet
✓ Ignored brand metadata (~5,800 missing)

Label Consolidation

✓ 4 classes → 3 classes (Positive, Negative, Neutral)
✓ Neutral = *No emotion* + *I can't tell*

Text Preprocessing

✓ Lowercasing
✓ Remove URLs, hashtags, mentions
✓ Remove punctuation, numbers, stopwords
✓ Lemmatization

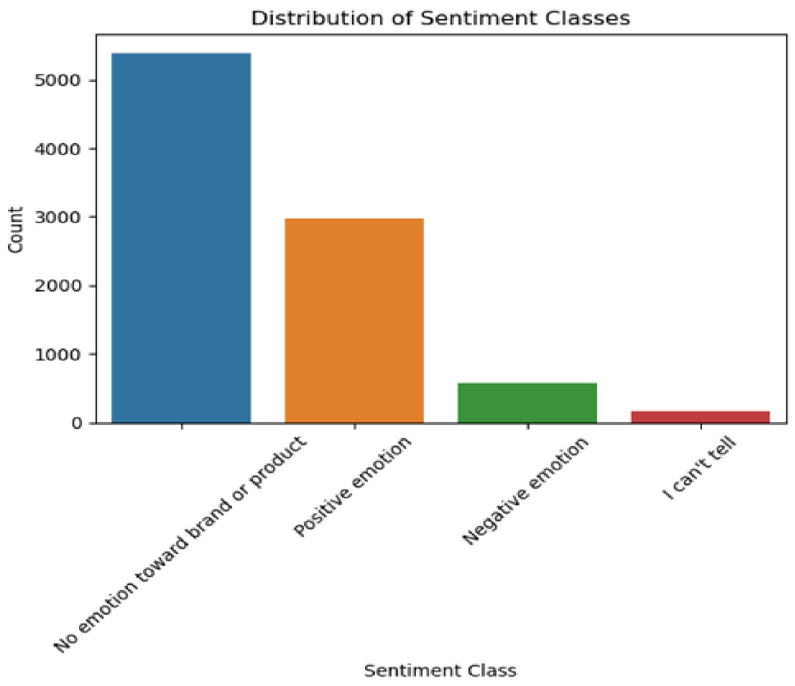
Modeling Readiness

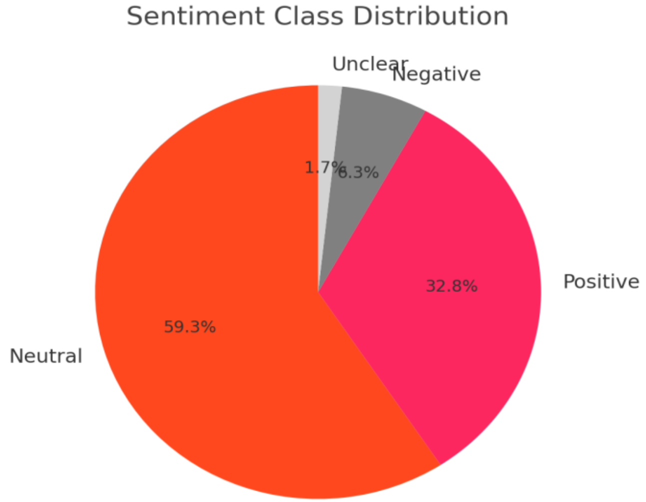
✓ Labels encoded numerically
✓ 80/20 train-test split (stratified)

Exploratory Data Analysis -EDA

Overview

- Neutral dominates (~60%), Positive ~33%, Negative ~6%.
- Negative underrepresented → detection harder.
- We will handle the inbalances via model parameters class weights and RandomOversampler method.

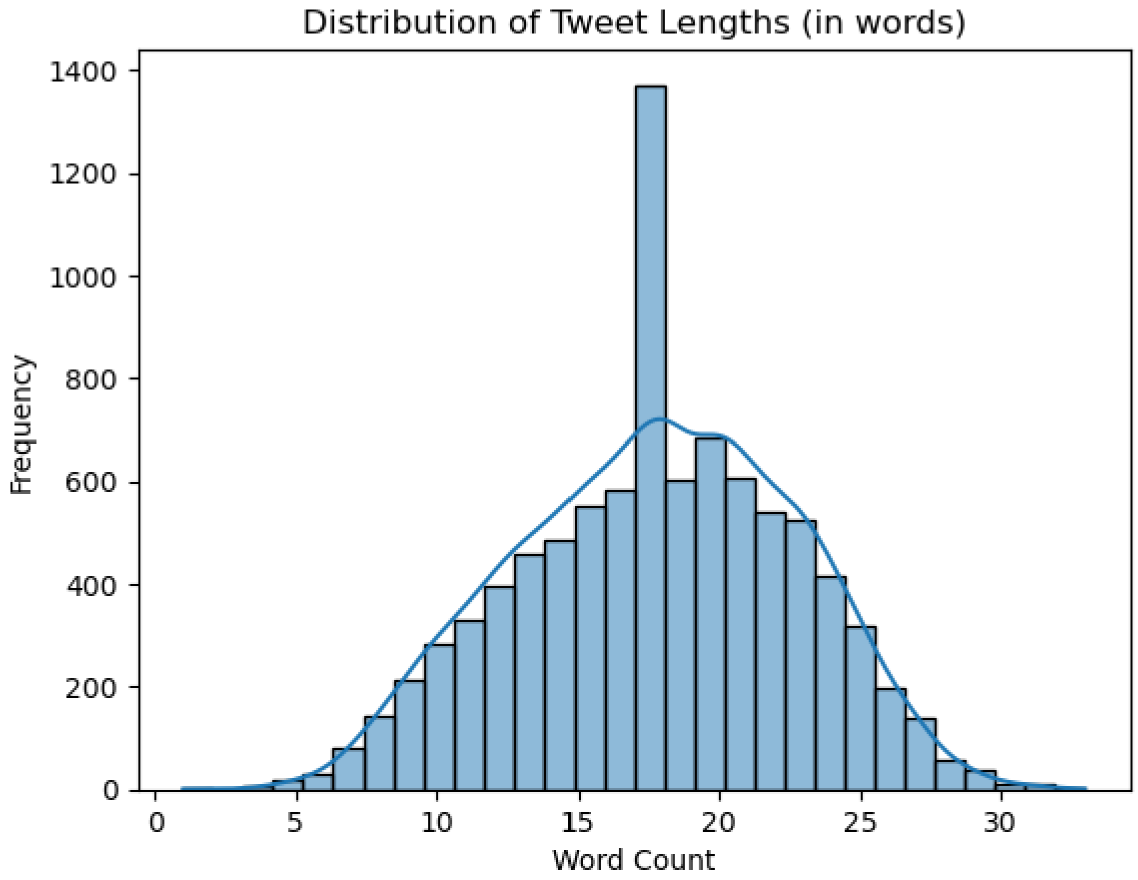




Exploratory Data Analysis -EDA

Data Set Overview

- Most tweets fall between 5–25 words.
- Reflects short, noisy nature of Twitter data.



Modelling For Binary Data

Best model(Before Tuning) : Logistic Regression with f1-score 0.71

Best model(After tuning)

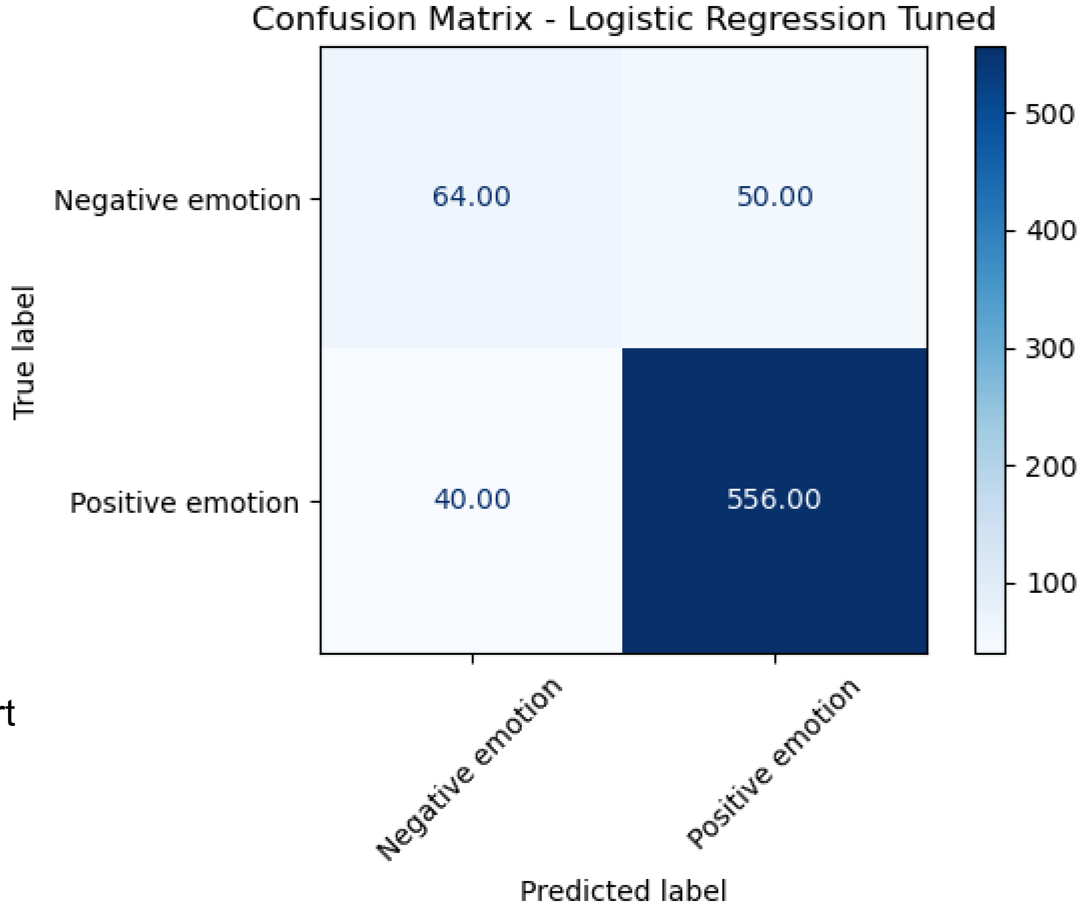
```
Test Accuracy: 0.8732

Classification Report:

                  precision    recall  f1-score   support

Negative emotion       0.62      0.56      0.59       114
Positive emotion       0.92      0.93      0.93       596

        accuracy                           0.87       710
       macro avg       0.77      0.75      0.76       710
    weighted avg       0.87      0.87      0.87       710
```



Accuracy = 87%

Per-Class Metrics
Negative: F1 = 0.59 → weakest, poor precision/recall due to low support (114).
Positive: F1 = 0.93 → good balance of precision/recall (596 examples).

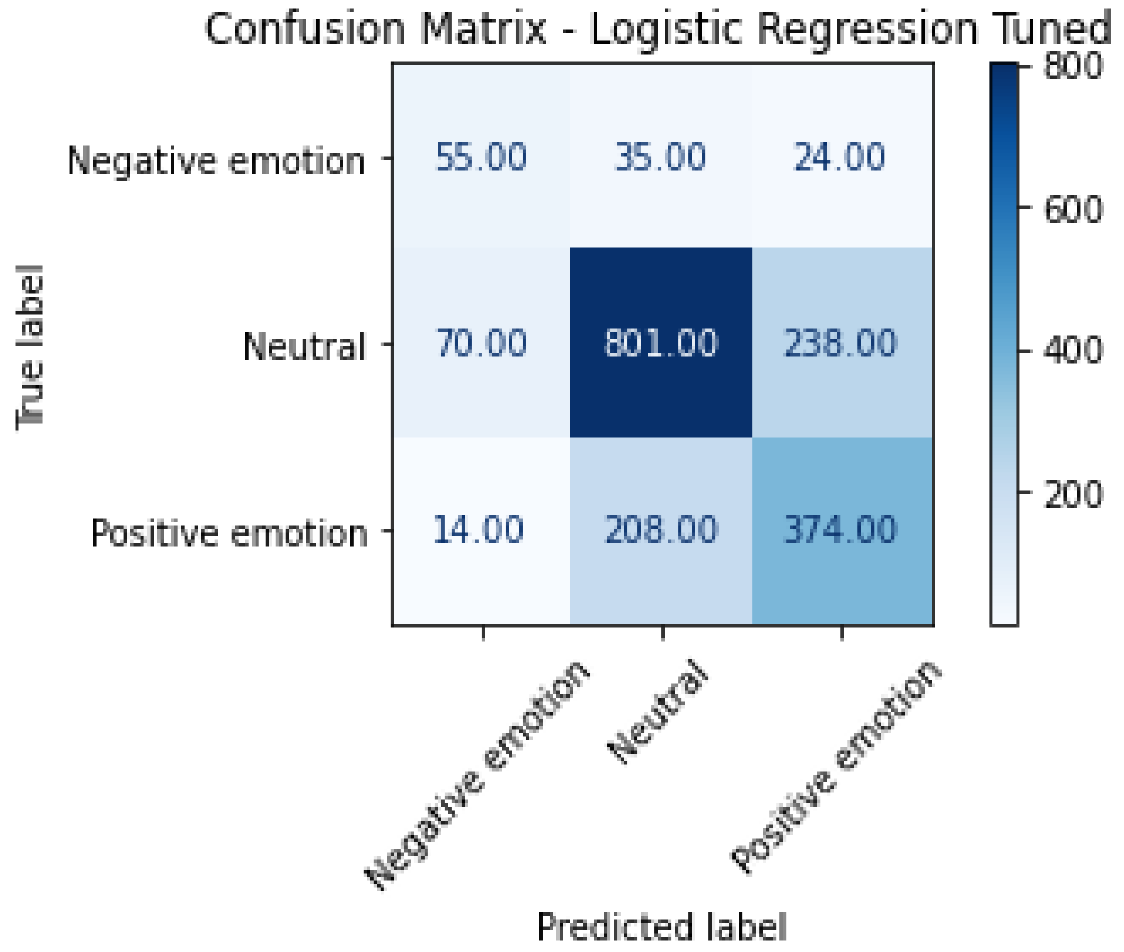
Modelling For Multi-Class Data

Best model(Before tuning): Logistic Regression with f1-score 0.55

Best model(After tuning)

Classification Report:

	precision	recall	f1-score	support
Negative emotion	0.40	0.48	0.43	114
Neutral	0.77	0.72	0.74	1109
Positive emotion	0.59	0.63	0.61	596
accuracy			0.68	1819
macro avg	0.58	0.61	0.60	1819
weighted avg	0.69	0.68	0.68	1819



Accuracy = 68%

Per-Class Metrics
Negative: F1 = 0.43 → weakest, poor precision/recall due to low support (114).
Positive: F1 = 0.61 → moderate, fair balance of precision/recall (596 examples).

Neutral: F1 = 0.74 → strongest, benefits from majority class size (1109 examples).

Model Comparison

Test Accuracy: 0.8732

Classification Report:

	precision	recall	f1-score	support
Negative emotion	0.62	0.56	0.59	114
Positive emotion	0.92	0.93	0.93	596
accuracy			0.87	710
macro avg	0.77	0.75	0.76	710
weighted avg	0.87	0.87	0.87	710

Classification Report:

	precision	recall	f1-score	support
Negative emotion	0.40	0.48	0.43	114
Neutral	0.77	0.72	0.74	1109
Positive emotion	0.59	0.63	0.61	596
accuracy			0.68	1819
macro avg	0.58	0.61	0.60	1819
weighted avg	0.69	0.68	0.68	1819

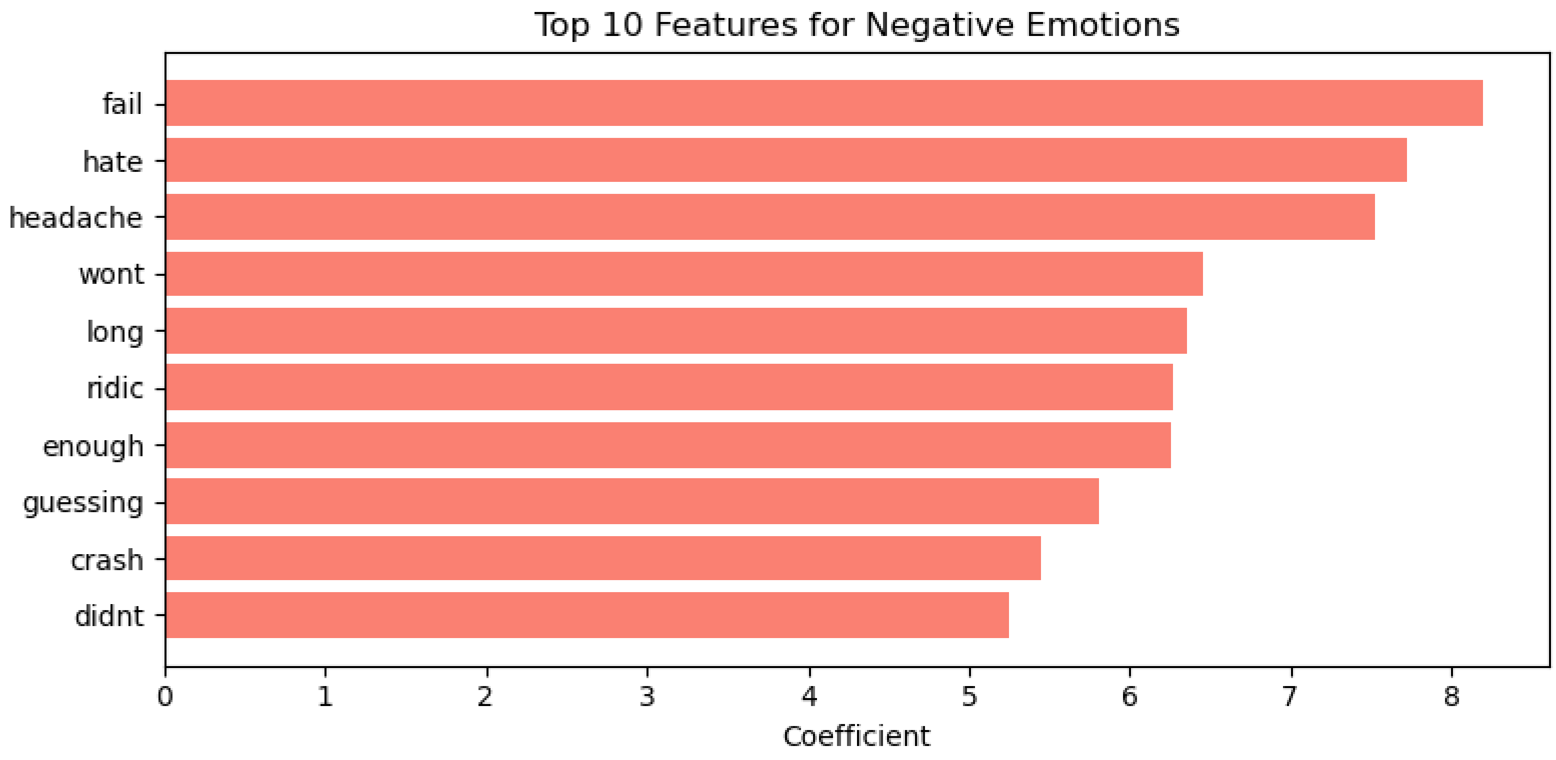
Model Benchmarking for the Binary Data

- Cross-validation performed across models.
- Logistic Regression performed best (Baseline F1 ~0.71).
- Tuned model achieved ~87% accuracy.
- Visualization: Logistic Regression outperformed others

Model Benchmarking for the Multiclass Data

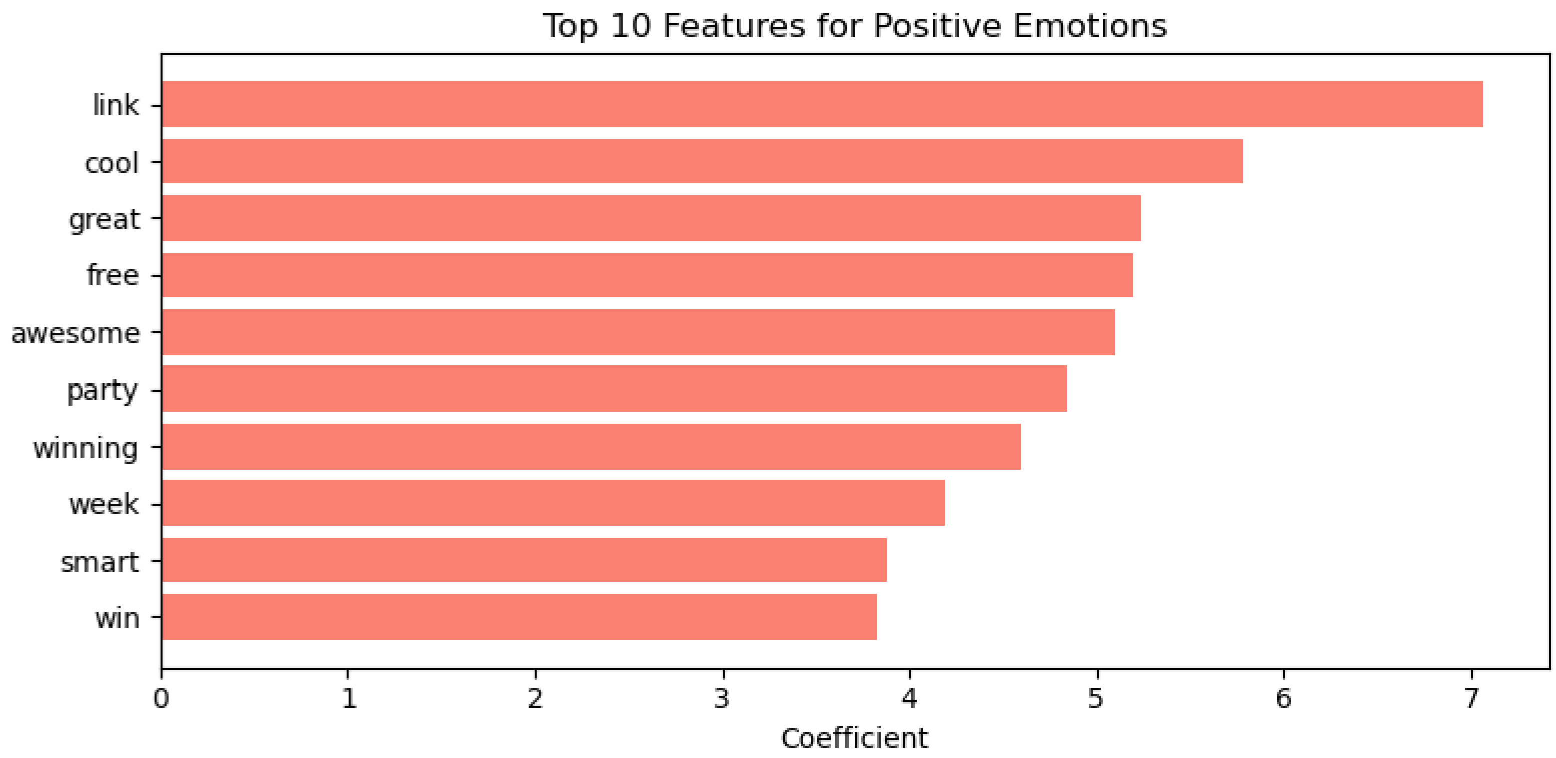
- Cross-validation performed across models.
- Best model achieved ~68% accuracy.
- Neutral class showed strongest performance (F1 ~0.74).
- Model struggled with minority classes, especially Negative (F1 ~0.43).
- Visualization: performance skewed toward Neutral due to class imbalance.

Features Driving Negative Emotions



- 'fail','hate','headache' drive negative classification.

Feature Driving Positive Emotions



- 'link',’cool’,’great' drive positive classification.

Key Insights & Recommendations

Key Insights

- Negative emotion signals: fail, hate, headache.
- Positive emotion signals: link, cool, great.
- Model strong on Positive and weaker on Negative class

Conclusion

- Real-Time Dashboard: Logistic Regression (~87% accuracy) can power a live dashboard for Marketing/Comms with trend alerts and CRM integration.
- Product Feedback: Negatives = crashes & usability; Positives = features & freebies. Fix recurring issues, build on positives.
- Customer Engagement: Track negative keywords (fail, hate, crash) and trigger alerts for proactive support.
- Marketing Leverage: Amplify positive tweets (cool, great, awesome) via retweets, testimonials, influencer/hashtag campaigns.
- Class Imbalance: Neutral dominates (~60%), Negative underrepresented (~16%). Add more labeled data and diversify sources (forums, app reviews).

Recommendations

- Model: TF-IDF + Logistic Regression, Accuracy ~84%, F1 = 0.87.Strong on Positive sentiment, weak on Negative due to imbalance.Next: Improve Negative detection (SMOTE, deep learning), deploy dashboard/API, expand to other brands, integrate CRM alerts.

Model Deployment

Link to the model: https://group-3-project-app-46ae69b19de0.herokuapp.com/

Thank You!

erastus@gmail.com
lucianandanu9@gmail.com
irenekibengo@gmail.com
bensonmwihia@gmail.com
danakwabi@gmail.com
sydneywere563@gmail.com