Twitter Sentiment Analysis

By Group 1:
Grace Wangui
Salma Mwende
Stephen Muema

Business understanding

Social media is a major channel for customers to share opinions and emotions. Platforms such as Twitter generate a large volume of unstructured text data. Monitoring sentiments helps technology companies like Apple and Google Improve products, enhance customer satisfaction and manage brand perception.

Business Problems

Negative sentiment about products or services can quickly harm brand perception, reduce customer loyalty, and impact sales. This project aims to address this problem by using Natural Language Processing (NLP) and machine learning techniques to automatically classify and analyze the sentiment expressed in tweets about Apple and Google products.

Key Questions

- How can a machine learning model classify tweets as positive or negative?
- How to clean and preprocess raw Twitter text data?
- How to explore and visualize sentiment distribution effectively?
- Which evaluation metrics ensure reliable model performance?

Sentiment Distribution

(Before Simplification)

Sentiment	Count	%
No emotion	5,389	59%
Positive	2,978	33%
Negative	570	6%
I can’t tell	156	2%

(After Mapping)

Sentiment	Count	%
Neutral	5,531	61%
Positive	2,970	33%
Negative	569	6%

The tables show the sentiment distribution, highlighting class imbalance with most tweets being neutral and far fewer negative or uncertain.

Data Preparation

To prepare the Twitter dataset for analysis, we first clean the tweets by removing URLs, user mentions, hashtags, and emojis. The text is then preprocessed through tokenization, stopword removal, and lemmatization to standardize the input for modeling. The dataset is split into training and testing sets using a stratified approach to maintain the original sentiment distribution

Modeling and Evaluation

After cleaning and preprocessing the tweets, machine learning models were trained to classify sentiment toward Apple and Google products. The models were evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score, to ensure reliable and consistent sentiment predictions.

Model Comparison Results

Model	Accuracy(%)	Macro Precision(%)	Macro Recall(%)	Macro F1 (%)
Baseline NB	65.49	59.63	40.43	39.21
Enhanced NB	67.48	59.09	51.92	54.06
Logistic Regression	64.55	54.61	61.43	56.49
Linear SVM	68.52	59.04	60.62	59.71
XGBoost	67.03	67.29	43.19	43.75

Linear SVM achieves the best overall performance, with the highest accuracy and Macro F1 score among all models.

Model Selection

Recommended Model for Deployment: Linear SVM

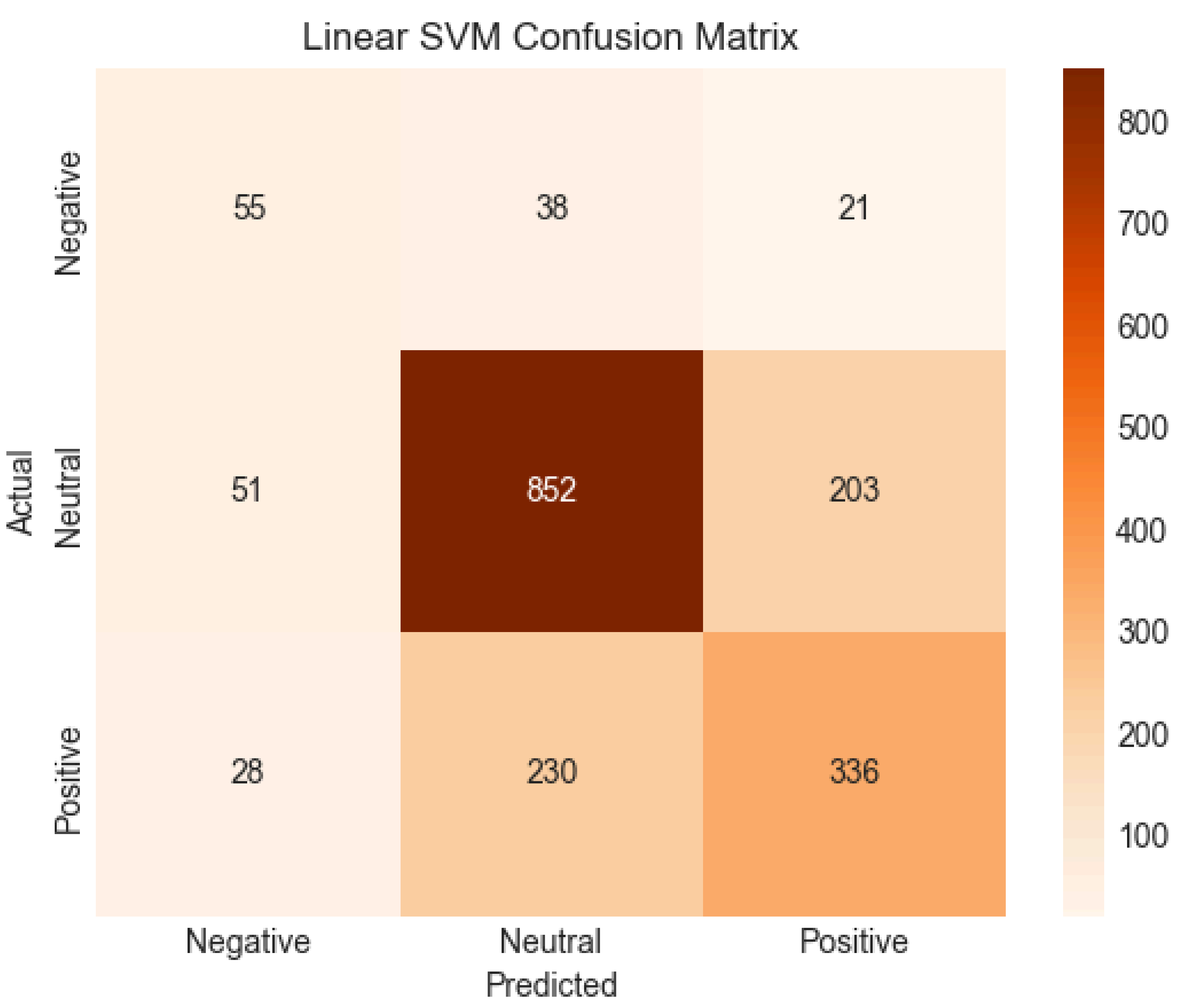
- **Top Performance**: Achieved the highest overall Accuracy (68.52%) and the strongest Macro F1-Score (0.5971).
- **Business Balance:** Provides the best "sweet spot" for Negative Sentiment Detection (0.4435 Negative F1), catching nearly half of all complaints while maintaining high precision.
- **Proven Stability:** 5-Fold Cross-Validation confirmed a high Mean F1 (0.5542) with very low variance, making it the most reliable model for production.

Negative Sentiment Class Performance

Model	Negative Recall(%)	Negative F1(%)
Baseline NB	0.88	1.72
Enhanced NB	22.81	30.59
Logistic Regression	56.14	38.79
Linear SVM	48.25	44.35
XGBoost	7.02	12.70

Linear SVM provides the best balance for detecting negative sentiment, achieving the highest Negative F1 score, while Logistic Regression attains the highest Negative Recall.

Confusion Matrix



The model correctly identifies 48% of negative tweets, classifies neutral sentiment with high accuracy, and mainly mislabels positive tweets as neutral, resulting in a strong balance between recall and precision.

Recommendations

1. **Operationalize Crisis Alerts**:
   - Integrate the Linear SVM with the Twitter API to trigger automated "high-priority" alerts when negative sentiment spikes.
2. **Customer Service Integration:**
   - Directly route tweets flagged as "Negative" to a specialized response team, reducing the detection-to-resolution window.
3. **Trend & Sentiment Dashboards:**
   - Visualize sentiment shifts during product launches to identify specific pain points or features that are driving positive engagement.
4. **Feedback Loop:**
   - Establish a process where customer service reps can flag misclassifications, creating a "Gold Standard" dataset to retrain and improve the model quarterly.

Conclusion

This project applied NLP and machine learning techniques to analyze sentiment in Twitter data related to Apple and Google products. The results highlight how automated sentiment analysis can help organizations understand customer opinions and support informed business decisions.

THANK YOU