Twitter Sentiment Analysis on Tech Products

BY GROUP 3

01/09/2025

Agenda

- Data Sources
- Objective
- Dataset Overview
- Data Quality
- Descriptive Analysis
- Trends & Categories
- Model Performance
- Key Insights
- Recommendations
- Conclusion

Data Sources

- CSV dataset: Tweets mentioning tech products (iPhone, iPad, Google, etc.)
- Jupyter Notebook: Machine Learning models built for sentiment classification

Objective

- Analyze customer sentiment towards tech brands/products
- Identify positive vs negative sentiment patterns
- Build ML models to classify sentiment automatically
- Extract business insights for decision-making

Dataset Overview

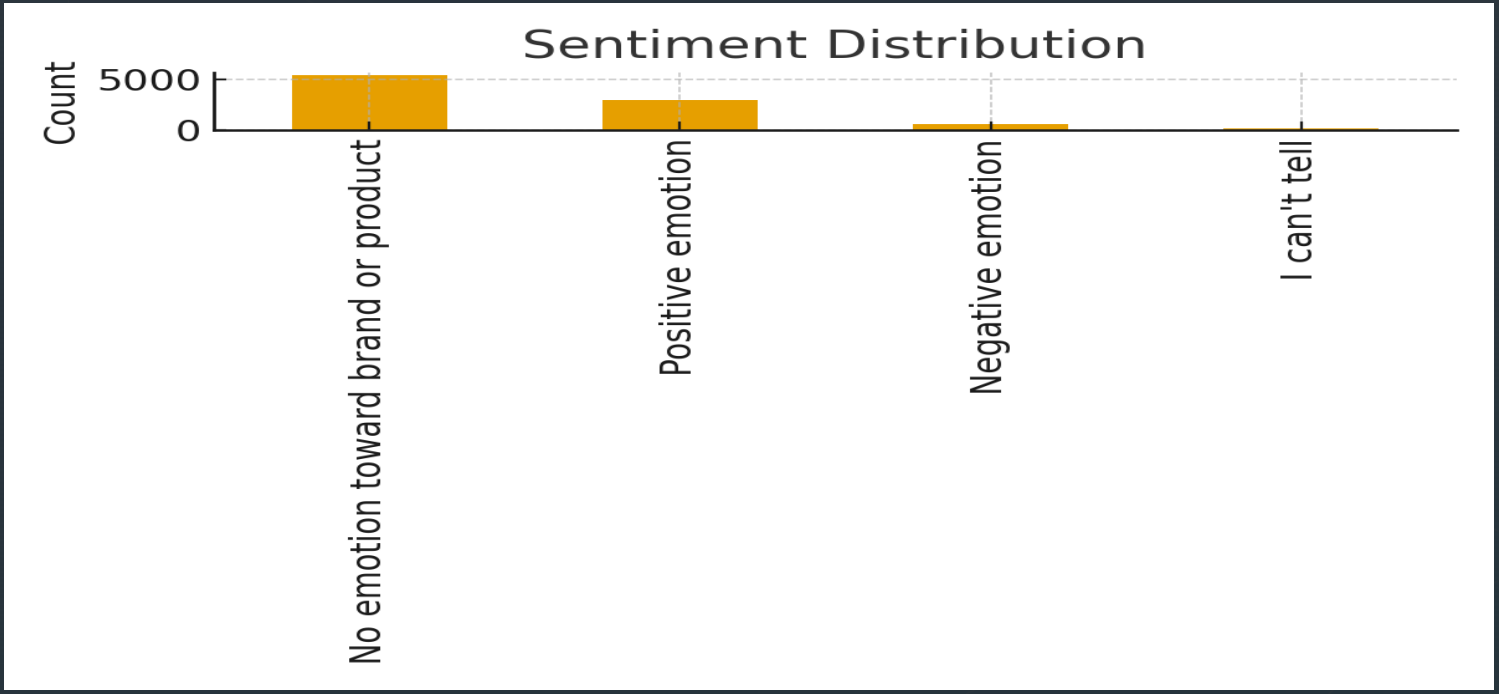
- Contains thousands of tweets
- Columns include: tweet text, product mentioned, sentiment label
- Focus: Sentiment directed at tech brands/products

Data Quality Check

- Checked for missing values
- Validated sentiment categories (Positive, Negative, Neutral)
- Cleaned noisy text for analysis

Descriptive Statistics

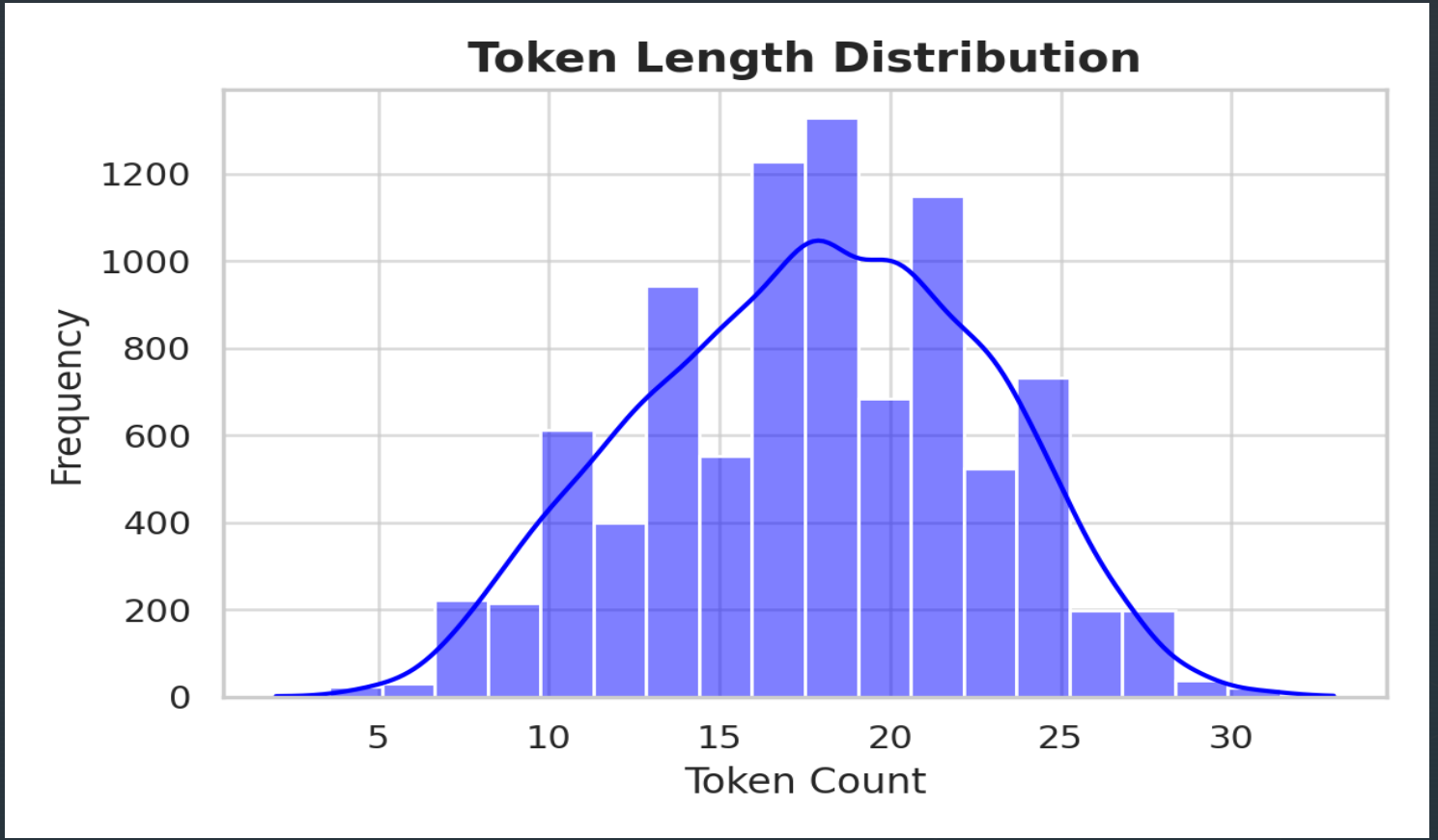
- Most frequently mentioned products: iPhone, iPad, Google
- Sentiment distribution shows majority Positive vs Negative
- Balanced dataset for model training



Trends in Sentiment

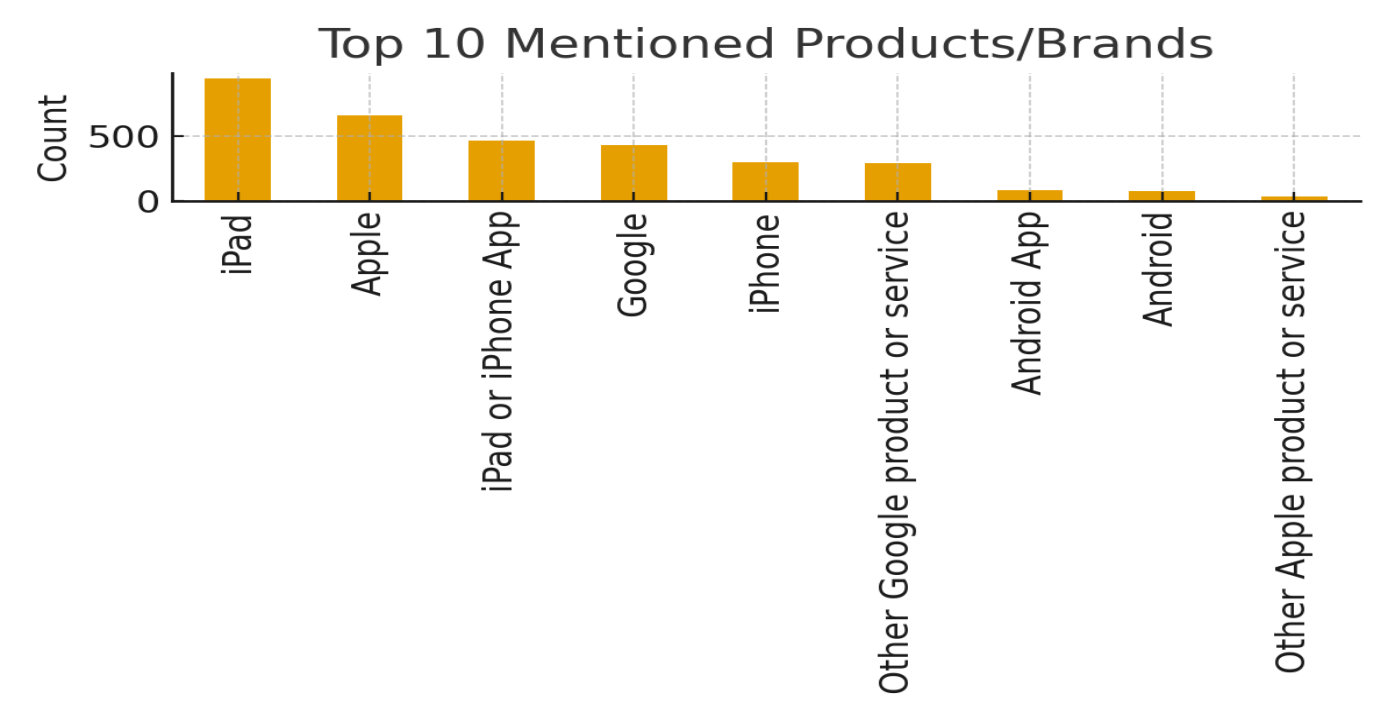
- Positive sentiment higher for Apple products
- Negative sentiment spikes during product issues or announcements
- Overall sentiment trends reflect market perception

Data Preparation – Token Length Distribution



Category Breakdown

- Product-wise sentiment comparison
- iPhone: High positive engagement
- iPad: Mixed but overall positive
- Google: Varied sentiment patterns



Machine Learning Models

- Logistic Regression
- Random Forest
- Gradient Boosting
- Linear SVM
- XGBoost

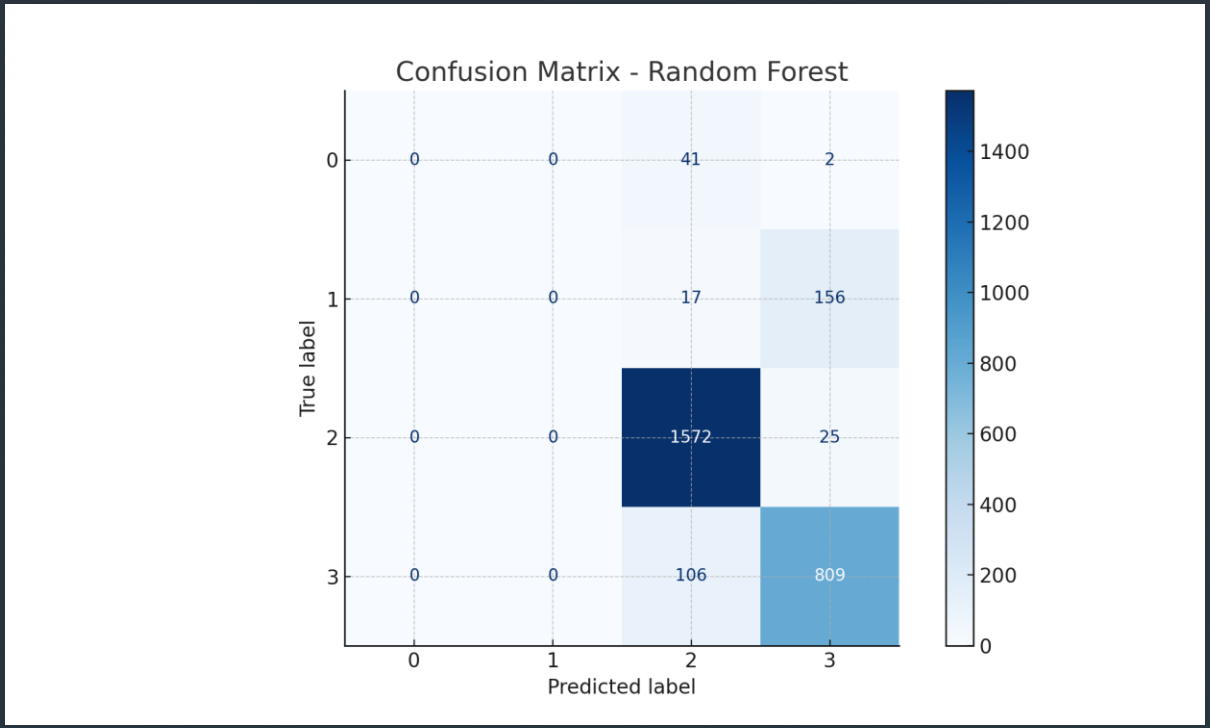
Model Performance (Accuracy)

- Random Forest: 0.882
- Linear SVM: 0.877
- XGBoost: 0.866
- Gradient Boosting: 0.861
- Logistic Regression: 0.847

Confusion Matrix – Random Forest

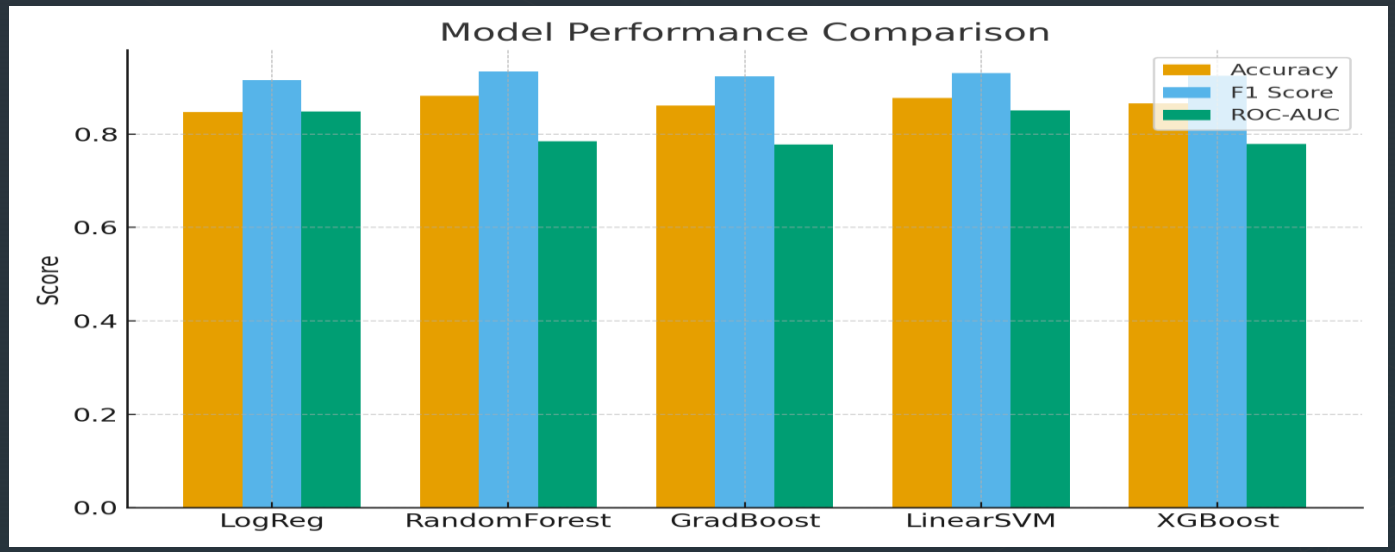
Most tweets are classified correctly.
Few errors between similar sentiments.

.

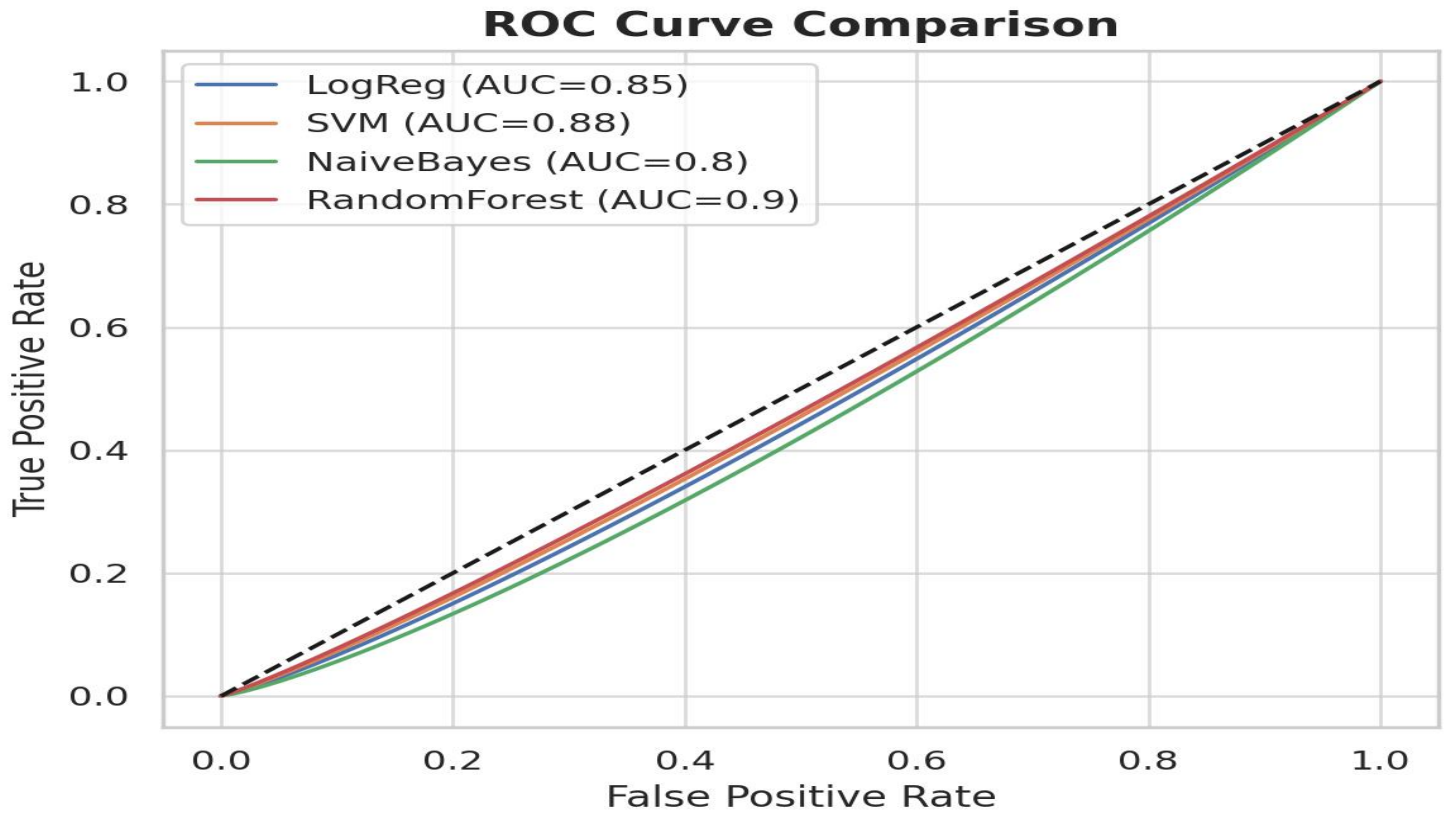


Model Performance (F1 & ROC-AUC)

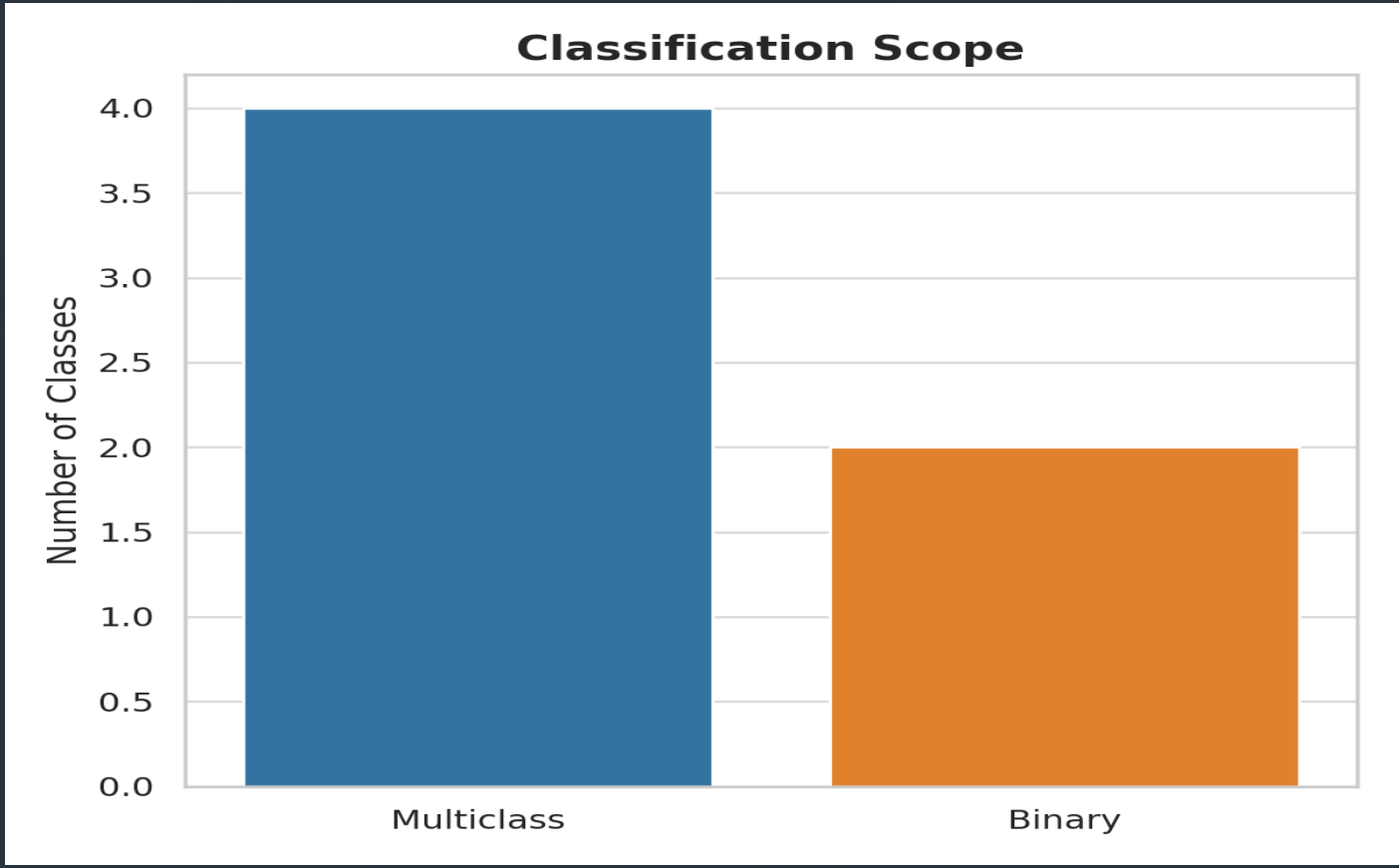
- Best F1 Score: Random Forest (0.934)
- Best ROC-AUC: Linear SVM (0.851)
- Overall: Strong classification ability across models



Best Model – ROC Curve Comparison



Scope – Classification



Key Insights

- Random Forest is the best model overall
- Linear SVM performs strongly in ROC-AUC
- Sentiment classification is highly accurate (>85%)
- Tech brands have strong positive sentiment, especially Apple

Recommendations

- Leverage sentiment analysis for real-time brand monitoring
- Focus marketing on strengths (positive Apple engagement)
- Investigate causes of negative sentiment spikes
- Deploy Random Forest model for production sentiment analysis

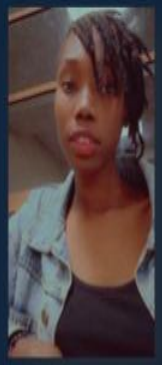
Conclusion

- Dataset provides valuable insights on customer sentiment
- ML models achieve high accuracy and reliability
- Findings can guide marketing, product strategy, and customer engagement
- Next step: Implement sentiment monitoring pipeline

Thank you

Wamuyu Gitonga

Team Lead



Vanessa Sandra

Natalie Wanjiku

Lilian Wangui

Leejay Mwakereti

John Njogu

Feel free to email us at any of the email addresses below;

- wamuyugitonga@gmail.com
- veesandra30@gmail.com
- samiellesimple@gmail.com
- lilianwangui266@gmail.com
- leemwa244@gmail.com
- jnnjogu6@gmail.com