Natural Language Processing Approach to Twitter Sentiment Analysis: Apple and Google products

Presented by Riché FLEURINORD – Akademi Education (2025)

OVERVIEW

Context

The rapid growth of social media has made it essential to understand public sentiment toward brands, events, and policies.

Objective

Build a model that can classify the sentiment of tweets as positive, negative, or neutral based on their textual content.

Method

Data collection from Twitter, text preprocessing using NLP techniques, feature extraction, and model training with machine learning algorithms.

Expected Impact

Provide real-time insights into public opinion to help organizations make informed, data-driven decisions and improve communication strategies.

Business Understanding

Problem

How can Apple and Google effectively extract and interpret public sentiment from the vast amount of data on Twitter to support strategic decision-making?

Stakeholders

Apple and Google's marketing and product development teams, shareholders, investors, and the broader consumer base.

Benefits

- Enhanced understanding of public perception and brand reputation
- Data-driven guidance for marketing campaigns and product innovation
- Early detection of negative trends or consumer dissatisfaction
- Improved strategic decision-making and customer engagement

Data Understanding

Data Source:

CrowdFlower dataset (accessible via data.world) containing over 9,000 manually labeled tweets related to Apple and Google.

Key Features:

- tweet_text – textual content of each tweet
- emotion/label – sentiment category (positive, negative, or neutral)
- brand – indicates whether the tweet refers to Apple or Google
- Additional metadata – may include user information or tweet ID for tracking

Dataset Size:

9,000+ records, providing sufficient volume for training and validation.

Main Challenge:

- Sentiment interpretation complexity due to sarcasm, ambiguity, or linguistic nuance
- Limited representativeness — may not capture all sentiment variations across the broader Twitter ecosystem

DATA PREPARATION

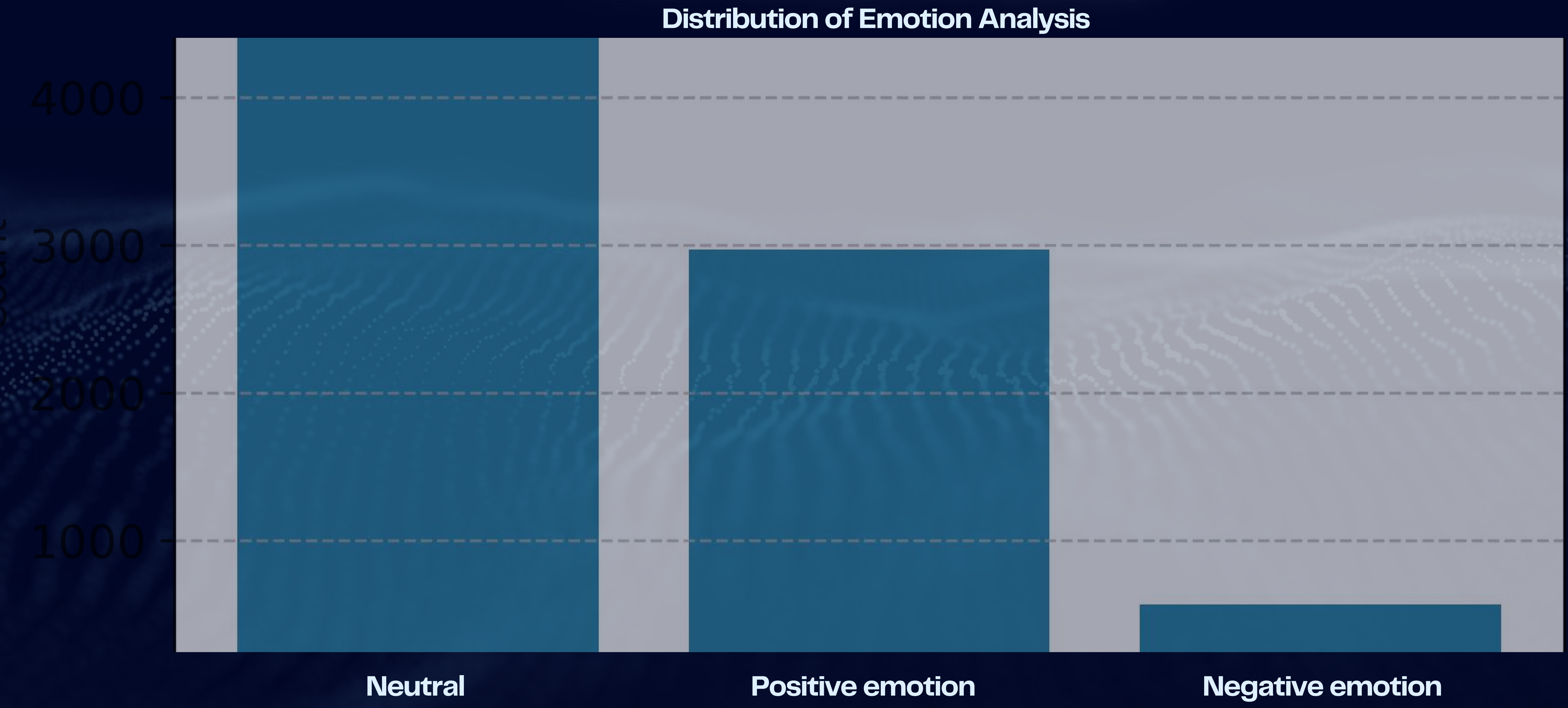
Key Steps

- **Dropped unnecessary columns** (emotion_in_tweet_is_directed_at).
- **Removed ambiguous categories** (“I can’t tell”, “No emotion toward brand or product”).
- **Kept only relevant sentiment labels** (Positive, Negative, Neutral).
- **Handled missing values** by removing tweets with no text.
- **Removed duplicate entries** to ensure data uniqueness.
- **Renamed the sentiment column** to target and standardized category names.
- **Cleaned tweet text:** removed URLs, hashtags, special characters, converted to lowercase, removed stop words, and applied tokenization and lemmatization.
- **Transformed cleaned text** into numerical form using TF-IDF vectorization.

Impact

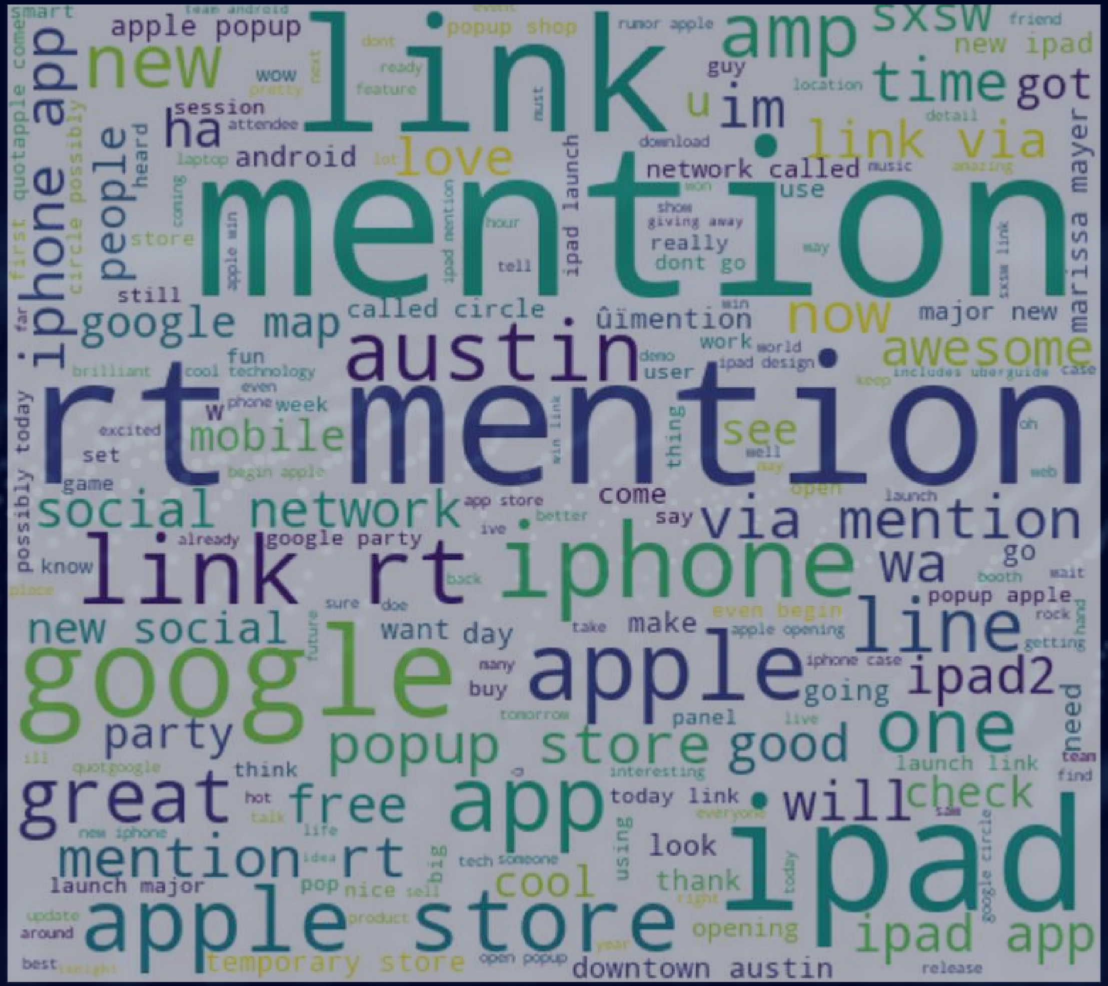
- Produced a clean, consistent, and balanced dataset.
- Ensured that textual data is normalized and ready for machine learning models.
- Improved model accuracy and reliability in sentiment classification.

(EDA)/EXPLORATORY DATA ANALYSIS

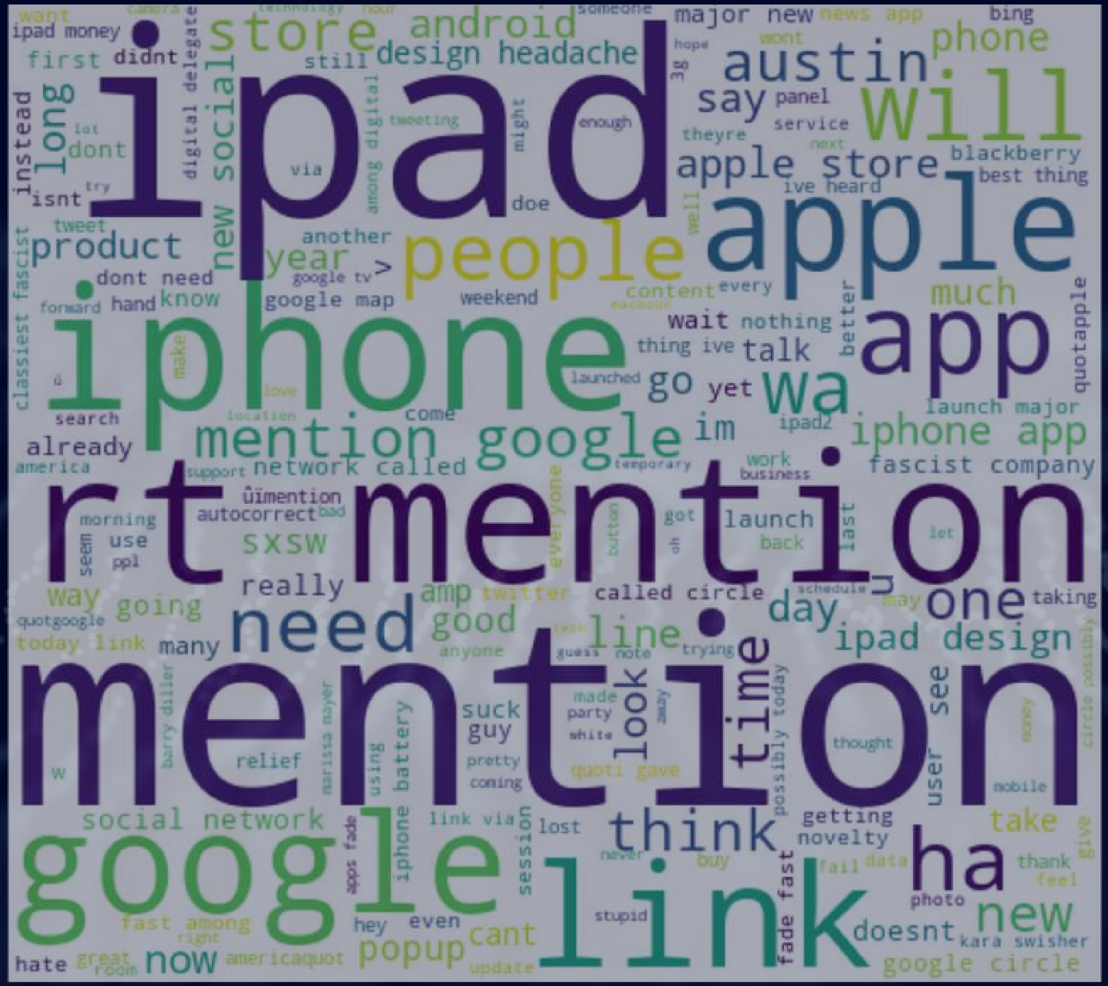


(EDA)/EXPLORATORY DATA ANALYSIS

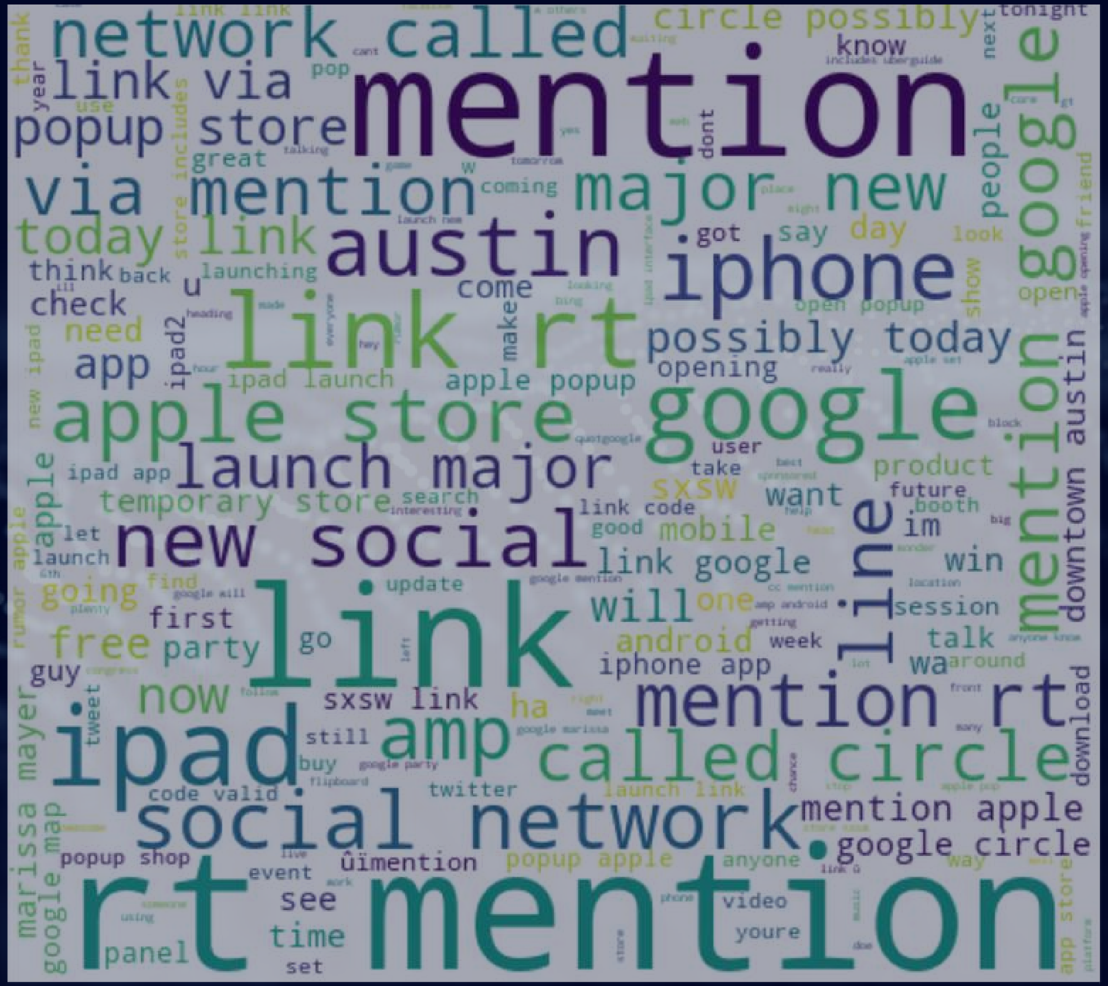
Word cloud



Positive tweets



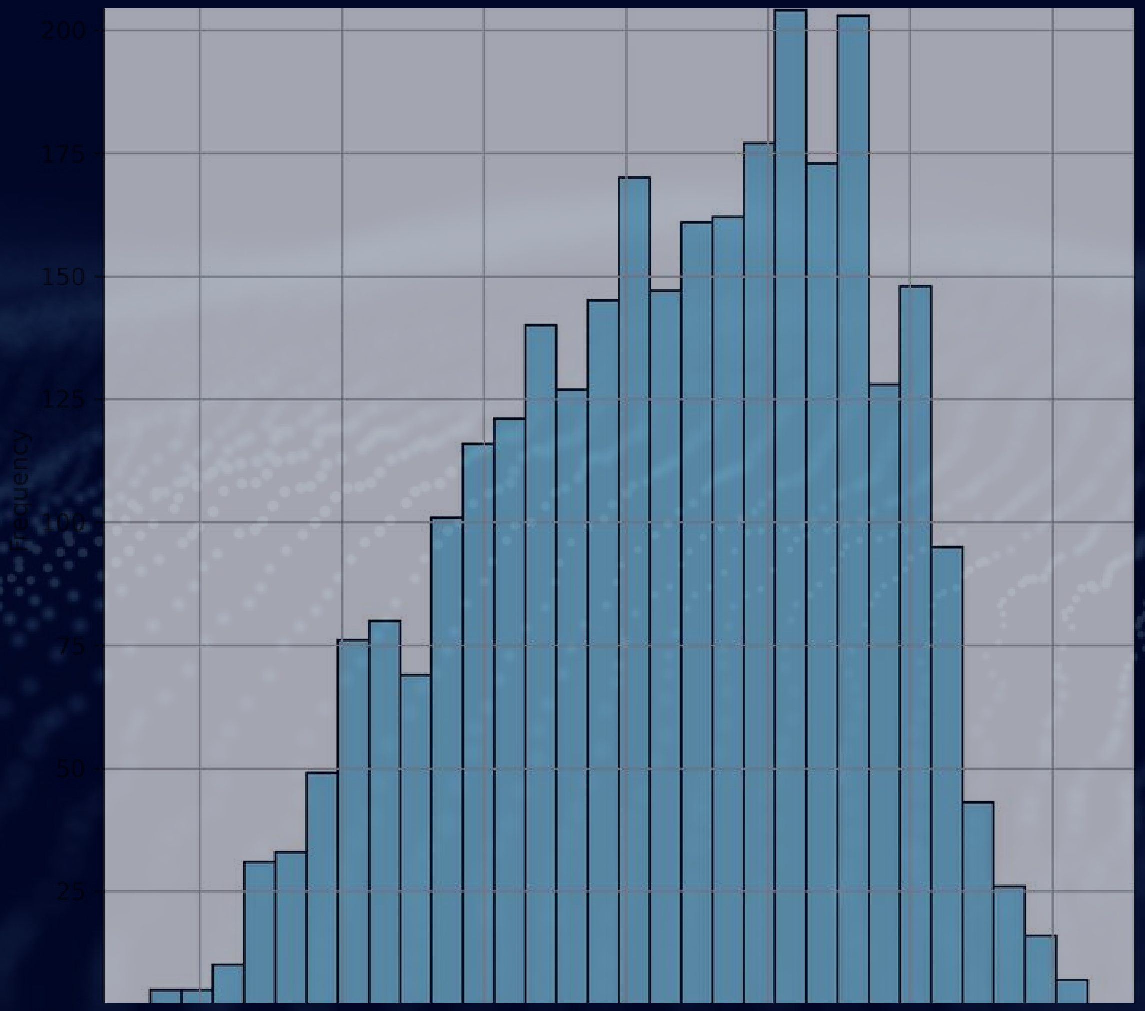
Negative tweets



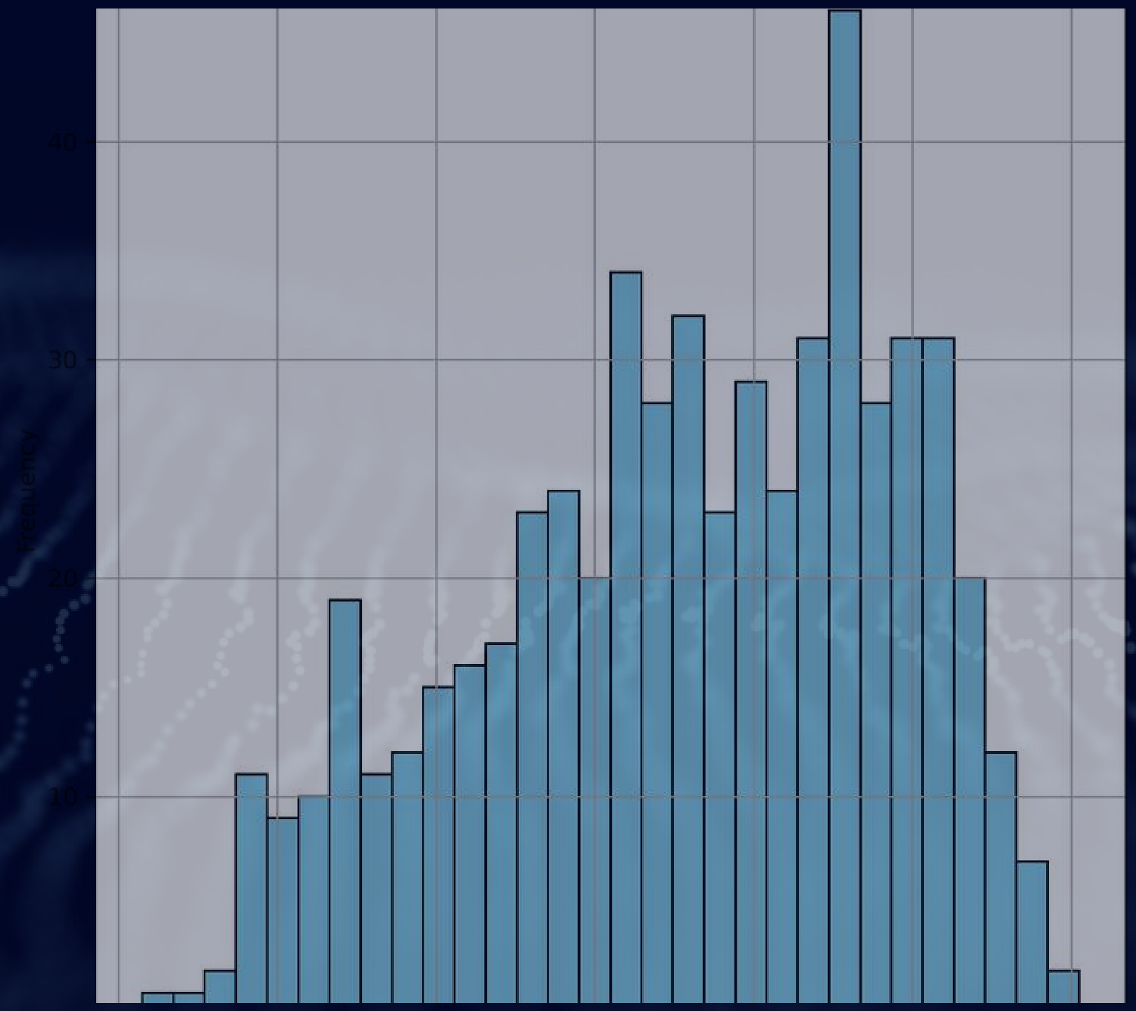
Neutral tweets

(EDA)/EXPLORATORY DATA ANALYSIS

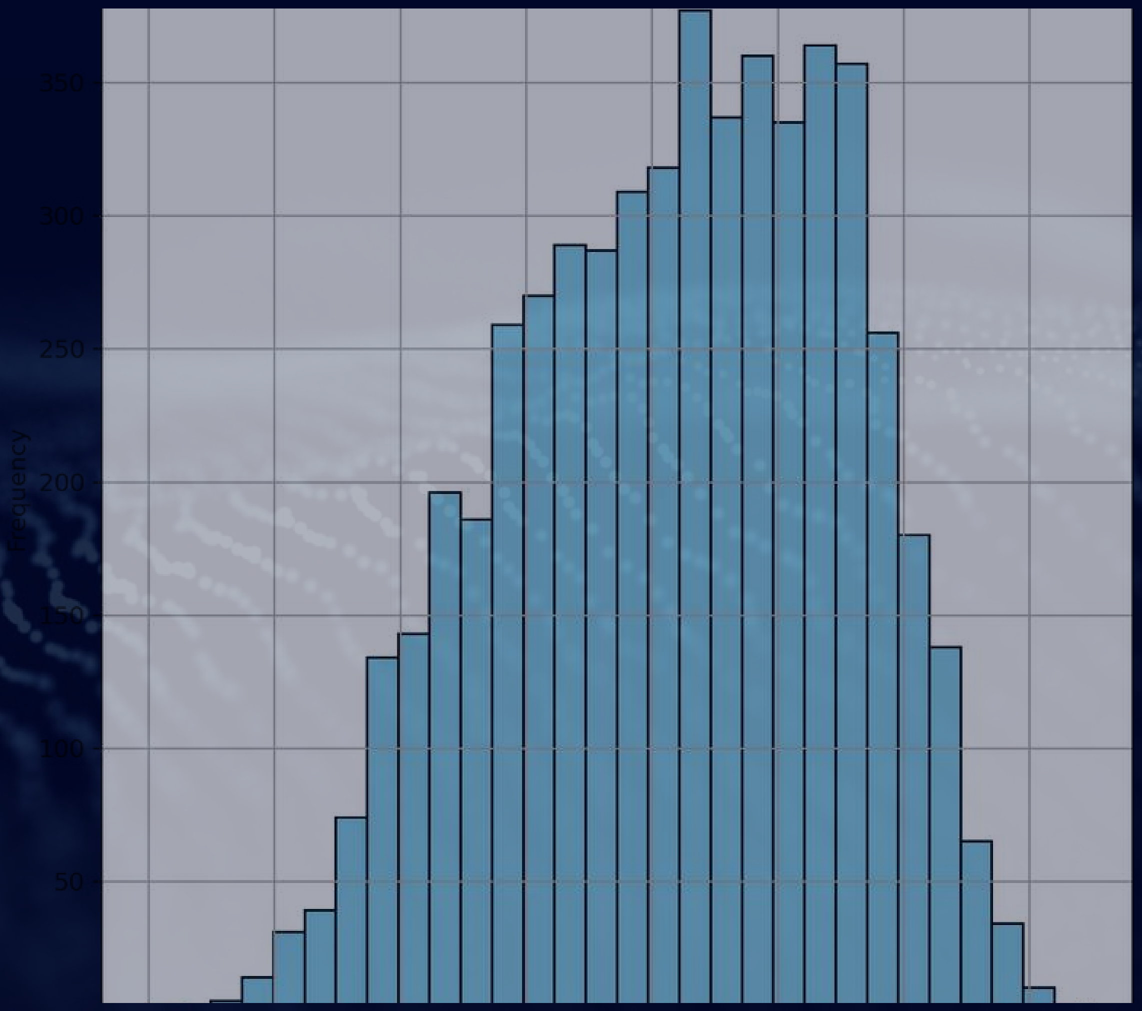
Tweet Lenght Distribution



Positive sentiment



Negative sentiment



Neutral sentiment

Modeling Approach

Models Tested

- Naive Bayes (Baseline for Binary Classification)
- Support Vector Machine (SVM)
- Random Forest
- Naive Bayes with Hyperparameter Tuning (GridSearchCV – Multiclass)
- Random Forest (Multiclass)
- Logistic Regression (Multiclass)

Approach

- **Binary Classification**: Initial focus on distinguishing positive vs negative sentiments.
- **Multiclass Extension**: Expanded to include neutral sentiment for a more comprehensive analysis.
- **Data Splitting**: Used train_test_split to ensure proper evaluation of model performance.
- **Label Encoding**: Converted textual sentiment labels into numeric values
- **TF-IDF Vectorization**: Transformed tweets into numerical vectors capturing word importance across the corpus.
- **Model Evaluation**: Employed metrics such as accuracy, precision, recall, F1-score, and confusion matrices to assess performance.

EVALUATION

Results Summary

- Naive Bayes (Baseline) → Accuracy ≈ 84%, efficient and interpretable.
- SVM → Accuracy ≈ 87.8%, improved recall for negative sentiments.
- Random Forest (Binary) → Accuracy ≈ 88.8%, strong overall performance.
- Naive Bayes (Tuned – Multiclass) → Accuracy ≈ 66%, balanced but moderate performance.
- Random Forest (Multiclass) → Accuracy ≈ 68%, strong neutral detection, limited negative recall.
- Logistic Regression (Multiclass) → Accuracy ≈ 69%, best overall multiclass performer.

Impact

- Provided a robust framework for both binary and multiclass sentiment analysis.
- Improved accuracy and interpretability across model iterations.
- Enabled the extraction of meaningful insights on public sentiment dynamics toward Apple and Google.

Recommendations

Business Recommendation 1

Highlight and Promote Positively Perceived Product Features

Business Recommendation 2

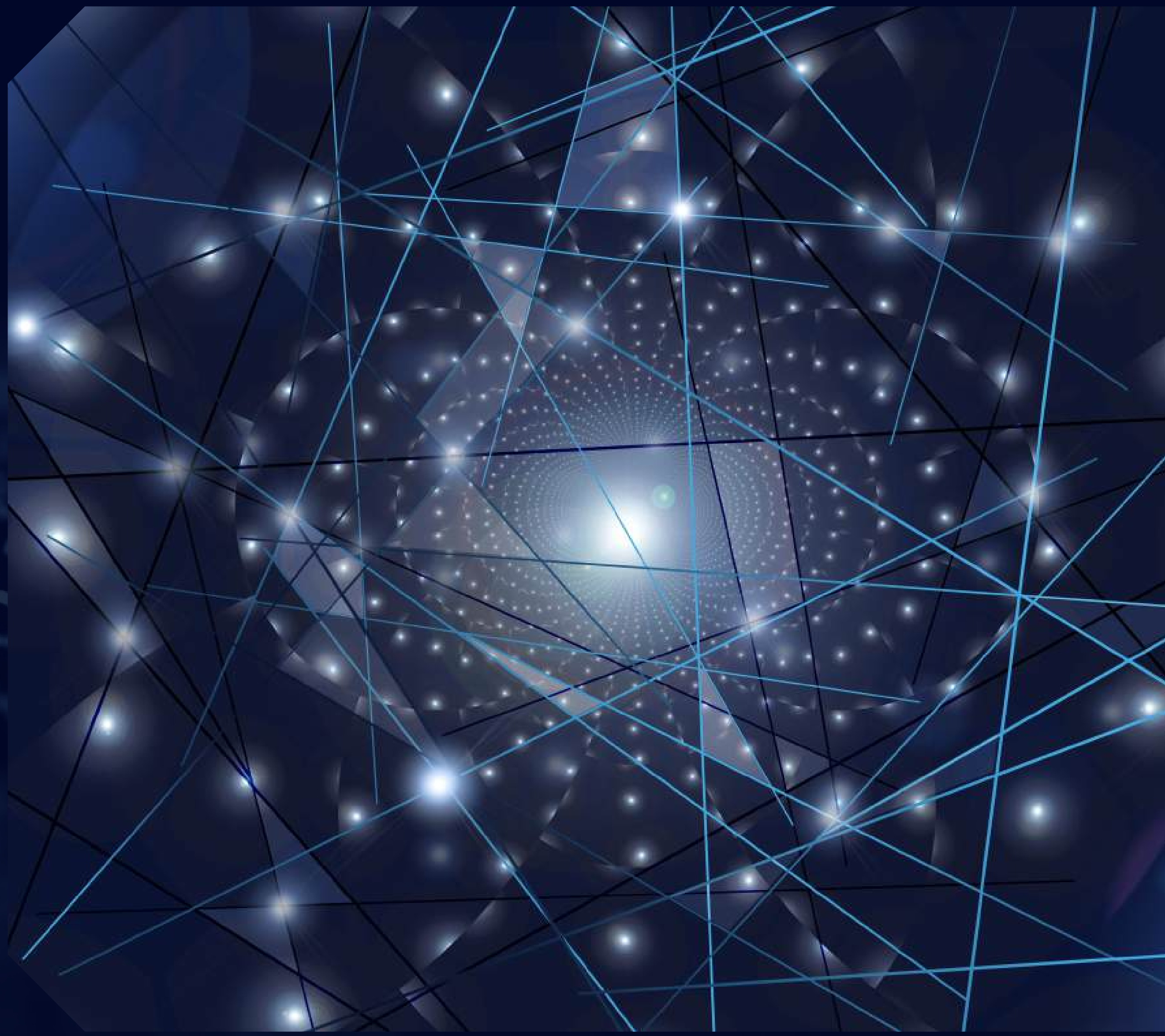
Monitor and Respond Quickly to Negative Feedback

Business Recommendation 3

Align Product and Marketing Strategies with Sentiment Trends

NEXT STEPS

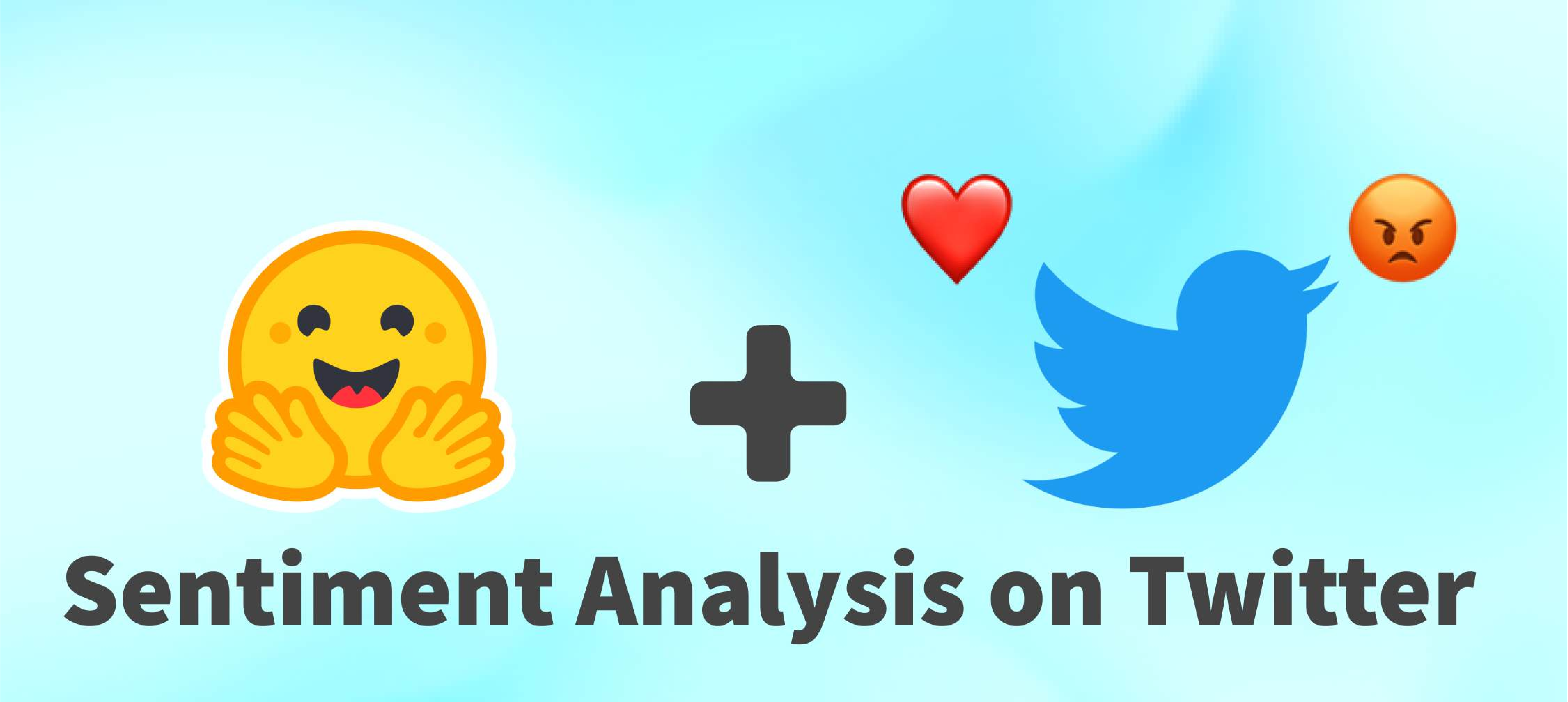
- **Model Deployment** : Integrate the best-performing sentiment analysis model into a real-time social media monitoring tool to track public opinions on Apple and Google products.
- **Continuous Learning** : Continuously retrain the model with new Twitter data to adapt to evolving language, trends, and user behavior.
- **Feature Expansion** : Incorporate additional variables such as hashtags, tweet metadata, and engagement metrics to improve prediction accuracy.
- **Business Intelligence Integration** : Create interactive dashboards for marketing and strategy teams to visualize sentiment patterns and identify key consumer insights.
- **Multilingual & Cross-Platform Extension** : Extend the analysis to other social media networks and multiple languages for a more global understanding of brand sentiment.



THANK YOU

Together, we can better understand public sentiment and transform data into meaningful insights.

- **Fourth Project: ADVANCED MACHINE LEARNING – Phase 4**
- **Instructors:** Wedter JEROME & Geovany Batista Polo LAGUERRE
- **GitHub Repository**: https://github.com/richefleuriord/Ds_Twitter_Sentiment_Analysis.git


Sentiment Analysis on Twitter