NLP SENTIMENT CLASSIFICATION PROJECT

A Data-Driven Apple & Google Product Sentiment Analysis

Table of contents

1

INTRODUCTION

Project Overview

Objective: Develop an NLP model that automatically classifies sentiment (Positive, Negative, Neutral) from tweets directed at Apple or Google products.

Data: Apple & Google Tweet Sentiment Dataset on Kaggle

Audience:

Product & Marketing Teams at Apple and Google
Customer Experience & Social Media Teams

Business Understanding

Stakeholder Need

Product, marketing & customer experience teams require early detection of public sentiment to:

- Track reactions to product launches
- Identify rising dissatisfaction
- Benchmark Apple vs Google perception

Business Understanding

Key Business Questions?

- What emotions dominate across Apple/Google tweets?
- Which products trigger positive or negative emotions?
- What keywords are associated with negative sentiment?

Success Criteria

A scalable and automated system enabling real-time tracking of brand sentiment

2

METHODOLOGY

Data Source

- The dataset used in this analysis is sourced from the Apple & Google Tweet Sentiment Dataset on Kaggle.
- It contains information on **9,093 tweets** collected for sentiment analysis.
- The dataset is structured for a **multiclass classification problem**, with the target variable sentiment label indicating emotion expressed on tweets.

Data Summary

Tweets collected: 9,093 (for sentiment analysis)

Columns: tweet_text, target product, sentiment label

Cleaned dataset: 3,282 samples (duplicates & missing removed)

Final sentiment classes:

- Positive: 2,664
- Negative: 518
- Neutral: 100

Key Characteristics

Tweets are short: most between 10–15 words and Apple dominates: 73% Apple vs. 27% Google.

Data Cleaning

- Removed **22 duplicates**
- Dropped **null rows**
- Renamed columns for clarity
- Converted product labels to lowercase
- Classified products into companies: Apple or Google
- Merged small class “I can’t tell” → Neutral
- Normalized text: Removed URLs, mentions, hashtags, removed stopwords and lemmatized words

Modeling

Feature Engineering:

- Text preprocessing
- TF-IDF vectorization (tweets → numerical representation)

Models Trained:

- Logistic Regression
- Random Forest Classifier
- Naive Bayes
- Support Vector Machine (SVM)

Evaluation Metrics:

- Accuracy
- Precision
- F1-Score

Tools

Python

Core Programming Language

Pandas & Numpy

Numerical Analysis & Data processing

NLP &Modelling

NLTK, TF_IDF, TensorFlow

Seaborn & Matplotlib

Visualization

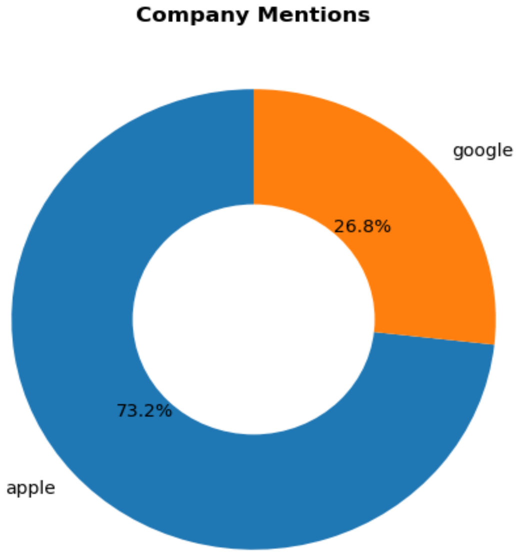
Jupyter Notebook

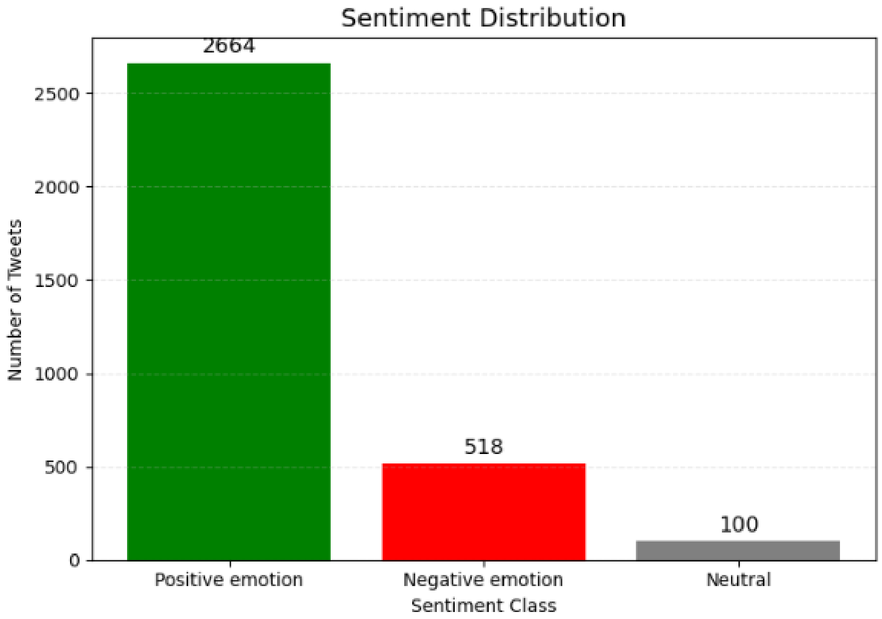
Interactive analysis and Documentation

3

RESULTS

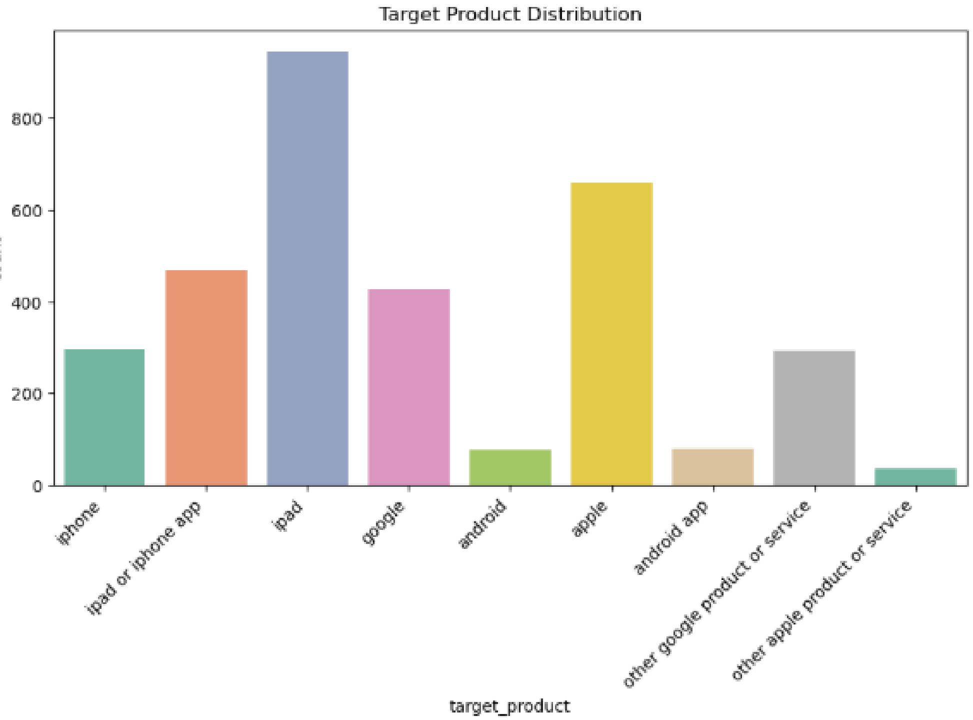
Dataset Distribution





- Apple dominates conversations, accounting for 73.2% of all tweets, while Google is mentioned in 26.8%.
- Sentiment is strongly skewed toward positivity.
- Overall, users talk far more about Apple, and the majority of tweets across both brands express positive sentiment.

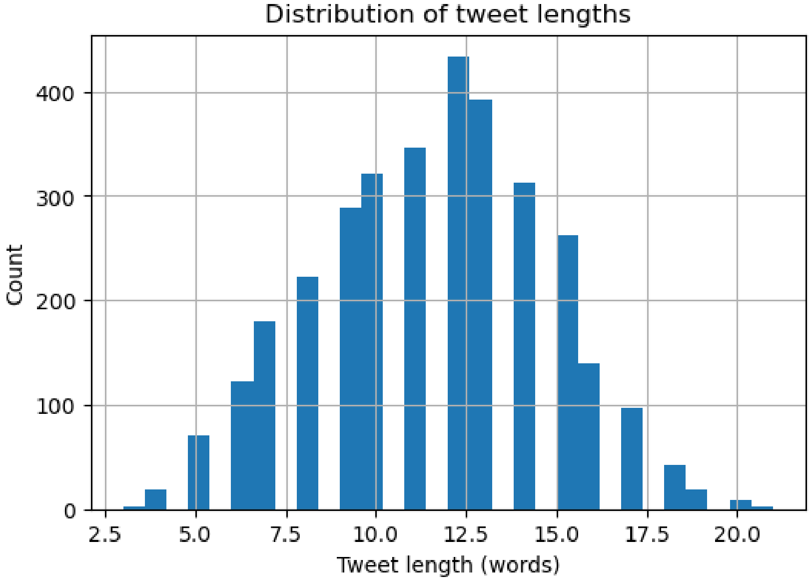
Target Product Distributions

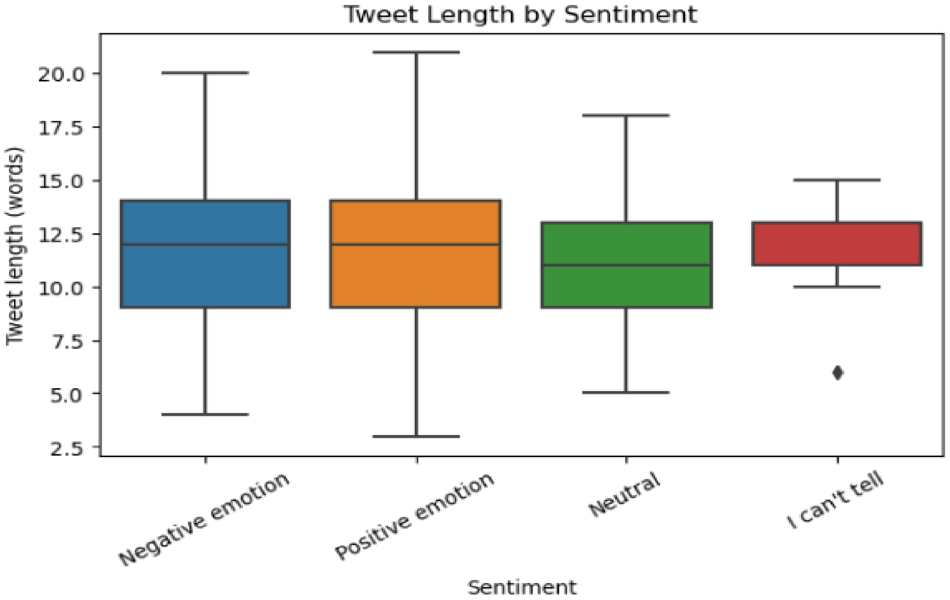


- **Most mentioned**: iPad
- **Followed by**: Apple, iPad/iPhone App, Google
- **Less frequent mentions:** Android, Android App, other Apple/Google services
- **Insight:**

User discussions are heavily centered on Apple's core products, with fewer references to Android or secondary services

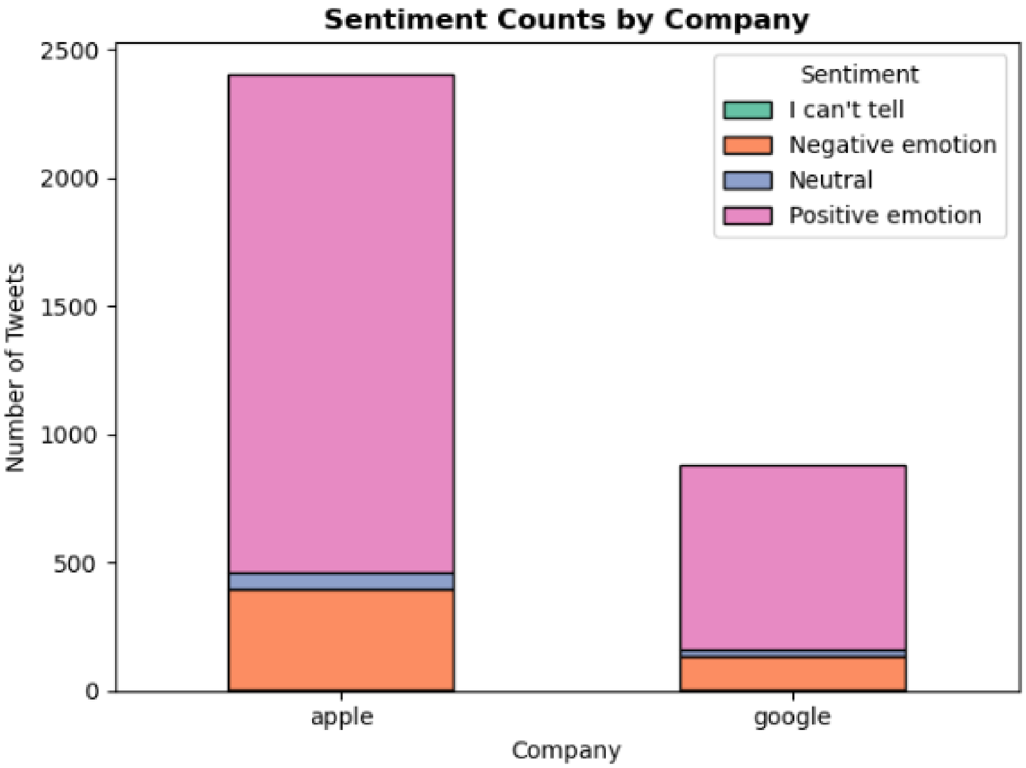
Tweet Length Analysis





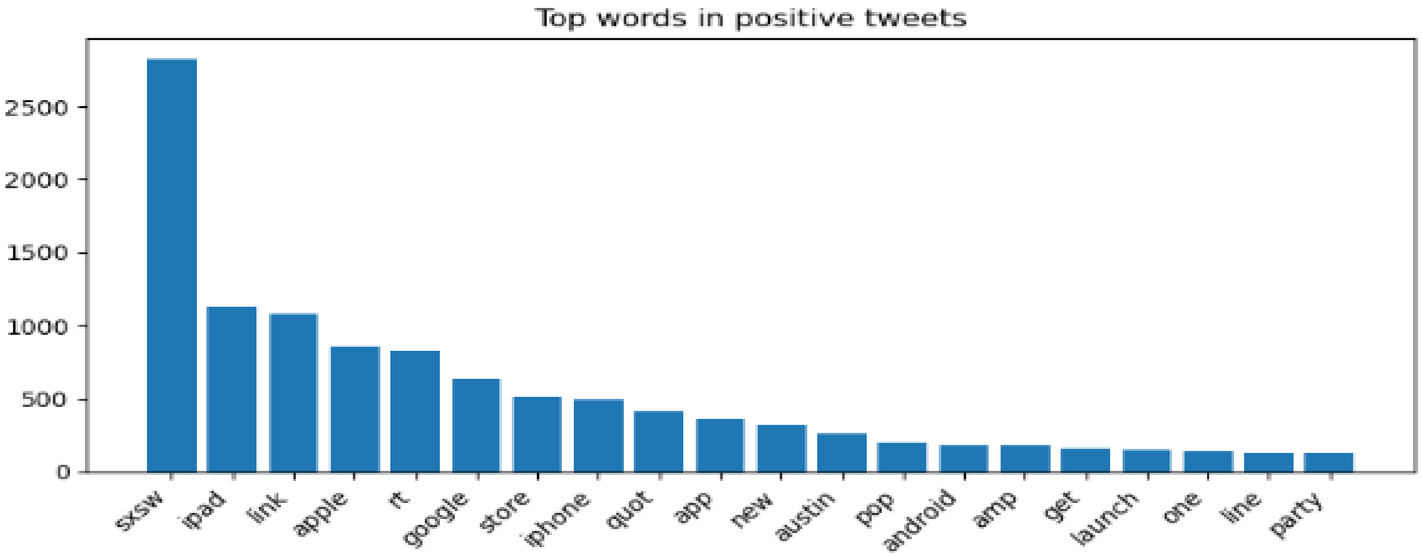
- Tweet lengths are similar across sentiments; **Neutral** tweets slightly shorter, "I can't tell" most consistent.
- **Sentiment does not strongly affect** how long users' messages are.
- Most tweets contain **10–15 words**, showing concise user expression
- Distribution is **roughly normal**, centered on medium-length tweets.

Sentiment by Company



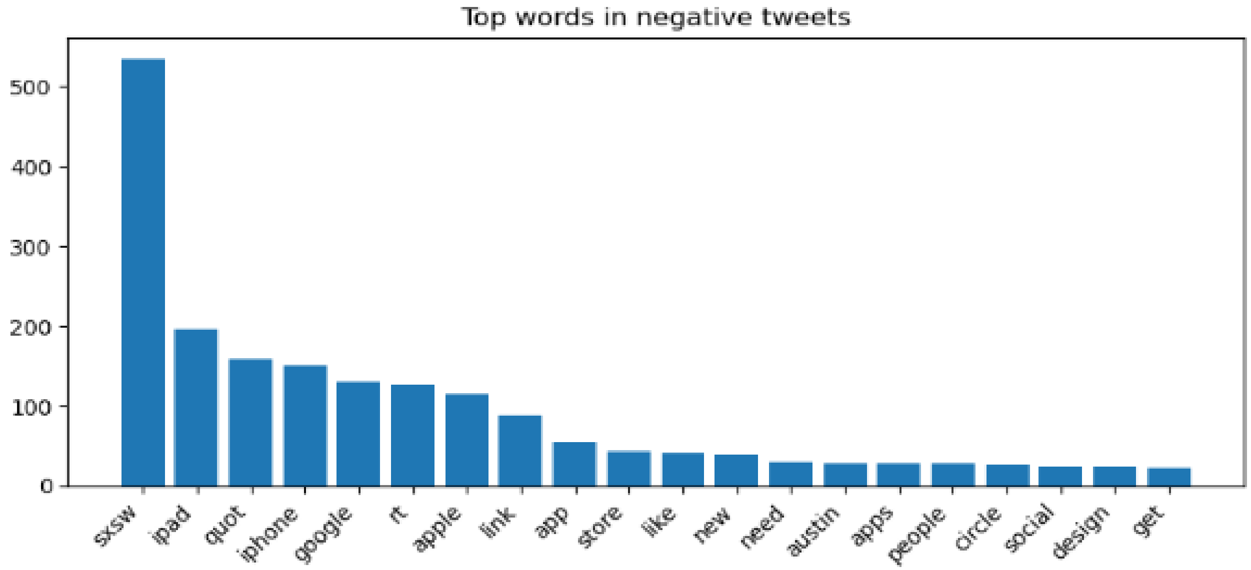
- **Positive sentiment dominates** overall, especially for Apple
- Apple receives **more tweets** than Google, with many favorable messages
- **Negative & neutral** sentiments are lower for both companies
- "I can't tell" category is minimal → most tweets show a clear emotional tone

Top Words in Positive Tweets



- Positive tweets highlight **sxsw, ipad, link, and apple**
- Indicates excitement around **events**, **products**, and **updates**
- Apple-related terms dominate positive discussions.
- Shows users’ tendency to **share enthusiasm and news**.

Top Words in Negative Tweets



- Negative tweets also mention ipad, iphone, and store, but in different contexts
- Positive tweets show a **broader variety** of frequent terms.
- Negative tweets have **fewer, more concentrated word patterns**.

Key Insights & Model Comparison

Best Overall Accuracy

- SVM (Combined Features) — 86.45%
- Logistic Regression (TF-IDF, GridSearch) — 86.19%

2 Best Macro F1 (Fairness Across Classes)

- SMOTE + Random Forest — 0.49
- Tuned Naive Bayes (α = 0.1) — 0.48
- SMOTE + XGBoost — 0.463

Key Insights & Model Comparison

Best Weighted F1 (Majority Performance)

- SVM, Logistic Regression, and Tuned Naive Bayes score ~0.83–0.86
- Strong performance on the dominant Positive

Neutral Class Remains the Weakest

- Recall stays between 0.00–0.10 for all models
- Caused by very few samples and overlapping language

Final Takeaways

- **Best Overall:** SVM (Combined Features)
- **Best Class-Balanced Model:** SMOTE + Random Forest
- **Best Simple Text Baseline:** Multinomial Naive Bayes ($\alpha = 0.1$)
- **Best Linear Model:** Logistic Regression (TF-IDF Only)

4

RECOMMENDATIONS

Leverage Positive Engagement

- Amplify themes around iPad/iPhone excitement
- **Use SXSW-style events to boost visibility(aligns with top-word insights).**

Strengthen Product Support

- Negative tweets frequently mention: battery, crash, slow, update issues
- **Improve update rollout, app stability, and troubleshooting guides.**

Monitor High-Risk Topics in Real Time

- Set alerts for spikes in negative words
- Auto-detect trending complaints using bigram analysis

Competitive Benchmarking

- Apple dominates sentiment volume - Google can study Apple-related positive triggers.
- Google should capitalize on its lower negativity levels per tweet.

THANK YOU

Prepared by: KIPROP BRIAN
Email: brian.kibor@student.moringaschool.com
Questions? I'm available for walkthroughs or further data analysis sessions.