



richefleurior / Ds_Twitter_Sentiment_Analysis



Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Set

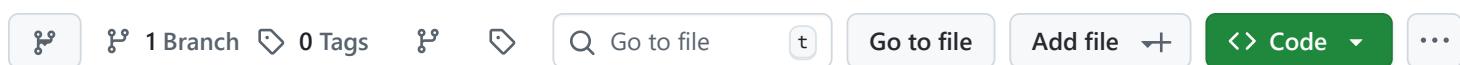


NLP project analyzing Twitter sentiment toward Apple and Google products using machine learning models (Naive Bayes, SVM, Random Forest, Logistic Regression). Includes data preprocessing, model evaluation, and business insights.

MIT license

0 stars 0 forks 0 watching Branches Activity Tags

Public repository



	richefleurior	Ajout de nouvelles modifications	eb72441 · 18 hours ago	
	.ipynb_checkpoints	Ajout de nouvelles modifications	18 hours ago	
	Data	Initial commit - Twitter Sentiment ...	2 days ago	
	images	Initial commit - Twitter Sentiment...	2 days ago	
	Final_Phase4.ipynb	Ajout de nouvelles modifications	18 hours ago	
	Final_Phase4.pdf	Ajout de nouvelles modifications	18 hours ago	
	Final_Presentation.pdf	Ajout de nouvelles modifications	18 hours ago	
	LICENSE	Initial commit	2 days ago	
	README.md	Update README.md	2 days ago	



Credit Card Fraud Detection: A Machine Learning Approach for Financial Institutions

Student: Riché FLEURINORD

Program: Data Science & AI – First Cohort (2025)

Project: ADVANCED MACHINE LEARNING - Phase 4 GitHub Repository:

https://github.com/richefleurinord/Ds_Twitter_Sentiment_Analysis.git

Instructors: Wedter JEROME & Geovany Batista Polo LAGUERRE

Submission Deadline: September 17, 2025

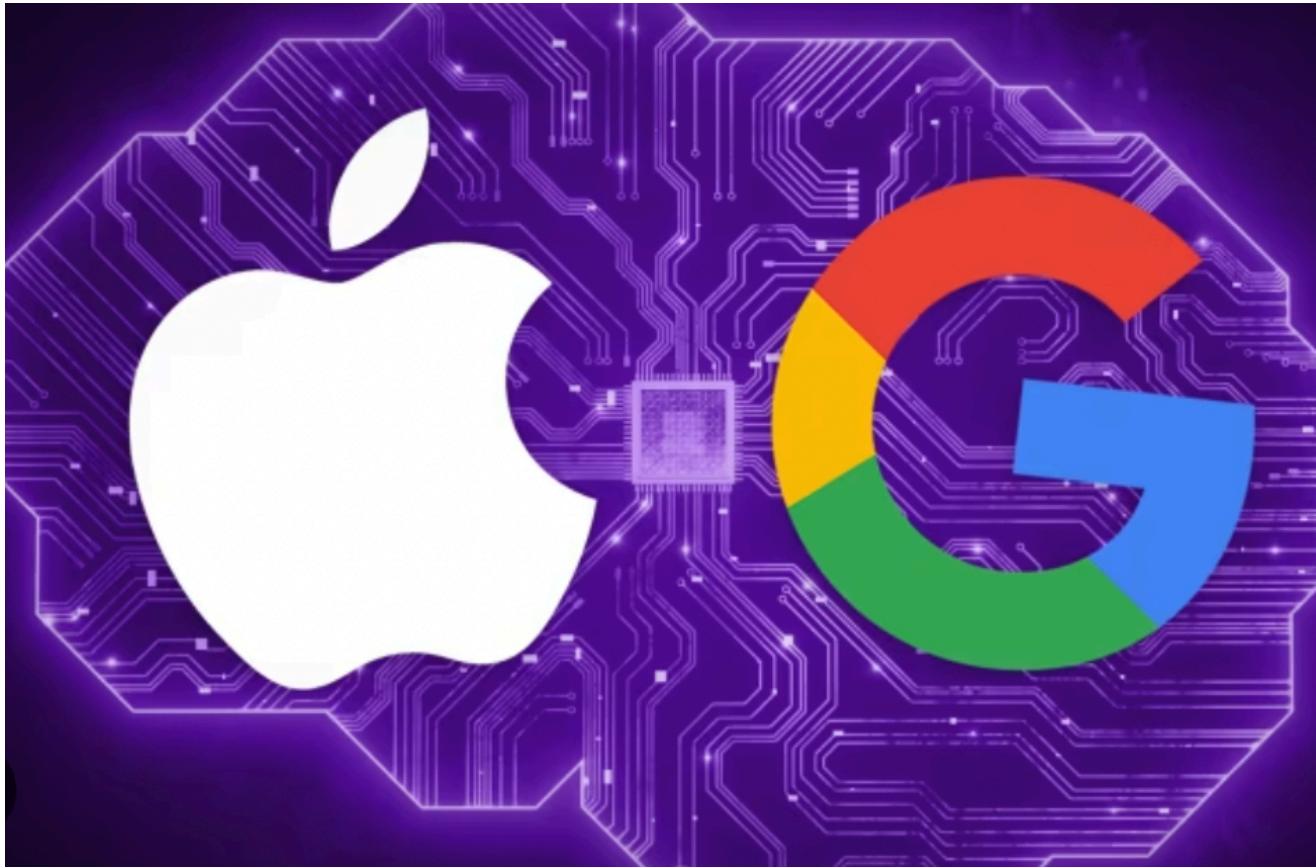
Natural Language Processing Approach to Twitter Sentiment Analysis: Apple and Google products



🔍 Overview

This project focuses on analyzing public sentiment toward Apple and Google products by leveraging Natural Language Processing (NLP) techniques. The primary objective is to develop a robust model capable of classifying tweets as positive, negative, or neutral based on their textual content. By doing so, the project aims to provide insights into consumer perceptions and opinions, offering a deeper understanding of public sentiment toward these leading technology companies.

Business Understanding



In the competitive landscape of technology giants, Apple and Google, public sentiment plays a crucial role in shaping corporate strategies and brand perception. This project focuses on analyzing the dynamics of sentiment expressed on Twitter regarding these companies. With a substantial global presence, both Apple and Google are influenced by various factors, including product launches, technological innovations, and market developments. In the previous fiscal year, Apple reported revenues of **383.3 billion dollars**, while Alphabet, Google's parent company, generated **307.4 billion dollars**. Effectively monitoring and interpreting public sentiment can provide valuable insights to guide marketing strategies, product development, and strategic decision-making.

The real-world problem addressed in this study is the challenge for Apple and Google to extract actionable insights from the vast volume of public sentiment available on Twitter. Understanding customer sentiment is critical, as it directly impacts product development, marketing effectiveness, and overall brand reputation. Negative sentiment can highlight areas requiring improvement, while positive sentiment can indicate successful strategies and initiatives.

Key stakeholders for this project include Apple and Google, along with their marketing and product development teams. Additionally, shareholders, investors, and the broader consumer base interested in these companies' performance can derive benefit from the insights generated. By analyzing Twitter sentiment, the project delivers tangible value, enabling data-driven decision-making, identifying potential challenges, and informing product and marketing strategy refinements.

Main Objective

The primary goal of this project is to comprehensively analyze Twitter sentiment data related to Apple and Google. This analysis aims to provide stakeholders within these organizations with valuable insights into the ebb and flow of public sentiment. Understanding when and why sentiment shifts occur can guide product development strategies, marketing campaigns, and brand management.

Specific Objectives

- Explore and preprocess the dataset, including handling missing values and transforming features.
 - Perform exploratory data analysis (EDA) to gain insights into the distribution and relationships between different features and the target variable.
 - Build binary and multiclass classification models and evaluate their performance using appropriate metrics.
 - Interpret the results of the models to understand sentiment patterns and key drivers.
 - Provide actionable recommendations to stakeholders based on the insights gained from the modeling process.
-

Data Preparation

In this stage, we prepared the dataset for sentiment analysis by:

- Removing irrelevant columns such as `emotion_in_tweet_is_directed_at`
- Dropping ambiguous sentiment categories ("I can't tell", "No emotion toward brand or product")
- Cleaning and standardizing text: removing URLs, hashtags, punctuation, converting to lowercase, and applying tokenization and lemmatization
- Handling missing values and duplicates
- Vectorizing text using **TF-IDF**, converting raw text into numerical format for model training



Sentiment Analysis on Twitter

Summary:

This process ensured that our data was clean, consistent, and machine-readable, forming a solid foundation for the modeling phase.

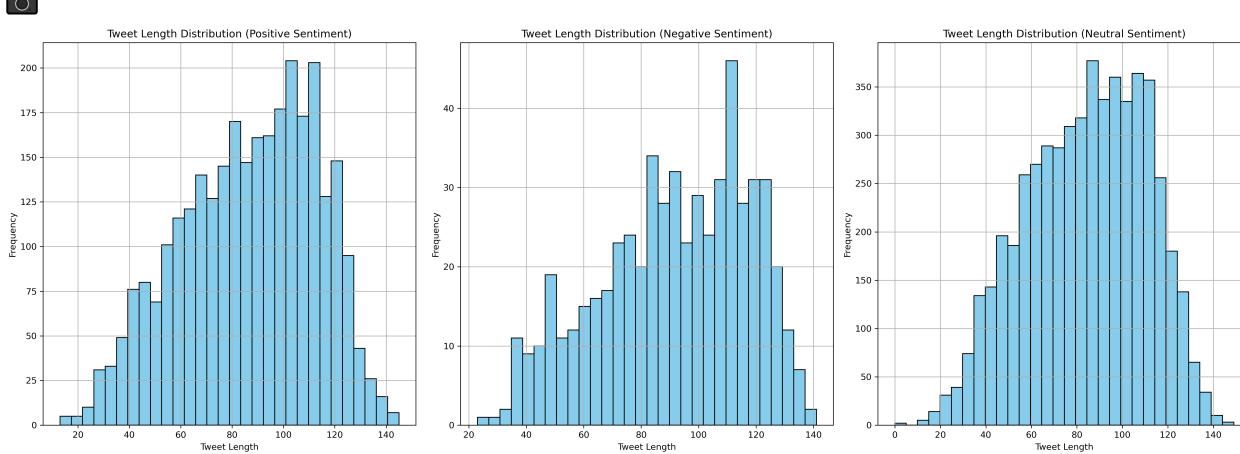
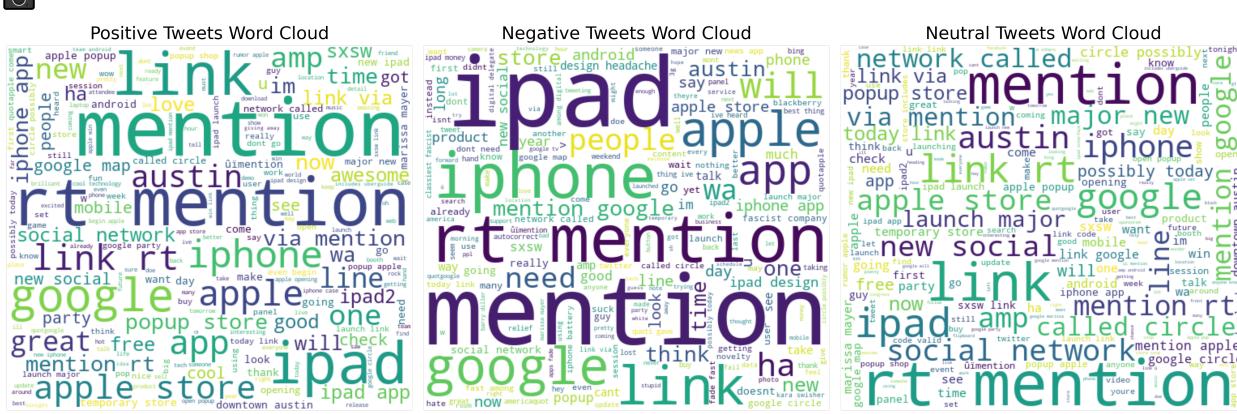
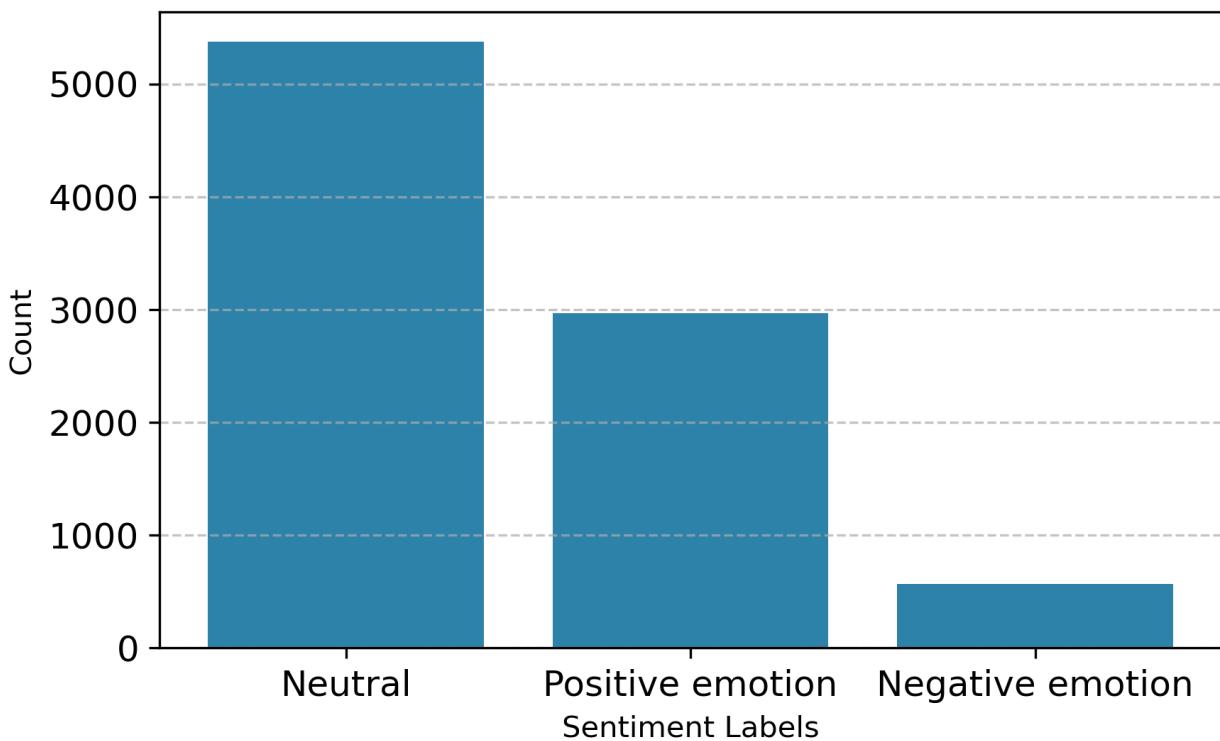
Exploratory Data Analysis (EDA)

We conducted an in-depth Exploratory Data Analysis to better understand the sentiment distribution and linguistic structure of tweets.

Key steps included:

- Visualizing sentiment distribution across positive, negative, and neutral classes
- Creating **word clouds** for each sentiment to reveal dominant words and themes
- Analyzing tweet lengths to study message density and tone

Distribution of Emotion Labels



Summary:

EDA provided both quantitative and qualitative insights into sentiment trends, word frequency, and message patterns, guiding our choice of preprocessing techniques and models.

Modelling

We employed several machine learning algorithms to classify tweet sentiment effectively:

- **Baseline Model – Naive Bayes:** Established a benchmark for performance.
- **SVM and Random Forest:** Used to handle class imbalance and improve recall for negative tweets.
- **Logistic Regression:** Evaluated for balanced accuracy across sentiment categories.
- **Hyperparameter Tuning (GridSearchCV):** Applied to optimize model parameters and improve results.

The models were tested in both **binary** and **multiclass** setups to ensure robustness and scalability.

Evaluation

Binary Sentiment Analysis:

The **Random Forest** model achieved the best results with an accuracy of **88.84%**, effectively balancing precision and recall between positive and negative sentiments.

Multiclass Sentiment Analysis:

Logistic Regression outperformed others with an accuracy of **69.1%**, demonstrating consistent and stable performance across sentiment categories.

Our findings highlighted key influencing factors such as frequent keywords, tone polarity, and engagement volume — offering actionable insights for marketing, product design, and investor analysis.

Recommendations

1. Highlight and Promote Positively Perceived Product Features

Positive tweets reveal the features and aspects of products that excite consumers. Apple and Google can leverage these insights by emphasizing these strengths in targeted marketing campaigns and social media posts, enhancing positive perception and driving product adoption.

2. Monitor and Respond Quickly to Negative Feedback

Although negative sentiments are less frequent, they are often focused on specific product issues. Companies should implement social listening systems to quickly identify recurring problems and respond proactively, whether through product improvements or personalized user engagement, to mitigate negative impact on brand reputation.

3. Align Product and Marketing Strategies with Sentiment Trends

Sentiment trends reveal consumer expectations and preferences. Integrating these insights into product development and marketing strategies allows Apple and Google to better align offerings with market needs. Features generating positive sentiment can be highlighted in future releases, while sources of dissatisfaction can be addressed or repositioned.



Next Steps

Future work will focus on enhancing the sentiment analysis framework and expanding its practical applications. Specifically:

- **Model Deployment:** Integrate the best-performing model into a real-time social media monitoring system to track public sentiment on Apple and Google products dynamically.
- **Continuous Learning:** Regularly retrain the models with new Twitter data to adapt to evolving language trends, emerging topics, and shifting public opinions, ensuring the system remains accurate and relevant.
- **Feature Expansion:** Explore additional data features such as tweet metadata, user engagement



Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

- Jupyter Notebook 100.0%
-