










Gladwellchebelyon /
Twitter-Sentiment-Analysis



[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)



 GPL-3.0 license










 0 stars  0 forks  1 watching  Branches  Tags  Activity




 Public repository

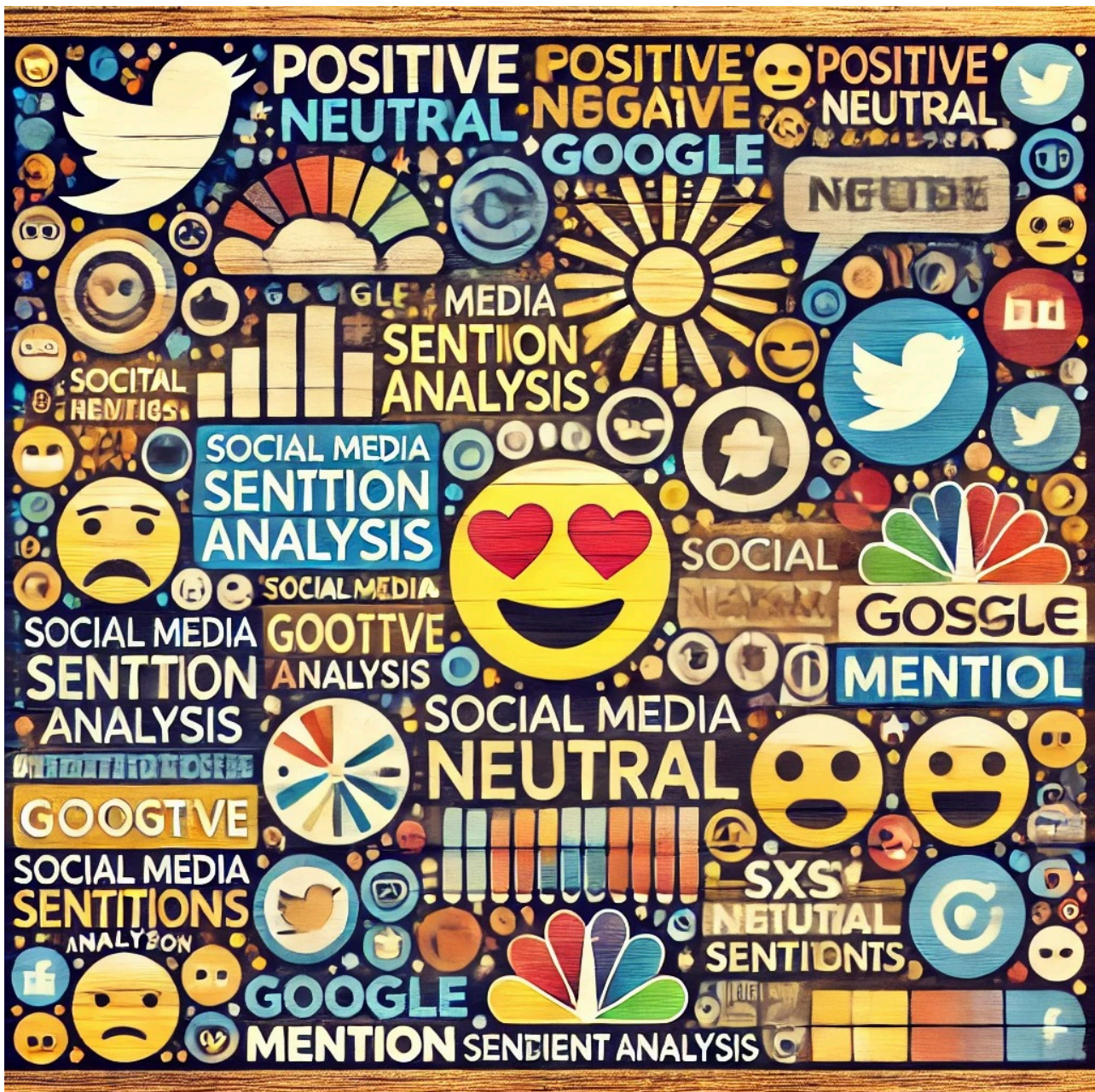
 5 Branches  0 Tags  



[Go to file](#) [+](#) [Add file](#) [Code](#) [...](#)

 Gladwellchebelyon presentation.pdf	7b5dea0 · 4 minutes ago	
 .gitignore	First commit	4 days ago
 LICENSE	Initial commit	4 days ago
 README.md	Read	1 hour ago
 judge-1377884607_tweet_pro...	Initial Commit	4 days ago
 presentation.pdf	presentation.pdf	4 minutes ago
 twitter_sentiment_analysis.ipynb	Read	1 hour ago
 twitter_sentiment_analysis.pdf	notebook.pdf	5 minutes ago

[README](#)  GPL-3.0 license  



Project Title: Sentiment Analysis of Tweets Related to Apple and Google Products

Table of Contents

1. Project Overview
2. Problem Statement
3. Dataset
4. Project Workflow
5. Models Used
6. Results
7. Conclusions and Recommendations
8. Installation



9. Usage

10. Authors

11. Acknowledgments

Project Overview

This project focuses on analyzing the sentiment of Tweets related to Apple and Google products using natural language processing (NLP) techniques. The goal is to classify the sentiment of each Tweet as positive, negative, or neutral, and understand the underlying opinions expressed in these posts.

Problem Statement

The main problem is to determine whether Tweets related to Apple and Google products carry a positive, negative, or neutral sentiment. This analysis aims to provide insights into public perception and help companies understand customer opinions about their products.

Dataset

Source: The dataset is from CrowdFlower and contains over 9,000 Tweets related to Apple and Google products.

Features: Includes Tweet_text and sentiment among other features.

Project Workflow

The project follows the typical steps of a data science workflow:

Business Understanding: Understanding the purpose of sentiment analysis for these tech products.

Data Understanding: Exploring the dataset and understanding its features and structure.

Data Preparation: Cleaning the dataset, handling missing values, and preparing text for modeling (e.g., vectorization).

Exploratory Data Analysis (EDA): Visualizing the data, including the use of WordClouds for each sentiment class.

Modeling: Training multiple models, including Logistic Regression, Decision Trees, Random Forest, XGboost classifier, and Neural Networks.

Evaluation: Evaluating the models based on accuracy, precision, recall, and F1 score.

Insights and Recommendations: Deriving conclusions and suggesting the best-performing model for deployment.

Models Used

Logistic Regression: Used as a baseline for both binary and multiclass classification without SMOTE and with SMOTE.

XGBoost: Both untuned and tuned improved performance

Decision Trees classifier: Achieved the best accuracy for multiclass classification(80.88%)

The models above were all used in Binary and Multiclass classification.

Results

For Binary Classification, XGBoost achieved the best accuracy of 87.79%.

For Multiclass Classification, Decision Trees achieved an accuracy of 60.24 %.

WordClouds were generated to visualize the distribution of positive, negative, and neutral sentiments.

Conclusions and Recommendations

Binary Sentiment Analysis: XGBoost is recommended for use due to its superior accuracy. It can be deployed for real-time sentiment analysis where a simpler binary outcome (positive/negative) is needed.

Multiclass Sentiment Analysis: Decision Trees performed best, but there is room for improvement. Further hyperparameter tuning and feature engineering may help enhance performance.

Future Work: Explore ensemble methods or hybrid approaches to potentially boost performance in multiclass classification.

Installation

To replicate this project, follow these steps:

1. Clone the repository.

bash terminal

```
git clone https://github.com/Gladwellchebelyon/Twitter-Sentiment-Analysis.git
```

3. Install the required dependencies.

bash terminal



Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Contributors 2



Ogam-Alvin



Gladwellchebelyon Gladwell Chepkorir

Languages

● Jupyter Notebook 100.0%