

Social Media Sentiment Analysis

Introduction

This project analyzes sentiment patterns in social media data to uncover public opinion and attitudes toward specific topics or brands. Using state-of-the-art Natural Language Processing (NLP) techniques, such as TF-IDF and sentiment analysis with VADER, this project provides insights into positive, negative, and neutral sentiments expressed in social media posts. The results are visualized to help stakeholders understand audience perceptions and track sentiment trends over time. This repository includes the code, dataset, and detailed documentation to guide you through the analysis process.

Table of Contents

- [Social Media Sentiment Analysis](#)
- [Introduction](#)
- [Table of Contents](#)
- [Features](#)
- [Getting Started](#)
 - [Installation](#)
- [File_Illustration](#)
 - [Data_Collection](#)
- [Dataset_Description](#)
 - [Column_Definitions](#)
 - [Data_cleaning](#)
 - [Data_Analysis](#)
 - [Sample_Outputs](#)
- [How_to_Contribute](#)
- [License](#)

Features

- EDA_Twitter
- Entity-Wise Sentiment Analysis
- Sentiment Distribution
- TF-IDF for keyword extraction

Getting Started

Installation

Follow the instructions in the setup guide to install dependencies and run the analysis. Check out the examples provided for guidance on replicating the analysis or customizing it for your own data.

To run this repo, be sure to install the following environment and library:

Python: 3.12.3

- **TensorFlow:** 2.18.0

- **Keras:** 3.6.0
- **Pandas:** 2.2.2
- **NumPy:** 1.26.4
- **h5py:** 3.12.1
- **scikit-learn:** 1.5.2

And also the **important libraries** for this project are

- [libraries.ipynb](#): Jupyter notebook listing libraries used in the project.

File_Illustration

Data_Collection

Data is collected from poloniex and parse to h5py file:

You can use the data uploaded in this repository [twitter_training.csv](#): or you can access the dataset used in this project on Kaggle: [Twitter Entity Sentiment Analysis Dataset](#).

Dataset_Description

Sample of the data set:

iD	entity	sentiment	text
0	2401	Borderlands	Positive
1	2401	Borderlands	Positive
2	2401	Borderlands	Positive
3	2401	Borderlands	Positive
4	2401	Borderlands	Positive

Column_Definitions

- **iD**: A unique identifier for each entry.
- **entity**: The entity being referred to, as example, "TomClancysRainbowSix, MaddenNFL, Borderlands.LeagueOfLegends,CallOfDuty,Verizon,CallOfDutyBlackopsColdWar,Facebook...."
- **sentiment**: The sentiment expressed in the text (Positive,Neutral,Negative).
- **text**: The user-generated text reflecting their opinion.

Data_cleaning

Before diving directly ton data analysis it is good to have readily available data fro further operations, so data cleaning python file is included in the repository.

- [loading_data_And_Data_Cleaning.ipynb](#): Jupyter notebook for loading and cleaning the data.

Data_Analysis

- Exploratory data analysis on Twitter data [EDA_Twitter.ipynb](#):

- sentiment analysis by entity [Entity-Wise Sentiment Analysis.ipynb](#)
- TF-IDF keyword extraction, [TF-IDF for keyword extraction.ipynb](#)

Sample_Outputs

This repository contains sample outputs of the data analysis

- Negative sentiment words [Negative Sentiment Words.png](#)
- Entity-wise sentiment analysis [outputEntity-Wise Sentiment Analysis.png](#)
- Positive sentiment words [Positive Sentiment Words.png](#)
- sentiment distribution [Sentiment Distribution.png](#)

How_to_Contribute

If you'd like to contribute to this dataset or improve the analysis, feel free to open a pull request or submit issues.

License

This dataset is available under the [MIT License](#).