

Social Media Sentiment Analysis: Analyzing Public Sentiment on Twitter Using NLP

Author: Akash Yadav

Date: 02/10/2024

Organization: Brainwave Matrix Solutions

Abstract

This report presents an analysis of public sentiment on Twitter using Natural Language Processing (NLP) techniques. Social media data is a valuable resource for understanding opinions towards specific topics, products, or events. The project aims to analyze tweets, preprocess the text data, extract sentiment scores, and visualize sentiment trends over time. Using sentiment analysis algorithms, we can categorize tweets as positive, negative, or neutral, and thus gain insights into the general public opinion. The findings highlight important trends, spikes in sentiment, and provide a deeper understanding of the prevailing opinions on the selected topic.

Introduction

Social media platforms like Twitter have become an integral part of society, serving as channels where people express their thoughts, opinions, and emotions on a wide range of topics. In this context, sentiment analysis—also known as opinion mining—has become an essential tool for extracting useful information from these data streams.

This project aims to analyze Twitter data to understand public sentiment toward. By applying Natural Language Processing (NLP) techniques, we can classify the sentiment in each tweet as positive, negative, or neutral. The key objective is to gain insights into how opinions shift over time and identify key events or triggers that cause a change in public perception.

Data Collection

The data for this project was collected from Twitter using the Twitter API, with a focus on [specific keywords, hashtags, or topics]. The **Twitter Developer App** named "PublicSentimentAnalyzer" was configured with the necessary API keys to fetch relevant tweets. We collected tweets over a period of [insert duration], applying filters to focus on tweets in [specific language] and excluding retweets to ensure originality of the content.

Each tweet in the dataset includes:

- Tweet content (text)
 - Timestamp of creation
 - Number of likes, retweets, and replies (optional)
-

Preprocessing

Before performing sentiment analysis, it is essential to clean and preprocess the collected data. The following preprocessing steps were applied:

- **Tokenization:** Splitting the text into individual words or tokens.
 - **Stopword Removal:** Removing common words (e.g., "is," "the," "and") that do not contribute to the sentiment of a sentence.
 - **Lemmatization:** Converting words to their root form (e.g., "running" to "run").
 - **Handling Special Characters:** Removing links, hashtags, user mentions, emojis, and punctuation marks.
 - **Lowercasing:** Converting all characters to lowercase for uniformity.
-

Methodology

Natural Language Processing (NLP) Techniques

NLP techniques were applied to the preprocessed text data to extract meaningful insights and sentiment. Specifically, the **VADER** (Valence Aware Dictionary and Sentiment Reasoner) algorithm was used to perform sentiment analysis, as it is well-suited for social media text. The algorithm assigns a sentiment polarity score to each tweet, classifying it as either positive, negative, or neutral based on the content.

Sentiment Scoring

Each tweet was assigned a sentiment score based on its textual content:

- **Positive:** If the tweet expresses favorable or optimistic views.
- **Negative:** If the tweet conveys unfavorable or pessimistic views.
- **Neutral:** If the tweet is neutral or does not express a strong sentiment.

The score for each tweet was calculated, and a sentiment label was attached accordingly.

Visualization

To better understand sentiment trends over time, we used data visualization tools like **Matplotlib** and **Seaborn**. Line charts were generated to track changes in sentiment over the selected period, while bar charts provided an overview of the overall sentiment distribution.

Results

Sentiment Distribution

The analysis revealed that out of [X number of tweets], the sentiment distribution was as follows:

- **Positive:** [Percentage] of tweets
- **Negative:** [Percentage] of tweets
- **Neutral:** [Percentage] of tweets

This distribution highlights that the majority of tweets are neutral/positive/negative, indicating overall sentiment.

Trends Over Time

The sentiment trends were analyzed over a period of [insert time period], revealing notable fluctuations. There were significant spikes in sentiment during key events, such as [specific event] on [specific date], which caused a surge in positive or negative sentiment.

Visualization:

- A line graph visualizing sentiment trends over time shows peaks corresponding to major events.
 - A bar chart displaying the sentiment distribution.
-

Discussion

Analysis of Findings

The analysis provided insight into the public sentiment surrounding [specific topic/event]. For example, the spike in negative sentiment during even] may indicate dissatisfaction, while a rise in positive sentiment during another event shows increased public approval.

Comparison to Baseline

The sentiment analysis results were consistent with initial expectations, as public opinion was largely shaped by specific factors. However, some unexpected trends were observed, such as anomalies or unexpected patterns.

Limitations

There were several limitations to the analysis:

- **Data Bias:** The dataset may be biased due to the use of specific keywords or hashtags, which can skew sentiment analysis results.
- **Language:** Tweets in languages other than [language] were excluded from the analysis.
- **Model Limitations:** The VADER algorithm, while effective for social media, may not capture complex sentiments or sarcasm effectively.

Conclusion

In conclusion, this project demonstrated the effectiveness of sentiment analysis in understanding public sentiment toward topic/product/event on Twitter. The analysis revealed key trends, shifts in sentiment, and provided valuable insights into public opinion. Future work could focus on expanding the dataset or applying more advanced machine learning models to improve sentiment classification.