

Sentiment Analysis of Twitter Dataset

Anup Kumar Choudhary, Mohit Sharma, Jyotiraditya Khatua, Anshu Verma

Department of Computer Science and Engineering, Mahatma Gandhi Mission's College Of Engineering & Technology

Abstract- Understanding public opinion from social media platforms has become increasingly important, and sentiment analysis plays a key role in this process. Although sentiment analysis tools such as VADER and TextBlob are widely used, they often require programming skills and lack accessible interfaces for exploratory analysis. This paper introduces the design and workflow of an interactive web-based application developed to simplify sentiment evaluation of English tweets. The system combines lexicon-based sentiment models with a Streamlit dashboard to offer an easy-to-use platform for analyzing individual texts, processing bulk tweet datasets (CSV files), and visualizing outcomes using dynamic plots and word clouds. By integrating established analysis techniques into an intuitive interface, the application makes sentiment analysis approachable for users without technical backgrounds.

Keywords: Sentiment Analysis, Natural Language Processing (NLP), Social Media Analytics, Streamlit, Lexicon-based Modeling, Data Visualization, VADER (Valence Aware Dictionary and sEntiment Reasoner).

I. INTRODUCTION

With platforms like Twitter generating vast amounts of user-created content every second, organizations now have unparalleled access to public opinion. For researchers, marketers, and analysts, this continuous flow of text offers valuable insights into public attitudes and emotions. Sentiment Analysis (SA) helps extract these insights by automatically determining whether a given piece of text expresses positive, negative, or neutral sentiment.

Despite the availability of reliable Python libraries such as VADER and TextBlob for analyzing English text, these tools are typically geared toward users with programming experience. This creates a practical challenge: individuals who could benefit the most from sentiment insights—such as business professionals or social media strategists—often lack the technical skills required to use them effectively. As a result, the potential of sentiment analysis remains underutilized outside technical settings.

To address this gap, we developed an interactive web application using Streamlit that transforms complex sentiment analysis methods into a simple, user-centered interface. The system enables users to:

- Evaluate the sentiment of short or long text instantly
- Upload and analyze large tweet datasets without coding

- Explore results through automated, interactive visualizations and charts

This platform delivers a complete pipeline from data ingestion to analysis and interpretation, illustrating how backend analytical tools can be made accessible through a user-friendly frontend.

II. LITERATURE REVIEW AND RELATED WORK

In developing this project, one of the primary considerations was selecting an appropriate approach for sentiment analysis. Broadly, two methodologies dominate the field: lexicon-based techniques and machine learning-based models. Since our objective was to design a system that is simple, fast, and accessible to non-technical users, the lexicon-based route was chosen.

Lexicon-Based Approaches

Lexicon-driven sentiment analysis operates on predefined word lists where each term is assigned a sentiment value. The sentiment of a sentence is determined by combining these values, allowing the system to interpret emotional tone without any training data. This rule-oriented structure makes the approach lightweight and easy to implement.

For this project, two well-known lexicon-based tools were incorporated:

- **VADER (Valence Aware Dictionary and sEntiment Reasoner):** VADER is tailored for online communication and excels at interpreting informal language. It accounts for features commonly found in social media posts—such as uppercase emphasis, emojis, punctuation intensity, and slang. This makes it particularly suitable for analyzing the expressive and fast-paced nature of Twitter content.
- **TextBlob:** TextBlob provides a straightforward method for computing sentiment by returning a polarity score ranging from -1.0 to +1.0. It is widely used due to its simplicity and reliability, and serves as a consistent baseline for general sentiment evaluation.

Rationale for Choosing These Tools: Both VADER and TextBlob are ready-to-use, require no training data, and perform efficiently even on large text inputs. Their ease of integration aligns well with the goal of making sentiment analysis accessible without technical expertise.

Comparison with Traditional Machine Learning Methods

Conventional machine learning approaches—such as Naive Bayes, Support Vector Machines (SVM), or logistic regression—treat sentiment detection as a supervised learning task. Although these models can achieve strong accuracy, they depend on several resource-intensive steps, including dataset collection, manual labeling, and time-consuming model training. This additional overhead makes them less suitable for users who need immediate insights without the complexity of managing machine learning workflows

III. PROBLEM STATEMENT

Social media sites such as Twitter have evolved into public spaces where individuals instantly express opinions about events, products, and experiences. As a result, these platforms hold vast amounts of authentic user sentiment. However, this data is extremely unstructured. Tweets commonly contain informal language, abbreviations, emojis, hashtags,

and multilingual expressions, making them difficult to interpret manually.

Because of this complexity, traditional text-analysis methods often fail to extract reliable sentiment from such noisy and inconsistent data. The sheer volume and informality of social media text highlight the need for an automated system specifically optimized for processing and understanding this type of content. Developing a model that can categorize tweets as positive, negative, or neutral can offer organizations meaningful insights into public views without requiring them to sift through thousands of posts manually.

IV. SYSTEM ARCHITECTURE AND METHODOLOGY

The system is designed as an end-to-end web application that enables easy sentiment analysis of English social media data. All components—data handling, analysis, visualization, and storage—are integrated within a unified interface.

Technology Stack Overview

The project uses Python due to its dominance in data science and its extensive library ecosystem, which simplifies development and accelerates prototyping.

User Interface Layer - Streamlit

Streamlit is used to build the full interface without requiring HTML, CSS, or JavaScript. It automatically generates input fields, upload components, and page layouts, allowing the developer to focus solely on analytical logic rather than frontend coding.

Analysis Layer – VADER and TextBlob

These two sentiment engines form the core of the system's analytical capabilities:

- **VADER:** Designed for social media content, it captures nuances such as capitalization, emoji usage, repeated punctuation, and internet slang.
- **TextBlob:** Provides simple polarity and subjectivity scores, functioning as a general-purpose sentiment evaluator.

Data Processing – Pandas and NumPy

Real-world datasets are messy, and these libraries help handle them efficiently:

- **Pandas:** Used for reading CSV files, cleaning data, and preparing structured tables.
- **NumPy:** Supports numerical operations behind the scenes, enhancing performance.

Visualization – Plotly and WorldCloud

To present insights effectively:

- **Plotly:** Creates interactive charts, enabling users to explore results by hovering or clicking.
- **WordCloud:** Illustrates common words visually, helping users identify frequently discussed topics.

Storage Layer - SQLite

SQLite is integrated to store analysis history locally. This lightweight database requires no server setup and ensures that users can revisit previous results.

V. IMPLEMENTATION AND FEATURES

The Streamlit application is organized into three primary modules, each addressing a specific user requirement.

Single Text Analysis

Users can input a sentence or short passage. After selecting an analysis method (VADER or TextBlob), the system displays sentiment scores, intensity levels, and polarity details.

Each analysis is stored automatically in the SQLite database.

Batch Dataset Analysis

Users may upload CSV files containing tweet collections. After selecting the column that contains the text, the application evaluates each entry—using VADER by default—and appends new sentiment and confidence columns. The processed file is then available for download.

Visualization Dashboard

For batch analyses, the system generates interactive visualizations, including:

- A pie chart showing sentiment distribution

- A bar graph of average confidence per sentiment category
- Word clouds highlighting common positive and negative expressions

These visual tools simplify interpretation and reveal patterns that might be missed in tabular data.

VI. RESULTS AND DISCUSSION

The key achievement of this work lies in transforming established sentiment analysis libraries into a user-friendly platform accessible to non-programmers. Instead of requiring expertise in Python or NLP, users can upload a dataset and obtain meaningful insights within seconds. The generated visualizations—pie charts, bar graphs, and word clouds—provide actionable summaries that can support decision-making in marketing, customer service, and research.

By storing analysis history using SQLite, the application serves as a complete workspace for tracking and comparing sentiment trends over time. This project demonstrates that advanced analytical tools can be democratized when paired with intuitive interfaces

VII. FUTURE SCOPE

Several improvements can enhance the system's capabilities in future iterations:

- **Advanced NLP Models:** Integrating transformer-based models like BERT or RoBERTa to handle contextual understanding.
- **Multilingual Support:** Extending analysis to mixed-language tweets and regional dialects.
- **Image and Meme Analysis:** Adding sentiment recognition for visual content commonly shared on social platforms.
- **Sarcasm and Irony Detection:** Developing modules to identify sarcastic expressions, which are challenging for lexicon-based methods.
- **Real-Time Monitoring:** Enabling live tracking of sentiment trends for fast decision-making.

VIII. CONCLUSION

This work introduces an interactive sentiment analysis tool built with Streamlit that simplifies the

evaluation of English social media text. By combining user-friendly interface design with powerful sentiment engines such as VADER and TextBlob, the system enables users to conduct single-text analysis, batch processing, and visualization within one platform. The application successfully reduces technical barriers and demonstrates how accessible interfaces can bring sophisticated NLP tools to a broader audience.

REFERENCE

1. VADER Sentiment: Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225. View on AAAI.
2. TextBlob Library: Loria, S. (2018). TextBlob: Simplified Text Processing. Documentation on ReadTheDocs.
3. Streamlit Framework: Streamlit Inc. (2023). Streamlit: The fastest way to build and share data apps. Official Site.
4. Sentiment Analysis Survey: Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ains Shams Engineering Journal*, 5(4), 1093-1113. View on ScienceDirect.
5. Twitter/X Analytics: Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, 3, 1277–1291. View on Springer.