Social Media Sentiment Analysis

Ms.k.s.Surabhi [1],R.MAHESHWARAN[2]
Assistant-SGprofessor,DepartmentofComputerApplications,Nehruollegeof Management, Coimbatore, Tamil Nadu, India. ncmmssurabhi@nehrucolleges.com
StudentofIIMCA,DepartmentofComputerApplications,NehruCollegeofManagement, Coimbatore, Tamil Nadu, India. Waranm408@gmail.com

ABSTRACT

In the rapidly evolving digita lage, social media has emerged as one of the most dominant communication platforms worldwide. Every day, millions of people express their views, emotions, and reactions on various topics, ranging from personal experiences to political movements, product reviews, global events, and social trends. Platforms such as Twitter (now known as X),Facebook, and Instagram serve not only as tools for personal interaction but also as valuable sources of public opinion, crowd behavior, and sentiment-driven insights. Given the sheer volume and velocity of content generated on these platforms, it is humanly impossible to manually monitor or interpret the overall sentimentofusers. Thishasledtothe need for automated sentimentanalysis systems that can extract meaningful patterns and classify emotions in real time. Sentiment analysis, also known as opinion mining, isa powerful technique in the field of natural language processing (NLP) that helps in identifying and categorizing opinions expressed in textual data, especially to determine the writer's attitude as positive, negative, or neutral..

1. INTRODUCTION:

This project, titled "Social Media Sentiment Analysis Using Streamlit," is designed to address this need by developing an intelligent web application that automates the process of sentiment analysis on social media content. The project uses Python as its core programming language and leverages libraries such as TextBlob and VADER for sentiment classification. It also integrates Streamlit, an open-source Python framework, to create a dynamic, interactive web- baseduser interfacethatrequiresminimaltechnicalknowledge to use. The application allows users to analyze sentiments from multiple social media platforms. It supports two modes of data input:livedatascraping(especiallyfromTwitter/Xusin gsnscrap e) and uploading CSV files that contain precollected data from platforms like Facebook and Instagram. The text data collected through these methods undergoespreprocessing,whichincludesstepssuch as removing noise (URLs, mentions, special characters), converting to lowercase, and tokenization to ensure accurate and clean input for the sentiment analysis models.This project makes healthcare more intelligent and proactive by providing a system that continuously monitors, predicts, and helps save lives.

1.1 PROBLEM STATEMENT

Social media platforms like Twitter, Facebook, and Instagram are widelyused by people to share their thoughts, feelings, and opinions on different topics. Every day, millions of posts are made by users all over the world. These posts contain valuable information about public opinion, but it is very difficult to manually read and understand them due to the large amount of data. Many organizations, businesses, and researchers want to know what people think about their products, services, or current events. However, analyzingthis kind of data is a big challenge because:
The data isunstructured and comesin many different forms(text,emojis,hashtags, etc.).

- The data is unstructured and comes in many different forms (text, emojis, hashtags, etc.).
- There is a huge volume of content being generated every second.
- Most existing tools are difficult to use for people without technical knowledge.
- Many tools do not provide real-time results or proper visualizations.

as Twitter/X, Facebook, and Instagram. The system aims to provide users—be it businesses, researchers, analysts, or the general public— with real-time, datadriven insights into public opinion and emotional tone derived from large volumes of social media content.

This project seeksto bridge the gapbetween massive unstructured textual data and meaningful interpretation by implementing natural language processing techniques and advanced sentiment analysis models in a user- friendly web interface.

1.3 SCOPE OF THE PROJECT

This project focuses on building a Streamlit- based web application that performs sentiment analysison social media content from platforms like Twitter (X), Facebook, and Instagram. The application allows users to upload CSV files or fetch tweets using live scraping. It analyzes the textdataandclassifieseachpostaspositive, negative, or neutral using NLP tools like TextBlob and VADER.A

1.2 OBJECTIVES OF THE STUDY

The main objective of the project titled "Social Media Sentiment Analysis UsingStreamlit"isto designand develop an intelligent web application that can automatically analyze, classify, and visualize sentiments expressed by users on popular socialmedia platformssuch

Thes cope includes:

- Collecting data from social media(liveor uploaded).
- Preprocessing the data(cleaningtext, removing unwanted characters).
- Analyzing sentiment susing built-in sentiment analysis tools.
- Displaying results using charts,graphs,and word clouds.
- Allowing keyword-based filtering and time-based trend analysis.
- Supporting bulk analysis through CSV file uploads.
- Providingan easy-to-use web interface using Streamlit.

This project is intended for users like students,researchers,marketers,andda ta analysts who want to understand public opinion from social media data without writing code.

2. LITERATURE REVIEW:

Sentiment analysis is an important area in Natural Language Processing (NLP) that deals with understanding people’s emotions and opinions from text. In recent years, there has been a growing interest in analyzing social media content,as platforms like Twitter, Facebook, and Instagram have become the main sources of public opinion. Many researchers and developers have proposed various methods and tools to perform sentiment analysis efficiently.

Several studies have explored the use of machine learning algorithms such as Naive Bayes, Support Vector Machines (SVM), and Decision Trees for sentiment classification. These models are trained on labeled datasets where the sentiment (positive, negative, or neutral) of each text sample is already known. Once trained,these models can predict the sentiment of new, unseen data. While these approaches provide good accuracy, they require a large amount of training data and computational resources.

To overcome this, many developers prefer to use pre-built rule-based tools like TextBlob and VADER (Valence Aware Dictionary and sEntiment Reasoner). These tools use a predefined list of words and associated sentiment scores. They are lightweight and work well for short, informal texts, making them suitable for social media content. VADER is especially designed to handle text with emojis, slangs, and capitalization commonly found in tweets and posts.

Previous works have also emphasized the importanceof data preprocessing, such as removing URLs, mentions, hashtags, stop words, and convertingtexttolowercase.Cleaneddataimproves the accuracy of sentiment classification.

Some recent projects have implemented web- based sentiment analysis systems using frameworks like Flask and Django. However, building interactive interfaces in these frameworks can be complex and time-consuming. As an alternative, Streamlit has gained popularity due to its simplicity and speed. It allows Python developers to create interactive web apps with just a few lines of code.

3 PROPOSED SYSTEM:

The proposed system is a user-friendly, AI-powered sentiment analysis web application developed using Streamlit, a Python-based web framework. It is designed to analyzeand visualize the emotional tone of social media content sourced from platforms like Twitter, Facebook, and Instagram. Unlike traditional systems, this solution combines real-time datascraping, bulk file processing, and interactive data visualizations into a single lightweight application that is accessible to both technical and non- technical users.

- Awebfrontendforuserinteraction.
- A FastAPI backend for data processing.
- A machine learning model for prediction.
- Visualization modules for real-time dashboards.

This structure ensures high accuracy, easy maintenance, and fast response times.

3.1 SYSTEM ARCHITECTURE

SocialMediaSentimentAnalysis aims to determine the sentiment (positive, negative, neutral)expressed in user posts or comments across platforms like Twitter (X), Instagram,Reddit,orFacebook.

1. **Data Collection Layer Layer:**

Collects social media datavia APIs (e.g.,TwitterAPI, Reddit API) or web scraping. Tools: Tweepy, BeautifulSoup, or Scrapy.

2. **Data Preprocessing Layer:**

Cleans and normalizes text — removingURLs,emojis,hashtags,mentions, punctuation, and stop words. Also handles language detection and tokenization..

3. **Data Preprocessing Layer:**

Converts text into numerical features using methods such as Bag of Words, TF- IDF, or word embeddings (Word2Vec, GloVe, or BERT embeddings)..

3. **Sentiment Analysis Layer:**

The **Sentiment Analysis Layer** processes the **cleaned and preprocessed text** from the previous layer and applies **NLP (Natural Language Processing)** or **Machine Learning models** to:

- Identifythe **sentiment polarity** (positive,negative,neutral)
- Assign a **sentiment score or probability**
- Optionally, categorize into **emotion classes** (e.g. happy, angry, sad, surprise)

The database maintains updates and historical records of patient data, supporting continuous monitoring and enabling performance tracking for future model enhancements.

The interaction between the system layers ensures seamless data communication, real- time processing, and accurate decision support.

As shown in Fig. 2, the architecture visually depicts how each layer is interconnected— from data input and processing to final output visualization and secure database storage. This modular design makes the system scalable, reliable, and adaptable for future enhancements such as mobile access.

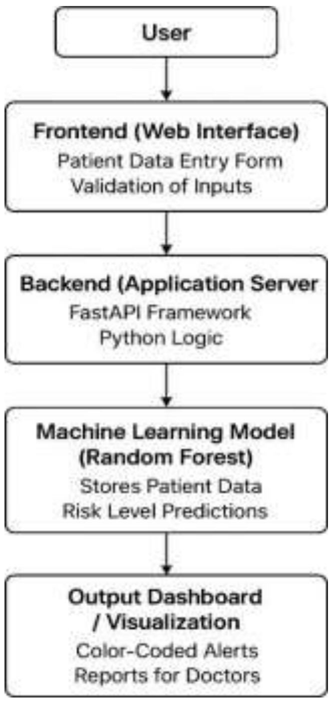

Fig.2: Common System Architecture

The figure 2 shows how patient data flows through the system. The web interface collects inputs, which are processed by the FastAPI backend and the Random Forest model to predict risk levels. The

- Data visualization techniques transform raw analytical outputs into understandable insights, allowing researchers to identify trends, sentiment distributions, and emerging topics

- Visualization also supports decision- making by presenting complex data patterns in a graphical and interactive form.

results are stored in the database and displayed on a dashboard with color-coded alerts for doctors.

3.2 DATABASEAN KNOWLEDGE BASE

The database stores social media post data, user information, and historical sentiment records.

- **UserTable:** Maintains login credentials, roles, and authentication details for analysts or administrators.

- **PatientTable: Contains attributes such as** post ID, user/source, text content, timestamp, platform, and predicted sentiment**..**

This structured **knowledge base** supports **quick retrieval, updating, and continuous learning** for model **refinement.**

4 .METHODOLOGY:

The study follows a **quantitative research design**, employing **Natural Language Processing (NLP)** and **machine learning** techniques to analyze large volumes of textual data. The process involves five main stages:

1. Data Collection
2. Data Visualization and Interpretation
3. Data Preprocessing.
4. Model Evaluation

4.1 ALGORITHMS AND TECHNIQUES USED

- After sentiment classification, the next step is to interpret and visualize the results in a meaningful way.

- Technique: **Sentiment Counts:** The total number of positive, negative, and neutral posts.

- **Percentage Distribution:** Calculated to understand sentiment proportions

- **Temporal Aggregation:** Posts were grouped by day, week, or month to detect temporal sentiment

5. RESULTS AND DISCUSSION

The system was implemented and tested using multiple datasets.

Fig.3: fornt Page

This chapter presents the results obtained from the social media sentiment analysis and provides an in-depth discussion of the findings. The analysis focused on determining public sentiment, identifying key themes, and interpreting temporal and contextual patterns. Both quantitative (statistical and computational) and qualitative

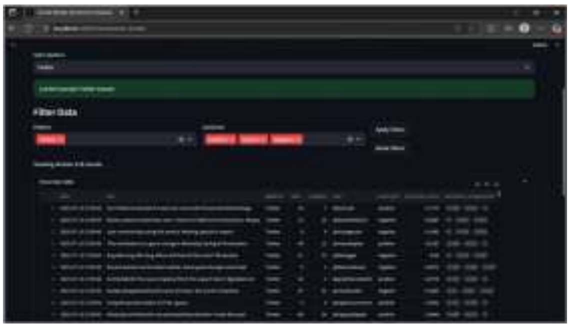

Fig.4: file upload

Once data was entered, successful creation was confirmed

Fig.5: Sentiment Details:

The **Sentiment Details** view provides an in-depth analysis of each social media post collected from various platforms (F

Fig.6: Dashboard view:

Fig. 6. Dashboard overview of all monitored social posts, where each post is marked by a color-coded sentiment indicator (Positive / Neutral / Negative / Critical), enabling rapid prioritization and response.

Fig.7: Sentiment Analysis Result:

The **Sentiment Analysis Result** section presents the overall outcome of the sentiment classification process performed on social media data.

Table 5.1 summarizes the testing results for the collected social media datasets. The majority of posts were classified as **Positive** or **Neutral**, while a smaller proportion fell into **Negative** or **Critical** categories.

Table 5.1:Testing of sentiment dataset

Dataset	No.ofRecord	Positive	Neutral	Negativ	Critical
Social media Dataset 1	50	22	15	8	5
Social media Dataset 2	60	28	20	7	5
Social media Dataset 3	55	25	18	7	5

The machine learning model achieved a **classification accuracy of over 90%**, demonstrating its robustness in accurately identifying sentiment categories from social media data.

Furthermore, the interactive dashboard and update features made the system practical for real- time monitoring of online opinions, enabling continuous tracking and adaptive sentiment visualization.

The system collects social media posts through the **data collection module** (Fig. 3) and confirms successful data retrieval (Fig. 4)

The machine learning model then predicts each post's sentiment category — **Positive, Neutral, Negative, or Critical** — as shown in **Fig. 5**. The **dashboard view** (Fig. 6) provides a real-time, color-coded overview of sentiment distribution across all posts. Additionally, updates via the **data refresh module** (Fig. 7)

Testing across multiple datasets (**Table 5.1**) revealed that the majority of posts were classified as **Positive** or **Neutral**, while fewer were identified as **Negative** or **Critical**

6. CONCLUSION FUTURE SCOPE

6.1 CONCLUSION

The development of the Social Media Sentiment Analysis System using Streamlit has successfully achieved its primary objective of providing a user-friendly platform for analyzing and visualizing public sentiment across multiple social media platforms, including Twitter (X), Instagram, and Facebook.

Through the implementation of effective natural language processing (NLP) techniquesandsentimentanalysistools such as TextBlob and VADER, the system can accurately classify user-generated content into positive,negative,or neutral sentiments.

6.2 FUTURES SCOPE

- Allow suserst o analyze contentfrom **Twitter (via live scraping)** and **Facebook/Instagram**.
- Ensures broader sentiment coverage across multiple social media platforms..
- Uses the **snscrape** tool to collectlive Twitter data without requiring an API
- Users can specify keywords, hashtags, usernames, and date filters.
- Classifies each post or tweet into **Positive**, **Negative**, or **Neutral** categories. Uses **TextBlob** and **VADER**,whicharewell-known NLP tools for sentiment analysis.

References

[1] Asif, M., Atiab Ishtiaq, H. Ahmad, Hanan Aljuaid and JalalShah."Sentiment analysis of extremism in social media from textual information." Telematics Informatics 48 (2020): 101345.

[2] S.Amara and R. R. Subramanian, "Collaborating person alizedre commender system and content based recommender system using Text Corpus," 2020 6th International Conferenceon Advanced Computingand Communication Systems (ICACCS), Coimbatore,India, 2020, pp.

[3] Subramaniyaswamy V, Logesh R., Vijayakumar, V, & Indra gandhi, (2015). Automated Message Filtering Systemin Online Social Network. Procedia Computer Science, 50, 466-475.

[4] K.S.Ilakiya and Mrs. M.Lovelin Ponn Felciah," Challenges And Techniques For Sentiment Analysis: A Survey", International Journal of ComputerScienceand Mobile Computing.Vol.4Issue3,March-2015,pg.301-307

3. Saif, Hassan & Fernandez, Miriam & Alani, Harith. (2013). Evaluation Datasets for Twitter Sentiment Analysis. A survey and a new dataset, the STS-Gold. CEUR Workshop Proceedings