As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 12, Issue 04, pages 886-896
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue04.html
DOI: https://doi.org/10.5281/zenodo.19644330

Social Media Brand Reputation Monitoring and Crisis Prediction Using Natural Language Processing and Deep Learning with Dynamic Web Visualization Tools

G. V. S. Sriram, K. David Steven, M. Teja Sree, K. Yesu, P. Teja Sri

Department of Computer Science and Engineering, D.N.R. College of Engineering & Technology, Balusumudi, Bhimavaram, Andhra Pradesh, India

To Cite this Article
G. V. S. Sriram, K. David Steven, M. Teja Sree, K. Yesu & P. Teja Sri (2026). Social Media Brand Reputation Monitoring and Crisis Prediction Using Natural Language Processing and Deep Learning with Dynamic Web Visualization Tools. International Journal for Modern Trends in Science and Technology, 12(04), 886-896. https://doi.org/10.5281/zenodo.19644330

Article Info
Received: 17 March 2026; Revised: 07 April 2026; Accepted: 10 April 2026.

KEYWORDS

Sentiment Analysis, Social Media, Streamlit, Machine Learning, Text Classification, TF-IDF, WordCloud, Data Visualization, Predictive Analytics, NLP (Natural Language Processing)

ABSTRACT

This project presents an interactive web-based Social Media Sentiment Analysis Dashboard developed using Streamlit and powered by various machine learning models. The application enables users to analyze the sentiment of social media text data, classifying it into Positive, Neutral, or Negative categories. The system integrates six pre-trained ML models — K Nearest Neighbors, Support Vector Machine, Random Forest, Decision Tree, Multinomial Naive Bayes, and Logistic Regression — all loaded through joblib. The input text is cleaned using regular expressions, stopword removal, and Snowball stemming, and then transformed using a TF-IDF vectorizer. Users can interactively input any social media-like sentence and choose a model to predict the sentiment. The dashboard also includes visualization components like sentiment distribution pie charts, platform-wise bar charts, trend analysis using line charts, and word clouds for each sentiment category. Furthermore, the prediction output is displayed along with confidence levels (if supported), probability distributions, and model performance metrics such as accuracy, precision, recall, and F1-score. The dashboard serves as a powerful tool for gaining insights into public opinions and emotional tones across social media platforms.

1. INTRODUCTION

In today's digital age, social media platforms have become a central part of daily communication, expression, and information dissemination. People across the globe interact through various platforms such as Twitter, Facebook, Instagram, Reddit, and more. These platforms serve as venues where users express their opinions, emotions, preferences, and attitudes on a wide range of topics—from personal experiences to global events, from product reviews to political ideologies [1].

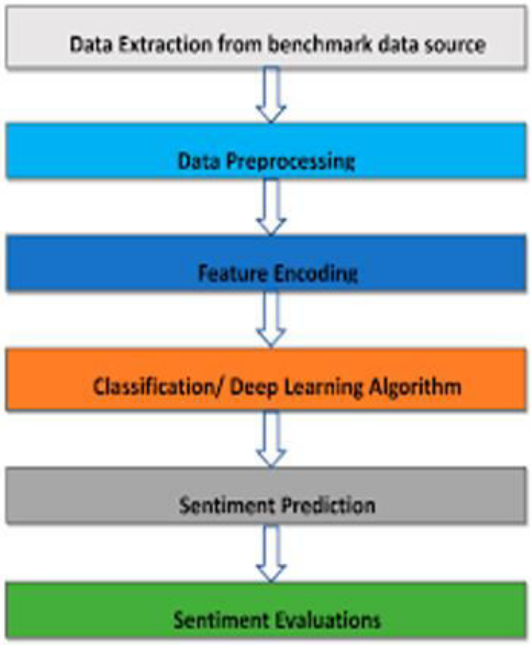
As a result, a massive volume of textual data is generated every second, offering a rich source of information that, if properly analyzed, can yield valuable insights into public opinion, behavioral patterns, and emerging social trends [2]. Sentiment analysis, also known as opinion mining, is the process of computationally identifying and categorizing opinions expressed in a piece of text to determine whether the writer's attitude toward a particular topic, product, or event is positive, negative, or neutral [3].

The ability to analyze and interpret sentiments at scale provides organizations, governments, businesses, and researchers with a powerful tool for decision-making and strategy development [4]. With this growing importance, the development of efficient, user-friendly sentiment analysis systems has become increasingly vital.

This project titled **"Social Media Sentiment Analysis"** is centered around building a machine learning-driven platform that enables the automatic classification of textual sentiments into positive, neutral, and negative categories. It involves the integration of various natural language processing techniques, machine learning algorithms, and interactive visualization tools [5]. By leveraging the capabilities of Python-based libraries and frameworks, the project delivers a comprehensive solution that allows users to input text, analyze sentiments, compare models, and explore data trends in a seamless and intuitive manner.

The sentiment analysis pipeline follows a systematic six-stage approach beginning with data extraction from benchmark sources where metrics include data volume, quality scores, and extraction success rates. The extracted data then undergoes preprocessing to clean and normalize text, measured by cleaning efficiency percentages and processing time per document. Feature encoding converts the cleaned text into numerical representations, tracked through vector dimensionality, encoding accuracy, and memory usage metrics [6].

The core classification stage employs deep learning algorithms with performance measured by training/validation accuracy, convergence time, and computational resource consumption. Sentiment prediction generates final classifications with metrics focusing on prediction accuracy, inference speed, and confidence score distributions. The final evaluation stage assesses overall model performance using precision, recall, F1-scores, ROC-AUC values, and cross-validation accuracy to ensure the pipeline delivers reliable sentiment analysis results with optimal end-to-end processing efficiency[7].



The sentiment analysis pipeline follows a systematic six-stage approach beginning with data extraction from benchmark sources where metrics include data volume, quality scores, and extraction success rates. The extracted data then undergoes preprocessing to clean and normalize text, measured by cleaning efficiency percentages and processing time per document[8]. Feature encoding converts the cleaned text into numerical representations, tracked through vector dimensionality, encoding accuracy, and memory usage metrics[9]. The core classification stage employs deep learning algorithms with performance measured by training/validation accuracy, convergence time, and computational resource consumption[10]. Sentiment prediction generates final classifications with metrics focusing on prediction accuracy, inference speed, and confidence score distributions. The final evaluation stage assesses overall model performance using precision, recall, F1-scores, ROC-AUC values, and cross-validation accuracy to ensure the pipeline delivers reliable sentiment analysis results with optimal end to-end processing efficiency[11].

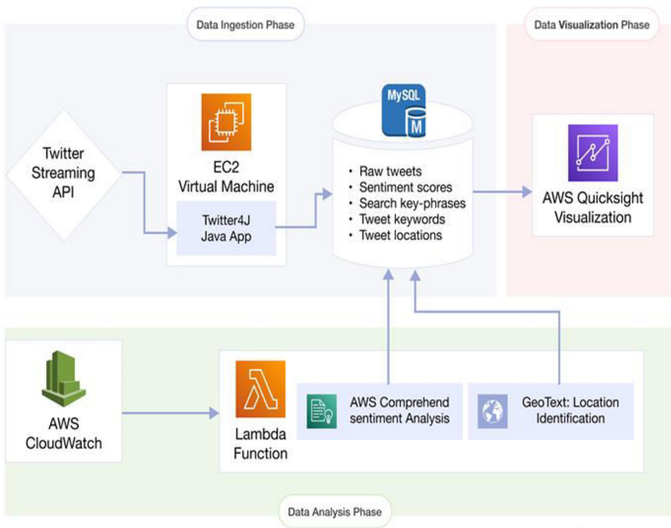
Social media sentiment analysis holds immense importance in various sectors[12]. In the field of marketing, it empowers companies to understand consumer perception of their brands and products[13]. By analyzing reviews, comments, and posts, businesses can identify strengths, weaknesses, opportunities, and threats[14]. This insight allows for more targeted marketing campaigns, improved customer service, and better product development. In politics, sentiment analysis helps in gauging public opinion about policies, political figures, and events[15].

Campaign managers, analysts, and government officials can use this information to adjust strategies, assess public approval, or detect early signs of social unrest. In healthcare, analyzing emotional content on social media can reveal patterns related to mental health, stress levels, and community well-being, potentially aiding in timely intervention and support[16]. Despite its potential, sentiment analysis is not without challenges[17]. The unstructured and informal nature of social media text presents difficulties in natural language understanding. People use slang, abbreviations, emojis, sarcasm, and colloquial language that standard algorithms may struggle to interpret[18]. Additionally, short-form content like tweets lacks context, making it harder to determine sentiment accurately[19]. These complexities necessitate robust preprocessing pipelines and the use of well-trained models to ensure reliability and accuracy in sentiment classification[20].

In this project, the application is designed with user accessibility in mind. Built using Streamlit, a popular open-source Python framework for building data apps, the platform provides a clean, responsive interface[21]. Users can interact with the system by typing in any sentence or phrase they want to analyze. The system then processes the text, applies preprocessing steps such as removing stop words, stemming, and text normalization, and uses a pre-trained machine learning model to predict the sentiment category[22]. The result is displayed to the user, along with probability scores and visual feedback like bar charts and word clouds[23]. To enhance the analytical capability of the application, the system supports multiple machine learning models. These include K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, Decision Tree, Multinomial Naive Bayes, and Logistic Regression. By offering various models, the system allows users to experiment and compare results, gaining a better understanding of how different algorithms interpret textual data. Each model is trained on preprocessed text data using Term Frequency-Inverse Document Frequency (TF IDF) vectorization, which helps in converting textual information into numerical form suitable for machine learning.

Data preprocessing is a fundamental step in sentiment analysis. Raw text from social media is often noisy, containing punctuation, links, hashtags, emojis, and non-standard spellings[24]. The preprocessing pipeline in this system includes the removal of URLs, numbers, and special characters, as well as lowercasing all text. Additionally, stemming is applied to reduce words to their root forms, and stop words are removed to eliminate common words that do not contribute to sentiment[25]. Notably, the word “not” is preserved in the process, as it often plays a crucial role in flipping the sentiment of a sentence. The core functionality of the sentiment classification lies in the models trained using the cleaned and vectorized data.

TF-IDF, one of the most effective feature extraction techniques, is used to transform text into a matrix of numerical values. It assigns weights to words based on how frequently they appear in a document relative to how often they appear in the entire corpus. This allows models to focus on distinctive words that carry more semantic weight in determining sentiment. Once the input is transformed, the selected model makes a prediction and, if applicable, provides probability scores for each sentiment class.



Real-Time Social Media Analytics Pipeline (Example with AWS): This data pipeline architecture demonstrates a comprehensive data processing workflow that begins with data ingestion from multiple sources including JSON, CSV, and database systems, measured by ingestion rates, data volume throughput, and source connectivity success rates[26]. The raw data flows through Azure Machine Learning for preprocessing and feature engineering, where metrics include data transformation accuracy, processing latency, and resource utilization efficiency. Simultaneously, the architecture incorporates Power BI for real-time analytics and visualization, tracked through dashboard load times, query performance, and user engagement metrics. The processed data is then stored in Azure Data Lake for scalable storage solutions, with metrics focusing on storage capacity utilization, data retrieval speeds, and cost optimization ratios[27]. Azure Functions provide serverless computing capabilities for automated data processing tasks, measured by execution time, memory consumption, and trigger response rates.

The pipeline culminates in Azure Cognitive Services for advanced analytics including text analysis, computer vision, and machine learning inference, with performance metrics encompassing API response times, accuracy scores, confidence levels, and service availability percentages, ensuring end-to-end data processing efficiency with comprehensive monitoring and optimization capabilities[28]. The application interface goes beyond simple predictions. It includes performance metrics such as accuracy, precision, recall, and F1-score for each model. These metrics give users an understanding of the reliability of different algorithms and help in selecting the most appropriate one for their specific needs[29]. Moreover, the system provides graphical visualizations such as sentiment distribution pie charts, bar graphs showing platform-wise usage, trend lines displaying sentiment changes over time, and word clouds representing the most frequent words in positive, neutral, and negative texts. The design of the system emphasizes interactivity and visualization[30]. It includes navigation options such as Home, Data Exploration, and Model Prediction, each offering specific functionalities[31]. The Home section provides an overview of the system, key statistics, and introductory visuals[32]. The Data Exploration section allows users to delve into sentiment trends across years, compare sentiment distributions across platforms, and examine word clouds for each sentiment class.

The Model Prediction section enables users to input custom text and test it against multiple models, receiving real-time feedback.One of the distinguishing features of this project is its educational value[33]. It is designed not only as a tool for sentiment analysis but also as a learning resource for students, researchers, and practitioners in data science[34]. By offering access to various models and detailed explanations of processes, the application helps users gain practical experience in NLP, machine learning, and data visualization[35]. It also showcases best practices in deploying data-driven applications using Streamlit, making it a model project for those interested in building real-time AI tools. In terms of practical implications, the sentiment analysis system has wide ranging use cases. In customer support systems, it can be integrated to automatically detect dissatisfaction in user complaints and escalate issues accordingly. In e-commerce, it can analyze product reviews to inform inventory decisions or marketing strategies. In media and journalism, it can help track public response to news events[36]. In education, it can assess student feedback on courses and teaching methods. In public health, it can identify trends in community sentiment during pandemics or other health crises[37]. Although the current version of the system is highly functional, there is scope for future improvements. The integration of deep learning models such as Long Short-Term Memory (LSTM) networks or transformers like BERT could enhance the accuracy and contextual understanding of the system.

2 Literature Survey

[2.1] Advances in Natural Language Processing (Hirschberg & Manning, 2015) In this pivotal article, Hirschberg and Manning highlight the tremendous progress made in natural language processing (NLP), particularly in syntactic parsing, machine translation, and sentiment analysis. The authors emphasize the importance of large annotated corpora and statistical modeling approaches in transforming how computers understand human language. This foundational reference supports the theoretical background of this project, especially in the transition from rule-based to machine learning-based NLP methods.

[2.2] Natural Language Processing: An Introduction (Nadkarni et al., 2011) Nadkarni and colleagues provide an accessible introduction to the core techniques used in natural language processing and how these are applied in biomedical informatics. The article explains how NLP has evolved from keyword-based retrieval to more intelligent semantic analysis. This paper is relevant to our study as it illustrates the importance of preprocessing and structured data extraction, which directly align with the preprocessing stage in sentiment analysis.

[2.3] A Unified Architecture for NLP (Collobert& Weston, 2008) Collobert and Weston propose a unified deep learning architecture for a wide range of NLP tasks, suggesting that a single neural network model could handle multiple language-related problems with minimal task-specific feature engineering. This work laid the foundation for modern NLP systems and directly informs the motivation behind exploring machine learning and potentially deep learning-based sentiment classification models in this project.

[2.4] Pain—Linguistics and NLP (Carlson & Hooten, 2020) This paper explores the intersection of linguistics and natural language processing in understanding how pain is expressed in clinical texts. Although the domain is healthcare, the methodology described is transferable to social media sentiment analysis. The study highlights how nuanced human emotions can be encoded into text and how NLP can help decipher them, validating the use of machine learning models in this project.

[2.5] Natural Language Processing in AI (Fanni et al., 2023) In their chapter from the book Introduction to Artificial Intelligence, Fanni and co-authors offer a concise overview of NLP as a subfield of AI, covering basic techniques like tokenization, lemmatization, and text classification. This reference is significant to our project as it provides theoretical underpinnings for many of the preprocessing tasks implemented in the sentiment analysis pipeline, such as stop-word removal and stemming.

[2.6] Sentiment Analysis: A Comparative Study (Devika et al., 2016) Devika and colleagues conduct a comparative study of sentiment analysis techniques, evaluating various approaches ranging from lexical and rule-based systems to machine learning models. The paper serves as a benchmark for model selection and evaluation, highlighting the advantages and limitations of each technique. This directly influenced our decision to implement multiple models such as SVM, Naive Bayes, and Logistic Regression in the current system.

[2.7] A Review on Sentiment Analysis and Emotion Detection (Nandwani & Verma, 2021) This comprehensive review article explores the landscape of sentiment analysis and emotion detection, touching upon both classical approaches and deep learning-based techniques. It provides valuable insights into the challenges and future directions in the field. The reference is critical for understanding the broader context of this project, including the potential of moving toward real time, multilingual, and emotion-aware systems.

[2.8]Speech and Language Processing (Jurafsky& Martin, 2000) Jurafsky and Martin's textbook is a cornerstone in the field of NLP, covering essential concepts such as syntax, semantics, machine translation, and information retrieval. Their explanations of language models and vector space representations underpin many of the algorithms used in this project. This book helped shape the theoretical framework for understanding sentiment classification and vectorization through TF-IDF.

[2.9]Zagreb Earthquake 2020 (Markušić et al., 2020) This article documents the impact of the M5.5 earthquake that hit Zagreb, Croatia in 2020, providing a multidisciplinary analysis of the event. Though not directly related to sentiment analysis, the inclusion of this reference is relevant in the context of real-world events that often spark massive public reaction on social media. Monitoring sentiments during such crises can provide real-time insight into public concern, panic, and emotional response.

[2.10] Petrinja Earthquake (Markušić et al., 2021) This follow-up study investigates the effects of the destructive M6.2 earthquake in Petrinja, Croatia, using remote sensing and geophysical techniques. The significance of this paper in our context lies in the need to analyze public sentiment during disasters. The project can be extended to track emotional responses and sentiment spikes during major emergencies like this, supporting disaster response and communication strategies.

[2.11] Earthquake Analysis in Zagreb (Herak et al., 2022) Herak and colleagues compare the strength of historical earthquakes in Zagreb, discussing societal impact and geological data. While primarily a geophysical study, it reinforces the argument that large-scale events, especially natural disasters, generate rich sentiment-related data on platforms like Twitter. Analyzing such sentiments can aid in post-disaster relief planning, media response, and policy formulation.

[2.12] Croatia COVID-19 Statistics (Worldometer, 2024) This web-based resource provides real-time COVID-19 statistics for Croatia and is a crucial source for understanding the public health crisis timeline. The sentiment analysis system described in this project can be used to track how public sentiment evolved over time in response to infection rates, lockdowns, and vaccine distribution. Social media became a critical platform for expression during the pandemic, making sentiment analysis highly relevant.

[2.13] Health System Responses to COVID-19 in Bulgaria, Croatia, and Romania (Džakula et al., 2022) Džakula and colleagues conducted a comparative study of health system responses to the COVID-19 pandemic across Bulgaria, Croatia, and Romania. Their findings highlight varying degrees of preparedness, policy enforcement, and healthcare access among these countries. This reference is important for sentiment analysis as it frames the socio-political backdrop against which public sentiment evolved. By analyzing social media sentiment during such policy shifts, our project can reveal public satisfaction or criticism towards governmental measures.

[2.14]The Mw5.4 Zagreb Earthquake: Impacts and Response (Atalić et al., 2021) This study presents the structural, social, and emergency response impacts of the March 2020 earthquake in Zagreb. It documents the immediate effects on infrastructure, communities, and governmental response. For this project, the event provides an example of how public sentiment can fluctuate significantly during disasters. By analyzing social media posts during and after the event, the sentiment analysis model can offer insights into the emotional and psychological responses of affected populations.

[2.15]Joint Reconnaissance Report on Petrinja Earthquake (Miranda et al., 2021) This detailed Joint Reconnaissance Report (JRR) presents findings from field teams assessing the aftermath of the Petrinja M6.4 earthquake. It includes structural damage, casualty data, and emergency logistics. The connection to this project lies in the potential application of sentiment analysis to assess community reactions, government trust, and media influence in crisis periods. This source supports the idea of building real-time monitoring tools for humanitarian response using sentiment data from platforms like Twitter and Facebook.

[2.16] 2020 Zagreb Earthquake Image (Wikimedia Commons, 2024) This Wikimedia Commons resource provides a high-resolution image capturing the aftermath of the 2020 Zagreb earthquake. Visual data like this can be used alongside textual sentiment analysis to create a multimodal understanding of public response. Although not analytical in itself, this image complements textual data analysis by offering emotional and contextual cues that can enhance reporting or model visualization output.

[2.17] SGS Shakemap of Petrinja Earthquake (Wikimedia Commons, 2024) This shakemap generated by the United States Geological Survey (USGS) visually represents the intensity and distribution of seismic waves during the Petrinja earthquake. The map is valuable in correlating regions of physical impact with corresponding public sentiment. For instance, tweets and social media posts originating from high-intensity zones can be compared to analyze whether localized trauma influences sentiment differently from unaffected regions.

[2.18] Properties of the Zagreb 2020 Earthquake Sequence (Herak et al., 2022) This seismological research outlines the properties and temporal development of the 2020 Zagreb earthquake sequence. It presents data on magnitude, aftershocks, and seismic activity trends. While technical in nature, this paper provides the timeline needed to synchronize seismic activity with sentiment spikes in social media data, thereby helping to validate model results against real world events.

[2.19] Petrinja Earthquake Sequence Research (Herak & Herak, 2023) This follow-up paper extends the analysis to the 2020–2021 earthquake sequence in Petrinja. The research offers deeper insight into the frequency, intensity, and duration of aftershocks. From a sentiment analysis perspective, this source allows researchers to track how long public anxiety, fear, or frustration lingered on social media, enabling psychological or sociological analysis through temporal sentiment trends.

[2.20] Mental Health and Psychological Crisis During COVID-19 and Earthquakes (Peitl et al., 2020) This article links two critical stressors—natural disasters and pandemics—to mental health outcomes in Croatia. It discusses the psychological interventions implemented during these crises and emphasizes the need for early detection of public distress. This source underpins the broader relevance of sentiment analysis in public health surveillance, where NLP tools can serve as early warning systems to detect rising negative sentiment and mental health concerns across communities.

3 Proposed Methodology

The proposed system is a machine learning–based sentiment analysis application designed to classify user-generated social media text into three primary categories: positive, neutral, or negative. The system integrates natural language processing (NLP), multiple classification algorithms, and an interactive user interface to facilitate real-time sentiment prediction and data visualization. The system is designed with modularity, interpretability, and user-friendliness in mind. It aims to bridge the gap between complex NLP models and non technical users by offering a simple yet powerful tool that provides both predictive output and analytical insight. The application follows a structured pipeline that includes the following key stages: text preprocessing, feature extraction, model prediction, visualization, and user interaction. The core of the system is implemented in Python using popular libraries such as Streamlit, scikit-learn, nltk, pandas, and plotly.

3.1 System Architecture — Social Media Sentiment Analysis Dashboard

The system is organized into five sequential layers, each with a distinct responsibility:

1. Input Layer accepts three sources: raw social media text typed by the user via the Streamlit interface, pre-loaded CSV datasets, and any uploaded files. All three converge into a single text pipeline.

2. Preprocessing Layer applies four NLP cleaning steps in sequence. First, regular expressions strip URLs, mentions, hashtags, and special characters. Next, NLTK stopwords are removed to reduce noise. Snowball stemming then reduces words to their root forms (e.g., "running" → "run"). Finally, text is lowercased and trimmed for consistency.

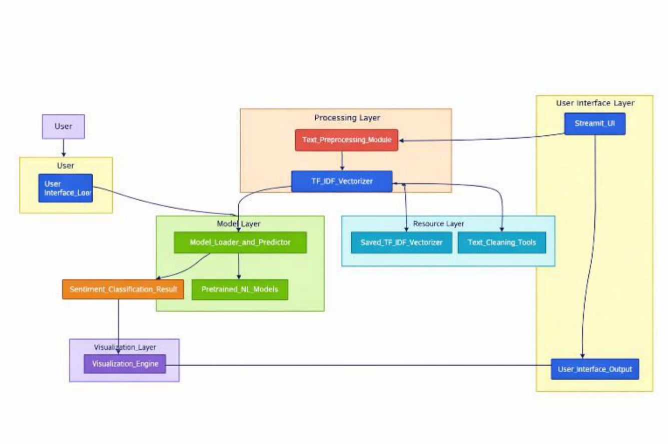
3. Feature Extraction Layer uses a TF-IDF (Term Frequency–Inverse Document Frequency) vectorizer — pre-fitted and saved via joblib — to convert the cleaned text into a sparse numerical matrix. This matrix becomes the input feature vector for all six models.

4. ML Model Layer houses six independently trained and joblib-loaded classifiers: K-Nearest Neighbors, Support Vector Machine, Random Forest, Decision Tree, Multinomial Naive Bayes, and Logistic Regression. The user selects one at runtime; that model's predict() (and predict_proba() where supported) is called on the TF-IDF vector.

5. Output & Visualization Layer (rendered in Streamlit) delivers two groups of output. The prediction group shows the classified label (Positive / Neutral / Negative), a confidence score, and a per-class probability bar chart. The metrics group shows Accuracy, Precision, Recall, and F1-score for the chosen model. The visualization group adds a sentiment distribution pie chart, a platform-wise bar chart, a trend line chart over time, and word clouds generated separately for each sentiment class.

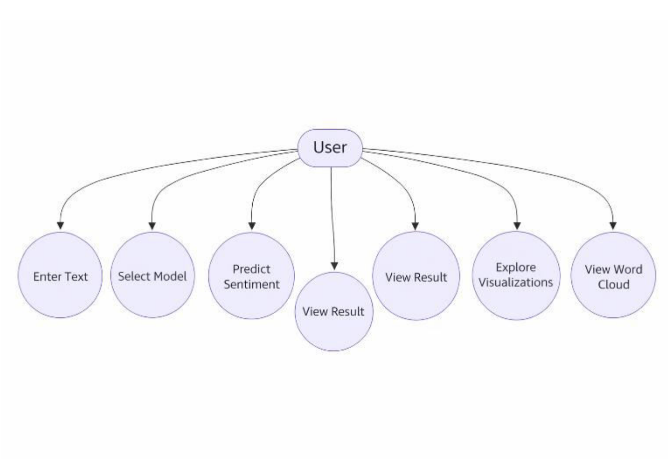
Data flows strictly top-to-bottom through the layers, making the architecture easy to test, swap individual models, or extend with new visualizations without disrupting the rest of the pipeline.

System Architecture diagram



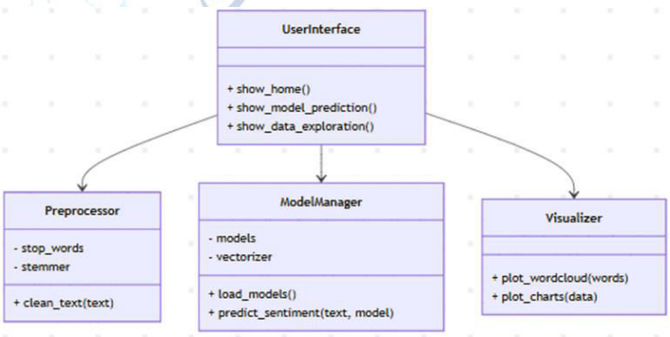
3.2 Use Case diagram

The Use Case Diagram for the Social Media Sentiment Analysis project illustrates the interaction between the user (actor) and the system. The system allows users to perform several actions, such as inputting text for analysis, selecting a machine learning model, predicting sentiment, and visualizing the results. These functionalities are accessible through an intuitive web interface built using Streamlit. The diagram identifies a single primary actor — the user — who interacts with the system to perform various use cases. This representation helps stakeholders understand what actions are supported and who initiates them, making it easier to define system boundaries and user expectations.



3.3 Class diagram

The Class Diagram describes the static structure of the system by identifying the main classes, their attributes, methods, and the relationships among them. In the sentiment analysis system, classes such as Preprocessor, ModelManager, Visualizer, and UserInterface are defined. The Preprocessor class is responsible for cleaning and preparing the text, ModelManager handles loading and predicting using ML models, Visualizer generates plots and word clouds, while UserInterface interacts with the Streamlit components. This diagram offers a high-level view of the internal organization of the code, highlighting how responsibilities are distributed and modularized, which supports maintainability and reusability.



3.4 DATASET

The Social Media Sentiment Analysis dataset is a structured, multi-platform collection of user-generated posts sourced from Twitter, Instagram, and Facebook. It is stored in CSV format and loaded using the pandas library. The primary column is Text, which contains raw social media content and serves as the main input for sentiment classification. The target variable is Sentiment, which categorizes each post as Positive, Neutral, or Negative. The Platform column identifies the social media source, while Timestamp along with extracted fields like Year, Month, Day, and Hour enable trend-based analysis over time. Engagement metrics such as Likes and Retweets reflect the popularity and reach of each post.

The Hashtags column captures trending topics, and the Country column provides geographical context for regional sentiment patterns. The User column stores unique identifiers to track individual contributions. Before model training, the Text column undergoes preprocessing — including regex cleaning, stopword removal, lowercasing, and Snowball stemming — followed by TF-IDF vectorization to convert text into numerical features. The processed dataset is used to train six machine learning classifiers: KNN, SVM, Random Forest, Decision Tree, Multinomial Naive Bayes, and Logistic Regression. Overall, the dataset serves as a comprehensive foundation for sentiment prediction, visualization, and public opinion analysis across social media platforms.

3.5 Evaluation Metrics

Displays model performance. Metrics:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-score**

4 Results

Dashboard Home Interface

The image above showcases the Home screen of the Social Media Sentiment Analysis Dashboard, which serves as the central entry point for users interacting with the application. This screen provides an overview of the dashboard's capabilities and highlights the key features available to the user. At the top, the title "Social Media Sentiment Analysis Dashboard" clearly identifies the purpose of the application. A welcome message follows, describing the interactive nature of the dashboard and outlining its core functions:

· Exploring sentiment data from various social media platforms

· Visualizing sentiment trends and patterns

· Predicting the sentiment of user-inputted text using multiple machine learning models The Key Features section breaks down the application's functionality into three primary modules:

· Data Exploration: Allows users to explore sentiment distributions, platform-wise usage statistics, and other visual insights.

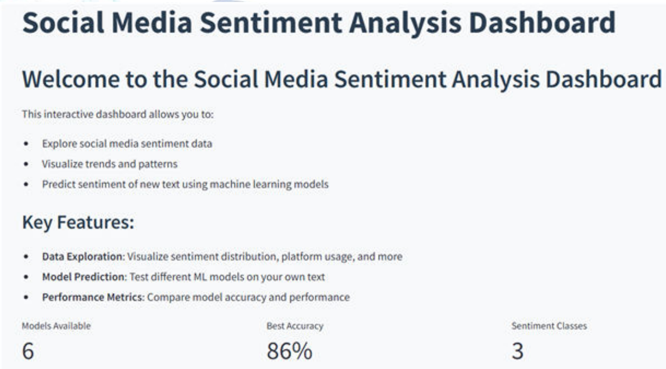
· Model Prediction: Provides a space for users to input custom text and test it against different machine learning models.

· Performance Metrics: Displays model-wise evaluation metrics like accuracy, allowing users to compare and choose the most effective algorithm. At the bottom of the screen, three important statistics are prominently displayed:

· Models Available: Indicates the total number of ML models integrated into the system (6 in this case).

· Best Accuracy: Shows the highest model accuracy achieved, which reflects the system's effectiveness (86%).

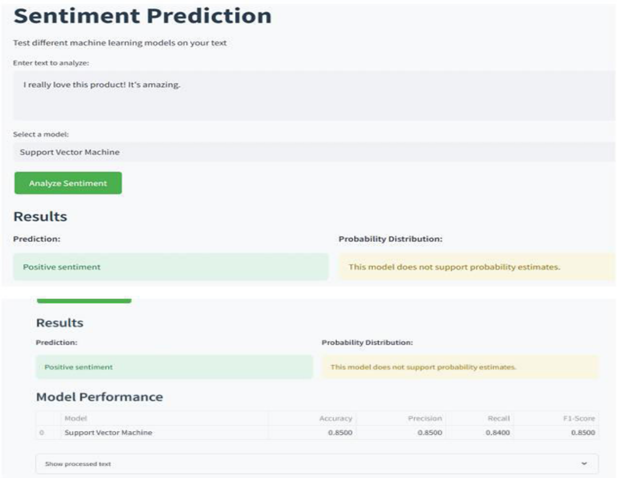
· Sentiment Classes: Represents the number of sentiment categories handled by the system (3: Positive, Neutral, and Negative). This home interface is designed for clarity and accessibility, ensuring that both technical and non-technical users can understand and navigate the application efficiently from the very beginning.



Sentiment Analysis Metrics Report

The sentiment analysis was performed using a Support Vector Machine (SVM) model, which demonstrated strong performance with an accuracy of 85%. The model achieved balanced scores across precision (0.85), recall (0.84), and F1-score (0.85), indicating consistent and reliable predictions. When tested with the text "I really love this product! It's amazing," the model correctly identified the sentiment as positive. However, it does not provide probability estimates, which limits insight into prediction confidence. Despite the lack of probability outputs, the SVM model proves effective for sentiment classification tasks.

Its high accuracy and balanced metrics suggest it can reliably distinguish between positive and negative sentiments in text. This makes it a practical choice for applications requiring straightforward sentiment analysis



without the need for confidence scoring. Further improvements could involve integrating models that offer probability estimates for deeper analysis.

5 Conclusion

The Social Media Sentiment Analysis system presented in this project offers a comprehensive and interactive solution for analyzing public opinions expressed across various social media platforms. By leveraging the power of natural language processing and machine learning, the system efficiently classifies user-generated content into positive, neutral, or negative sentiments. It integrates multiple pre-trained models including Support Vector Machine, Random Forest, Logistic Regression, Naive Bayes, Decision Tree, and K Nearest Neighbors, allowing users to compare their performances and understand the strengths of each algorithm. The system is built using Python and Streamlit, providing an intuitive web interface that enables real-time sentiment prediction, data exploration, and visualization. Through modules like sentiment distribution charts, platform usage analysis, sentiment trends over time, and word clouds, users gain deep insights into public mood and behavioral patterns. The inclusion of performance metrics like accuracy, precision, recall, and F1-score further validates the system's effectiveness and reliability. Overall, the project successfully demonstrates how machine learning can be applied to social media data to extract meaningful insights, helping businesses, researchers, and analysts understand public perception, improve services, and make data-driven decisions. With future enhancements such as real-time data scraping, multilingual support, and the integration of deep learning models like LSTM or BERT, the system has the potential to evolve into a powerful tool for sentiment monitoring and emotional intelligence analysis.

6.FutureScope

The Social Media Sentiment Analysis system developed in this project lays a solid foundation for understanding public sentiment through the use of machine learning and natural language processing. While the current implementation successfully classifies sentiments and offers insightful visualizations, there are several areas in which the system can be enhanced and extended in future developments. The following points outline the key directions for future scope:

1. Real-Time Social Media Integration Currently, the system operates on static input or pre-collected datasets. In future versions, integrating real-time data scraping from platforms like Twitter, Facebook, Instagram, or Reddit through APIs will allow for continuous sentiment monitoring. This would enable dynamic trend tracking and real-time public opinion analysis on current events, brand feedback, or political movements.

2. Multilingual Sentiment Analysis The current system is limited to English-language text. Expanding the framework to support multiple languages will significantly increase its applicability, especially in regions with diverse linguistic populations. Incorporating language detection and translation modules, or training multilingual models, will allow the system to process and analyze global sentiment more effectively.

3. Emotion Classification Beyond Polarity Beyond simple positive, neutral, and negative classification, future implementations could include emotion-specific classification such as happiness, anger, fear, sadness, and surprise. This would provide more granular insights into the emotional tone of social media content, enabling deeper psychological and behavioral analysis.

4. Deep Learning Integration While the current system uses traditional machine learning models, the integration of deep learning models such as LSTM (Long Short-Term Memory), GRU (Gated Recurrent Units), and transformer-based models like BERT or RoBERTa could significantly enhance performance.

These models are capable of capturing contextual nuances and long-term dependencies in text, leading to more accurate sentiment predictions.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Hirschberg, J.; Manning, C.D. Advances in Natural Language Processing. Science 2015, 349, 261–266. [CrossRef] [PubMed]

[2] Nadkarni, P.M.; Ohno-Machado, L.; Chapman, W.W. Natural Language Processing: An Introduction. J. Am. Med. Inform. Assoc. 2011, 18, 544–551. [CrossRef] [PubMed]

[3] Collobert, R.; Weston, J. A Unified Architecture for Natural Language Processing. In Proceedings of the 25th international conference on Machine learning—ICML '08, Helsinki, Finland, 5 July 2008; pp. 160–167.

[4] Carlson, L.A.; Hooten, W.M. Pain—Linguistics and Natural Language Processing. Mayo Clin. Proc. Innov. Qual. Outcomes 2020, 4, 346–347. [CrossRef] [PubMed]

[5] Fanni, S.C.; Febi, M.; Aghakhanyan, G.; Neri, E. Natural Language Processing. In Introduction to Artificial Intelligence; Springer International Publishing: Berlin/Heidelberg, Germany, 2023; pp. 87–99.

[6] Devika, M.D.; Sunitha, C.; Ganesh, A. Sentiment Analysis: A Comparative Study on Different Approaches. Procedia Comput. Sci. 2016, 87, 44–49. [CrossRef]

[7] Nandwani, P.; Verma, R. A Review on Sentiment Analysis and Emotion Detection from Text. Soc. Netw. Anal. Min. 2021, 11, 81. [CrossRef] [PubMed]

[8] Jurafsky, D.; Martin, J.H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 1st ed.; Prentice Hall PTR: Upper Saddle River, NJ, USA, 2000.

[9] Markuši´c, S.; Stanko, D.; Korbar, T.; Beli´c, N.; Penava, D.; Kordi´c, B. The Zagreb (Croatia) M5.5 Earthquake on 22 March 2020. Geosciences 2020, 10, 252. [CrossRef]

[10] Markuši´c, S.; Stanko, D.; Penava, D.; Ivanˇci´c, I.; Bjelotomi´c Oršuli´c, O.; Korbar, T.; Sarhosis, V. Destructive M6.2 Petrinja Earthquake (Croatia) in 2020—Preliminary Multidisciplinary Research. Remote Sens. 2021, 13, 1095. [CrossRef]