International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(1) (2026)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

GENAI POWERED SENTIMENT ANALYSIS FOR SOCIAL MEDIA

[1] R Uma, [2] K Keerthi, [3] M Keerthana, [4] B Poojitha, [5] M Naga Sri
[1]AssistantProfessor, [2345]Students
Department of AIML
Siddhartha Institute of Technology & Sciences, Narapally

umakola_cse@siddhartha.co.in, 24tq1a6677@siddhartha.co.in, 24tq1a6693@siddhartha.co.in, 24tq1a66b0@siddhartha.co.in, 25tq5a6610@siddhartha.co.in

Abstract

The "YouTube Sentiment Analysis Dashboard" is an intelligent system developed to analyze user opinions expressed in YouTube comments using Artificial Intelligence (AI), Machine Learning (ML), and Generative AI (GenAI) techniques. With the rapid growth of digital platforms, YouTube has become a major source of public interaction where users share feedback, opinions, and emotions through comments.

This project focuses on extracting comments from YouTube videos and analyzing them using Natural Language Processing (NLP) techniques. The system preprocesses textual data by removing noise such as stopwords, punctuation, and irrelevant symbols. It then applies machine learning algorithms and transformer-based models to classify comments into positive, negative, and neutral sentiments.

A key feature of this project is the development of an interactive dashboard using Streamlit, which visually represents sentiment distribution, trending keywords, and user engagement patterns. The dashboard enables users to understand audience reactions in real-time and make informed decisions.

The system improves accuracy by leveraging GenAI models that understand context better than traditional approaches. This project demonstrates the importance of sentiment analysis in content strategy, marketing, and social media analytics.

I. Introduction

In today's digital world, social media platforms generate massive amounts of user-generated content every day. Among these platforms, YouTube is one of the most widely used services where millions of users watch videos, share opinions, and interact through comments, likes, and discussions. These comments contain valuable information about audience emotions, opinions, satisfaction levels, and engagement patterns. Understanding these opinions is highly important for businesses, content creators, marketers, and researchers to analyze public perception and improve decision-making processes.

Sentiment analysis is a Natural Language Processing (NLP) technique used to identify and classify opinions expressed in textual data as positive, negative, or neutral.

Traditional sentiment analysis methods mainly relied on rule-based systems and classical machine learning algorithms, which often struggled to understand context, sarcasm, slang, emojis, and complex human language patterns. With the advancement of Artificial Intelligence and Generative AI technologies, sentiment analysis systems have become more intelligent, accurate, and context-aware. Modern Large Language Models and deep learning techniques can understand semantics, emotional tone, and contextual relationships within social media text more effectively.

The GENAI Powered Sentiment Analysis for Social Media project is developed to analyze user sentiments from YouTube comments using advanced AI and NLP technologies. The system automatically extracts comments from YouTube videos, preprocesses the textual data, classifies sentiments, and visualizes the analysis results through an interactive dashboard. The project aims to help users understand audience reactions, trending opinions, and emotional responses related to video content in real time.

II. Literature Survey

Sentiment analysis has become one of the most important research areas in Artificial Intelligence and Natural Language Processing (NLP), especially with the rapid growth of social media platforms such as YouTube, Twitter, Facebook, and Instagram. Social media platforms generate massive amounts of unstructured textual data in the form of comments, reviews, opinions, and discussions. Analyzing this data manually is extremely difficult due to its large volume, dynamic nature, and informal writing styles. Sentiment analysis helps in identifying emotional tone and public opinion by classifying textual content into positive, negative, or neutral categories. Over the years, sentiment analysis techniques have evolved significantly from simple rule-based methods to advanced deep learning and Generative AI models capable of understanding context, semantics, and emotions more accurately.

In the early stages of sentiment analysis research, lexicon-based approaches were widely used. These methods relied on predefined dictionaries containing positive and negative words to determine sentiment polarity. For example, words such as “good,” “excellent,” or “amazing” were categorized as positive, while words like “bad,” “poor,” or “terrible” were considered negative. Lexicon-based systems were simple to implement and required minimal training data. However, these approaches lacked contextual understanding and struggled to interpret sarcasm, slang, negation, and complex sentence structures commonly found in social media comments.

Later, researchers introduced machine learning approaches for sentiment analysis to improve classification accuracy and adaptability. Algorithms such as Naive Bayes, Support Vector Machines (SVM), Decision Trees, and Logistic Regression became widely used for text classification tasks. These models required labeled datasets for training and used feature extraction techniques such as Bag-of-Words and TF-IDF to represent textual information numerically. Machine learning methods provided better accuracy compared to lexicon-based systems and could adapt to different domains more effectively. However, these models still depended heavily on manual feature engineering and had limitations in understanding long-range contextual relationships within text.

The emergence of deep learning techniques significantly transformed sentiment analysis research. Models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks improved the ability of AI systems to capture sequential information and contextual dependencies within textual data. These architectures were particularly effective for analyzing long social media comments and conversational text because they could remember previous words and contextual relationships while processing sequences. Deep learning models reduced the need for manual feature engineering and improved sentiment classification accuracy considerably.

Recent advancements in Transformer-based architectures such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have further improved sentiment analysis capabilities. These models use self-attention mechanisms to understand deep contextual relationships between words and sentences. Transformer models can interpret sarcasm, contextual meanings, emojis, and informal social media language more effectively than traditional methods. Research studies show that transformer-based models outperform earlier machine learning and deep learning approaches in terms of accuracy, contextual understanding, and generalization across diverse datasets.

Several NLP tools and frameworks are also commonly used in sentiment analysis systems. Tools such as VADER (Valence Aware Dictionary and sEntiment Reasoner) are specifically designed for social media sentiment analysis and perform well on short informal text containing emojis, abbreviations, and slang. Other popular libraries include TextBlob and NLTK, which provide functionalities for text preprocessing, tokenization, sentiment scoring, and natural language understanding. These tools simplify the development of sentiment analysis applications and improve accessibility for researchers and developers.

Research findings indicate that sentiment analysis accuracy improves significantly with the use of deep learning and context-aware transformer models. Modern AI systems provide better understanding of semantics, emotions, and user intent compared to traditional rule-based approaches. Studies also highlight the importance of visualization techniques such as graphs, dashboards, and sentiment trend analysis for improving decision-making and user interpretation. Furthermore, real-time sentiment analysis has emerged as an important research direction, enabling organizations to monitor audience reactions and public opinions dynamically as social media interactions occur.

The GENAI Powered Sentiment Analysis for Social Media project builds upon these advancements by integrating Generative AI, NLP, machine learning, and visualization techniques into a unified intelligent platform for analyzing YouTube comments. The system aims to provide accurate, scalable, and real-time sentiment analysis capabilities while transforming large-scale social media data into meaningful insights for businesses, researchers, and content creators.

III. System Analysis

The GENAI Powered Sentiment Analysis for Social Media system is designed to analyze and classify user opinions from social media platforms using Artificial Intelligence and Natural Language Processing technologies. The system focuses on extracting comments from YouTube videos and identifying whether user opinions are positive, negative, or neutral. Social media platforms generate massive amounts of unstructured textual data every day, making manual analysis difficult and time-consuming. The proposed system automates this process using machine learning, deep learning, and Generative AI models to improve accuracy and contextual understanding. The application performs text preprocessing tasks such as tokenization, stop-word removal, stemming, and cleaning before sentiment classification. Advanced AI models analyze contextual meaning, emotional tone, and semantic relationships within comments to improve sentiment prediction accuracy. The system also provides visualization dashboards using charts and graphs to display sentiment distribution and audience reactions clearly. Real-time analysis capabilities help businesses, content creators, and researchers understand public opinion dynamically. The modular architecture supports scalability and future integration with multiple social media platforms and multilingual sentiment analysis systems. The project demonstrates how Generative AI can transform large-scale social media data into meaningful insights for decision-making and audience analysis.

Existing System

In the existing system, sentiment analysis mainly relied on traditional Natural Language Processing techniques and rule-based classification methods. Early systems used lexicon-based approaches where predefined dictionaries of positive and negative words were used to determine sentiment polarity. These methods were simple and easy to implement but lacked contextual understanding and failed to handle sarcasm, slang, emojis, and complex sentence structures effectively. Later, machine learning algorithms such as Naive Bayes and Support Vector Machines were introduced for sentiment classification using labeled datasets. Although these models improved classification accuracy compared to rule-based systems, they still depended heavily on manual feature engineering and limited contextual interpretation. Existing systems also struggled with processing large-scale real-time social media data efficiently. Many traditional platforms provided only static analysis results without interactive dashboards or visual insights. Existing sentiment analysis models often produced inaccurate results when handling informal language and multilingual social media content. Earlier systems also lacked scalability and adaptability to evolving social media trends and user behaviors. These limitations created the need for more advanced Generative AI-powered sentiment analysis systems capable of understanding contextual semantics and performing real-time analysis dynamically.

Disadvantages of Existing System

- Limited contextual understanding of text.
- Difficulty handling sarcasm and slang.
- Poor interpretation of emojis and informal language.
- Dependence on predefined dictionaries and rules.
- Manual feature engineering requirements.
- Reduced accuracy on complex sentence structures.
- Limited real-time processing capabilities.

- Difficulty handling large-scale social media data.

Proposed System

The proposed GENAI Powered Sentiment Analysis for Social Media system is designed to provide intelligent, accurate, and real-time sentiment analysis using Generative Artificial Intelligence and advanced Natural Language Processing techniques. The system automatically extracts YouTube comments and processes them through multiple NLP stages including text cleaning, tokenization, stop-word removal, stemming, and normalization. Advanced AI models such as transformer-based architectures and Generative AI systems analyze textual context, semantics, and emotional tone to classify comments into positive, negative, or neutral sentiments accurately. Unlike traditional rule-based systems, the proposed solution can understand sarcasm, informal language, emojis, and contextual relationships within social media comments. The application also provides real-time dashboards and visualizations using graphs, charts, and sentiment distribution reports for better audience understanding. The system supports scalable processing of large volumes of social media data and can adapt to evolving online communication patterns dynamically. The modular architecture enables future enhancements such as multilingual sentiment analysis, fake review detection, trend prediction, and recommendation systems. The proposed system improves audience insight generation, customer feedback analysis, and marketing strategy development for businesses and content creators. Overall, the system provides a scalable, intelligent, and AI-driven solution for modern social media sentiment analysis.

Advantages of Proposed System

- Improved contextual understanding and sentiment accuracy.
- Handles sarcasm, slang, and emojis effectively.
- Real-time sentiment analysis and monitoring.
- Automated extraction and processing of comments.
- Scalable for large-scale social media data.
- Interactive dashboards and visualization support.
- Reduced manual effort and processing time.
- Supports advanced AI and NLP integration.
- Flexible for future multilingual analysis extensions.
- Better audience insight and decision-making support.

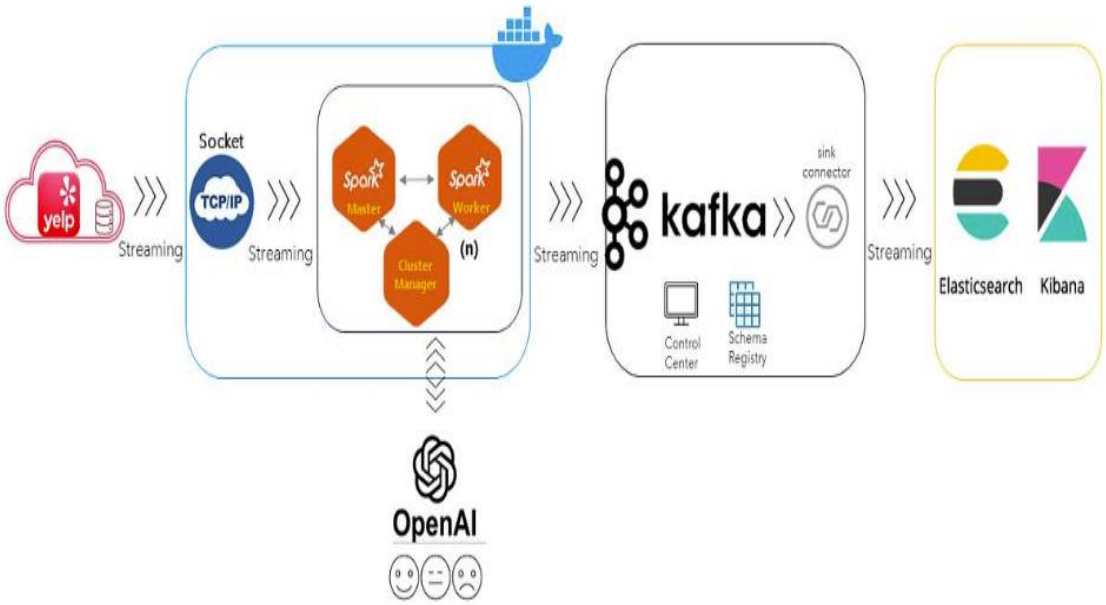
IV. Methodology

The development methodology of the GENAI Powered Sentiment Analysis for Social Media system includes data collection, preprocessing, model training, sentiment classification, visualization, testing, and deployment phases. Initially, YouTube comments and social media textual data were collected using APIs and data extraction techniques. The collected data underwent preprocessing tasks such as text cleaning, tokenization, stop-word removal, stemming, normalization, and noise reduction to prepare it for analysis. Feature extraction and vectorization methods were applied to represent textual information effectively for machine learning and AI processing. Advanced sentiment analysis models including deep learning architectures and

transformer-based Generative AI models were trained to classify sentiments accurately. The system then categorized comments into positive, negative, and neutral classes based on contextual understanding and semantic analysis. Visualization tools and dashboard modules were developed using charts, graphs, and statistical representations to display sentiment trends and audience reactions interactively. Evaluation metrics such as accuracy, precision, recall, and F1-score were used to measure model performance and classification quality. Optimization techniques improved processing speed and real-time analysis capabilities. Finally, the complete application was deployed as a web-based sentiment analysis platform. The methodology ensures scalability, maintainability, and efficient AI-powered social media analytics functionality.

System Architecture

The system architecture of the GENAI Powered Sentiment Analysis for Social Media follows a layered architecture consisting of data collection, preprocessing, AI processing, visualization, backend, and database layers. The data collection layer extracts comments and social media text from YouTube and other social media platforms using APIs or scraping tools. The preprocessing layer performs text cleaning, tokenization, stop-word removal, stemming, normalization, and feature extraction for efficient sentiment analysis. The AI processing layer integrates machine learning, deep learning, and transformer-based Generative AI models to analyze contextual meaning, semantics, and emotional tone within comments dynamically. The sentiment classification module categorizes comments into positive, negative, and neutral sentiments. The visualization layer provides interactive dashboards, charts, pie graphs, bar graphs, and sentiment trend analysis for better interpretation of audience reactions and engagement patterns. The backend layer handles application logic, API communication, real-time processing, and user request management operations. The database layer securely stores collected comments, processed text, sentiment results, user interaction data, and analytical reports for future reference and analysis. Security modules ensure safe data handling and API management. The modular architecture also supports future integration of multilingual analysis, recommendation systems, trend prediction, and cross-platform social media analytics. Overall, the architecture provides a scalable, intelligent, and efficient framework for AI-powered social media sentiment analysis systems.



V. Result and Output

VI. Conclusion

The GENAI Powered Sentiment Analysis for Social Media project successfully demonstrates the application of Generative Artificial Intelligence, Natural Language Processing, and deep learning techniques in analyzing public opinions and audience emotions from social media platforms such as YouTube. The system effectively extracts user comments, processes textual data, classifies sentiments, and presents meaningful insights through interactive dashboards and visualizations.

The proposed system overcomes many limitations of traditional sentiment analysis methods by utilizing advanced AI models capable of understanding contextual semantics, emotional tone, sarcasm, slang, and informal language commonly found in social media interactions. By integrating transformer-based architectures and Generative AI technologies, the system improves sentiment classification accuracy and provides more reliable audience analysis compared to rule-based and classical machine learning approaches.

The implementation of real-time dashboards and visualization tools enhances user understanding by presenting sentiment trends, audience engagement patterns, and statistical summaries through graphs and charts. This allows businesses, content creators, marketers, and researchers to analyze public reactions, customer feedback, and user engagement more efficiently. The system also reduces manual effort and enables faster decision-making based on large-scale social media data analysis.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.