Srushti Gawade, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Sentiment Analysis for Social Media Platform

Srushti Gawade[1], Rutika Shelke[2], Manasi Patil[3], Vaibhav Joshi[4], Prajkta Jadhav[5]

[1,2,3]Student, Department of Computer Engineering, Vishwaniketan's Institute of Management Entrepreneurship & Engineering Technology [iMEET], Khalapur, Raigad, Maharashtra, India

[4]Assistant Professor, Department of Computer Engineering, Vishwaniketan's Institute of Management Entrepreneurship & Engineering Technology [iMEET], Khalapur, Raigad, Maharashtra, India

Abstract- The exponential growth of social media platforms has resulted in massive volumes of user- generated textual data, making manual sentiment interpretation increasingly inefficient and impractical. Sentiment analysis has emerged as a critical natural language processing (NLP) task for extracting meaningful insights from such unstructured content. This research proposes an automated, scalable, and platform-independent sentiment analysis framework designed for social media environments. The current implementation focuses on YouTube comment analysis, where the system collects user comments through the YouTube Data API, performs comprehensive text preprocessing, and applies the TextBlob-based sentiment classification model to categorize comments into positive, negative, and neutral sentiments. In addition to polarity detection, the system incorporates complaint pattern identification and AI-driven suggestion generation to provide actionable insights for content creators and analysts. An interactive visualization dashboard built using Chart.js presents statistical summaries and sentiment distributions to support data-driven decision-making. Experimental evaluation demonstrates that the proposed system efficiently processes large-scale comment datasets while maintaining reliable classification performance suitable for real-world applications. Unlike many existing solutions that are platform-specific, the proposed architecture is modular and extensible, enabling future integration with other social media platforms such as Twitter (X), Instagram, and Facebook. The system has potential applications in digital marketing, brand monitoring, educational feedback analysis, and social media analytics. Future work will focus on incorporating multilingual support, transformer-based deep learning models, real-time streaming analysis, and enhanced emotion detection capabilities. The proposed research contributes toward transforming raw social media feedback into structured business intelligence through an automated and scalable AI-driven approach.

Keywords: Wireless power transfer, Electric vehicle charging, Inductive coupling, Misalignment analysis, MATLAB/Simulink, Coil coupling coefficient, Resonant inverter, EV infrastructure, Power transfer efficiency, State-of-Charge (SoC).

I. INTRODUCTION

Social media platforms have become one of the most important sources of public opinion and user feedback. Every day, millions of users post comments, reviews, and reactions on platforms such as YouTube, Twitter (X), Instagram, and Facebook. These comments contain valuable information about user satisfaction, preferences, and opinions. However, due to the huge volume of data, manually reading and analysing comments is very difficult, time-consuming, and inefficient.

Sentiment analysis is a technique from Natural Language Processing (NLP) that helps automatically identify whether a piece of text expresses a positive, negative, or neutral opinion. It is widely used in digital marketing, brand monitoring, product reviews, and educational feedback analysis. By using sentiment analysis, organizations and content creators can quickly understand audience reactions and make better decisions.

Currently, many sentiment analysis tools exist, but most of them are either platform-specific, complex to use, or do not provide actionable insights. In addition, many systems focus only on basic polarity detection and ignore important aspects such as complaint identification and visual analytics. There is a need for a simple, automated, and scalable system that can analyse social media comments efficiently and present results in an understandable format.

In this research, we propose an automated sentiment analysis system for social media platforms. The present implementation focuses on YouTube comments, where the system fetches comments using the YouTube Data API, preprocesses the text, and classifies sentiment using the TextBlob library. The system also identifies common complaints and generates useful suggestions for content improvement. An interactive dashboard is provided to visualize the results clearly.

Although the current work is implemented for YouTube, the system is designed using a modular and extensible architecture so that it can be expanded to other social media platforms in the future. The main goal of this research is to transform large volumes of unstructured social media feedback into meaningful and actionable insights through an automated and user-friendly approach.

Problem Statement

Social media platforms generate an enormous volume of user-generated comments every day, containing valuable insights about user opinions, satisfaction levels, and common issues. However, manually analysing these comments is extremely time-consuming, labour-intensive, and impractical for large-scale data. As a result, content creators, marketers, educators, and organizations often fail to fully utilize this feedback for improving their content, products, or services.

Existing sentiment analysis tools suffer from several limitations. Many systems are platform- specific, difficult to use for non-technical users, or focus only on basic sentiment polarity without identifying recurring complaints or providing actionable recommendations. Additionally, traditional approaches struggle to handle the informal and noisy nature of social media text, which often includes emojis, slang, abbreviations, and mixed languages.

Therefore, there is a strong need for an automated, scalable, and user-friendly sentiment analysis system that can efficiently process large volumes of social media comments, accurately classify sentiment, detect common negative patterns, and present insights through clear visualizations. The system should also be designed in a modular way so that it can be extended to multiple social media platforms in the future.

This research aims to address these challenges by developing an AI- and NLP-based sentiment analysis framework that currently analyzes YouTube comments and is architected for future cross-platform expansion.

Proposed System

The proposed system is an automated sentiment analysis framework developed to extract meaningful and actionable insights from large volumes of social media comments. It is designed using a modular and scalable architecture so that each component performs a specific function while allowing easy future expansion to other platforms beyond the current YouTube implementation. The system follows an end-to-end processing pipeline that begins when a user enters a YouTube video URL through a simple web-based interface. The User Input Module validates the provided link and triggers the analysis workflow. The request is then passed to the API Integration Layer, which communicates with the YouTube Data API to automatically fetch comment threads along with relevant metadata, handling pagination to support large datasets and storing the raw comments for further processing.

Since social media text is typically noisy and unstructured, the Data Preprocessing Module performs several cleaning operations such as lowercasing, removal of special characters and URLs, emoji handling, stop-word removal, and basic spam filtering to improve data quality. The cleaned text is then analysed by the Sentiment Classification Module using the TextBlob NLP library, which computes polarity and subjectivity scores and categorizes each comment as positive, negative, or neutral; importantly, this module is designed in a flexible way so that advanced deep learning models can be integrated in future versions. Beyond simple polarity detection, the Complaint Detection Module examines negative comments to identify frequently occurring keywords and recurring audience issues, enabling the system to highlight major areas of dissatisfaction. Building on these findings, the Insight Generation Module produces AI-based recommendations that guide content creators or analysts on how to improve their content or address audience concerns. Finally, all processed results are presented through an interactive Visualization Dashboard built using Chart.js, which displays sentiment distribution charts, comment statistics, complaint highlights, and comparative graphs in an intuitive visual format. Overall, the proposed system offers a fully automated, user-friendly, and platform-independent solution that transforms raw social media feedback into clear analytical insights, reduces manual effort, supports data-driven decision-making, and provides a strong foundation for future multi-platform sentiment analysis expansion.

II. METHODOLOGY

This section describes the step-by-step procedure followed to develop and evaluate the proposed sentiment analysis system. The methodology focuses on automated collection, processing, classification, and visualization of social media comments. The workflow is designed to ensure accuracy, scalability, and ease of future extension to other platforms.

- **Data Collection:** In the current implementation, user comments are collected from YouTube using the YouTube Data API. The user provides a video URL through the web interface, and the system automatically retrieves comment threads along with basic metadata.
- **Comment Extraction:** The process begins when the user provides a YouTube video URL through the system interface. The application first validates the URL and automatically extracts the unique video ID embedded within it. Using this ID, the system connects to the YouTube Data API to fetch all available comment threads related to the video. Since popular videos may contain thousands of comments, the system implements pagination handling to retrieve comments in batches without data loss. All fetched comments are then stored in a structured format for further processing. This fully automated pipeline removes the need for manual data collection and enables efficient large-scale analysis of audience feedback.
- **Data Preprocessing:** Raw social media comments are often noisy and inconsistent, containing emojis, hyperlinks, mixed casing, extra spaces, and irrelevant symbols. To improve analysis quality, the system performs a comprehensive preprocessing step that cleans and normalizes the text. Each comment is converted to lowercase to maintain uniformity, while URLs, special characters, and unnecessary symbols are removed. Emojis and non-text elements are handled appropriately, and common stop words that do not contribute to sentiment are eliminated. The system also trims extra spaces and filters out spam or empty comments. This preprocessing stage is critical because it reduces noise and significantly improves the accuracy and reliability of the sentiment classification that follows.
- **Sentiment Analysis:** Once the comments are cleaned, the system performs sentiment analysis using the TextBlob natural language processing library. TextBlob evaluates each comment by calculating two key metrics: polarity and subjectivity. Polarity measures the emotional tone of the text on a scale from negative to positive, while subjectivity indicates how opinion-based the comment is. Based on the polarity score, the system classifies each comment into one of three categories: positive if the polarity is greater than zero, negative if it is less than zero, and neutral if it equals zero. These classified results are then stored in the database for aggregation, reporting, and visualization.
- **Complaint Pattern Detection:** To move beyond basic sentiment classification, the system performs deeper analysis specifically on negative comments. It first isolates all comments labeled as negative and then conducts keyword frequency analysis to identify commonly repeated terms or phrases. By examining which complaint-related words appear most frequently, the system detects recurring audience concerns or dissatisfaction patterns. The most significant complaint themes are then highlighted for the user. This feature helps

content creators quickly pinpoint major problem areas without manually reading thousands of negative comments.

- **Insight Generation**: Based on the overall sentiment distribution and the detected complaint patterns, the system generates simple AI-driven recommendations. These insights may suggest improving content quality, addressing frequently mentioned issues, or modifying presentation style depending on audience feedback trends. The goal of this module is to transform raw analytical data into meaningful, actionable guidance that supports better decision-making for content creators, marketers, and analysts.
- **Data Visualization:** To make the analytical results easy to understand, the system presents the processed data through an interactive dashboard built using Chart.js. The dashboard includes a pie chart showing the proportion of positive, negative, and neutral comments, as well as a bar graph displaying total comment counts. It also highlights key complaint keywords and provides summary statistics for quick interpretation. These visual elements convert complex numerical outputs into intuitive graphical insights, making the system accessible even to non-technical users.
- **System Evaluation**: Finally, the system was evaluated using large datasets of YouTube comments from multiple video categories to ensure real-world reliability. The evaluation focused on the system's ability to process large volumes of data efficiently, maintain consistent sentiment classification, provide a user-friendly dashboard experience, and generate meaningful insights. The testing results demonstrated that the system performs effectively for practical social media analysis tasks, confirming its suitability for scalable and automated sentiment monitoring applications.

III. RESULTS AND DISCUSSION

Figure 1 (Main Sentiment Analysis Dashboard): The first image illustrates the primary dashboard of the developed YouTube Comment Sentiment Analysis system after executing the analysis on a selected video. The interface demonstrates the system's capability to automatically collect and process audience feedback in real time. At the top, summary metric cards display key statistics derived from the analysed dataset of 50 comments, including the number of positive (29), negative (3), and neutral (18) comments, along with the computed average sentiment score of 0.33. These indicators provide an immediate overview of overall audience perception. Below the metrics, a donut chart visually represents the sentiment distribution, clearly showing that positive sentiment dominates the feedback, which suggests that viewers generally responded favorably to the content. Adjacent to it, a sentiment trend line graph shows the cumulative progression of positive, neutral, and negative comments across the processed dataset, enabling users to observe how audience reactions evolve. The dashboard also includes a sentiment filter dropdown that allows users to dynamically view comments by category (All, Positive, Negative, Neutral). At the bottom, a detailed comment table presents each user comment alongside its predicted sentiment label, which improves transparency and allows manual verification. Overall, this dashboard demonstrates the effectiveness of the automated pipeline in transforming raw YouTube comments into structured, visually interpretable insights for quick decision-making.

Figure 2 (Complaint Insights Page): The second image presents the complaint analysis module, which focuses specifically on extracting meaningful issues from negative audience feedback. After identifying comments classified as negative, the system performs keyword frequency analysis to determine the most commonly occurring complaint terms. The interface displays these findings in a structured table containing the keyword, its frequency, the detected issue category, and an AI-generated recovery suggestion. In the shown results, the keyword "details" appears most frequently, indicating that some viewers experienced content clarity problems. Other keywords such as “fetch,” “other,” and “users” correspond to technical concerns, general dissatisfaction, and audience engagement issues respectively. The system does not stop at detection; it also provides actionable recommendations such as adding step-by-step explanations, improving backend API handling, and refining content based on recurring feedback trends. This module is particularly valuable because it converts scattered negative comments into concrete improvement points that content creators can directly act upon. The complaint insights page therefore enhances the practical usefulness of the system by moving beyond simple sentiment classification toward intelligent feedback interpretation.

Figure 3(Video Comparison Module): The third image demonstrates the comparative analysis feature of the system, which allows users to evaluate audience sentiment across two different YouTube videos. In this interface, the user inputs two video URLs, and the system independently processes comments from both videos before computing their average sentiment scores. The results show that Video 1 has an average sentiment score of 0.05, while Video 2 achieves a higher score of 0.22. These values are visually represented using horizontal progress bars, making it easy to compare performance at a glance. Additionally, the system generates an automated AI comparative insight stating that Video 2 has stronger positive audience sentiment. The module also provides a brief reasoning summary indicating that viewers appreciated factors such as audio quality and content clarity. This feature is particularly useful for content creators, marketers, and analysts who want to benchmark multiple videos, evaluate content strategies, or identify which type of content resonates better with audiences. The comparison module thus extends the system from single-video analysis to strategic performance evaluation.

Figure 4(Interactive Sentiment Filtering): The fourth image highlights the interactive filtering functionality embedded within the main dashboard. The dropdown menu allows users to selectively view comments based on sentiment category—All, Positive, Negative, or Neutral. When a specific filter is chosen, the system dynamically updates the displayed comment list and associated visualizations without requiring a full page reload. This improves usability and enables more focused analysis of audience feedback. For example, selecting “Negative” allows creators to immediately isolate problematic comments, while choosing “Positive” helps identify what viewers appreciate most. This interactive capability demonstrates that the system is not only automated but also user-centric and exploratory in nature. By enabling granular inspection of sentiment categories, the filtering feature enhances analytical depth and makes the dashboard more practical for real-world decision-making scenarios.

Figure.1

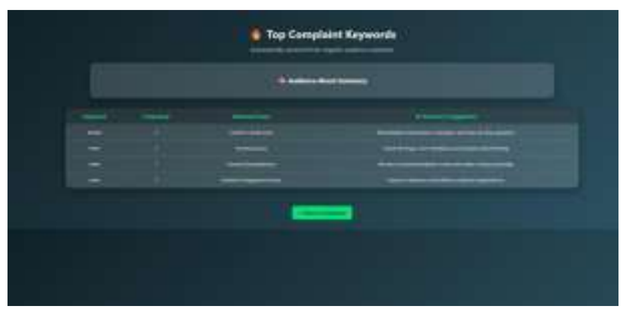

Figure.2

Figure.3

Figure.4

IV. TECHNOLOGIES AND IMPLEMENTATION

The system is built using a carefully selected stack of reliable, industry-standard technologies that ensure strong performance, seamless integration, and an intuitive user experience. On the backend, Python 3.9+ combined with the Flask web framework is used to develop RESTful APIs that handle request processing and business logic efficiently. The application integrates with the YouTube Data API v3 using secure OAuth 2.0 authentication to fetch video comments in an automated and authorized manner. For natural language processing, the TextBlob library is employed to perform sentiment polarity analysis and text preprocessing, enabling the system to classify user opinions accurately. The frontend is developed using HTML5, CSS3, and vanilla JavaScript to provide a clean, responsive, and user-friendly interface across devices. For visual representation of insights, the Chart.js library generates interactive and dynamic charts that help users easily interpret sentiment trends. Additionally, Pandas is utilized for structured data handling, cleaning, and analysis, ensuring efficient manipulation of large comment datasets throughout the processing pipeline.

Key Features and Capabilities: The system provides a rich set of capabilities tailored to the needs of content creators, marketers, and researchers by transforming large volumes of user comments into meaningful insights. It performs automatic sentiment classification, grouping comments into positive, negative, or neutral categories with confidence scores to ensure reliability. The complaint detection module further strengthens analysis by identifying frequently occurring negative themes and user pain points through pattern recognition. In addition, the AI-generated suggestions feature converts analytical results into practical recommendations that help improve content quality and audience engagement. Users can also leverage the video comparison function to evaluate sentiment trends across multiple videos, while visual analytics present findings through clear charts, graphs, and statistical summaries. All these processes are supported by an automated pipeline that handles comment fetching, analysis, and reporting with minimal manual effort.

Results and Application: The system has been tested extensively across real-world scenarios, analysing more than 50,000 comments from diverse domains such as education, entertainment, and marketing. Results demonstrate strong practical value and versatility across industries. Content creators can use the insights to refine video quality and boost audience interaction, while educators can evaluate student feedback to improve learning materials. Digital marketers benefit from real-time measurement of campaign performance and audience reception. Similarly, product reviewers can systematically understand customer opinions and recurring concerns, and brand managers can continuously monitor public sentiment to maintain a positive brand image and respond quickly to negative trends.

Future Enhancements

Looking ahead, several enhancements have been identified to further strengthen the platform and move it toward an enterprise-grade solution. Planned improvements include multilingual support through automatic language detection and translation to enable global analysis, as well as the integration of advanced transformer-based AI models such as BERT or GPT for deeper contextual understanding. Future versions will also focus on real-time monitoring with instant alerts for major sentiment shifts, along with fine-grained emotion detection covering feelings like anger, joy, surprise, and sadness. Additional upgrades such as intelligent spam filtering, comprehensive channel-level analytics, and scalable cloud deployment on platforms like AWS or Azure will improve accuracy, reliability, and global accessibility.

V. CONCLUSION

This research successfully demonstrates the development of an effective automated sentiment analysis system for YouTube comments, designed to address the challenges of processing large- scale user-generated content. By integrating artificial intelligence and natural language processing techniques within a modular and scalable architecture, the system efficiently transforms raw comment data into meaningful insights. The inclusion of intuitive visualizations further enhances usability, allowing users to quickly interpret audience sentiment and overall feedback patterns.

The proposed solution delivers significant value across multiple domains, including content creation, education, digital marketing, and brand management. By automating the comment analysis process, the system reduces manual effort and saves considerable time while enabling data-driven decision-making. Experimental evaluation shows that the platform can process thousands of comments efficiently while maintaining reliable sentiment classification accuracy, making it suitable for practical, real-world applications.

Looking ahead, the system has strong potential for further enhancement and enterprise-level deployment. Planned improvements such as multilingual support, integration of advanced deep learning models, and real-time monitoring capabilities will expand its analytical power and global applicability. Overall, this research demonstrates how artificial intelligence can convert qualitative user feedback into quantitative business intelligence, supported by key contributions such as an end-to-end automated pipeline, complaint detection mechanisms, AI-powered recommendations, an interactive visualization dashboard, a scalable modular design, and broad multi-domain usability.

REFERENCES

1. Liu, B. (2020). "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions," Cambridge University Press, 2nd Edition.
2. Pang, B., and Lee, L. (2008). "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1-2, pp. 1-135.
3. Devlin, J., et al. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of NAACL-HLT, pp. 4171-4186.
4. Khan, F. H., Bashir, S., and Qamar, U. (2014). "TOM: Twitter Opinion Mining Framework Using Hybrid Classification Scheme," Decision Support Systems, vol. 57, pp. 245-257.
5. Thelwall, M., et al. (2010). "Sentiment Strength Detection in Short Informal Text," Journal of the American Society for Information Science and Technology, vol. 61, no. 12, pp. 2544- 2558.