Prof Supriya G Purohit, 2025, 13:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

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

Spamshield Sentiment Analysis on Youtube Comments

Prof.Supriya G Purohit, Ms.G Keerthi, Ms.Hafsa Zareen , Ms. Ayesha Hasan Osmani
Dept Computer science and engineering, Navodaya institute of technology, Raichur

Abstract - The exponential growth of social media platforms has resulted in an overwhelming volume of user-generated textual content, making effective content moderation and opinion analysis increasingly challenging. YouTube, as one of the most popular video-sharing platforms, receives millions of comments daily, which include genuine feedback as well as spam, promotional messages, and emotionally charged content. Manual analysis of such large-scale data is inefficient, time-consuming, and prone to inconsistencies. Therefore, there is a growing need for automated systems capable of analyzing user comments and extracting meaningful insights in real time. This paper presents SpamShield, a web-based automated system designed to analyze YouTube comments using Natural Language Processing (NLP) techniques. The proposed system retrieves real-time comments from YouTube videos using the YouTube Data API and performs comprehensive text preprocessing to remove noise and normalize the data. Preprocessing steps include text normalization, removal of special characters and URLs, tokenization, and stop-word elimination, ensuring that the comments are suitable for reliable analysis. The sentiment analysis process is implemented using the Natural Language Toolkit (NLTK) along with the VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon, hich is specifically optimized for analyzing social media text. Each comment is evaluated based on lexical and contextual features, and the sentiment polarity is classified into positive, negative, or neutral categories. The system further aggregates sentiment results to generate comprehensive sentiment reports that provide a clear overview of audience opinions and engagement patterns. Experimental evaluation demonstrates that the proposed system effectively analyzes informal and unstructured social media text, offering reliable sentiment classification and intuitive visualization of results. The modular and scalable architecture of SpamShield enables efficient processing of large volumes of comments and supports real-time analysis. The proposed approach provides a practical solution for content creators, marketers, and researchers to understand audience sentiment, enhance user engagement, and support data-driven decision- making in social media environments.

Keywords - YouTube Comment Analysis, Spam Detection, Sentiment Analysis, Natural Language Processing (NLP), Social Media Analytics.

I. INTRODUCTION

Social media platforms have become an integral part of modern communication, enabling users to share opinions, experiences, and feedback on a global scale. Among these platforms, YouTube stands out as one of the largest video-sharing services, hosting a vast amount of user-generated content and receiving millions of comments daily. These comment sections play a vital role in enhancing user engagement and interaction between content creators and audiences. However, the rapid growth of comment volume has also introduced challenges such as spam messages, promotional advertisements, abusive language, and highly negative or misleading content, which degrade the quality of online discussions.

Manual moderation of YouTube comments is increasingly impractical due to the scale and speed at which new content is generated. Content creators often struggle to analyze audience feedback efficiently while maintaining a positive and constructive community. As a result, automated solutions that can process large volumes of textual data and extract meaningful insights are essential for effective content moderation and audience analysis.

Sentiment analysis is a key application of Natural Language Processing (NLP) that focuses on identifying the emotional polarity of textual content. It typically classifies text into positive, negative, or neutral categories based on the expressed sentiment. In the context of social media, sentiment analysis provides valuable insights into user opinions, satisfaction levels, and engagement trends. Alongside sentiment analysis, automated text processing techniques help in handling unstructured and informal language commonly found in online comments.

Recent advancements in NLP and machine learning have significantly improved the ability to analyze large-scale text data with higher accuracy and efficiency. Lexicon-based and machine learning-based sentiment analysis methods have been widely adopted to handle social media content due to their adaptability and performance. These techniques enable automated systems to understand user emotions and patterns without requiring extensive manual intervention.

This paper proposes SpamShield, an automated system designed to analyze YouTube comments and provide sentiment-based insights. The system retrieves real-time comments using the YouTube Data API, performs text preprocessing to remove noise, and applies NLP-based sentiment analysis techniques to classify comments. By generating sentiment reports and visual summaries, the proposed system assists content creators and organizations in understanding audience feedback, improving engagement quality, and supporting data-driven decision-making. The modular design of the system ensures scalability and provides a foundation for future enhancements in social media analytics.

II. LITERATURE REVIEW

The rapid growth of user-generated content on social media platforms has attracted significant research interest in the areas of spam detection and sentiment analysis. Automated analysis of online comments is essential for maintaining healthy digital environments and understanding user opinions at scale.

Hayoung Oh proposed a YouTube spam comment detection scheme using a cascaded ensemble machine learning model [1]. The study focused on combining multiple classifiers to improve detection accuracy and demonstrated that ensemble-based approaches outperform single classifiers in identifying spam comments. However, the computational complexity of ensemble models can limit their applicability in real-time systems.

Abdulah et al. presented a comparative analysis of common YouTube comment spam filtering techniques [2]. Their work evaluated several machine learning algorithms, including Naive Bayes and Support Vector Machines, and concluded that probabilistic classifiers are efficient and suitable for large-scale text-based spam detection. The study emphasized the importance of feature selection and preprocessing but did not address sentiment analysis.

Muhammad et al. explored sentiment analysis of YouTube comments using the Naïve Bayes Support Vector Machine (NBSVM) classifier [3]. Their research demonstrated that hybrid classifiers can improve sentiment classification accuracy by combining probabilistic and margin-based learning techniques. However, the study focused solely on sentiment polarity and did not consider spam filtering as a preprocessing step.

Sah and Parmar proposed an approach for malicious spam detection in email systems using multiple machine learning classifiers [4]. Their results showed that supervised learning models significantly

outperform rule-based systems in detecting spam. Although the domain was email communication, the techniques and findings are applicable to social media spam detection due to similarities in textual patterns.

Gunawan et al. conducted sentiment analysis on mobile application reviews using Naive Bayes classification combined with text normalization techniques [5]. The study highlighted the importance of preprocessing steps such as lemmatization and normalization in improving classification performance. However, the work was limited to review-based sentiment analysis and did not address real-time social media data.

From the existing literature, it is evident that both spam detection and sentiment analysis have been extensively studied as independent research problems. However, limited work has focused on integrating both tasks into a single automated system for real-time content moderation. The proposed SpamShield system addresses this gap by combining Naive Bayes-based spam detection with SVM-based sentiment analysis in a unified framework, providing efficient moderation and meaningful audience insights for YouTube content creators.

III. SYSTEM ARCHITECTURE OVERVIEW

The system architecture illustrates the overall workflow of the proposed SpamShield Sentiment Analysis system, highlighting the process of collecting, cleaning, analyzing, and generating sentiment reports from YouTube comments. The architecture is designed to enable automated extraction and analysis of user- generated content in a structured and efficient manner.

Initially, the system interacts with the YouTube platform, where comments associated with a specific video are retrieved. The YouTube Data API is used to extract real-time comments from the selected YouTube video. These comments are stored temporarily for further processing and analysis.

Once the comments are collected, they are passed to the Data Cleaning module, which performs preprocessing operations to remove noise from the raw textual data. This step includes eliminating unnecessary symbols, URLs, special characters, and redundant text, ensuring that the comments are suitable for accurate analysis.

After preprocessing, the cleaned comments are forwarded to the Comment Analysis module, where the textual data is examined to determine sentiment characteristics. Natural Language Processing techniques are applied to understand the structure and meaning of the comments.

The system then applies NLTK and VADER Lexicon-based sentiment analysis techniques to evaluate the emotional polarity of each comment. Based on the lexical scores obtained, comments are classified into sentiment categories such as positive, negative, or neutral.

Finally, the analyzed results are used for the Generation of Sentiment Reports, which present an overall summary of audience sentiment. These reports provide valuable insights into user opinions and engagement patterns, helping content creators understand feedback and improve content quality. The modular design of the architecture ensures scalability, ease of maintenance, and effective handling of large volumes of YouTube comments.

The major components include:

- Data Retrieval Module: Fetches comments from YouTube using the YouTube Data API.
- Text Preprocessing Module: Cleans raw comments by removing noise such as URLs, emojis, and special characters.
- Spam Detection Module: Uses a Multinomial Naive Bayes classifier to identify spam comments.
- Sentiment Analysis Module: Applies a Support Vector Machine (SVM) to classify sentiment polarity.
- Visualization Module: Displays sentiment distribution and spam statistics using charts and tables.

The sequential data flow ensures spam comments are filtered before sentiment analysis, improving result reliability.

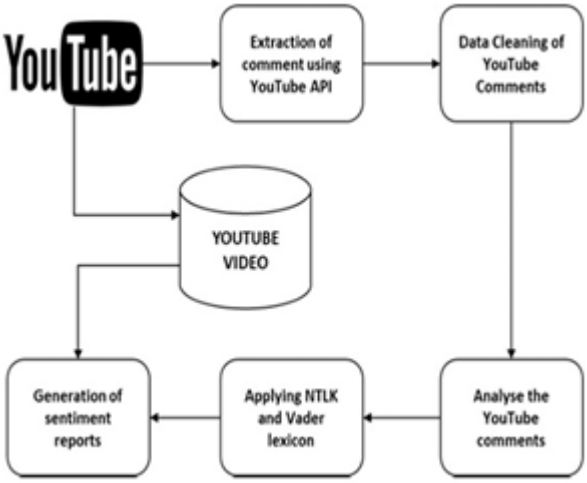

Fig: System Architecture of YouTube Comment Sentiment Analysis

IV. METHODOLOGY

The proposed methodology outlines the step-by-step process adopted in the SpamShield Sentiment Analysis system to analyze YouTube comments and generate sentiment-based reports. The methodology focuses on automated data collection, preprocessing, sentiment evaluation, and result generation using Natural Language Processing techniques.

Comment Extraction Using YouTube Data API

The first step of the methodology involves collecting user comments from YouTube videos. The system utilizes the YouTube Data API to retrieve real-time comments associated with a selected video. By providing a valid video ID, the API fetches comment threads efficiently, enabling automated data acquisition without manual intervention. The retrieved comments are temporarily stored for further processing.

Data Cleaning and Preprocessing

Raw YouTube comments often contain noise such as URLs, emojis, special characters, and unnecessary whitespace. To ensure accurate analysis, the collected comments undergo a data cleaning process. This includes removing unwanted symbols, converting text to lowercase, and eliminating redundant information.

Preprocessing helps standardize the textual data and improves the reliability of sentiment evaluation.

Comment Analysis

After preprocessing, the cleaned comments are analyzed to understand their textual structure and contextual meaning.

This stage prepares the data for sentiment evaluation by organizing the text into a suitable format for Natural Language Processing tools. The analysis phase ensures that each comment is processed individually to maintain classification accuracy.

Sentiment Analysis Using NLTK and VADER Lexicon

The sentiment analysis process is carried out using the Natural Language Toolkit (NLTK) along with the VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon. VADER is a rule-based sentiment analysis technique that is particularly effective for social media text.

Each comment is assigned a sentiment score based on lexical features, and the polarity is classified as positive, negative, or neutral.

Generation of Sentiment Reports

In the final stage, sentiment classification results are aggregated to generate sentiment reports. These reports provide a summarized view of audience sentiment and help identify overall trends in user feedback.

The output is presented in an easily interpretable format, enabling content creators to gain meaningful insights into viewer opinions and engagement levels.

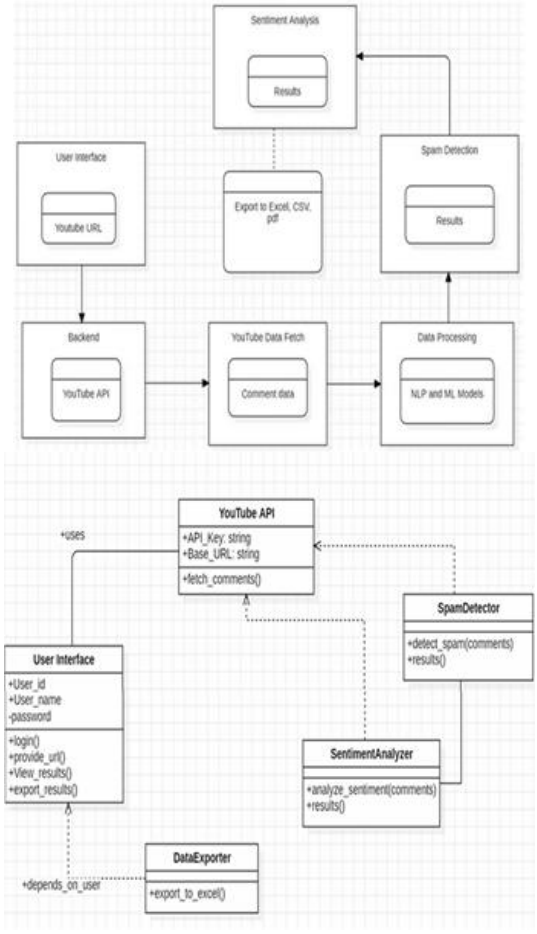

Fig: Master Class Diagram

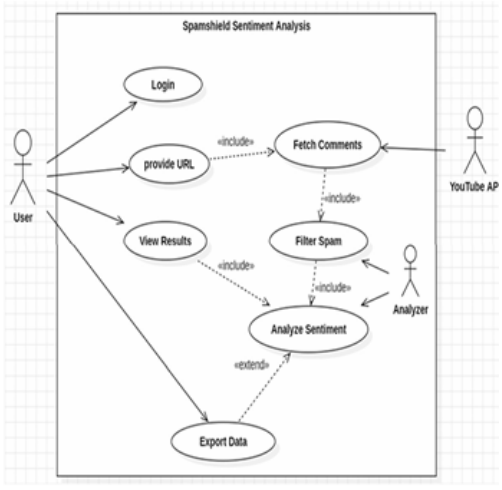

Fig : User-Interface Diagram

Dataset Collection and Preprocessing

Dataset Collection

The dataset used in the proposed SpamShield Sentiment Analysis system consists of user comments collected from YouTube videos. The comments are retrieved dynamically using the YouTube Data API, which allows programmatic access to publicly available comment threads associated with a given video. By providing a valid YouTube video ID, the system automatically fetches comments in real time, eliminating the need for manual data collection.

The collected dataset includes a diverse range of user- generated comments such as opinions, feedback, promotional messages, and informal social media text. This diversity helps in effectively analyzing real-world sentiment patterns and user behavior. The dataset is stored temporarily in a structured format for further processing and analysis. Only English-language comments are considered in the current implementation to maintain consistency and improve analysis accuracy.

Data Preprocessing

Raw comments obtained from social media platforms often contain noise and unstructured elements that can negatively impact sentiment analysis performance. Therefore, a preprocessing stage is applied to clean and normalize the collected data before analysis.

The preprocessing steps include the following:

- Text Normalization: All comments are converted to lowercase to ensure uniform representation and avoid case-sensitive inconsistencies.
- Removal of Special Characters and URLs: Unnecessary symbols, numbers, emojis, and web links are removed using regular expressions to reduce noise.
- Tokenization: Each comment is broken down into individual words or tokens to enable lexical analysis.
- Stop-word Removal: Commonly used words that do not contribute to sentiment, such as “the”, “is”, and “and”, are removed to reduce dimensionality.

- Whitespace Handling: Extra spaces and repeated characters are eliminated to improve text clarity.

After preprocessing, the cleaned comments are transformed into a format suitable for sentiment analysis using NLP tools. This preprocessing phase significantly enhances the reliability and accuracy of the sentiment classification process by ensuring that only meaningful textual information is analyzed.

Training & Implementation Workflow

The training and implementation workflow of the proposed SpamShield Sentiment Analysis system describes the sequence of steps involved in preparing the data, applying sentiment analysis techniques, and generating meaningful outputs. The workflow ensures efficient processing of YouTube comments and accurate classification of sentiment.

Data Preparation

The workflow begins with the preparation of the dataset obtained through the YouTube Data API. The collected comments undergo preprocessing steps such as text normalization, removal of special characters and URLs, tokenization, and stop-word elimination. These steps convert raw, unstructured text into a clean and consistent format suitable for Natural Language Processing tasks.

Sentiment Analysis Configuration

After preprocessing, the cleaned text is passed to the sentiment analysis module. The system utilizes the NLTK library along with the VADER sentiment lexicon, which is specifically designed to handle social media text. VADER assigns polarity scores to each comment based on lexical features and contextual intensity, enabling accurate sentiment interpretation.

Sentiment Classification

Each preprocessed comment is evaluated using the VADER scoring mechanism, which produces positive, negative, neutral, and compound scores. Based on predefined threshold values, the compound score is used to classify comments into positive, negative, or neutral sentiment categories. This rule-based classification eliminates the need for extensive supervised training while maintaining reliable performance for real-world social media data.

System Implementation

The complete system is implemented using Python, with backend processing handled through the Flask framework. The implementation integrates data collection, preprocessing, sentiment evaluation, and result generation into a single automated pipeline. Each module operates independently, allowing efficient execution and easy maintenance.

Output Generation and Visualization

In the final stage, sentiment classification results are agregated to generate reports. The system produces graphical representations such as pie charts and bar charts to visualize sentiment distribution. These outputs provide content creators with clear insights into audience opinions and engagement trends.

Performance Evaluation

The performance of the proposed SpamShield Sentiment Analysis system was evaluated based on its ability to accurately analyze YouTube comments and generate meaningful sentiment insights. The evaluation focused on the effectiveness of data preprocessing, sentiment classification reliability, system efficiency, and overall usability.

Effectiveness of Preprocessing

Text preprocessing plays a crucial role in improving sentiment analysis performance. The removal of noise such as URLs, emojis, special characters, and redundant symbols significantly enhanced the quality of the input data. Normalization and tokenization ensured consistency in text representation, which contributed to more reliable sentiment interpretation. The preprocessing stage reduced ambiguity and improved the overall robustness of the analysis pipeline.

Sentiment Classification Performance

The system utilizes the NLTK library with the VADER sentiment lexicon, which is specifically designed for analyzing social media text. VADER effectively captured sentiment intensity, negation, and contextual emphasis commonly found in YouTube comments. The classification into positive, negative,

and neutral categories demonstrated reliable performance for informal and short-text data, which is typical of user-generated content.

System Efficiency

The automated workflow enabled efficient processing of large volumes of comments without manual intervention. Real-time comment retrieval using the YouTube Data API and streamlined text processing ensured timely sentiment evaluation. The lightweight nature of the NLP-based approach allowed the system to operate efficiently on standard computing resources.

Visualization and Interpretability

The performance of the system was further enhanced through clear visualization of results. Graphical outputs such as pie charts and bar graphs enabled quick interpretation of sentiment distribution and audience feedback trends. These visual representations improved the usability of the system by allowing content creators to easily understand sentiment patterns.

Overall System Performance

The integrated architecture of SpamShield successfully combines data collection, preprocessing, sentiment analysis, and result visualization into a unified framework. The system demonstrated stable and consistent performance when applied to real-time YouTube comments. The results indicate that the proposed approach is effective for supporting content moderation and analyzing audience sentiment in social media environments.

Results and Discussions

This section discusses the outcomes obtained from implementing the proposed SpamShield Sentiment Analysis system on YouTube comment data. The results demonstrate the effectiveness of the system in analyzing user sentiment and providing meaningful insights into audience feedback.

The system was tested using real-time comments retrieved from YouTube videos through the YouTube Data API. The collected comments included a wide variety of textual expressions such as opinions, informal language, and feedback commonly found on social media platforms. After preprocessing, the cleaned comments were processed through the sentiment analysis module.

The sentiment classification results showed that the system was able to successfully categorize comments into positive, negative, and neutral classes. The use of the VADER sentiment lexicon proved effective in handling informal expressions, capitalization, and intensity commonly present in YouTube comments. Positive comments typically reflected appreciation and approval, while negative comments captured dissatisfaction and criticism. Neutral comments represented factual statements or emotionally balanced responses.

The visualization of results played a significant role in interpreting system performance. Pie charts and bar graphs provided a clear overview of sentiment distribution, enabling easy identification of dominant sentiment trends. These visual insights allowed content creators to quickly assess overall audience response without manually reviewing individual comments.

The discussion of results indicates that preprocessing significantly improved sentiment analysis reliability by removing noise and standardizing text. The automated nature of the system reduced the need for manual moderation and enabled scalable analysis of large comment volumes. Overall, the results confirm that the proposed system effectively supports sentiment-based analysis of YouTube comments and assists content creators in understanding audience engagement.

Real Life Applications

The proposed SpamShield Sentiment Analysis system has significant practical relevance in real-world social media environments where large volumes of user-generated content are produced continuously. With platforms like YouTube receiving millions of comments daily, automated systems such as SpamShield play an important role in managing, analyzing, and interpreting audience feedback efficiently.

One of the primary real-life applications of the system is automated comment moderation for YouTube content creators. Manually reviewing thousands of comments is time- consuming and often impractical. The proposed system enables creators to automatically analyze comments and understand overall audience sentiment, helping them identify positive feedback, recurring complaints, and negative reactions. This improves engagement quality and supports informed decision- making regarding content improvement.

The system is also highly useful in brand reputation monitoring and digital marketing analysis. Companies and organizations frequently use YouTube to promote products, services, and campaigns. By analyzing sentiment in viewer comments, marketers can evaluate public response, measure customer satisfaction, and detect potential issues early. These insights assist businesses in refining marketing strategies and improving customer engagement.

Another important application lies in spam and abusive content management. Social media platforms often suffer from spam comments, irrelevant promotions, and abusive language. Automated sentiment and comment analysis systems help identify such harmful content, supporting platform administrators and moderators in maintaining a safer and more constructive online environment.

The proposed system can also be applied in the education and e-learning domain. Educational institutions and instructors can analyze student comments on lecture videos and online courses to assess learner satisfaction and content effectiveness. Sentiment analysis helps educators understand student feedback and improve teaching methods and learning materials.

In addition, the system can support public opinion analysis and social research. Researchers can analyze sentiment trends related to social issues, awareness campaigns, and public discussions using YouTube comments as a data source. The scalable nature of SpamShield makes it suitable for large-scale studies involving real-time social media data.

Overall, the SpamShield system provides a versatile and scalable solution for sentiment analysis in social media platforms, making it valuable for content creators, businesses, educators, researchers, and platform administrators.

Future Scope

The proposed SpamShield Sentiment Analysis system provides a strong foundation for automated analysis of YouTube comments; however, there are several opportunities for future enhancement to improve its accuracy, scalability, and functionality. These extensions can further strengthen the system's applicability in real-world social media environments.

One major area for future improvement is the integration of advanced machine learning and deep learning models. While the current system employs lexicon-based sentiment analysis techniques, future versions can incorporate models such as Long Short-Term Memory (LSTM), Bidirectional LSTM, and Transformer-based architectures like BERT. These models can better capture contextual information, sarcasm, and complex sentence structures commonly found in social media comments, thereby improving sentiment classification accuracy.

Another important extension is the support for multilingual sentiment analysis. Currently, the system focuses on English- language comments. Future enhancements can include language detection and translation mechanisms to analyze comments written in multiple regional and international languages. This would make the system more inclusive and applicable to a global audience.

The system can also be expanded to include integrated spam, sentiment, and toxicity detection within a single framework. By using multi-label classification techniques, future implementations can simultaneously identify spam comments, abusive language, and emotional polarity, providing a more comprehensive content moderation solution.

Scalability can be further improved through cloud-based deployment. Migrating the system to cloud platforms would enable real-time processing of massive comment volumes and ensure high availability for popular channels with large audiences. Cloud integration would also support distributed processing and faster response times.
Another promising direction is the development of interactive real-time dashboards. Advanced visualization tools can be used to display live sentiment trends, historical comparisons, and engagement analytics. Such dashboards would greatly enhance usability for content creators, marketers, and analysts.

Future work may also explore emotion-level analysis beyond basic sentiment categories. Detecting emotions such as joy, anger, frustration, and sadness would provide deeper insights into audience behavior and reactions. This finer-grained analysis could be particularly useful for marketing and psychological studies.

Overall, these future enhancements would transform SpamShield into a more intelligent, scalable, and comprehensive social media analytics platform capable of addressing complex challenges in online content analysis and moderation.

V. CONCLUSION

This paper presented SpamShield, an automated system for analyzing YouTube comments using Natural Language Processing techniques. The proposed system effectively retrieves user comments through the YouTube Data API, performs data preprocessing to remove noise, and applies sentiment analysis to classify comments into positive, negative, and neutral categories. The integrated workflow enables efficient handling of large volumes of user-generated content without manual intervention.

The use of NLP-based sentiment analysis techniques proved effective in interpreting informal and unstructured social media text. Preprocessing significantly enhanced the quality of the input data, resulting in more reliable sentiment classification. The visualization of results further improved interpretability by providing clear insights into audience sentiment trends.

The experimental outcomes demonstrate that the SpamShield system can assist content creators and organizations in understanding viewer feedback, improving engagement quality, and supporting moderation decisions. The modular architecture ensures flexibility, scalability, and ease of future enhancements.

Overall, the proposed system offers a practical and efficient solution for sentiment analysis of YouTube comments and serves as a strong foundation for advanced research in social media analytics and automated content moderation.

REFERENCES

1. HAYOUNG OH, A YouTube Spam Comments Detection Scheme Using Cascaded Ensemble Machine Learning Model, National Research Foundation of Korea through 'Convergence and Open Sharing System".
2. [6] Abdulah O. Abdulah, Mashhood Ali, Murat Karabatak, Abdulkadir Sengur, A Comparative Analysis of Common YouTube Comment Spam Filtering Techniques.
3. Abbi Nizar Muhammad, d Saiful Bukhori, Priza Pandunata, Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes Support Vector Machine (NBSVM) Classifier, Proc. ICOMITEE 2019, October 16th-17th 2019, Jember.
4. Sah, U. K., & Parmar, N. (2017). An approach for Malicious Spam Detection in Email with comparison of different classifiers
5. F. Gunawan, M. A. Fauzi dan P. P. Adikara, "Sentiment analysis on mobile application reviews using Naïve Bayes and Levenshtein Distance-based word normalization (Case study of BCA mobile applications)," SYSTEMIC, pp. 1-6, 201