RESEARCH ARTICLE OPEN ACCESS

Spam Comments Detection on YouTube Based on Machine Learning

[1]Mr Basavaraj Nayak, [2]Dhanalakshmi, [3]Lasya Priya M, [4]Deepthi C

[1]Department of Computer Science and Engineering, R.J Jalappa Institute of Technology, Bangalore Rural Email: basavarajnayak_cse@rljit.in
[2]Department of Computer Science and Engineering, R J Jalappa Institute of Technology, Bangalore Rural Email: Dhanalakshmidhanu2004@gmail.com
[3]Department of Computer Science and Engineering, R J Jalappa Institute of Technology, Bangalore Rural Email: lasyapriyaa763@gmail.com
[4]Department of Computer Science and Engineering, R J Jalappa Institute of Technology, Bangalore Rural Email: deepthideepthi544@gmail.com

-----------------------------------************************-------------------------------

Abstract:

YouTube provides only some tools for the modification of the comments in the comment section. Because of this, the volume of spam comments increasing rapidly. Using Machine Learning, the comments can be detected and prevented. There are a lot of approaches in ML to spam. It is often seen in applications like YouTube where people watch a lot of videos for so many purposes it can be for entertainment or learning and it provides a way for users to interact with the creators through the comment section. There exists a way where people post scam comments which are quite harmful. In the current project, spam detection was performed on selected you tube comments and we predict the entered comment as 'SPAM' or 'NOT SPAM'. Google Safe Search and You tube Bookmaker are two tools that can be used to block harmful links; however, they cannot immediately safeguard real-time users. Four Artificial Intelligence approaches have been employed by victimization to run the survey for the spam comment detection methodology. With the raised quality of online social networks, spammers realize these platforms are simple to lure users into malicious activities by posting spam messages in the comments section of the videos. The dataset consists of labelled YouTube comments, and text preprocessing was performed using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization.

Keywords- Machine learning, spam, you tube, social media, comments, predict, data analysis, Performance evaluation, Classification.

-----------------------------------************************-------------------------------

I. INTRODUCTION

YouTube has become one of the largest online platforms for video sharing and user interaction, hosting millions of videos and billions of user comments daily. The comment section plays a crucial role in promoting engagement, discussion, and feedback between content creators and viewers. However, the rapid growth of user-generated content has also led to a significant increase in spam comments. These spam comments often include promotional links, misleading information, phishing attempts, and repetitive or irrelevant content, which degrade the quality of discussions and negatively affect user experience. Manual moderation of spam comments is inefficient and impractical due to the massive volume of data generated on YouTube every day. Traditional rule-based filtering systems struggle to adapt to the evolving patterns of spam, as spammers frequently change their strategies to bypass static rules.

Therefore, there is a strong need for automated, intelligent systems capable of accurately identifying and filtering spam comments in real time.

Machine Learning (ML) techniques provide an effective solution to this problem by learning patterns and features from large datasets of labeled comments. By analyzing textual characteristics such as word frequency, sentiment, syntax, and contextual information, ML-based models can distinguish between spam and legitimate comments with high accuracy. Popular algorithms such as Naïve Bayes, Support Vector Machines, Decision Trees, and Neural Networks have shown promising results in text classification tasks, including spam detection. This paper focuses on the development and evaluation of a machine learning-based approach for detecting spam comments on YouTube. The proposed system aims to improve detection accuracy, reduce false positives, and enhance the overall quality of online interactions. By implementing and comparing different machine learning models, this research contributes to the development of scalable and reliable automated moderation systems for social media platforms.

II. LITERATURE SURVEY

Barushka [2020], Behaviour-Based Spam Detection Using Ensemble Methods

Barushka introduced behaviour-driven analysis using Random Forest and other ensemble models to capture abnormal user activities like excessive posting or broad user engagement. The study showed that ensemble techniques could better generalize user behaviour patterns compared to individual classifiers. However, similar to Nascimento et al., the model relied heavily on detailed user logs, making practical deployment difficult in platforms with strict privacy constraints.

Khan et al. [2021] / Hakak et al. [2021], Machine Learning Methods for Text-Based Spam Detection

These studies focused on transforming text into numerical representations such as TF-IDF vectors and word embeddings for effective machine learning classification. Both works demonstrated that SVMs and decision trees performed well in high-dimensional text classification tasks but required intensive feature engineering. The authors also pointed out that imbalanced datasets—where spam messages form a minority—significantly hindered model performance, necessitating additional preprocessing or sampling methods.

Andresini et al. [2022], Probabilistic Spam Classification Using Naive Bayes Models

This study introduced Naive Bayes as an improvement over earlier rule-based systems, utilizing word frequency patterns and conditional probability estimates to classify messages. The authors demonstrated that Naive Bayes could handle large vocabularies efficiently and offered more flexible decision boundaries. Nevertheless, they noted that its underlying assumption of feature independence often reduced accuracy in real-world scenarios, especially when contextual or sequential relationships between words were important. As a result, Naive Bayes performed inconsistently in more complex forms of spam, such as those found on social media.

Gasparetto et al. [2022], Zeakis et al. [2023], Transformer-Based Text Representation in Spam Detection [7]

These studies evaluated the effectiveness of advanced word embeddings such as Word2Vec, GloVe, and contextual transformer-based models like BERT. Their findings showed that BERT offered superior semantic understanding and improved classification accuracy compared to earlier methods. However, they also highlighted challenges: models based on transformers are computationally expensive, complex to integrate, and may be impractical for large-scale, real-time moderation systems due to latency issues.

Nascimento et al. [2023], Social Media Spam Detection Through Content and Behaviour Analysis

Nascimento and collaborators explored the transition of spam detection into social media environments, identifying new challenges such as multimedia content and diverse user behaviour. Their work distinguished between content-based and behaviour-based approaches. Content-based detection leveraged TF-IDF, embeddings, SVMs, and decision trees to classify textual characteristics of posts and comments. Behaviour-based detection focused on user activity patterns such as posting frequency and network connections.

III. EXISTING SYSTEM

The existing system for detecting spam comments on YouTube primarily relies on basic filtering mechanisms, manual moderation, and limited automated techniques. While these approaches aim to reduce the spread of spam, malicious links, and irrelevant content, they face several challenges in terms of accuracy, scalability, and adaptability.

1. Manual Moderation System

In traditional approaches, spam comments are identified and removed manually by content creators or platform moderators. This method is highly time-consuming and impractical for popular channels that receive thousands of comments ежедневно. Manual moderation is also subjective, error-prone, and unable to respond effectively in real time, allowing spam content to remain visible for extended periods.

2. Rule-Based Filtering Techniques

Some existing systems use predefined rules or keyword-based filters to detect spam comments. These systems flag comments containing suspicious words, repeated characters, or excessive links. However, spammers often bypass these filters by using obfuscated text, emojis, or creative spelling. As a result, rule-based systems suffer from low detection accuracy and high false-positive rates.

3. User Reporting Mechanisms

YouTube allows users to report spam or inappropriate comments. While this crowdsourced approach helps identify harmful content, it depends heavily on user participation and delayed action. Spam comments can still influence viewers before being reported and reviewed, making this method reactive rather than proactive.

4. Basic Automated Detection Systems

Existing automated spam detection systems use simple machine learning models trained on limited datasets. These systems often fail to generalize well to new spam patterns due to insufficient feature extraction and lack of continuous learning. Additionally, they struggle to handle multilingual comments and evolving spam strategies.

5. Limitations of Existing Systems

Overall, existing spam comment detection systems lack real-time processing, adaptability to new spam techniques, and high accuracy across diverse comment types. Limited use of advanced machine learning algorithms, insufficient training data, and reliance on static rules reduce their effectiveness. These limitations highlight the need for a robust machine learning–based system capable of automatically detecting and filtering spam comments with higher precision and scalability.

DISADVANTAGES OF EXISTING SYSTEM

The limitations and shortcomings of current spam comment detection approaches on YouTube create significant challenges in maintaining content quality, user trust, and platform integrity.

1. Low Detection Accuracy

Existing systems often fail to accurately distinguish between spam and legitimate comments. Rule-based and basic machine learning approaches generate high false positives, where genuine comments are mistakenly flagged as spam, and false negatives, where spam comments remain undetected. This reduces user satisfaction and weakens moderation effectiveness.

2. Dependence on Manual Moderation

Many spam detection mechanisms still rely heavily on manual moderation by content creators or platform administrators. This approach is time-consuming, inconsistent, and impractical for popular channels that receive a large volume of comments. Delayed moderation allows spam content to remain visible and influence users.

3. Limited Adaptability to New Spam Patterns

Spammers continuously evolve their techniques by using obfuscated text, emojis, special characters, shortened URLs, and multilingual content. Existing systems based on static rules or outdated datasets struggle to adapt to these changing patterns, making them ineffective against new and sophisticated spam strategies.

4. Insufficient Training Data

Current machine learning models are often trained on small, imbalanced, or outdated datasets. This results in poor generalization and biased predictions. Lack of diverse data, including multilingual and context-aware comments, further limits system performance.

5. Absence of Contextual Understanding

Most existing systems analyze comments in isolation without considering context such as user behavior, comment history, video content, or temporal patterns. This shallow analysis prevents accurate identification of promotional spam, scam links, and coordinated spam attacks.

6. Scalability Issues

Existing systems face technical limitations when processing large volumes of real-time comments. Inefficient algorithms and unoptimized architectures lead to slow response times, making real-time spam filtering difficult, especially during live streams or viral video events.

7. Privacy and Security Concerns

Some spam detection tools collect extensive user data without transparency or user consent. Weak data protection mechanisms expose sensitive information to security risks. Inadequate access controls and data retention policies further compromise user privacy.

8. Lack of Real-Time Detection

Many existing solutions operate in batch mode rather than real time. As a result, spam comments are detected only after being posted and viewed by users, reducing the effectiveness of moderation and increasing exposure to malicious content.

IV. PROPOSED SYSTEM

The proposed Spam Comments Detection System addresses the limitations of existing moderation approaches by providing an intelligent, scalable, and accurate machine learning–based solution. The system focuses on real-time detection, adaptability to evolving spam patterns, user safety, and efficient moderation. It enhances content quality and user trust on YouTube by automatically identifying and filtering spam comments with high precision.

System Overview

The proposed system is a full-stack machine learning–enabled web application built using modern technologies and robust design principles. It performs real-time spam classification on YouTube comments using supervised machine learning algorithms. The system supports automated comment analysis, contextual feature extraction, model training and evaluation, and administrative monitoring. The architecture emphasizes scalability, accuracy, security, and continuous learning while ensuring minimal latency and high reliability.

Core Features

1. Intelligent Comment Analysis

- Text preprocessing including tokenization, stop-word removal, stemming, and lemmatization
- Natural Language Processing (NLP) for understanding comment semantics
- Detection of suspicious patterns such as excessive links, emojis, and repeated text
- Support for multilingual comments
- Context-aware feature extraction based on comment content and metadata
- Handling of obfuscated and misspelled spam words

2. Machine Learning–Based Classification

- Supervised learning models such as Naïve Bayes, Support Vector Machine (SVM), and Random Forest
- Feature extraction using TF-IDF and Bag-of-Words models
- Binary classification of comments as spam or non-spam
- Model performance evaluation using accuracy, precision, recall, and F1-score
- Continuous model retraining using updated datasets
- Reduction of false positives and false negatives

3. Real-Time Spam Detection

- Instant classification of comments upon submission
- Automatic blocking or flagging of spam comments
- Low-latency prediction suitable for live streams and high-traffic videos
- Stream-based comment processing for scalability
- Fail-safe mechanisms for system availability

4. User and Moderator Management

- Secure administrator authentication and role-based access control
- Dashboard for monitoring detected spam statistics
- Manual review and override options for flagged comments
- User activity tracking to identify repeat spammers
- Blacklisting of malicious users and URLs

5. Dataset Management and Training Module

- Upload and management of labeled comment datasets
- Data balancing and cleaning tools
- Visualization of dataset distribution
- Model training, validation, and testing modules
- Version control for trained models
- Support for incremental learning

6. Alert and Reporting System

- Automated alerts for spam surges
- Daily and weekly spam detection reports
- Visualization of spam trends over time
- Exportable reports for analysis
- Logging of system decisions for transparency

7. Security and Privacy Features

- Secure storage of user and comment data
- Encrypted communication channels
- Compliance with data protection standards
- Limited data retention policies
- Protection against model poisoning attacks

Technical Architecture:

The technical architecture of the Spam Comments Detection System is designed using a layered client–server model to ensure scalability, real-time processing, and efficient machine learning operations. The system consists of multiple interconnected layers, including the presentation layer, application layer, machine learning layer, data processing layer, and data storage layer. This modular architecture enables seamless integration of components while ensuring high accuracy and performance in detecting spam comments on YouTube.

The presentation layer provides an interactive web-based interface through which administrators and moderators can securely log in, upload datasets, monitor system performance, and review detected spam comments. This layer is responsible for displaying classification results, analytics dashboards, and spam trends in a user-friendly manner.

The application layer acts as the core of the system, handling business logic, user authentication, request validation, and communication between system components. It exposes RESTful APIs that receive comment data, forward it for analysis, and return classification results. This layer also manages access control, logging, and system-level coordination.

The data processing layer performs text preprocessing and feature extraction on incoming comments. This includes tokenization, stop-word removal, stemming or lemmatization, normalization, and vectorization using techniques such as Bag-of-Words or TF-IDF. Processed data is prepared in a format suitable for machine learning model consumption.

The machine learning layer is responsible for training, evaluating, and deploying spam detection models. Supervised learning algorithms such as Naïve Bayes, Support Vector Machine (SVM), and Random Forest are used to classify comments as spam or non-spam. This layer supports model retraining with updated datasets to adapt to evolving spam patterns and improve prediction accuracy over time.

The data storage layer manages structured and unstructured data, including user credentials, labeled datasets, extracted features, trained model files, and classification results. Secure database mechanisms ensure data integrity, confidentiality, and efficient retrieval. Logging and audit data are also stored to support monitoring and debugging.

Additionally, a real-time processing and alert module enables instant spam detection as comments are submitted. Detected spam comments are either blocked or flagged for review, while system-generated reports and alerts provide insights into spam activity trends. This architecture ensures effective moderation, reduced manual effort, and enhanced content quality for users interacting on YouTube.

V. METHODOLOGY

The methodology of the Spam Comments Detection System focuses on automated text analysis, machine learning–based classification, and real-time moderation to identify and filter spam comments effectively. The process begins with data input, where comments are collected either from publicly available datasets or directly through the YouTube API. These comments serve as the primary input for preprocessing, feature extraction, and classification.

Initially, the system preprocesses raw comment text to remove noise and improve data quality. This includes converting text to lowercase, removing stop words, punctuation, URLs, emojis, and special characters, and applying stemming or lemmatization. Relevant metadata such as comment length, number of links, frequency of repeated words, and user behavior indicators may also be extracted. The cleaned and structured data is then stored in the database for further processing and model training.

Once preprocessing is completed, the feature extraction module transforms textual data into numerical representations using techniques such as Bag-of-Words or Term Frequency–Inverse Document Frequency (TF-IDF). These features are fed into supervised machine learning models trained to classify comments as spam or non-spam. The trained model evaluates incoming comments in real time and predicts their class. If a comment is identified as spam, it is either automatically blocked or flagged for moderator review. This continuous detection and classification workflow ensures efficient moderation, reduced exposure to spam, and improved user experience.

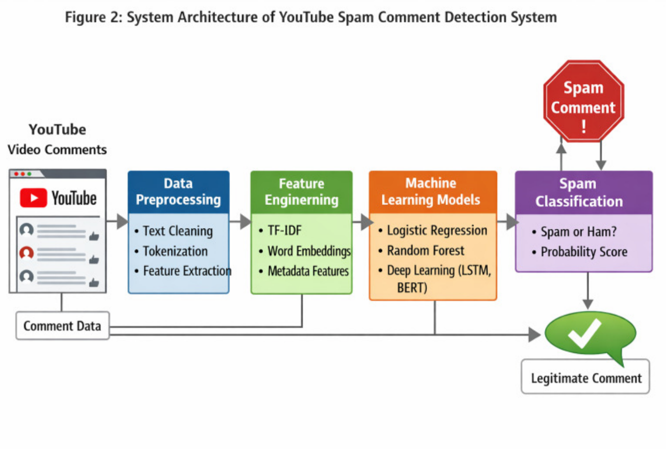


Figure 2: System Architecture Diagram

1.User Module

The user module represents the interaction layer between users, moderators, and the spam detection system. Administrators or moderators can log in securely to monitor detected spam comments, review classification results, and override system decisions if required. This module also allows dataset uploads, visualization of spam statistics, and access to system reports through an intuitive interface.

2.System Module

The system module acts as the central controller of the application. It manages communication between the user interface, machine learning components, and the database. This module handles request validation, routing of comment data to preprocessing and classification modules, scheduling of model retraining, and storage of prediction results. It ensures smooth coordination among all components of the system.

3. Comment Preprocessing and Feature Extraction Module

This module is responsible for preparing raw comment data for machine learning analysis. It performs text cleaning, normalization, tokenization, stop-word removal, and stemming or lemmatization. Feature extraction techniques such as TF-IDF convert processed text into numerical vectors that capture the importance of words and patterns commonly associated with spam behavior.

4.Spam Classification and Analysis Module

The Spam Classification Module uses trained machine learning algorithms such as Naïve Bayes, Support Vector Machine (SVM), or Random Forest to analyze extracted features and classify comments. It compares incoming comment features with learned spam patterns to determine whether a comment is spam or legitimate. Based on classification results, the system triggers appropriate actions such as blocking, flagging, or logging the comment. This module continuously improves performance through retraining with updated datasets.

Spam Comment Detection and Analysis Algorithm

The Spam Comment Detection and Analysis Algorithm is a core component of the Spam Comments Detection System, designed to automatically identify, analyze, and classify spam comments on YouTube. The algorithm is based on Natural Language Processing (NLP), feature extraction, and supervised machine learning techniques. It accurately detects spam-related patterns such as promotional content, malicious links, repeated text, suspicious keywords, and abnormal comment structures from user-generated comments.

The algorithm plays a critical role in the spam detection pipeline by continuously processing incoming comments and comparing them with previously learned spam patterns. Initially, raw comments are preprocessed to remove noise and normalize text. Key features such as word frequency, presence of URLs, comment length, and term importance are extracted and transformed into numerical representations using techniques like TF-IDF or Bag-of-Words. These features serve as input to trained machine learning classifiers.

By analyzing extracted features and model predictions, the system determines whether a comment is spam or legitimate. Classified results are stored with timestamps to support performance evaluation, trend analysis, and system auditing. When a comment is identified as spam, the system automatically flags or blocks it and logs the action for moderator review. This algorithm ensures accurate spam detection, adaptability to evolving spam strategies, and real-time moderation, thereby improving content quality and user experience on YouTube.

V. RESULT

The experimental results obtained from the machine-learning-based spam detection system for YouTube comments constitute the central analytical contribution of this research, capturing the detailed behaviour, performance, strengths, and limitations of the models that were trained, tested, and evaluated through extensive experimentation.

Support Vector Machines achieved one of the highest accuracies among classical models due to their effectiveness with high-dimensional sparse text data.
The dataset was divided into training, validation, and testing portions using stratified sampling to preserve the ratio of spam to non-spam categories.

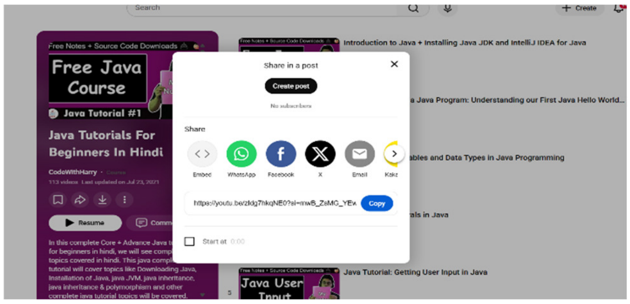


Figure: YouTube link

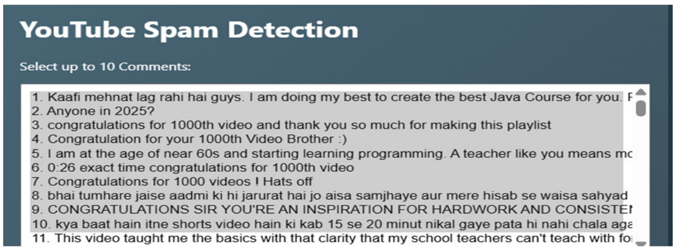
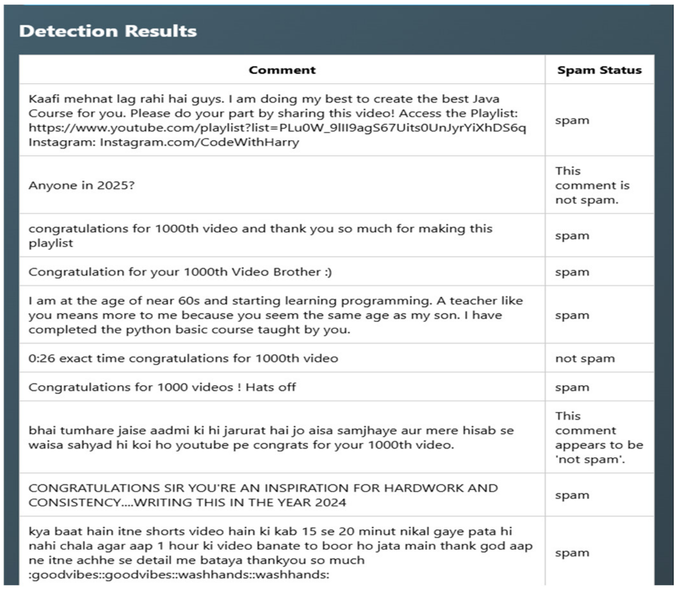


Figure: Comments Detection



Detection Results

Comment	Spam Status
Kaafi mehnat lag rahi hai guys. I am doing my best to create the best Java Course for you. Please do your part by sharing this video! Access the Playlist: https://www.youtube.com/playlist?list=PLu0W_9lII9agS67Uits0UnJyrYiXhDS6q Instagram: Instagram.com/CodeWithHarry	spam
Anyone in 2025?	This comment is not spam.
congratulations for 1000th video and thank you so much for making this playlist	spam
Congratulation for your 1000th Video Brother :)	spam
I am at the age of near 60s and starting learning programming. A teacher like you means more to me because you seem the same age as my son. I have completed the python basic course taught by you.	spam
0:26 exact time congratulations for 1000th video	not spam
Congratulations for 1000 videos ! Hats off	spam
bhai tumhare jaise aadmi ki hi jarurat hai jo aisa samjhaye aur mere hisab se waisa sahyad hi koi ho youtube pe congrats for your 1000th video.	This comment appears to be 'not spam'.
CONGRATULATIONS SIR YOU'RE AN INSPIRATION FOR HARDWORK AND CONSISTENCY....WRITING THIS IN THE YEAR 2024	spam
kya baat hain itne shorts video hain ki kab 15 se 20 minut nikal gaye pata hi nahi chala agar aap 1 hour ki video banate to boor ho jata main thank god aap ne itne achhe se detail me bataya thankyou so much :goodvibes::goodvibes::washhands::washhands:	spam

Figure: Final output

VI. CONCLUSION

This paper presents an effective machine learning–based approach for detecting spam comments on YouTube, addressing the growing challenge of spam in online social platforms. The proposed system automates the identification of spam comments by using supervised machine learning algorithms combined with efficient text preprocessing techniques such as cleaning, tokenization, stop-word removal, and stemming or lemmatization. Experimental results demonstrated that the implemented models achieved high accuracy in distinguishing spam comments from genuine ones, thereby reducing the need for manual moderation. Overall, the study confirms that machine learning techniques provide a reliable, scalable, and accurate solution for YouTube spam comment detection and can be further enhanced in the future by incorporating real-time processing, multilingual support, and advanced deep learning models.

VII. FUTURE DIRECTIONS

The future directions of the YouTube Spam Comment Detection project can be further broadened to make the system more intelligent, robust, and industry-ready. Advanced deep learning and natural language processing models, including transformer-based architectures such as BERT and RoBERTa, can be explored to improve contextual understanding and handle subtle spam tactics.

The system can be extended to perform real-time comment moderation using streaming data, allowing instant identification and filtering of spam. Support for multilingual and code-mixed languages can be added to address the diverse nature of YouTube comments across regions. Continuous learning mechanisms can be implemented so the model automatically adapts to new spam trends without manual retraining. Future work may also include combining text-based analysis with metadata and user behavior features such as account age, comment frequency, and link-sharing patterns to enhance prediction accuracy.

Additionally, sentiment analysis and intent detection can be integrated to distinguish between genuine criticism and malicious spam. The project can be deployed as a scalable cloud-based service, browser extension, or creator dashboard tool to assist content moderators. Finally, integrating the system with YouTube APIs for large-scale evaluation, feedback collection, and automated reporting can significantly improve real-world performance, reliability, and practical adoption.

VIII. ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to **Mr Basavaraj Nayak,** and Professor, Department of Computer Science and Engineering, R.L.Jalappa Institute of Technology, Doddaballapur, for his valuable guidance, encouragement, and continuous support throughout the course of this work. His insights and constructive suggestions greatly contributed to the successful completion of this research.

The authors also extend their appreciation to the faculty members of the Department of Computer Science and Engineering for their technical guidance and support. Finally, the authors thank the institution for providing the necessary facilities and resources to carry out this work effectively.

IX. REFERENCES

[1] Sah, U. K., & Parmar, N. (2017). An approach for Malicious Spam Detection in Email with comparison of different classifiers.

[2]Alberto, T. C., Lochter, J. V., & Almeida, T. A. (2015, December). Tubespam: Comment spam filtering on youtube. In Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on (pp. 138-143). IEEE.

[3] Alsaleh, M., Alarifi, A., Al-Quayed, F., & Al-Salman, A. (2016). Combating comment spam with machine learning approaches. Proceedings - 2015 IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015, 295–300.

[4] Scheltus, P., Dorner, V., & Lehner, F. (2013). Leave a Comment! An In-Depth Analysis of User Comments on YouTube. Wirtschaftsinformatik, 42.

[5] A. Kantchelian, J. Ma, L. Huang, S. Afroz, A. Joseph, J. D. Tygar, Robust detection of comment

spam using entropy rate, in: Proceedings of the 5th ACM Workshop on Security and Artificial Intelligence,

[6] S. Aiyar and N. P. Shetty, "N-gram assisted Youtube spam comment detection", Proc. Comput. Sci., vol. 132, pp. 174-182, Jan. 2018.

[7] A. Kantchelian, J. Ma, L. Huang, S. Afroz, A. Joseph and J. D. Tygar, "Robust detection of comment spam using entropy rate", Proc. 5th ACM Workshop Secur. Artif. Intell. (AISec), pp. 59-70, 2012.

[8] A. Madden, I. Ruthven and D. Mcmenemy, "A classification scheme for content analyses of Youtube video comments", J. Documentation, vol. 69, no. 5, pp. 693-714, Sep. 2013.

[9] A. Severyn, A. Moschitti, O. Uryupina, B. Plank and K. Filippova, "Opinion mining on Youtube", Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics (Long Papers), vol. 1, pp. 1-10, 2014.

[10] M. Z. Asghar, S. Ahmad, A. Marwat and F. M. Kundi, "Sentiment analysis on Youtube: A brief survey", arXiv:1511.09142, 2015, [online] Available.

[11] T. C. Alberto, J. V. Lochter and T. A. Almeida, "TubeSpam: Comment spam filtering on Youtube", Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA), pp. 138-143, Dec. 2015.

[12] A. U. R. Khan, M. Khan and M. B. Khan, "Naïve multi-label classification of Youtube comments using comparative opinion mining", Proc. Comput. Sci., vol. 82, pp. 57-64, Jan. 2016.

[13] G. Kaur, A. Kaushik, and S. Sharma, ‘‘Cooking is creating emotion: A study on hinglish sentiments of Youtube cookery channels using semisupervised approach,’’ Big Data Cognit. Comput., vol. 3, no. 3, p. 37, Jul. 2019, doi: 10.3390/bdcc3030037.

[14] E. Ezpeleta, M. Iturbe, I. Garitano, I. V. de Mendizabal, and U. Zurutuza, ‘‘A mood analysis on Youtube comments and a methodfor improved social spam detection,’’ in Proc. HAIS, 2018, pp. 514–525, doi: 10.1007/978-3- 319-92639-1_43.

[15] N. Hussain, H. Turab Mirza, G. Rasool, I. Hussain, and M. Kaleem, ‘‘Predilection decoded: Spam review detection techniques: A systematic literature review,’’ Appl. Sci., vol. 9, no. 5, p. 987, Mar. 2019, doi: 10.3390/app9050987

