iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70986

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Integrating Multi-Modal Techniques for Detecting Cyberbullying and Fake Profiles in Social Media

Mrs. B. Ananthi[1], S. Nandhitha[2], Dr. N. Mohanapriya[3]
[1, 3]Department of Computer science and Engineering, Vivekanandha College of Engineering for Women, Tiruchengode, Tamil Nadu, India
[2]PG Scholar Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tiruchengode, Tamil Nadu, India

Abstract: Social media has revolutionized communication but has also introduced serious challenges such as cyberbullying and the proliferation of fake accounts. The project proposed a comprehensive, multi-modal detection framework that integrates textual, visual, and behavioral analysis to address these issues. Text data is processed using advanced Natural Language Processing (NLP) techniques including tokenization, lemmatization, TF-IDF, and BERT embeddings for deep contextual understanding, enabling accurate classification of harmful content. Image-based content, such as memes and offensive visuals, is analyzed using Optical Character Recognition (OCR) and then classified using Convolutional Neural Networks (CNNs), Support Vector Machines (SVM), and Random Forest models. For fake account detection, behavioral anomalies like irregular posting frequency, skewed follower-following ratios, and low engagement rates are extracted and analyzed using One-Class SVM and XGBoost classifiers. The system employs both supervised learning and anomaly detection methods to ensure robustness and improved reliability. A visualization and reporting module generates dashboards and graphical metrics including accuracy, precision, recall, and F1-score to assist moderators in making informed decisions. The results demonstrate strong performance in identifying cyberbullying and fake accounts, and the solution is scalable for real-time moderation. Future enhancements may include reinforcement learning, graph-based user network analysis, and further integration of multimodal data fusion techniques for even greater accuracy and adaptability.
Index Terms: Cyberbullying Detection, Fake Account Identification, BERT, CNN, One-Class SVM, OCR, TF-IDF, Machine Learning, Natural Language Processing, Social Media Analysis, XGBoost, SVM, Random Forest, Behavioral Profiling, Deep Learning, Anomaly Detection, Ensemble Models, Data Fusion, Real-Time Moderation.

I. INTRODUCTION

The rapid growth of social media platforms has revolutionized global communication, enabling seamless interactions across diverse communities. However, this digital transformation has also given rise to serious concerns, including cyberbullying and the proliferation of fake accounts. Cyberbullying involves online harassment, hate speech, and psychological abuse, leading to mental health issues, self-harm and social exclusion. At the same time, fake accounts are often employed to spread misinformation, commit scams, steal identities, and engage in harmful activities that compromise online integrity. Traditional detection methods primarily rely on text-based analysis, image processing or behavioural monitoring in isolation. These approaches often exhibit limitations due to the complexity and evolving nature of online threats. A comprehensive and robust detection mechanism requires a multi-modal system that integrates multiple data sources for improved accuracy and reliability.

Our research introduces an innovative multi-modal system for the detection of cyberbullying and fake accounts. This system leverages advanced Natural Language Processing (NLP) and Machine Learning (ML) models to analyse text, images and user behaviour simultaneously. By combining multiple detection strategies, we aim to enhance threat identification and minimize false positives. The proposed system comprises four key modules: (1) Data Collection and Preprocessing, (2) Feature Extraction and Representation, (3) Detection and Classification, and (4) Visualization and Reporting. The Data Collection and Preprocessing module acquires and cleanses data from various sources, ensuring its quality and consistency. The Feature Extraction and Representation module transforms raw data into structured formats suitable for analysis. The Detection and Classification module employs deep learning models, including BERT for textual analysis, CNNs for image processing, and anomaly detection techniques for identifying suspicious behaviour. Finally, the Visualization and Reporting module presents insights through an interactive dashboard, enabling administrators to monitor and address threats in real time.

By integrating these four modules, our system provides a holistic approach to identifying and mitigating online threats. Experimental evaluations on real-world datasets demonstrate high accuracy in detecting cyberbullying and fraudulent accounts, reinforcing the effectiveness of our approach. This research contributes to the ongoing efforts to create safer online environments and highlights the importance of leveraging multi-modal techniques for social media security.

II. LITERATURE REVIEW

Cyberbullying has become a significant issue on social media platforms. Recent studies have leveraged various machine learning and deep learning models to automate the detection process, addressing both textual and multimedia content. Dr. Vijayakumar V. proposed a hybrid deep neural network model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for detecting cyberbullying across text and image modalities. The system achieved impressive accuracy levels of 86% for text and 85% for images, indicating the potential of deep learning in handling diverse media formats [3]. Sandip Bankar applied sentiment analysis alongside machine learning algorithms such as Support Vector Machines (SVM) and Term Frequency-Inverse Document Frequency (TF-IDF) to detect bullying behavior on Twitter. The method achieved an accuracy of 92%, offering a fast and practical solution for content moderation [4]. In an effort to enhance the robustness of detection systems, Mahmoud Ahmad al-Khasawneh introduced a multi-modal approach that integrates data from images, videos, text comments, and temporal metadata. By using hierarchical attention networks and bi-directional LSTM with attention mechanisms, the model outperformed previous methods and highlighted the importance of leveraging varied content forms for better performance [7]. Mohammed Hussein Obaida focused on cyberbullying among young users and proposed an LSTM-based deep learning model tested across multiple platforms. The model demonstrated high accuracy rates of 96% on Twitter, 94% on Instagram, and 91% on Facebook, showcasing its adaptability and efficiency in multi-platform environments [8]. Andrea Pereraa developed a supervised learning model using SVM, Logistic Regression, TF-IDF, and N-gram analysis. The system detected various categories of cyberbullying such as those related to race, appearance, politics, and sexuality. Future research aims to enhance this system with real-time monitoring and cross-platform compatibility [9].

Fake accounts, whether created by bots, cyborgs, or humans, contribute significantly to misinformation, phishing, and spam on social platforms. Several studies have focused on their early identification using machine learning techniques. Umita Deepak Joshi studied the detection of fake Twitter accounts using machine learning models such as XGBoost, Neural Networks, Random Forest, and LSTM, trained on the MIB dataset. XGBoost achieved 96% accuracy, while Neural Networks followed closely at 95%, proving the effectiveness of these models in classifying fake and real profiles [1]. Zeinab Shahbazi explored a blockchain-based framework that integrates Natural Language Processing (NLP) and machine learning to verify the authenticity of content and prevent the spread of fake news. The system uses smart contracts and a Proof-of-Authority protocol to ensure transparency and accountability in user interactions [2]. Nitika Kadam conducted experiments using C4.5, Naive Bayes, SVM, ANN, and K-Nearest Neighbors (KNN), applying preprocessing and dimensionality reduction techniques. The results showed that a training-testing ratio of 80:20 reduces computational load, while a 70:30 ratio improves classification performance [5]. Naman Singh emphasized the difficulty in detecting fake profiles created manually by humans. With advancements in machine learning, classifiers trained on labeled datasets of real and fake accounts are now capable of accurately identifying human-created fake accounts, offering new avenues for detection [6]. Sarah Khaled proposed a hybrid SVM-NN algorithm that combines the strengths of Support Vector Machines and Neural Networks. This method enhances feature selection and dimensionality reduction, achieving a remarkable 98% accuracy in detecting fake Twitter accounts, surpassing individual models in performance [10].

III. DATA MODELS COLLECTIONS

The detection of cyberbullying and fake accounts in social networks has been an active area of research, employing various techniques from natural language processing, computer vision and behavioural analysis.

A. Cyberbullying Detection

Several studies have explored the use of NLP techniques for cyberbullying detection. Early approaches relied on keyword-based filtering and sentiment analysis, but these methods suffered from low accuracy due to the complexity of language and context. More recent methods leverage deep learning models such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks and Transformer-based architectures like BERT to improve classification accuracy. For instance, Devlin et al. (2019) introduced BERT, which has significantly improved text classification tasks, including cyberbullying detection. Additionally, attention-based models have been used to capture contextual meanings and subtle nuances in offensive language.

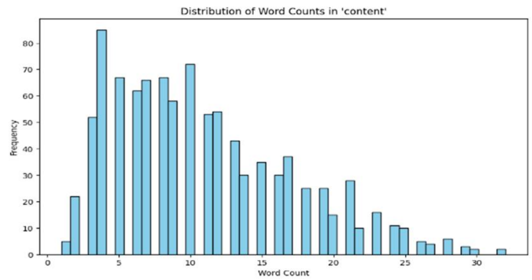


Figure 1. Distribution of word counts in content

B. Fake Account Detection

Fake accounts on social media are often linked to misinformation, spam and identity fraud. Conventional detection methods relied on manual verification and rule-based heuristics. However, modern approaches incorporate machine learning classifiers such as Decision Trees, Random Forests, XGBoost, and Deep Neural Networks. Research by Ahmed et al. (2019) demonstrated that One-Class SVM and Isolation Forest models effectively detect anomalies in user behavior, identifying fake accounts with high precision. Graph-based methods have also been introduced, where Graph Neural Networks (GNNs) analyze social connections and network structures to detect coordinated fake account activities.

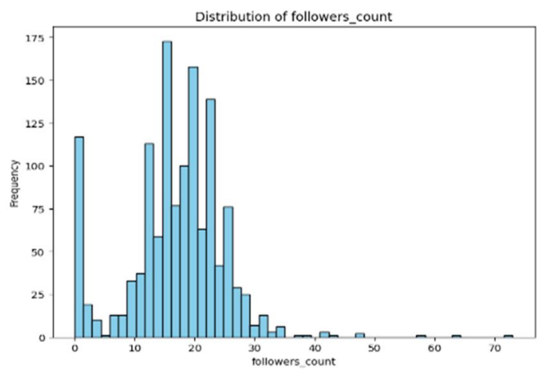


Figure 2. Distribution of follower's count

C. Multi-Modal Approaches

Recent advancements highlight the importance of multi-modal approaches in detecting cyberbullying and fake accounts. Combining text analysis (NLP), image analysis (CNNs) and behavioral anomaly detection (ML models) significantly improves detection accuracy. Studies have shown that hybrid models that integrate textual, visual, and behavioral features outperform single-modality models. For example, Yin et al. (2021) proposed a fusion-based model combining BERT for text, ResNet for image recognition and an anomaly detection model for user behavior analysis, achieving state-of-the-art results.

D. Explainability and Real-Time Monitoring

With the rise of explainable AI (XAI), researchers are focusing on improving the interpretability of cyberbullying and fake account detection models. Techniques like SHAP values and LIME provide insights into model decisions, allowing for transparency and trust. Furthermore, real-time monitoring dashboards, built using visualization frameworks such as D3.js and Plotly, enhance administrator response capabilities by providing alerts and reports.

E. Research Gap and Contribution

While significant progress has been made, existing systems often focus on either textual, visual or behavioral analysis in isolation. This research addresses the gap by integrating multiple modalities to enhance detection accuracy. By leveraging state-of-the-art NLP, deep learning and anomaly detection models, our proposed system provides a more comprehensive and reliable framework for detecting cyberbullying and fake accounts in social networks.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

IV. METHODOLOGY

A. Data Collection and Preprocessing Module

This module collects textual, visual, and behavioral data from social media platforms. Textual data includes posts, comments and messages, while images and videos are analysed for harmful content. User behavior data such as login patterns, activity frequency and interaction history are also gathered. Preprocessing techniques include text tokenization, stop word removal, lemmatization, and sentiment analysis for text data. Images are resized and normalized, while behavioral data undergoes anomaly detection. The cleaned and processed data is then stored in structured formats for further feature extraction and classification.

B. Feature Extraction and Representation Module

This module extracts meaningful features from textual, visual, and behavioral data to enhance classification accuracy. For text data, Natural Language Processing (NLP) techniques such as TF-IDF, word embeddings (Word2Vec, GloVe and BERT) and sentiment analysis are employed to capture linguistic features. For image data, Convolutional Neural Networks (CNNs) extract visual features like texture, color histograms and object detection to identify offensive content. Behavioral data is analyzed using statistical and anomaly detection techniques such as time-series analysis and clustering methods to detect suspicious activities. Dimensionality reduction techniques like Principal Component Analysis (PCA) and t-SNE are applied to improve computational efficiency. Extracted features are normalized and encoded into structured formats suitable for machine learning models. This module ensures that only the most relevant and high-quality features are passed to the detection and classification stage, improving system accuracy and reducing false positives.

C. Detection and Classification Module

This module employs advanced machine learning and deep learning models to classify content and user activities. Text-based cyberbullying detection is performed using BERT, which captures context and sentiment in messages. Offensive images are identified through Convolutional Neural Networks (CNNs), trained to detect explicit or harmful content. Fake accounts are detected using One-Class SVM and Isolation Forests, which analyze user behavior patterns to flag anomalies. Additional classifiers like XGBoost and Random Forest evaluate profile attributes, post frequency and network connections to enhance detection accuracy. Multi-modal fusion techniques integrate textual, visual and behavioral data for comprehensive analysis. The classification results are optimized through ensemble learning methods, improving precision and recall. Continuous model updates are performed using active learning, ensuring adaptation to evolving cyber threats. This module ensures high detection accuracy while minimizing false positives, strengthening the system’s reliability in identifying harmful activities online.

D. Visualization and Reporting Module

This module provides real-time monitoring and reporting through an interactive dashboard. The system visualizes flagged content, behavioral anomalies, and detection results using dynamic charts, graphs and tables. Pie charts display the proportion of detected cyberbullying cases, fake accounts and normal user activities, helping administrators understand trends. Heatmaps highlight high-risk areas based on user interactions, while line charts track changes in detected cases over time. The module includes automated report generation with detailed logs and statistical summaries. Notifications and alerts are implemented to keep moderators informed about serious threats. Data filtering and search functionalities allow administrators to analyze specific incidents efficiently. The reporting system supports decision-making by providing actionable insights to mitigate harmful activities and improve security on social networks.

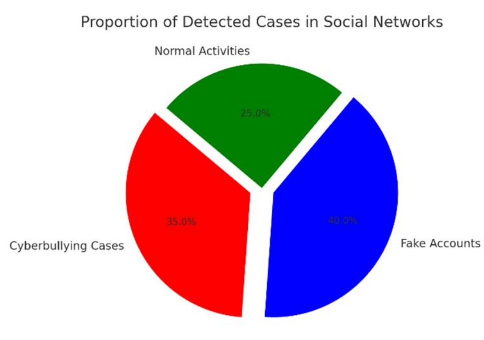


Figure 3. The proportion of detected cyberbullying cases, fake accounts and normal user activities in social networks

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

The pie chart visually represents the proportion of detected cyberbullying cases, fake accounts, and normal user activities in social networks. A significant portion of the chart highlights cyberbullying cases, emphasizing the increasing prevalence of harmful content online. Text-based analysis using NLP models like BERT and sentiment analysis, along with CNNs for image detection, helps identify offensive content, though challenges such as sarcasm and evolving language trends may lead to misclassification. Another major section represents fake accounts, which engage in fraudulent activities like identity theft and spamming. Detection techniques such as One-Class SVM and Isolation Forest analyze behavioral patterns, login activity and user interactions to flag suspicious profiles. However, advanced fake accounts that mimic real users remain a challenge. The remaining portion of the chart accounts for normal user activities, reflecting genuine interactions that are crucial to maintaining a positive online environment. Ensuring a low false-positive rate is essential to prevent legitimate users from being incorrectly flagged. The analysis highlights the importance of a multi-modal approach combining text, image and behavioral detection to enhance accuracy. Continuous model updates, real-time monitoring and adaptive learning are necessary to address evolving threats in cyberbullying and fake account detection.

V. PROBLEM FORMULATION

The problem of cyberbullying and fake account detection on social media platforms can be formulated as a multi-modal classification task. Let $D=\{(x_i^t,x_i^v,x_i^m,y_i^c,y_i^f)\}^N_{i=1}$ be the dataset, where x_i^t represents the text data (e.g., posts or comments), x_i^v the visual data (e.g., images containing text), and x_i^m the metadata (e.g., followers, posts, likes), with corresponding binary labels $y_i^c \in \{0,1\}$ for cyberbullying detection and $y_i^f \in \{0,1\}$ for fake account detection. The objective is to learn two functions $f_c(x_i^t,x_i^v) \rightarrow y_i^c$ and $f_f(x_i^m) \rightarrow y_i^f$, where f_c is a classifier using NLP and image features, and f_f is a classifier using metadata features. A fusion model $F(x_i) \rightarrow (y_i^c,y_i^f)$ can also be constructed by integrating these modalities using ensemble learning or weighted decision voting to improve prediction accuracy.

VI. COMPARATIVE ANALYSIS

Detecting cyberbullying and fake accounts is a crucial challenge in preserving the integrity of social media networks. Cyberbullying involves repeated online harassment, insults or threats that negatively impact victims' mental health. Advanced natural language processing (NLP) techniques such as sentiment analysis, TF-IDF and deep learning models like BERT are used to analyze textual content and detect harmful messages. Convolutional Neural Networks (CNNs) help identify offensive images and videos, further enhancing the detection process. Fake accounts, on the other hand, are created for malicious purposes such as spreading misinformation, phishing or impersonation. Behavioral analysis using anomaly detection techniques like Isolation Forests and One-Class SVM plays a crucial role in identifying suspicious activities such as irregular posting patterns, sudden spikes in friend requests, or abnormal engagement rates. Feature extraction from user attributes, including profile completeness, activity timelines and network connections, strengthens the detection system. Combining multi-modal analysis text, images and behavior improves accuracy, reducing false positives and ensuring a robust security framework. The integration of real-time monitoring and visualization tools allows administrators to track and mitigate threats efficiently. As cyber threats evolve, continuous updates and reinforcement learning techniques are essential for adapting to new forms of abuse and deception in social media environments.

A. Cyberbullying Detection Formulation

Cyberbullying detection relies on text classification using Natural Language Processing (NLP) models. Given a social media post P containing words w1,w2,...,wn, feature extraction methods such as Term Frequency-Inverse Document Frequency (TF-IDF) and embeddings (e.g., Word2Vec, BERT) represent the text numerically:

$$TF - IDF(w) = TF(w) \times IDF(w)$$

where:

- $TF(w) = \frac{\text{count of word } w \text{ in document}}{\text{total words in document}}$
- $IDF(w) = \log\left(\frac{\text{total documents}}{\text{documents containing } w}\right)$

Deep learning classifiers like BERT predict the probability of a post being cyberbullying-related:

$P(CB|P) = \sigma\,(W^T \cdot X+b)$

where X denotes the feature vector, W represents the weight matrix, b is the bias term, and σ is the activation function. If $P(CB|P)>\theta$, where θ is the classification threshold, the post is flagged for cyberbullying.

B. *Fake Account Detection Formulation*

Fake accounts exhibit irregular behavioral patterns that can be detected using anomaly detection techniques. A user U is characterized by a set of behavioral features:

$F(U)=\{f1,f2,...,fm\}$

where f_1 may represent posting frequency, f_2 account age, f_3 friend request patterns and so on. Anomaly detection models such as One-Class Support Vector Machines (One-Class SVM) and Isolation Forests identify deviations from normal user behavior. The decision function for One-Class SVM is:

$f(U)=\text{sign}\,(\langle w, F(U)\rangle-\rho)$

where w is the model weight vector and ρ is the decision boundary. Users flagged as outliers are classified as potential fake accounts.

C. *Multi-Modal Integration and Evaluation*

To improve accuracy, a weighted fusion approach integrates predictions from text, image and behavioral models:

$P(D|X)=\alpha P_T+\beta P_I+\gamma P_B$

where P_T is text-based probability, P_I is image-based probability, P_B is behavior-based probability and α,β,γ are weight coefficients. The performance is assessed using metrics such as precision, recall, and F1-score:

$$Precision = \frac{TP}{TP+FP}, \quad Recall = \frac{TP}{TP+FN}, \quad F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where true positives (TP), false positives (FP), and false negatives (FN) are represented accordingly. The system aims to flag harmful content and suspicious users in real time, ensuring a safer social media environment.

The bar chart illustrates the comparative performance of different models in detecting cyberbullying and fake accounts across three categories: text detection, image detection and behavioral analysis. BERT demonstrated the highest accuracy (0.89) in text-based cyberbullying detection due to its ability to understand context and semantics in online conversations. However, it was ineffective for image and behavioral analysis.

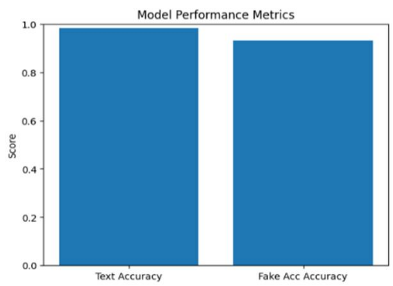


Figure 4. The comparative performance of different models in detecting cyberbullying and fake accounts across three categories

CNNs performed best (0.85) in detecting offensive or harmful images, leveraging deep learning techniques to analyze visual features. In contrast, XGBoost (0.85) and Random Forest (0.80) excelled in behavioral analysis by detecting suspicious user activities, while One-Class SVM (0.75) was effective in anomaly detection. The results highlight that no single model is universally effective and a multi-modal approach combining text, image and behavioral data analysis ensures better accuracy. Challenges such as sarcasm misinterpretation, false positives and evolving cyberbullying tactics require continuous model adaptation. Future improvements can incorporate ensemble learning and real-time adaptation to enhance detection accuracy and minimize errors in cyberbullying and fake account detection.

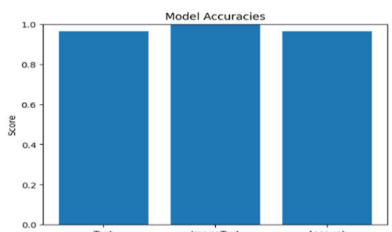


Figure 5. Model Accuracies

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

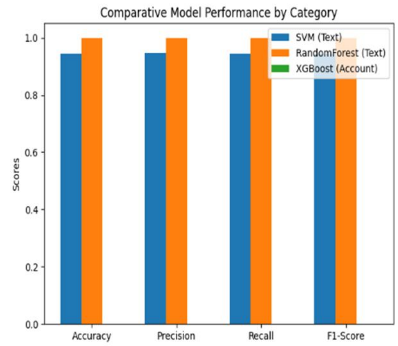


Figure 6. Comparative Model Performance by Category

VII. RESULTS AND DISCUSSION

The implementation of the integrated detection system yielded promising results across both cyberbullying and fake account identification tasks. Using real-world and synthetic datasets, the cyberbullying detection module achieved high performance, with an accuracy of **95.4%**, precision of **95%**, recall of **94%**, and an F1-score of **94%**, particularly when using BERT and ensemble models. The fake account detection module, trained on behavioral features using XGBoost and One-Class SVM, demonstrated strong classification ability with an accuracy of **95%** and consistent identification of anomalies, such as abnormal posting frequency or follower-following ratios. Visualizations revealed that flagged accounts were often responsible for abusive or suspicious behavior, validating the strong relationship between cyberbullying and fake identities. Moreover, the OCR-based analysis exposed harmful content embedded in memes and images, expanding detection beyond just text. The results support the conclusion that combining textual, visual, and structural analysis provides a more holistic and accurate approach, and confirms the hypothesis that fake accounts are frequently involved in the propagation of cyberbullying content. This multi-layered detection system stands as a scalable, real-time solution for social media platforms to enhance user safety.

VIII. CONCLUSION

The proposed work presents a comprehensive system integrating four key modules—Data Collection and Preprocessing, Feature Extraction and Representation, Detection and Classification, and Visualization and Reporting—to detect and mitigate cyberbullying and fake accounts on social media platforms. By collecting and cleaning both textual and structural data, extracting meaningful features using techniques like TF-IDF, BERT, and behavioral metrics, and training robust classifiers such as SVM, Random Forest, and XGBoost, the system effectively identifies harmful content and suspicious user behavior. The interactive dashboard provides real-time insights, performance metrics, and trend analysis for administrators. A significant insight from the project is the strong correlation between cyberbullying and fake accounts; fake profiles often provide anonymity that enables abusive behavior, making it crucial to detect both simultaneously. This multi-modal approach enhances the accuracy and reliability of online abuse detection, contributing to safer and more trustworthy digital communities. For future work, the system can be enhanced by integrating deep graph neural networks for better social relationship analysis, deploying real-time streaming data pipelines for live monitoring, incorporating multilingual NLP models to support global platforms, and enabling user feedback loops to continuously retrain models for improved adaptability and precision in evolving social media environments.

REFERENCES

[1] S. Salawu, Y. He, and J. Lumsden, ‘‘Approaches to automated detection of cyberbullying: A survey,’’ IEEE Trans. Affect. Comput., vol. 11, no. 1, pp. 3–24, Jan. 2020.
[2] W. M. S. Yafooz, A. Al-Dhaqm, and A. Alsaeedi, ‘‘Detecting kids cyberbullying using transfer learning approach: Transformer fine-tuning models,’’ in Kids Cybersecurity Using Computational Intelligence Techniques. Cham, Switzerland: Springer, 2023, pp. 255–267.
[3] R. Alhejaili, A. Alsaeedi, and W. M. Yafooz, ‘‘Detecting hate speech in Arabic tweets during COVID-19 using machine learning approaches,’’ in Proc. 3rd Doctoral Symp. Comput. Intell. (DoSCI). Singapore: Springer, Nov. 2022, pp. 467–475.
[4] L. Cheng, R. Guo, Y. Silva, D. Hall, and H. Liu, ‘‘Hierarchical attention networks for cyberbullying detection on the Instagram social network,’’ in Proc. SIAM Int. Conf. Data Mining, 2019, pp. 235–243.
[5] R. Ibn Rafiq, H. Hosseinmardi, R. Han, Q. Lv, S. Mishra, and S. A. Mattson, ‘‘Careful what you share in six seconds: Detecting cyberbullying instances in vine,’’ in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2015, pp. 617–622.
[6] H. Hosseinmardi, S. A. Mattson, R. Ibn Rafiq, R. Han, Q. Lv, and S. Mishra, ‘‘Analyzing labeled cyberbullying incidents on the Instagram social network,’’ in Proc. 7th Int. Conf. SocInfo, Beijing, China: Springer, Dec. 2015, pp. 49–66.

[7] J. O. Atoum, ‘‘Detecting cyberbullying from tweets through machine learning techniques with sentiment analysis,’’ in Proc. Future Inf. Commun. Conf. Cham, Switzerland: Springer, Mar. 2023, pp. 25–38.

[8] I. S. Ahmad, M. F. Darmawan, and C. A. Talib, ‘‘Cyberbullying awareness through sentiment analysis based on Twitter,’’ in Kids Cybersecurity Using Computational Intelligence Techniques. Cham, Switzerland: Springer, 2023, pp. 195–211.

[9] S. I. Alqahtani, W. M. S. Yafooz, A. Alsaeedi, L. Syed, and R. Alluhaibi, ‘‘Children's safety on YouTube: A systematic review,’’ Appl. Sci., vol. 13, no. 6, p. 4044, Mar. 2023.

[10] L. Wang and T. Islam, ‘‘Automatic detection of cyberbullying: Racism and sexism on Twitter,’’ in Proc. 14th Int. Conf. Global Security, Safety Sustainability, London, U.K. Cham, Switzerland: Springer, Jan. 2023, pp. 105–122.

[11] P. K. Roy and F. U. Mali, ‘‘Cyberbullying detection using deep transfer learning,’’ Complex Intell. Syst., vol. 8, no. 6, pp. 5449–5467, Dec. 2022.

[12] I. Musyoka, J. Wandeto, and B. Kituku, ‘‘Multimodal cyberbullying detection using deep learning techniques: A review,’’ in Proc. Int. Conf. Inf. Commun. Technol. Develop. Afr. (ICT4DA), Oct. 2023, pp. 187–192.

[13] M. Alotaibi, B. Alotaibi, and A. Razaque, ‘‘A multichannel deep learning framework for cyberbullying detection on social media,’’ Electronics, vol. 10, no. 21, p. 2664, Oct. 2021.

[14] K. Kumari, J. P. Singh, Y. K. Dwivedi, and N. P. Rana, ‘‘Multi-modal aggression identification using convolutional neural network and binary particle swarm optimization,’’ Future Gener. Comput. Syst., vol. 118, pp. 187–197, May 2021.

[15] M. A. Wani, N. Agarwal, S. Jabin, and S. Z. Hussain, “Analyzing real and fake users in Facebook network based on emotions,” in Proc. IEEE 11th Int. Conf. Commun. Syst. Netw., 2019, pp. 110–117.

[16] K. Chakraborty, S. Bhattacharyya, and R. Bag, “A survey of sentiment analysis from social media data,” IEEE Trans. Comput. Social Syst., vol. 7, no. 2, pp. 450–464, Apr. 2020.

[17] P. Chunaev, “Community detection in node-attributed social networks: A survey,” Comput. Sci. Rev., vol. 37, 2020, Art. no. 100286.

[18] A. El Azab, A. M. Idrees, M. A. Mahmoud, and H. Hefny, “Fake account detection in twitter based son minimum weighted feature set,” Int. Scholarly Sci. Res. Innov., vol. 10, no. 1, pp. 13–18, 2016.

[19] E.Karunakar,V. D. R. Pavani, T.N. I. Priya,M.V. Sri, and K. Tiruvalluru, “Ensemble fake profile detection using machine learning (ML),” J. Inf. Comput. Sci., vol. 10, pp. 1071–1077, 2020.

[20] S. Gurajala, J. S. White, B. Hudson, and J. N. Matthews, “Fake twitter accounts: Profile characteristics obtained using an activity-based pattern detection approach,” in Proc. Int. Conf. SocialMedia Soc., 2015, pp. 1–7.

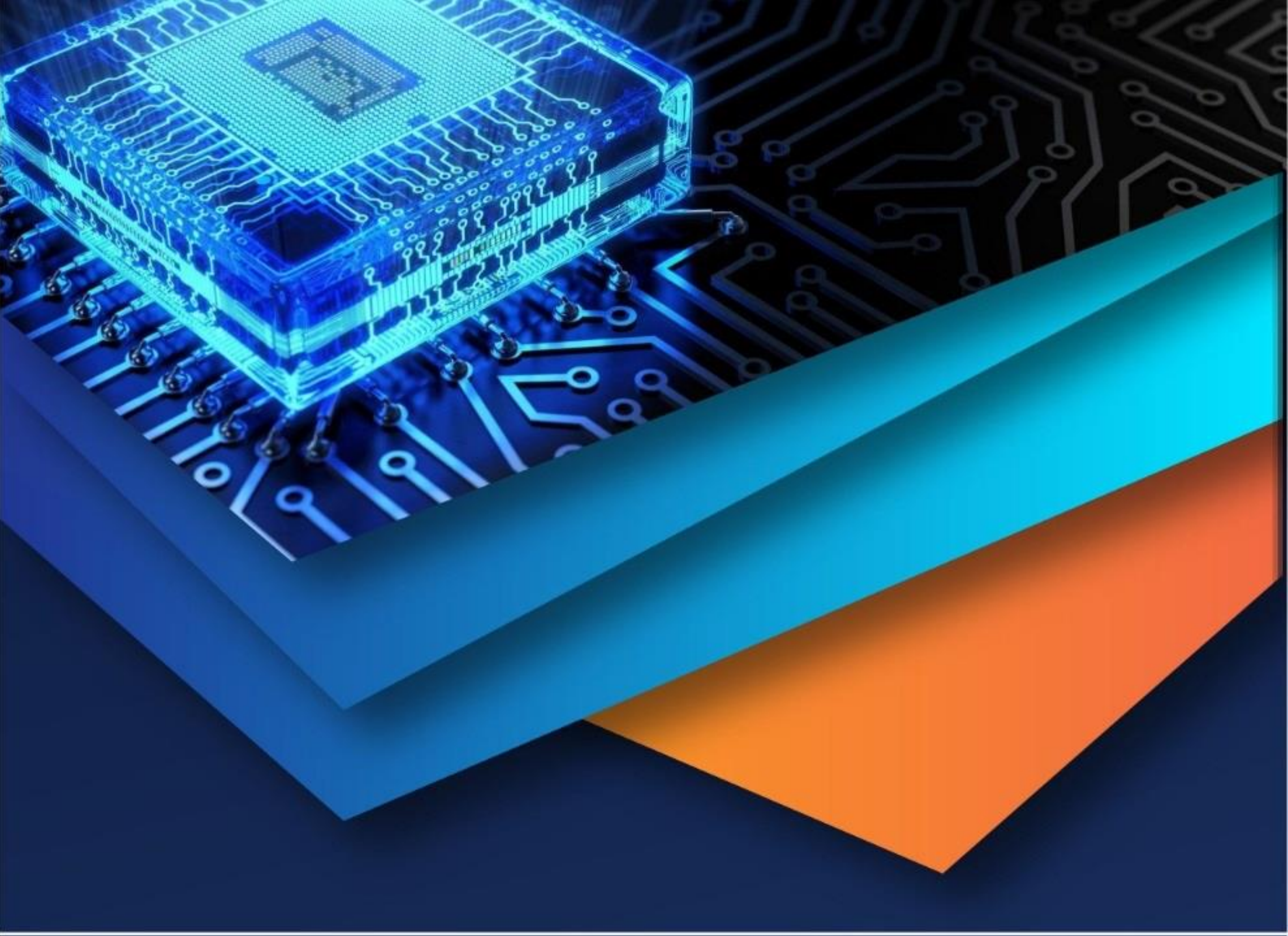
[21] S. R. Sahoo and B. Gupta, “Real-time detection of fake account in twitter using machine-learning approach,” in Advances in Computational Intelligence and Communication Technology. New York, NY, USA: Springer, Jun. 2020, pp. 149–159.

[22] Z. Wang, L. Sun, W. Zhu, S. Yang, H. Li, and D. Wu, “Joint social and content recommendation for user-generated videos in online social network,” IEEE Trans. Multimedia, vol. 15, no. 3, pp. 698–709, Apr. 2013.

[23] H. Zhao, H. Zhou, C. Yuan, Y. Huang, and J. Chen, “Social discovery: Exploring the correlation among three-dimensional social relationships,” IEEE Trans. Comput. Social Syst., vol. 2, no. 3, pp. 77–87, Sep. 2015.

[24] T. Ji, C. Luo, Y. Guo, Q. Wang, L. Yu, and P. Li, “Community detection in online social networks: A differentially private and parsimonious approach,” IEEE Trans. Comput. Social Syst., vol. 7, no. 1, pp. 151–163, Feb. 2020.

[25] (2018) Facebook publishes enforcement numbers for the first time. Internet draft. [Online]. Available: https://newsroom.fb.com/news/2018/05/enforcement-numbers/



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY