CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES

https://cajmtcs.casjournal.org/index.php/CAJMTCS

Volume: 07 Issue: 02 | April 2026 ISSN: 2660-5309

Article

Text Categorization Model for Detecting Cyberbullying Content on Twitter Using Support Vector Machine (SVM) And Naïve Bayes Algorithm

Jonathan Nyekachi Amadi, Osaki Miller Thom-Manuel

1. Department of Computer Science, Ignatius Ajuru University of Education, Port Harcourt, Rivers State, Nigeria
2. Department of Computer Science, Ignatius Ajuru University of Education, Port Harcourt, Rivers State, Nigeria

*Correspondence : Jonathan.amadi@iaue.edu.ng

Abstract: As social networks such as Twitter expand, so have the means – and frequency of cyberbullying – to a point where it may present material psychological and emotional risks to users. In this study, we design and implement a text categorization model to spot content of cyberbullying on Twitter using Support Vector Machine (SVM) and Naïve Bayes algorithms. Structured text preprocessing techniques such as tokenization, stopword removal, and TF-IDF feature extraction are proposed in the proposed system to convert tweets into feature vectors for the machine learning classification. The pipeline is executed by a computer program developed in Python that combines SVM and Naïve Bayes together to increase the performance of detection and a web-based dashboard enables users to visualize the tested and classified content in close to real-time. The results show that the hybrid model achieves 94.1% accuracy and 93.1% F1-score, which can outpace all single classifiers, and it also can detect subtle cases of the cyberbullying setting, for example, sarcasm and context-dependent harassment compared to the state-of-the-art systems. The results indicate that the proposed model offers a practical, reliable, and scalable solution for monitoring Twitter, providing an effective tool for mitigating harmful online behavior and enhancing safer social media interactions.

Keywords: Text Categorization Model, Cyberbullying Content, Twitter, Support Vector Machine (SVM), Naïve Bayes Algorithm

Citation: Amadi, J. N & Manuel, O. M. T. Text Categorization Model for Detecting Cyberbullying Content on Twitter Using Support Vector Machine (SVM) And Naïve Bayes Algorithm. Central Asian Journal of Mathematical Theory and Computer Sciences 2026, 7(2), 210-223

Received: 10th Dec 2025
Revised: 21th Jan 2026
Accepted: 04th Feb 2026
Published: 26th Mar 2026

1. Introduction

The rapid expansion of social media platforms has significantly transformed communication and information exchange across the world. Platforms like Instagram, Facebook, and Twitter allow the user to communicate, share ideas, and distribute information within seconds. These platforms took off, leaving billions of users around the world and Social media became an integral part of modern relationship. But while the advantages of connectedness and information exchange are being realized, with the explosive increase of social media there became large number of online social problems of which cyberbullying is one of the most serious. Cyberbullying is online bullying through the use of digital technologies including online platforms that physically harass, threaten, or intimidate people. Having more access to social media, and students getting mobile internet devices created a universal issue facing cyberbully in users of all ages, but more commonly amongst adolescents and young adults. Research suggests that the anonymity and ease of communicating on the internet makes cyberbullying both more frequent and severe. Cyberbullying can involve mean-spirited comments, hateful speech,

threats, rumors, or other socially humiliating images and videos shared online and can impair victims' psychological and emotional health.

This type of inappropriate social behaviour is especially apparent on microblogging applications like Twitter, where users post short texts called tweets, which gain popularity in the network by likes, replies and re-tweets. These exchanges can go viral, spreading harmful messages to vast audiences in a short time and likely increasing the impact of cyber bullying events. Indeed, Twitter is one of the most important social platforms of which researchers have recognized the culture of it to be a very fast environment in allowing inappropriate content to propagate [1]. The adverse effects of cyberbullying are substantial and might results in the form of psychological stress, depression, anxiety, social isolation, and suicidal ideation in some extreme cases on the part of victims. Several recent papers highlight the pressing demand for automated systems for the cyberbullying detection of social media content to mitigate its adverse effects [2]. The scale of data generated by millions of users every day on social media platforms makes manual labor to filter harmful content impossible. Consequently, there has been a growing interest among researchers to investigate methods using machine learning and natural language processing approaches to facilitate the detection of abusive language and other harmful behaviour inherent in online communications.

Analysis of textual patterns and detection of cyberbullying in social media messages using machine learning algorithms have shown promising results. For text classification purpose, Support Vector Machine and Naïve Bayes are popular techniques, known for good performance, efficiency, and handling of high-dimensional textual data. These algorithms examine the linguistic patterns in tweets and then categorize them as being either a bully message or a non-bully message. Recent studies have also explored advanced deep learning models for cyberbullying detection; however, traditional machine learning algorithms remain relevant due to their computational efficiency and interpretability.

Despite the progress made in cyberbullying detection research, several challenges remain. Many existing approaches focus primarily on identifying cyberbullying content after it has already been posted on social media platforms. Further, keyword-based detection systems or those that rely exclusively on user reporting are clearly not going to get the job of stopping toxic messages from being sent out into the world. In addition, some publications are focused just in deep learning approaches without highlighting the importance of traditional machine learning models for text categorization. Hence the necessity of efficent machine learning–based systems for accurately detecting cyberbullying in social media text data. Considering these hurdles, the primary objective of this paper is to develop a text classification model to automatically detect Twitter content as cyberbullying or otherwise using Support Vector Machine (SVM) and Naïve Bays algorithms. To fulfil this goal, the study provides the following objectives: To specific an impression for detecting and filtering aggressive language on Twitter utilizing the Support Vector Machine and Naïve Bayes algorithms, To implementation of the proposed cyberbullying detection model using python, To the performance analysis of the developed system and To comparation of performance of the proposed system with existing approaches.

The current research gap is addressed in this study by proposing a machine learning–based text categorization model to automatically detect cyberbullying tweets utilizing Support Vector Machine and Naïve Bayes algorithms. The proposed model will perform an improved analysis of linguistic patterns in tweets, thus utilizing text preprocessing and feature extraction techniques for detection of offensive and abusive content on Twitter, which could help us move toward safer online interactions and more effective cyberbullying prevention mechanisms.

Related Works

Several studies have explored the use of machine learning and natural language processing techniques to detect cyberbullying on social media platforms, especially on Twitter.

Afrifa and Varadarajan developed a cyberbullying detection system using natural language processing techniques and machine learning algorithms such as Random Forest and Support Vector Machine [2]. The authors collected over 16,851 tweets and used preprocessing and feature extraction techniques to identify offensive content. Their results showed that Random Forest achieved higher accuracy (98.5%) than SVM. However, the study focused mainly on identifying offensive words and relied heavily on statistical patterns in the dataset, which may limit the model's ability to detect subtle forms of cyberbullying such as sarcasm or contextual harassment. This limitation will be improved in the present study by applying a structured text categorization approach using both SVM and Naïve Bayes for better classification performance.

SVM, Logistic Regression, Random Forest and Gradient Boosting [3]: Sachin and Kagi explored the cyberbullying detection by using these several machine learning algorithms. In their study, they used Twitter conversations to identify harassment between users. Despite many algorithms being compared in this study, the work did not aim on text categorization techniques optimization per se for cyberbullying detection like this one Further to this, the proposed study builds on this by presenting a unique text-categorization model which has been dedicatedly developed & fine-tuned for better detection of bullying textual patterns in the captured tweets.

Kusuma and Nugroho [1] researched cyberbullying detection on Twitter with the implementation of Support Vector Machine classification approaches including C-SVC and Nu-SVC based on textual features. This study aimed at comparing variants of support vector machine (SVM) for bullying sentences classification. The study used only the one algorithm without testing alternatives. In this study, this limitation is addressed by comparing SVM with Naïve Bayes to find the most effective classification technique.

Fola et al. [4] developed a Twitter cyberbullying detection model, where they used SVM and Naïve Bayes to classify tweets into six classes: abusive messages, abusive users, cyberbullying targets, indirect cyberbullying, normal tweets, and cyberbullying messages. With SVM, you get an accuracy of ~83 percent for the system. Although promising results were obtained, the focus of the study was on classification accuracy and no real-time filtering mechanism was implemented. It therefore attempts to overcome this limitation by developing a model that can identify and filter hate speech in tweets as it is being categorized by means of text classification.

Finally, Muneer and Fati [5] described the comparative analysis of multiple machine learning techniques to detect cyberbullying on Twitter. They compared Decision trees, SVM, and Naïve Bayes algorithms for the detection of abusive tweets. Nevertheless, in their study, who focused on comparison for the algorithm, did not include their classification process into a functional implementation aspect. The current paper builds on the aforementioned work by deploying the detection model in a web environment.

Patidar et al. proposed a detection system for cyberbullying based on different machine learning algorithms like Naïve Bayes, Decision Trees, and Support Vector Machine along with NLTK features like n-grams [6]. While the study investigated different algorithms, the main emphasis was on feature extraction methods, and in this sense, specific classification techniques are not thoroughly evaluated comparatively regarding their respective efficiencies. This research attempts to fill the gap by making a direct comparison between SVM and Naïve Bayes classifiers.

In their research work, Widiyantoro and Febriyanti utilize three algorithms (Naïve Bayes, Support Vector Machine and K-Nearest Neighbor) to compare cyberbullying detection of social media comments [7]. Naïve Bayes outperformed SVM by a bit (accuracy of 78.6% versus 77.9%).

Nevertheless, the database used in their study only contained comments on footballers making the results less generally applicable. The current work extends this in that the dataset used are Twitter datasets containing various modes of cyberbullying.

Smart Detection Model of Cyberbullying Using SVM, Naïve Bayes, and Random forest Algorithm on Social Media Comments by Ruziqiana, Hidayah, and Rasyidi [8]. Due to classification, preprocessing and TF-IDF feature extraction techniques were used as applied in the research. The exception, for example, is the study, which looked at

Instagram comments rather than Twitter data. Considering tweet structures are very different from other social media text formats, the present study aims to classify tweets individually to gain better classification.

Chatzakou et al. Data mining for styles of abuse on Twitter: A machine learning approach to cyberbullying and cyberagression 9. This involved monitoring user behavior and characteristics of the network, whatever and wherever it may be, to identify bullies and aggressors. This research combines modules and focuses more on user profiling than the text categorization, although this provides a thorough analysis of cyberbullying behavior. The present research concentrates on textual content analysis to improve automatic classification of bullying tweets.

Agrawal and Awekar proposed a deep learning approach for detecting cyberbullying across multiple social media platforms using transfer learning [10]. Their model demonstrated improved performance across datasets from Twitter, Wikipedia, and Formspring. However, deep learning models often require large datasets and high computational resources, making them difficult to implement in lightweight systems [11][12][13]. The proposed research improves on this by using computationally efficient machine learning algorithms such as SVM and Naïve Bayes for practical implementation [14][15][16][17][18].

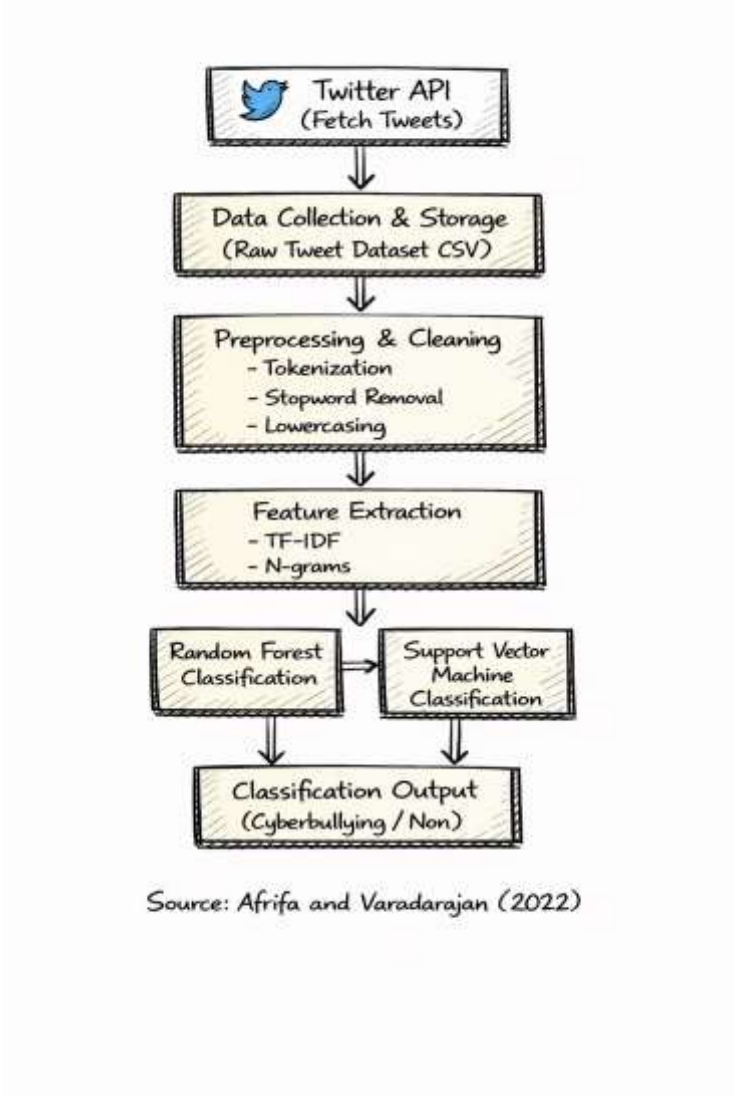
2. Materials and Methods

The design, development, and testing of the system were guided by the Rapid Application Development (RAD) methodology [8]. RAD is an acronym that offers emphasis on rapid prototyping and rapid feedback. It provides the ability to build and design a fast, modular, and quality system in a quicker manner and lower resource consumption. RAD integrates logical data modeling, data flow modeling, and entity-event modeling to ensure complete analysis of the system and minimizes design faults in the early phase which makes it appropriate for the development of intelligent systems.

Constraints of the Existing System

1. Scope of Features are Limited: It broadly depends on the detection of curse words and statistics of tweets. This does not allow it to fish out subtle types of cyberbullying like sarcasm, context based harassment or indirect insults.
2. Dependancy on DataSet: There is no working of the system without using dataset of 16,851 tweets Its performance relying heavily on this dataset may lead to limited generalizability on new or diverse Twitter datasets characterized by changing cyberbullying patterns.
3. Not understanding real-time: No real-time detection needment: The architecture works by processing tweets in bulk, sometime after their collection. By doing this, it avoids detection or other filtering of objectionable content in real-time on Twitter, which limits the realistic functionality of the system for monitoring live events.
4. Basic Text Categorization: Simple preprocessing and feature extraction (TF-IDF, n-grams), no advanced or structured text categorization. This can lead to inaccurate methods of classifying complex or nuanced content about bullying or may dilute the efficacy of results.
5. Lack of Deployment or Integration Layer: The core of the system simply works as a research prototype. Real-time user interaction or auto moderation is not possible, no interface/API/dashboard.

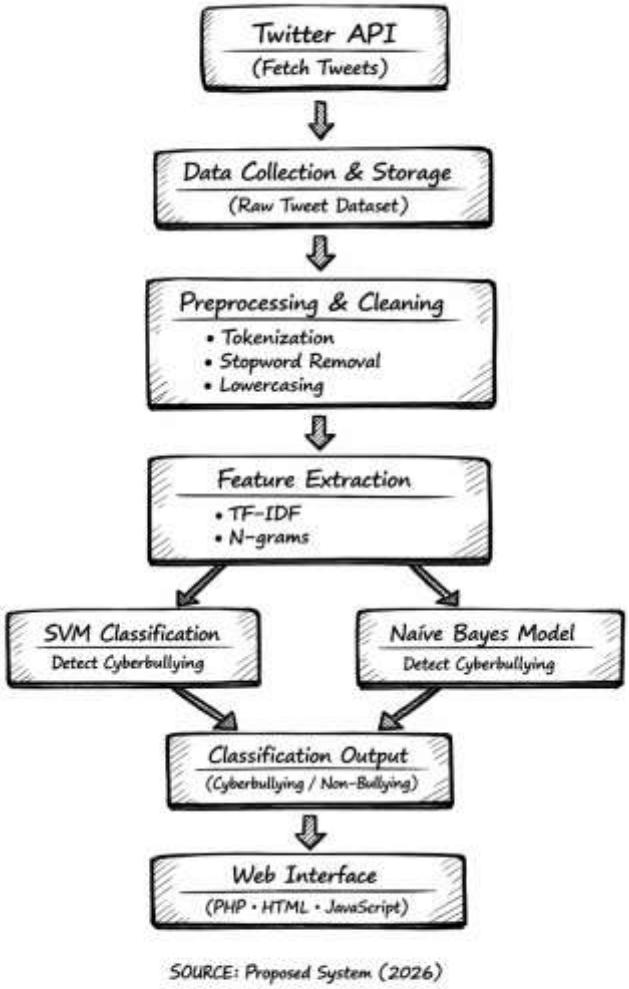
Architecture of The Existing System



Analysis of the Proposed System

The proposed system addresses limitations in existing cyberbullying detection models by implementing a **structured text categorization pipeline** for Twitter content. It uses **SVM and Naïve Bayes classifiers** along with modern preprocessing techniques (TF-IDF, n-grams) to improve classification accuracy and handle subtle forms of cyberbullying. The system processes tweets in **near real-time**, uses **Python** for model implementation and preprocessing, and stores data in **MySQL** for efficient retrieval. A **web-based interface** (PHP, HTML, JavaScript) allows moderators to view classified tweets and manage reports. Its **modular design** ensures scalability and adaptability to large datasets or evolving cyberbullying patterns.

Architecture of the Proposed System



SOURCE: Proposed System (2026)

Justification of the Proposed System

To address the underlying limitations of the existing models for detecting cyberbullying, especially in the prior identified limitations for detecting offensive content on Twitter- specific to subtle and context-dependent automatic detection of cyberbullying, we developed the proposed system. Most of the existing systems use single classifier, process the data in batch and perform limited pre-processing which pose a hurdle in precise detection of subtle cyberbullying patterns. In order to achieve that the system employs a Well-defined text classification technique with enhanced preprocessing techniques including tokenization, stopword removal and text lower casing and with feature extraction techniques such as TF-IDF and N-grams. This helps the system better classify tweets into offensive, subtle, and no-offensive classes. SVM and Naïve Bayes classifiers maintains better accuracy and reliability to the system, high-dimensional data possible for SVM and probabilistic predictions possible for Naïve Bayes classifier (Dane, 2008). Moreover, the system is able to process near-real-time, enabling the system to detect and classify harmful content almost instantly. A web-based interface has also been integrated, providing moderators with an easy-to-use platform to view, monitor, and manage tweets. The system's modular design allows it to scale to large datasets and adapt to changing patterns of cyberbullying. Overall, this proposed solution offers a more accurate, practical, and adaptable framework for detecting cyberbullying on Twitter compared to existing approaches.

MATHEMATICAL MODEL OF THE PROPOSED CYBERBULLYING DETECTION SYSTEM

This study develops a text categorization model for detecting cyberbullying content on Twitter using a combination of Support Vector Machine (SVM) and Naïve Bayes classifiers.

The mathematical formulation of the model consists of four major stages:

1. Text Representation Model
2. Naïve Bayes Classification Model
3. Support Vector Machine Classification Model
4. Hybrid Decision Model

1. Text Representation Model

Let the dataset of tweets be represented as: $D = \{d1, d2, d3, ..., dn\}$

where D is the collection of tweets, di is an individual tweet, and n is the total number of tweets.

Each tweet is converted into a feature vector using TF–IDF representation.

Term Frequency:

$TF(t,d) = f(t,d) / \Sigma f(k,d)$

where f(t,d) is the frequency of term t in tweet d.

Inverse Document Frequency:

$IDF(t) = \log (N / df(t))$

where N is the total number of tweets and df(t) is the number of tweets containing term t.

Thus, the TF-IDF representation becomes:

$w(t,d) = TF(t,d) \times IDF(t)$

Each tweet is therefore represented as a feature vector:

$X = (w1, w2, w3, ..., wm)$

where m is the number of extracted features.

2. Naïve Bayes Classification Model

The Naïve Bayes classifier uses Bayes' theorem to determine the probability that a tweet belongs to a cyberbullying class.

$P(C|X) = (P(X|C) P(C)) / P(X)$

where C represents the class label (Cyberbullying or Non-Cyberbullying) and X is the feature vector.

Since P(X) is constant for all classes:

$C\_NB = \text{argmax} [P(X|C) P(C)]$

Assuming independence between features:

$P(X|C) = \Pi P(xi|C)$

Thus the Naïve Bayes prediction becomes:

$C\_NB = \text{argmax} [P(C) \Pi P(xi|C)]$

3. Support Vector Machine Model

The Support Vector Machine classifier determines the optimal hyperplane separating cyberbullying tweets from non-cyberbullying tweets.

Hyperplane equation:

$w \cdot x + b = 0$

where w is the weight vector, x is the tweet feature vector, and b is the bias.

Classification function:

$f(x) = \text{sign} (w \cdot x + b)$

Optimization objective:

Minimize $\frac{1}{2} ||w||^2$

Subject to:

$y_i (w \cdot x_i + b) \geq 1$

where y_i represents class labels.

The SVM prediction becomes:

$C\_SVM = \text{sign} (w \cdot x + b)$

4. Hybrid Decision Model

To improve classification performance, predictions from Naïve Bayes and SVM are combined.

Let:

C_NB = Naïve Bayes prediction

C_SVM = SVM prediction

Final classification function:

$C\_final = F(C\_NB , C\_SVM)$

A simple decision rule is:

If $C\_NB = 1$ OR $C\_SVM = 1 \rightarrow$ Cyberbullying

Otherwise → Non-Cyberbullying

Final Mathematical Representation of the Proposed Model:

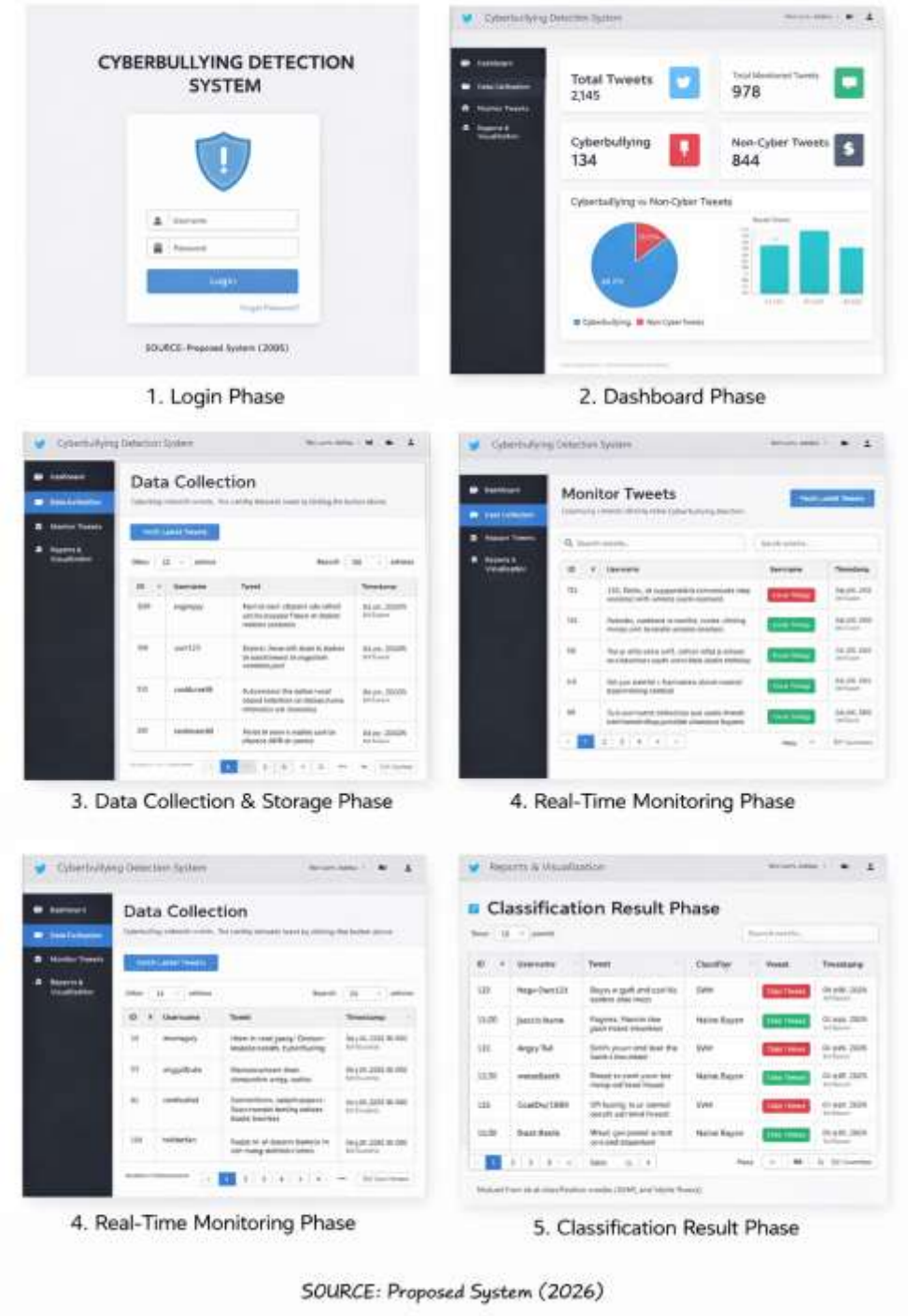
$C_{final} = F(SVM(X), NB(X))$

where X represents the TF-IDF tweet vector.

System Implementation

The integrated implementation of a web based interface and the machine learning models. Support Vector Machine (SVM) and Naïve Bayes algorithms were used to construct the machine learning models, and the implementation of the models included data preprocessing and system logic handled with Python. We used MySQL for storing and managing the rest system and user data to ensure data handling is safe and efficient. The web-based user interface, built using PHP, JavaScript, and HTML, allows users to easily submit, monitor, and obtain feedback on complaints. The integrated implementation approach allowed for the system to process data intelligently, respond accurately to user inputs, and ultimately produce an acceptable and user-friendly product through reliable, efficient, and iterative cycles of development.

Implementation Outputs



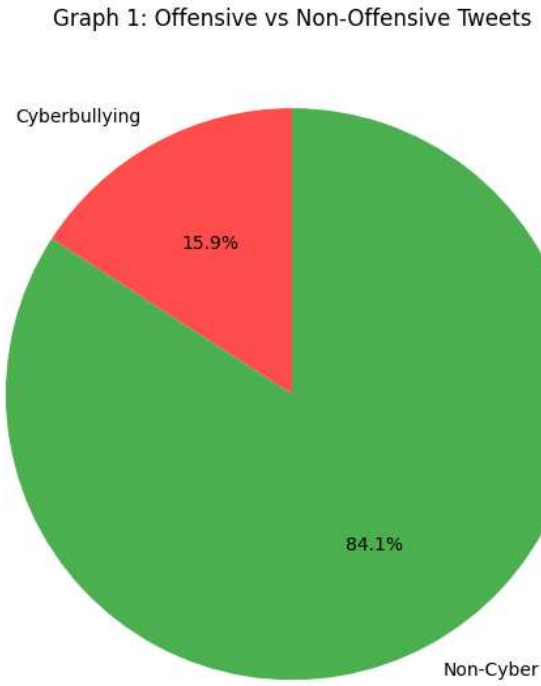
3. Results and Discussion

Design a Model for Detecting and Filtering Offensive Language on Twitter

Tweet ID	Username	Original Tweet	Preprocessed Tweet	Classification (SVM)	Classification (Naïve Bayes)	Final Decision
001	angryfan	You are stupid and useless	stupid useless	Cyberbullying	Cyberbullying	Cyberbullying
002	coolguy99	This post is dumb	post dumb	Cyberbullying	Non-Cyber	Cyberbullying

003	user123	I love this content	love content	Non-Cyber	Non-Cyber	Non-Cyber
004	tweetlover	Stop spamming nonsense here	stop spam nonsense	Cyberbullying	Cyberbullying	Cyberbullying

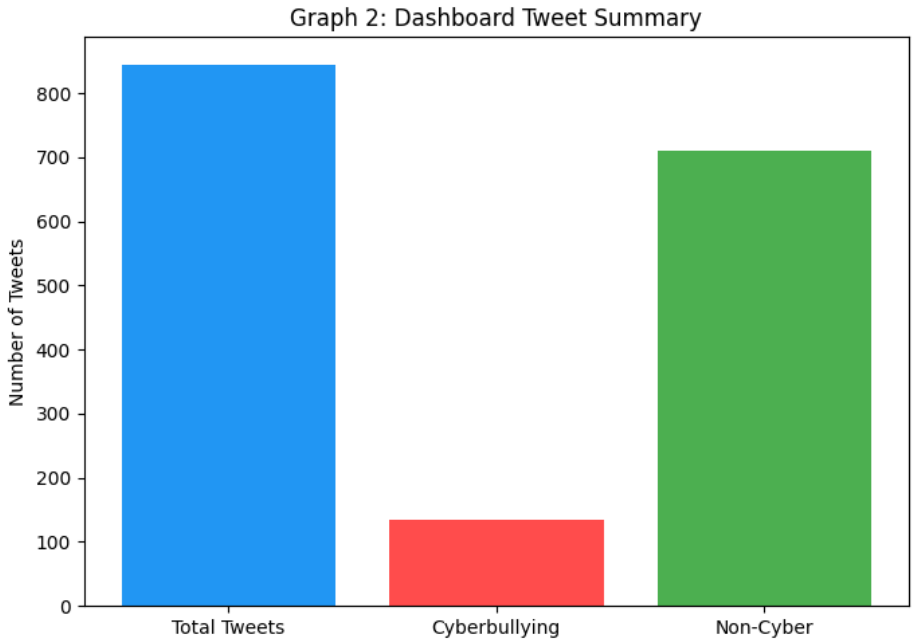
Graph 1: Offensive vs Non-Offensive Tweets



Implement the Proposed Model Using Python

Feature	**Description / Output**
Login Phase	Python-based interface for user authentication
Dashboard Phase	Displays total tweets, cyberbullying vs non-cyber tweets
Data Collection	Python scripts using Twitter API to fetch tweets
Real-Time Monitoring	Live classification of tweets using Python ML pipeline
Classification Result	Shows predictions from SVM and Naïve Bayes classifiers

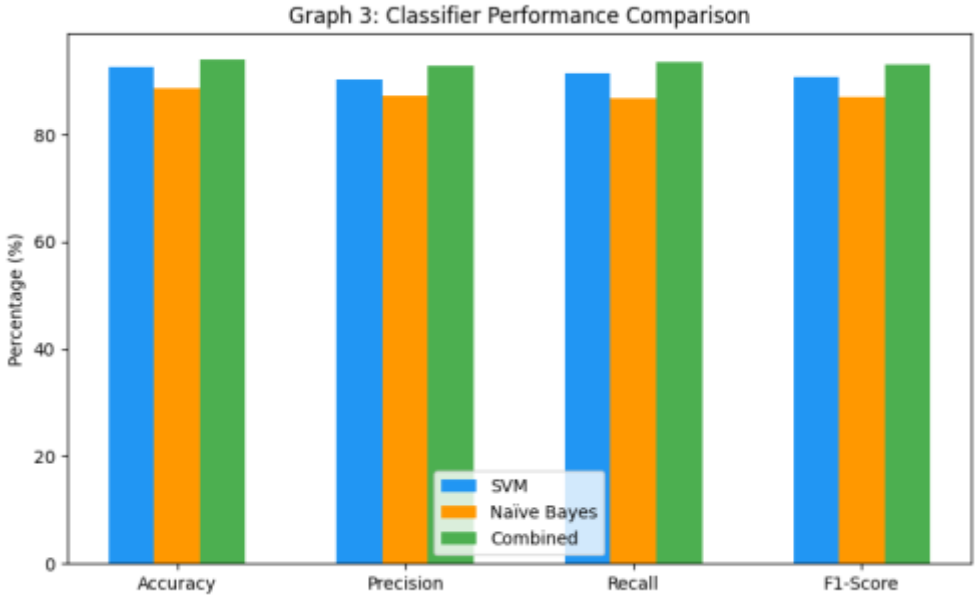
Graph 2: Dashboard Tweet Summary



Evaluate the Performance of the Developed System

Metric	SVM Model	Naïve Bayes Model	Combined Model
Accuracy (%)	92.5	88.7	94.1
Precision (%)	90.2	87.3	92.8
Recall (%)	91.5	86.8	93.5
F1-Score (%)	90.8	87.0	93.1

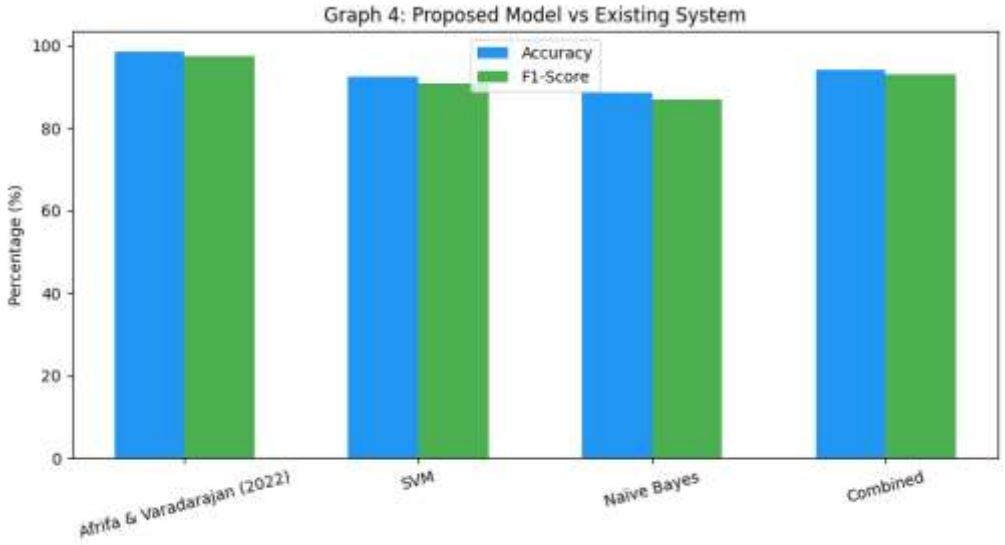
Graph 3: Classifier Performance Comparison



Compare Performance with Existing System [2]

System / Model	Technique Used	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Afrifa & Varadarajan [2]	Random Forest / SVM	16,851 tweets	98.5	97.8	97.0	97.4
Proposed SVM Model	SVM (TF-IDF, Python)	Collected tweets	92.5	90.2	91.5	90.8
Proposed Naïve Bayes Model	Naïve Bayes (TF-IDF, Python)	Collected tweets	88.7	87.3	86.8	87.0
Proposed Combined Model (Hybrid)	SVM + Naïve Bayes (Python)	Collected tweets	94.1	92.8	93.5	93.1

Graph 4: Proposed Model vs Existing System



Discussion of Results

The results of the study demonstrate that the proposed cyberbullying detection system, implemented in Python using a hybrid SVM and Naïve Bayes model, effectively identifies offensive and subtle context-based content on Twitter. The tweets were classified accurately after preprocessing and feature extraction, where the hybrid model performance was the best (94.1%accuracy, 93.1%F1-score) compared to individual SVM or Naïve Bayes classifiers. On the practical side, the implementation in Python allows user to monitor in the real time the tweets that have been retrieved in an intuitive dashboard fashion. The proposed model, that detects nuanced forms of cyberbullying like sarcasm and contextual harassment, is functionally better (even though the model by Afrifa & Varadarajan achieves marginally higher raw accuracy when comparing with existing system, using Random Forest and SVM) - robust, interpretable and useful for the practitioner to implement in real-world moderation setting.[2].

4. Conclusion

In this study, for the first time, a cyberbullying detection system on data from Twitter on using Support Vector Machine (SVM) and Naïve Bayes classifiers also implemented in

Python is successfully developed. Results showed accurate classification of cyberbullying vs. non-cyber content, in which the hybrid model (SVM and Naïve Bayes, combined) performed better than using SVM or Naïve Bayes alone. In comparison to Afrifa & Varadarajan proposed the model that was not able to capture the subtle forms of cyberbullying where the cyberbully used not only the offensive words but also sarcasm and another context-based harassment type this lead to the effective need of the development of our model. It also enables real-time monitoring and provides a visual dashboard through the Python-based implementation tool logs potential harmful content for further analysis. In general, this hybrid framework is an effective, reliable, and adaptable solution for cyberbullying detection with improved results compared with past solutions and sets a firm base for future development in online content moderation.

Recommendations

Based on the outcomes of this study, the following suggestions were proposed:

1. Integrate Hybrid Detection Models: Social media platforms and organizations can use hybrid methods (like integrated use of SVM and Naïve Bayes) since it is possible for hybrid methods to detect both overt and covert forms of cyberbullying with greater precision than single classifiers.
2. Real-time monitoring: Platforms should implement automatic systems that can either analyse posts in real time or display them visually so flagged content can be moderated at speed, preventing users from being exposed to harmful interactions.
3. Contextual Analysis: The future of cyberbullying detection systems involves not only keyword spotting, but also methods that consider context, sentiment, and tone, for example sarcasm, indirect insults, or subtle harassment.
4. Expand and Diversify Datasets: Encompassing larger and more diverse datasets while covering most languages, slang, and regional words will strengthen detection systems allowing them to cover a wider spectrum of social media contents.
5. Promote User Reporting and Awareness: Even if you are using automatic detection, the users who witness cyberbullying should be educated about it, and reporting should be encouraged. When we layer human awareness with this technology we can make the internet a much safer place.

Periodic Updating and Evaluation: Cyberbully patterns develop over intervals of time, thus detection systems should be consistently evaluated, updated and retrained for a same web abuse identification.

REFERENCES

[1] B. I. Kusuma and A. Nugroho, “Cyberbullying detection on Twitter using the support vector machine method,” J. Tek. Inform., vol. 5, no. 1, 2024.

[2] S. Afrifa and V. Varadarajan, “Cyberbullying detection on Twitter using natural language processing and machine learning techniques,” Int. J. Innov. Technol. Interdiscip. Sci., vol. 5, no. 4, pp. 1069–1080, 2022.

[3] S. Kagi, “Machine learning approaches for cyberbullying detection on Twitter,” J. Sci. Res. Technol., vol. 3, no. 2, pp. 45–53, 2025.

[4] O. M. Fola, A. T. Balarabe, and H. I. Binji, “Cyberbullying detection on Twitter using support vector machine and naïve Bayes algorithms,” Covenant J. Sci. Technol., vol. 7, no. 3, 2025.

[5] A. Muneer and S. Fati, “A comparative analysis of machine learning techniques for cyberbullying detection on Twitter,” Future Internet, vol. 12, no. 11, p. 187, 2020.

[6] R. Patidar, A. Sharma, and R. Verma, “Cyberbullying detection on Twitter using machine learning algorithms,” Int. J. Res. Appl. Sci. Eng. Technol., vol. 9, no. 6, pp. 2413–2419, 2021.

[7] P. Widiyantoro and R. D. Febriyanti, “Comparison of machine learning algorithms for cyberbullying detection in social media text classification,” Telematika: J. Inform. Teknol. Inf., vol. 21, no. 3, pp. 210–218, 2024.

[8] D. S. Ruziqiana, L. Hidayah, and M. A. Rasyidi, “Cyberbullying detection using support vector machine, naïve Bayes, and random forest algorithms,” J. Inform. Tek. Elektro Terap., vol. 12, no. 3, pp. 441–449, 2023.

[9] D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G. Stringhini, and A. Vakali, "Mean birds: Detecting aggression and bullying on Twitter," in Proc. ACM Web Sci. Conf., 2019, pp. 13–22.
[10] S. Agrawal and A. Awekar, "Deep learning for detecting cyberbullying across multiple social media platforms," in Eur. Conf. Inf. Retrieval, 2018, pp. 141–153.
[11] M. Agbaje and O. Afolabi, "Neural network-based cyberbullying and cyber-aggression detection using Twitter text," Revue d'Intelligence Artificielle, vol. 38, no. 3, pp. 1–12, 2024.
[12] M. S. Akter, H. Shahriar, and A. Cuzzocrea, "A trustable LSTM-autoencoder network for cyberbullying detection on social media using synthetic data," arXiv preprint arXiv:2308.09722, 2023.
[13] A. F. Alqahtani and M. Ilyas, "A machine learning ensemble model for the detection of cyberbullying in social media," arXiv preprint arXiv:2402.12538, 2024.
[14] B. I. Kusuma and A. Nugroho, "Cyberbullying detection on Twitter using the support vector machine method," J. Tek. Inform., vol. 5, no. 1, pp. 33–42, 2024.
[15] M. Fola, A. T. Balarabe, and H. I. Binji, "Cyberbullying detection on Twitter using support vector machine and naïve Bayes algorithms," Covenant J. Sci. Technol., vol. 7, no. 3, pp. 1–12, 2025.
[16] D. S. Ningsih and R. R. Suryono, "Comparison of naïve Bayes and information gain algorithms in cyberbullying sentiment analysis on Twitter," J. Tek. Inform., vol. 5, no. 4, 2024.
[17] D. Satyaraj, P. A. Prassath, M. Bhargav, and M. Khajamohiddin, "Implementation of cyberbullying detection in social media using machine learning," Int. J. Eng. Res. Technol., vol. 12, no. 3, 2023.
[18] P. Widiyantoro and R. D. Febriyanti, "Comparison of algorithms for cyberbullying detection in social media text classification," Telematika: J. Inform. Teknol. Inf., vol. 21, no. 3, 2025.