



AN ENSEMBLE-ORIENTED HYBRID XNS APPROACH FOR DETECTING ABNORMAL BEHAVIOR IN SOCIAL NETWORKS

Sarfaraz Alam¹, Mohammad Faisal²

^{1,2}Department of Computer Application, Faculty of Engineering & IT, Integral University, Lucknow, India.

¹<http://orcid.org/0000-0002-7846-293X>, ²<http://orcid.org/0000-0002-6120-5259>

Email: sarfarazalam2006@gmail.com, mdfaisal@iul.ac.in

ARTICLE INFO

Article History

Received: March 27, 2026

Reviewed: August 3, 2026

Accepted: September 3, 2026

Published: September 30, 2026

Keywords:

Anomaly Detection,
Online Social Networks,
Ensemble Learning,
Cybersecurity,
XGboost.

ABSTRACT

The rapid rise of online social networks (OSNs) has significantly reshaped how people communicate and share information, but it has also brought growing concerns around user security, privacy, and overall well-being. Identifying abnormal behavior—such as the spread of fake news, sharing malicious links, or engaging in harmful online activities—has become increasingly important for ensuring safer digital environments. To address these challenges, this study introduces an ensemble-oriented hybrid XNS framework that combines XGBoost, Support Vector Machine (SVM), and Naïve Bayesian Classifier (NBC) in a cascaded architecture, designed to improve detection accuracy and robustness. The proposed framework draws on a rich set of features extracted from both user profiles and user-generated content, allowing for a more comprehensive understanding of user behavior patterns. It is trained and tested using a combination of synthetic and real-world datasets, ensuring that the model remains adaptable and reliable across diverse scenarios. In addition to conventional anomaly detection, the framework also places special emphasis on identifying users who may show signs of depression or suicidality, thereby supporting early intervention and promoting mental health awareness. Experimental results demonstrate that the model achieves an accuracy of approximately 97%, outperforming traditional methods and highlighting its potential as a scalable and efficient solution for enhancing both cybersecurity and user well-being in modern social networks.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

In today's digital age, social networks have become pivotal in people's daily lives, serving as global communication, interaction, and knowledge exchange platforms. Within these networks, users can connect, collaborate, and build personal and professional relationships regardless of where they are located, thus bringing like-minded people who share similar interests, values, and points of view together [1]. But, beyond socializing, users also turn to social media, to find, create, and share information that pertains to their personal, academic, or professional needs. Yet, more engagement brings more to the table regarding the sharing of their sensitive personal data, which brings to light paramount considerations regarding privacy, security, and appropriate use of information. Social networks, through greater access to information than ever before, underscore the complex nature of networks that connect people globally, while treating the implications of an ecosystem steeped in ethical, technological, and social considerations as just another aspect of human context [2]. The exponential growth of online social networks (OSNs) has been accompanied by a significant rise in malicious activities, including various forms of anomalies (point, contextual, collective) and spam [3]. These threats, which range from disseminating harmful content to targeted harassment, are particularly potent within the trusted context of social circles, making them more effective than traditional email spam [4]. The vast volume of user-generated data, while valuable, also facilitates the evolution and propagation of these attacks [5]. Consequently, detecting these malicious actors—such as spammers, predators, and fraudsters—using advanced machine learning techniques has become a critical and pressing challenge in cybersecurity research. Social media has emerged as a major channel for cyberbullying, posing severe risks to vulnerable adolescents, including heightened tendencies toward self-harm and suicide [6].

Malicious users, such as Sybil accounts and impersonators, exploit these platforms to inflict psychological harm through hostile interactions [7]. Detecting such anomalous behavior relies on computational analysis of social network data to uncover patterns of malicious intent. To address this challenge, researchers increasingly employ machine learning classifiers, which have become central to distinguishing harmful actors from legitimate users and enhancing social media safety [8]. Machine learning (ML) is an essential component of modern Intrusion Detection Systems (IDS) that drastically improves the ability to categorize network traffic as either normal or malicious. While the standalone classifiers like Support Vector Machine (SVM), Naive Bayes, Random Forest, and Logistic Regression have proven usefulness, research has repeatedly shown that hybrid and ensemble-based approaches offer the best performance [9]. These are methods that leverage multiple algorithms to increase accuracy, robustness, and adaptability.

Performance can also be enhanced even further with sophisticated methods for feature selection, such as chi-square analysis, which serve to determine the most discriminative attributes for classification, strengthening detection capabilities and offering computational efficiency. Ultimately, the trajectory of ML-based IDS has moved toward intelligent, adaptive frameworks which can secure complex, large-scale network environments [10]. Distinguishing anomalous actors from legitimate users in social networks is more complex than traditional intrusion detection due to the nuanced, dynamic, and implicit nature of social interactions [11]. Unlike protocol-level anomalies, social network anomalies manifest in behavioral, linguistic, and relational patterns, posing challenges such as high-dimensional feature spaces, elevated false-positive rates, biased or unrepresentative training data, and the heavy computational demands of real-time analysis. This article presents a new cascaded hybrid model for detection of anomalies that aims to address the shortcomings of monolithic classifiers in the context of social network security.

The proposed architecture, employs a sequential framework that has three stages: first, a XGboost[12] classifier performs a coarse-grained anomaly detection; next, a Support Vector Machine[13] analyzes this subgroup to identify users who show significant disappointment; and finally, a Naive Bayesian Classifier[14] carries out a fine-grained risk analysis to differentiate between distress (non-suicidal) and high-risk (potential suicidal) behavior. This sequence transitions from coarse-grained filtering to fine-grained psychological assessment and yields a system that achieves higher accuracy, efficient computation, and actionable identification of at-risk users than monolithic single or parallel hybrid models. The incorporation of k-means clustering with more advanced classifiers like XGboost and Support Vector Machine presents a powerful hybrid model for dealing with two well-known limitations of k-means clustering, the problem of class dominance and the problem of forced assignment when the user-specified k is lower than the true number of natural groupings in the data. Building on this work:

- The first contribution of this study focuses on detecting anomalous users in social networks through the application of a XGboost classifier, thereby mitigating both forced assignment and class dominance challenges during the classification of normal and anomalous behaviors.
- The second contribution advances this effort by introducing a novel vulnerable user detection framework, termed XSN Ensemble, which employs a cascade of machine learning algorithms to achieve superior accuracy in identifying at-risk social media users. Beyond vulnerability detection, this model is also applicable in anomaly detection tasks such as high-performance spammer identification and suicidal behavior recognition, demonstrating its broader utility in social network analysis.
- The third contribution is the study of the proposed XSN Ensemble classifier, which takes advantage of XGboost, Support Vector Machine, and Naive Bayes Classifier in three variations of hierarchical level, and verifies the proposed model considering six performance metrics, including accuracy, on the performance of the classifier. Empirical findings also confirm that this hybrid classifier outperforms the existing classifiers by providing improved precision and robustness. To establish the credibility of the proposed model, extensive experiments, using both synthetic and actual datasets, have verified that the proposed model is scalable, adaptable and effective in solving complex issues related to anomaly detection in social networks.

This paper is organized into several sections. Following this introduction, we present a review of relevant literature in the Related Work section and establish the necessary theoretical background in the Preliminaries of classifier section. The proposed methodology for predicting vulnerable social media users is then detailed. Subsequently, the Experimental Analysis section presents a comprehensive evaluation of the method's performance. The paper concludes with a summary of findings and a discussion of potential future work.

II. RELATED WORK

This section reviews existing machine learning models designed for spammer detection in Online Social Networks (OSNs). Given the critical role of OSNs in modern communication and their inherent vulnerability to malicious activities like spam, the development of robust, automated detection mechanisms is essential [15]. This necessity has driven sustained research into machine learning solutions to preserve platform integrity and user trust. The evolution of online social networking from early email systems to modern platforms such as Facebook, Twitter, and Instagram have enabled large-scale communication and community building but also facilitated the rise of malicious activities, particularly spamming. To address this, researchers employ machine learning techniques that leverage behavioral features to distinguish legitimate users from spammers. These anomaly detection methods, now widely applied across email, web systems, and social networks, have become essential for enhancing platform security and reliability [11].

This section outlines the application of diverse machine learning methodologies for spam detection in social networks, encompassing both unsupervised and supervised paradigms. Unsupervised approaches, such as Random Forest ensembles, utilize a combination of content-based (e.g., URL ratio, keywords) and behavioral features (e.g., posting frequency) to identify anomalies [16]. Supervised techniques, notably Support Vector Machines are trained on multifaceted feature sets integrating linguistic and content-specific markers to achieve high accuracy [17]. While the literature demonstrates a predominant focus on content-based models, which analyze user-generated text and metadata, the review suggests a need for more holistic, multi-modal frameworks that integrate content, behavioral, and graph-based features to advance detection capabilities [18]. Hybrid machine learning approaches have proven effective for spammer detection in social networks by leveraging the complementary strengths of multiple algorithms.

For example, models combining Naive Bayes, clustering, and decision trees integrate probabilistic classification, behavioral grouping, and interpretable rules to capture diverse spam patterns, while frameworks that merge unsupervised clustering with supervised methods such as decision trees and SVM refine classification through both structure discovery and discriminative learning. These hybrid strategies offer a robust solution to the dynamic and large-scale challenges of spam detection on platforms like Facebook [19]. Recent studies on anomaly detection in social networks have strived to create new frameworks to manage the scale, dynamics, and complexity of social media data. The seminal research contributions include a scalable, real-time architecture using Apache Kafka for streaming the data and HBase for storing the data, while using new Political Independence Index metrics to identify threats during in-flight ingestion. Another contribution is an Online Fusion of Experts model to deal with concept drift in data streams.

This model combines linear-order and Gaussian-order algorithms to generate an adaptive forecasting of anomalies. Finally, a new statistical model is proposed that uses an infinite feature cascade, consisting of a normal modeling phase that provides a formalization of feature dynamics, as well as an anomaly detecting component utilizing an Infinite Factorial Hidden Markov Model (IFHMM), which identifies deviations from very complex cascading dependencies captured through nodes [20]. This paper describes an unsupervised, parameter-free anomaly detection method for multilayer social networks. By examining structural and behavioral features at multiple layers of the network, the framework detects users behaving anomalously, all while being completely unlabeled. It produces a ranking of anomalousness to enable identifying ranked individuals for intervention, and it provides a simple yet effective, scalable solution for assessing and uncovering suspicious acts occurring in large, complicated social settings.

Contemporary research prioritizes behavior-based anomaly detection to classify unusual activities in social media by analyzing dynamic user actions rather than static content [21]. This approach identifies anomalies by quantifying statistical deviations from established normative behavioral patterns. Methodologies often employ unsupervised learning algorithms, such as K-means clustering on multi-dimensional feature sets, to isolate outliers [22]. To enhance transparency, these systems integrate advanced visualization techniques, providing interpretable outputs that make algorithmic results accessible and actionable for security analysts, thereby bridging data mining and practical cybersecurity response [23]. This study tackles two major problems in online social networks: i) detecting malicious behavior, and ii) identifying vulnerable users [24].

In terms of malicious actor detection, the study proposes social influence-based analysis of behavior, using unsupervised clustering on fine-grained engagement data (e.g., social media responses) to locate shifts in typical interaction behavior. This enables detection of coordinated inauthentic behavior, without relying solely on analyzing content. However, the paper also points out a major gap in the protection of user privacy and well-being in the mentioned context. Specifically, traditional machine learning often puts social media users' confidentiality at risk, and vulnerable user detection has not received sufficient attention during a significant global mental health crisis. To address the gap, the study suggests a new, privacy-sensitive cascading machine learning framework to accurately predict users who are at risk of psychological distress, or self-harm - adding enhanced precaution and preventative intervention capabilities to the malicious tension.

III. PRELIMINARIES OF CLASSIFIER

III.1 INFORMATION THEORY

Information-theoretic measures provide a mathematical foundation for quantifying uncertainty, variability, and informational content within datasets, making them essential for developing and evaluating anomaly detection models. Key metrics such as entropy, conditional entropy, relative entropy (Kullback–Leibler divergence), information gain, and information cost enable researchers to characterize data patterns, assess feature significance, and identify irregularities. These tools collectively enhance the design of robust anomaly detection systems by measuring uncertainty, distributional deviations, and the efficiency of data encoding.

- Entropy, a fundamental concept in information theory, serves as a quantitative measure of uncertainty or randomness within a dataset and is widely used to detect unexpected patterns or anomalies. In the context of anomaly detection, entropy provides a way to evaluate how well the data conforms to expected distributions, with higher entropy indicating greater unpredictability and potential irregularities. For a dataset D , where each data item belongs to a class $y \in C_D$, the entropy of D relative to the $|C_D|$ -wise classification is defined as:

$$H(D) = - \sum_{y \in C_D} p(y) \log_2 p(y)$$

Where $p(y)$ is the probability of a data item belonging to class y . This formula calculates the average information needed to describe the class membership of a randomly chosen data item D .

- Conditional entropy is an extension of the entropy concept that measures the remaining uncertainty in a dataset D when some additional information Y is known. While standard entropy quantifies the overall unpredictability of a dataset, conditional entropy evaluates how much uncertainty persists about one variable given knowledge of another, making it particularly useful for understanding dependencies between features or classes. Formally, for a joint probability distribution $P(x,y)$ of two random variables X and Y , the conditional entropy of D given Y is defined as:

$$H(D|Y) = - \sum_{x \in X} \sum_{y \in Y} p(y) \log_2 p(x|y)$$

Where $P(x|y)$ is the conditional probability of x given y . This measure captures the average uncertainty remaining about X once the value of Y is known.

- Relative entropy, also known as Kullback–Leibler (KL) divergence, is a fundamental concept in information theory used to quantify the difference between two probability distributions defined over the same set of outcomes. Given a true probability distribution

$p(x)$ and an approximate or reference distribution $q(x)$, relative entropy measures how much information is lost when $q(x)$ is used to approximate $p(x)$. Mathematically, it is defined as:

$$D_{KL}(p||q) = - \sum_{x \in C} p(x) \log_2 \frac{p(x)}{q(x)}$$

where C represents the set of all possible outcomes. A higher value of relative entropy indicates a greater divergence between the two distributions, implying that the approximation $q(x)$ poorly represents the true distribution $p(x)$.

- Relative conditional entropy is an extension of relative entropy that quantifies the divergence between two joint probability distributions while considering conditional dependencies between variables. Given two distributions $p(x,y)$ and $q(x,y)$ defined over the same sets $x \in C_D$ and $y \in C_D$, relative conditional entropy measures how much information is lost when $q(x,y)$ is used to approximate $p(x,y)$ while accounting for the conditional relationship of X given Y . It is formally defined as:

$$D_{KL}(p||q | Y) = - \sum_{y \in C_D} \sum_{x \in C_D} p(x,y) \log_2 \frac{p(x|y)}{q(x|y)}$$

where $p(x|y)$ and $q(x|y)$ are the conditional probabilities of x given y under the respective distributions.

- Information gain is a fundamental metric in information theory that quantifies the amount of information obtained about a dataset by observing a particular attribute or feature. Specifically, for a dataset D and an attribute Y , information gain measures how much knowledge of Y reduces the uncertainty or entropy of the dataset, effectively indicating the attribute's relevance for distinguishing between different classes or patterns. Formally, information gain is computed as the difference between the overall entropy of the dataset and the conditional entropy of the dataset given the attribute:

$$IG(D|Y) = H(D) - H(D|Y)$$

where $H(D)$ represents the entropy of the dataset, and $H(D|Y)$ represents the conditional entropy of D given Y .

III.2 XGBOOST

XGBoost (Extreme Gradient Boosting) is a powerful and scalable machine learning framework widely used for structured or tabular data due to its high predictive performance and computational efficiency. As an ensemble learning method, it builds successive models to predict residuals of previous models and aggregates their outputs for robust predictions. Its effectiveness is enhanced by L1 and L2 regularization to prevent overfitting, along with computational optimizations such as parallel processing, tree pruning, and cache-aware memory management. Supporting distributed and out-of-core processing, XGBoost scales to large datasets and accommodates various tasks, including regression, classification, and ranking. These features have made it a preferred choice in both research and real-world applications, consistently delivering state-of-the-art results.

III.3 SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine (SVM) is a strong supervised learning algorithm that is used extensively for classification and regression problems. SVM finds a best hyperplane that maximizes the margin between the classes, using key data points referred to as support vectors, which makes SVM more resistant to outliers. For non-linearly separable data, SVM utilizes kernel functions like the Gaussian (RBF) kernel to transform the inputs into higher feature spaces where the model can handle complex decision boundaries in an efficient manner. Its integration of margin maximization, sparsity in support vectors, and kernel-based flexibility renders SVM extremely powerful and popular in various applications.

III.4 NAIVE BAYES CLASSIFIER (NB)

The Naive Bayes approach is a probabilistic supervised learning method used to classify social network users based on posterior probabilities of observed features. In mental health applications, it can distinguish users exhibiting distress into suicidal and non-suicidal groups, leveraging features derived from specialized lexical resources, such as a suicidal dictionary. By assuming feature independence, Naive Bayes enables efficient computation even with high-dimensional data. Its simplicity, interpretability, and computational efficiency make it an effective framework for identifying vulnerable users and supporting timely interventions.

IV. THE PROPOSED MODEL FOR IDENTIFYING VULNERABLE USERS IN SOCIAL NETWORKS

The proposed system architecture is founded on a cascaded analytical framework that synergistically integrates three distinct machine learning algorithms to optimize the detection of anomalous activities within social networks. This multi-tiered methodology, depicted in Figure 1, employs a sequential filtering process wherein each classifier is strategically selected and deployed according to its inherent computational strengths and performance characteristics. The initial stage of the pipeline leverages a XGboost algorithm, which is specifically chosen for its high interpretability, efficiency in handling high-dimensional data, and demonstrated proficiency in achieving robust classification accuracy on large-scale datasets. This stage performs a coarse-grained filtration of the entire user population, effectively segregating anomalous entities from normative behavioral profiles, thereby establishing a refined subset for subsequent, more granular analysis.

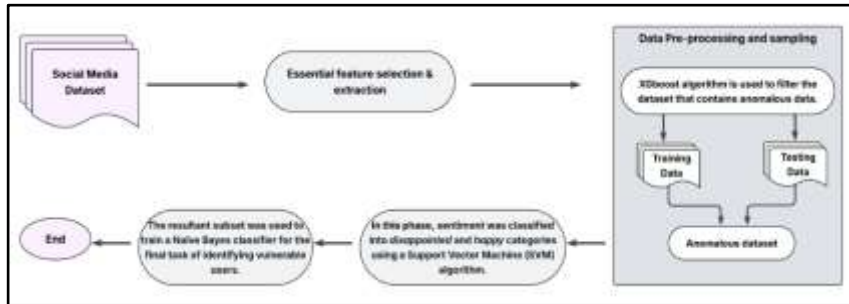


Figure 1: Work flow of Proposed Anomaly Detection Model.
Source: Authors, (2026).

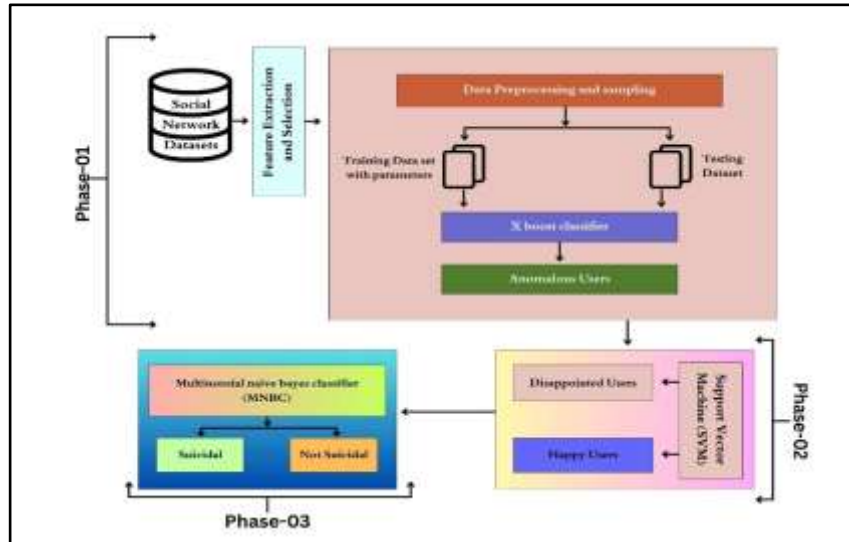


Figure 2: Proposed Framework.
Source: Authors, (2026).

So, the operational pipeline of the proposed system, as illustrated in Figure 2, employs a cascaded classification strategy designed to progressively refine the detection process. The initial phase processes the complete dataset through a XGboost algorithm, capitalizing on its efficiency with large-scale data and strong feature selection capabilities to perform a broad filtration of anomalous entities. This refined subset is subsequently analyzed by a Support Vector Machine (SVM), which functions as an optimal binary classifier to precisely identify users exhibiting behavioral indicators of disappointment, leveraging SVM's ability to maximize marginal separation between classes in a high-dimensional space. The final stage applies a Naïve Bayesian Classifier (NBC) to this highly targeted group; selected for its high performance and probabilistic robustness on smaller, well-defined datasets, the NBC performs a fine-grained assessment to distinguish users presenting potential suicidal risk. This hierarchical architecture visually summarized in the decision workflow of Figure 3 ensures that each algorithmic stage operates on a progressively reduced and more relevant subset of data, enhancing both computational efficiency and discriminative accuracy at each step of the analytical cascade.

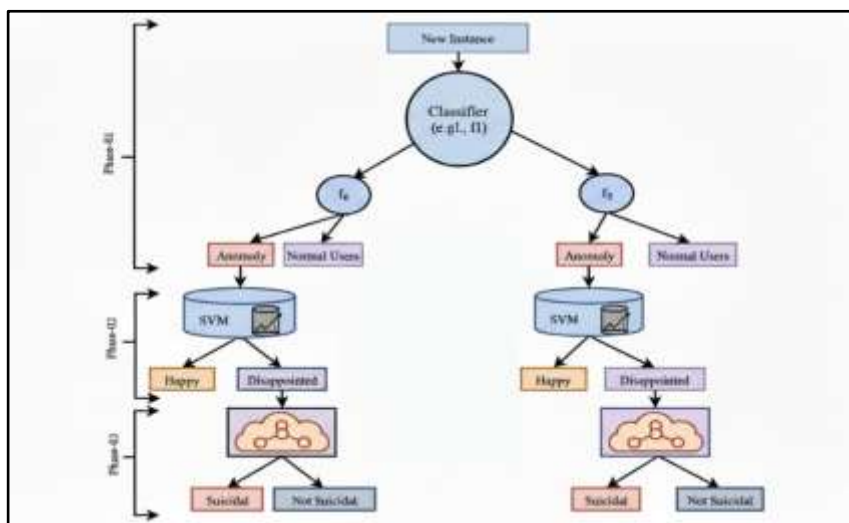


Figure 3: Workflow of the decision-making process of Proposed Framework.
Source: Authors, (2026).

IV.1 FEATURE ANALYSIS FOR ANOMALY DETECTION IN SOCIAL NETWORKS

The effectiveness of machine learning approaches for detecting anomalous users in social networks largely depends on the careful selection of features that best differentiate abnormal behavior from legitimate user activity. Unlike normal users, anomalous actors often pursue commercial or malicious objectives, which manifest through distinctive patterns in their profiles and content. To capture these differences, features can be broadly categorized into profile-based and content-based types. Profile-based features focus on user attributes and behavioral indicators, such as the number of followers, frequency of wall posts within a given time frame, and the size of a user's friend network, all of which provide insights into account activity and social engagement. These features, commonly employed in prior studies, offer a structured view of how user behavior deviates from typical patterns. Complementing these are content-based features, which examine the textual and multimedia elements of user-generated posts, including tags, comments, URLs, likes, hashtags, and the presence of spam-related keywords. Such content indicators are critical for uncovering unusual or malicious activity embedded within interactions on platforms like Facebook. By integrating both profile-based and content-based features, researchers can build a more comprehensive and discriminative representation of user behavior, thereby enhancing the accuracy and robustness of anomaly detection models.

IV.2 IMPORTANT FEATURES USED FOR TRAINING THE MACHINE LEARNING MODEL

The discriminative capability of the proposed model is underpinned by a meticulously engineered feature set that constructs a holistic behavioral profile, integrating dimensions of social connectivity, linguistic content, and temporal dynamics to effectively differentiate between normative and anomalous user activities. Social graph metrics, including the total number of friends (f1) and followers (f2), serve as foundational indicators, with spammers often exhibiting impoverished social networks [25]. Content-based features perform granular lexical and semantic analysis, quantifying the prevalence of negative emoticons (f3), embedded URLs (f4) often associated with phishing, suspicious inclusions like phone numbers (f5), and the use of hashtags (f6) for maximizing visibility. User engagement is rigorously assessed through metrics such as average likes (f7), comments per post (f13), and post shares (f14), which collectively reveal characteristically low interaction rates for inauthentic accounts.

Furthermore, specific linguistic markers—including the frequency of money-related terms (f8) and symbols (f9), alongside psychologically indicative lexicons of disappointment (f15) and suicidality (f16)—provide critical signals for identifying both malicious intent and severe psychological distress. Temporal patterns, including message size (f10), share frequency (f11), and the average time interval between posts (f12), are incorporated to detect the burst-like posting behavior symptomatic of automated scripts or coordinated campaigns. This multidimensional feature architecture is explicitly designed to support the cascaded classification process, enabling the precise detection of a spectrum of anomalies, from orchestrated spam and fraud to individuals exhibiting high-risk psychological distress. The proposed anomaly detection framework, as illustrated in Figure 3, is architecturally structured into three sequential and logically distinct phases, each designed to progressively refine the analysis and enhance detection specificity.

The initial phase employs a Xgboost algorithm to perform broad-spectrum filtering of the entire dataset, leveraging its computational efficiency and robust handling of high-dimensional data to isolate a subset of users exhibiting generalized anomalous behavior. The second phase applies a Support Vector Machine (SVM) to this pre-filtered subset, utilizing its superior performance as a binary classifier to precisely identify users demonstrating specific linguistic and behavioral markers associated with disappointment. The final phase utilizes a Naïve Bayesian Classifier (NBC) to conduct a granular, probabilistic risk assessment on this highly refined cohort, effectively discriminating between non-suicidal distress and high-risk suicidal indicators. This cascaded analytical design ensures that each stage operates on a progressively smaller, more relevant subset of data, thereby optimizing computational efficiency while maximizing classification accuracy at each step of the hierarchical detection process.

V. EMPIRICAL ANALYSIS AND MODEL EVALUATION

The proposed model employs a three-phase classification framework integrating XGboost, SVM, and Naive Bayes algorithms. In the first phase, XGboost with the information gain method is applied, with performance evaluated using a 70:30 training-to-testing split. The second phase utilizes a polynomial kernel-based SVM to capture complex non-linear patterns, while the third phase incorporates Naive Bayes for probabilistic classification. Together, these phases combine interpretability, robust pattern recognition, and probabilistic reasoning to achieve comprehensive detection accuracy. To evaluate the effectiveness of the proposed approach, the true positive rate (TPR) and false positive rate (FPR) are calculated across different threshold values to measure the model's ability to accurately detect anomalous users while minimizing incorrect classifications. These metrics provide a balanced view of performance by capturing both detection capability and error tendency, thereby serving as key indicators for analyzing the overall outcomes of the model. The performance of the proposed model is quantitatively evaluated using a confusion matrix shown in the table 1, which provides a structured representation of classification results by categorizing predictions into four distinct outcomes: True Positives (TP), correctly identified anomalous instances; True Negatives (TN), accurately classified normal cases; False Positives (FP), normal instances misclassified as anomalies; and False Negatives (FN), anomalous cases erroneously labeled as normal. This matrix facilitates a granular analysis of the model's discriminative power, enabling the calculation of key performance metrics such as precision, recall, and accuracy, while critically highlighting the trade-offs between detection sensitivity and specificity in the identification of targeted behaviors.

Table 1: Confusion matrix

	Condition Positive	Condition Negative
Class 0	756	1
Class 1	21	22

Source: Authors, (2026).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Where TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative

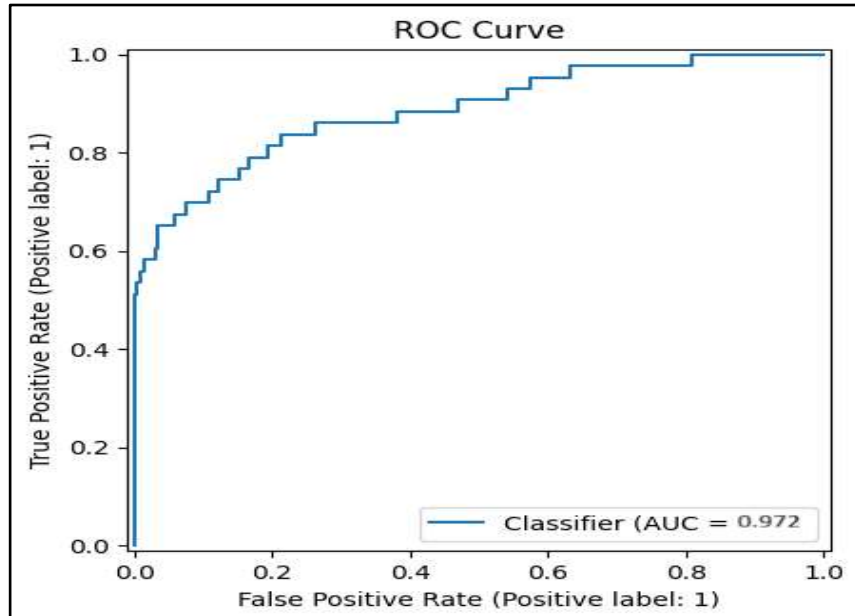


Figure 4: ROC Curve of the Proposed Framework.

Source: Authors, (2026).

The Receiver Operating Characteristic (ROC) Curve is a widely used graphical representation that illustrates the diagnostic performance of a test across various decision thresholds by plotting the true positive rate (sensitivity) against the false positive rate (specificity). This curve visually demonstrates the inherent trade-off between sensitivity and specificity, as improving the ability of a test to correctly identify true positives often comes at the cost of increasing false positives, thereby reducing specificity. An ideal diagnostic test would achieve high sensitivity while maintaining a low false positive rate, which is reflected in a curve that closely follows the left-hand border of the ROC space and then rises sharply along the top border, indicating both strong discriminatory power and precision. In practice, the ROC curve provides critical insight into how well a test distinguishes between conditions, allowing researchers and clinicians to compare alternative diagnostic tools, select optimal cut-off points, and evaluate the overall accuracy of prediction models. The foundational phase of the hybrid XSN ENSEMBLE framework utilizes a XGboost algorithm, selected for its optimal balance of high accuracy and computational efficiency in large-scale anomaly detection. Empirical results in table 2 demonstrate that XGboost surpasses alternative classifiers, including KNN, NBC, SVM and Random Forest across standard performance metrics such as Precision, Recall, and Accuracy on both synthetic and real-world datasets. Although SVM exhibits competitive accuracy, XGboost achieves superior scalability with reduced processing time, making it ideally suited for the initial coarse-grained filtration stage. This strategic selection ensures robust preliminary identification of anomalous users while establishing an efficient and effective data pipeline for subsequent specialized classification stages.

Table 2: Accuracy Comparison for Anomalous User Detection on Dataset1.

Algorithms	Recall	Precision	F-Measure	Accuracy
KNN	0.955	0.9468	0.9483	0.955
NBC	0.9175	0.9475	0.9293	0.9175
SVM	0.96	0.954	0.954	0.96
Random Forest	0.95	0.9523	0.9511	0.95
XSN ENSEMBLE Model	0.9733	0.9726	0.9698	0.972

Source: Authors, (2026).

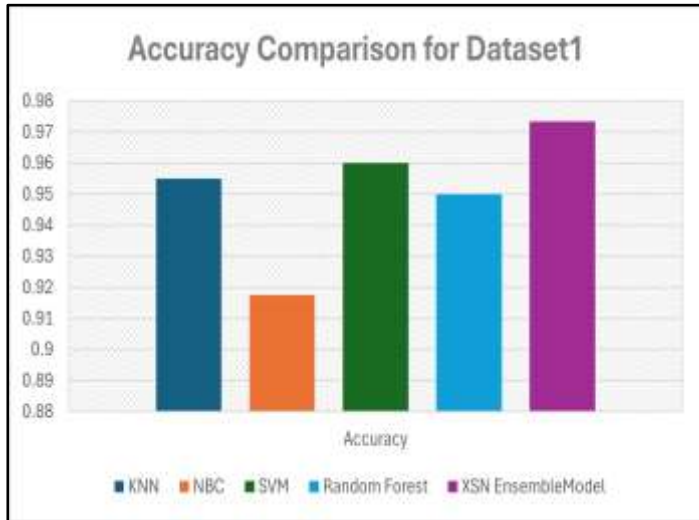


Figure 5: Accuracy comparison for Dataset1.
Source: Authors, (2026).

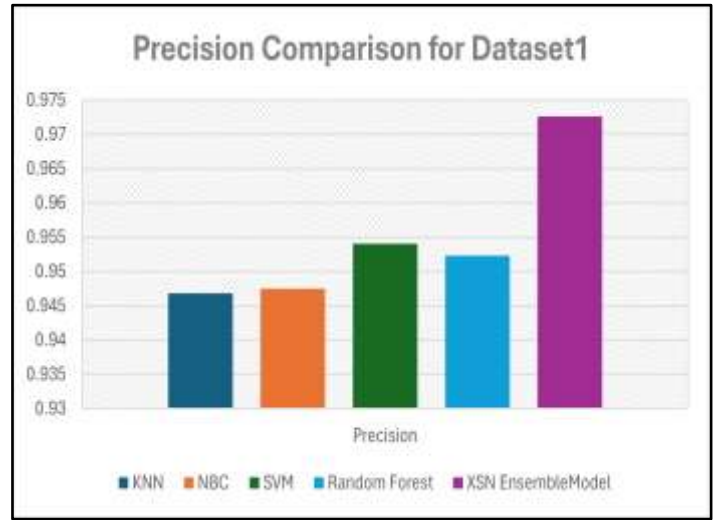


Figure 6: Precision comparison for Dataset1
Source: Authors, (2026).

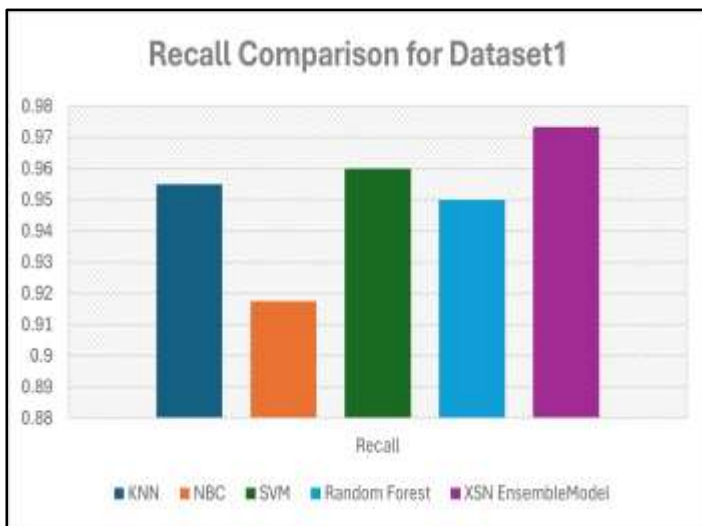


Figure 7: Recall comparison for Dataset1
Source: Authors, (2026).

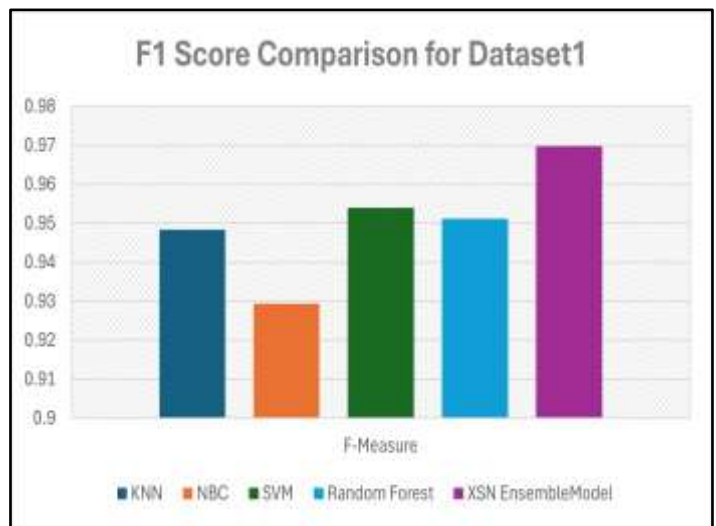


Figure 8: F1-Score comparison for Dataset1
Source: Authors, (2026).

The detection accuracy of the proposed XSN ENSEMBLE model for identifying anomalous suicidal users is quantitatively demonstrated in Figure 5 to Figure 8, which compares its performance measured through Accuracy, Precision, Recall and F1-Score on Dataset 1 against established algorithms including KNN, SVM, and NBC. The results indicate that the hybrid XSN ENSEMBLE framework achieves a notably superior accuracy of approximately 97%, significantly outperforming all comparative methods. This evaluation confirms that the cascaded architecture, which synergistically combines decision tree, support vector machine, and naïve Bayesian classifiers, provides enhanced discriminatory capability in identifying both general anomalous behavior and specific high-risk psychological profiles within social networks. The findings substantiate the efficacy of the proposed approach over conventional single-model detection systems.

VI. CONCLUSION

In response to the growing critical need for advanced cybersecurity in online social networks (OSNs), this paper presents a comprehensive machine learning-based framework for anomaly detection, with a specific focus on identifying both malicious actors and users exhibiting high-risk suicidal behaviors. We introduce a novel hybrid model, termed XSN ENSEMBLE, which integrates a Xgboost, Support Vector Machine and Naïve Bayesian Classifier in a cascaded architecture to sequentially refine detection accuracy. Distinct from prior works, our solution employs a meticulously curated set of content-based and behavioral features, derived through statistical analysis and manual selection, to capture multifaceted user activities. A significant contribution of this approach is its demonstrated capability to generalize to previously unseen anomaly patterns, enhancing its practicality for real-world OSN environments. The efficacy of the proposed system is rigorously validated through extensive comparative experiments on both real and synthetic datasets, where it outperforms existing techniques, thereby offering a robust and scalable solution for simultaneous detection of anomalous and psychologically vulnerable users. Future work will prioritize the development of a real-time, adaptive detection system that dynamically models evolving user behavior to improve contextual and temporal anomaly recognition. This will involve creating online learning mechanisms to continuously address concept drift and emerging threats in streaming social network data, enhancing the framework's practical deployment in live cybersecurity applications.

VII. ACKNOWLEDGEMENT

This work is acknowledged under Integral University manuscript No IU/R&D/2026-MCN0004413.

VIII. REFERENCES

- [1] Partelow, S., & Manlosa, A. O. (2023). Commoning the governance: a review of literature and the integration of power. *Sustainability Science*, 18(1), 265-283, doi: 10.1007/s11625-022-01191-2.
- [2] Fazil, A. W., Hakimi, M., Akrami, K., Akrami, M., & Akrami, F. (2024). Exploring the role of social media in bridging gaps and facilitating global communication. *Studies in Media, Journalism and Communications*, 2(1), pp 13-21, doi: 10.32996/smjc
- [3] Aljabri, M., Zagrouba, R., Shaahid, A., Alnasser, F., Saleh, A., & Alomari, D. M. (2023). Machine learning-based social media bot detection: a comprehensive literature review. *Social Network Analysis and Mining*, 13(1), 20, doi: 10.1007/s13278-022-01020-5
- [4] Bikkur, T., Biyyapu, N. S., Sekhar, J. C., Kumar, M. K., Nokerov, S. M., & Pratap, V. K. (2024). The Social Network Dilemma: Safeguarding Privacy and Security in an Online Community. *International Journal of Safety & Security Engineering*, 14(1), doi: <https://doi.org/10.18280/ijss.140112>.
- [5] Saura, J. R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2025). Privacy concerns in social media UGC communities: Understanding user behavior sentiments in complex networks. *Information Systems and e-Business Management*, 23(1), 125-145, doi: 10.1007/s10257-023-00631-5.
- [6] POLAT, O., & CANSIZ, A. (2024). All Aspects of Adolescent Suicides. *ISPEC International Journal of Social Sciences & Humanities*, 8(1), 92-117, doi: <https://doi.org/10.5281/zenodo.10804736>.
- [7] Shafik, W. (2024). Cyber attacker profiling and cyberbullying overview. In *Cyber space and outer space security* (pp. 125-149). River Publishers, doi: 10.1201/9781003558118-5.
- [8] Bhattacharya, M., Roy, S., Chattopadhyay, S., Das, A. K., & Shetty, S. (2023). A comprehensive survey on online social networks security and privacy issues: Threats, machine learning-based solutions, and open challenges. *Security and Privacy*, 6(1), e275, doi: 10.1002/spy2.275.
- [9] Shaikh, T. A., Rasool, T., Verma, P., & Mir, W. A. (2024). A fundamental overview of ensemble deep learning models and applications: systematic literature and state of the art. *Annals of Operations Research*, 1-77, doi: 10.1007/s10479-024-06444-0.
- [10] E. Dritsas and M. Trigka, "Machine Learning in Intelligent Networks: Architectures, Techniques, and Use Cases," in *IEEE Access*, vol. 13, pp. 102724-102745, 2025, doi: 10.1109/ACCESS.2025.3577968.
- [11] Xing, L., Li, S., Zhang, Q., Wu, H., Ma, H., & Zhang, X. (2024). A survey on social network's anomalous behavior detection. *Complex & Intelligent Systems*, 10(4), 5917-5932, doi: 10.1007/s40747-024-01446-8.
- [12] M. Abdul et al., "A Multimodal AI-Driven System for Preoperative Difficult Airway Assessment," in *IEEE Access*, vol. 13, pp. 164630-164640, 2025, doi: 10.1109/ACCESS.2025.3610860.
- [13] J. Dhammananda, I. Budi and A. B. Santoso, "Sentiment Analysis and Topic Modeling of E-Grocery Application Reviews Using Naive Bayes and Support Vector Machine: A Case Study of Segari Data Review on the Google Play Store," 2023 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS), Yogyakarta, Indonesia, 2023, pp. 13-18, doi: 10.1109/ICE3IS59323.2023.10335206.
- [14] Liu, Q., Ren, Y., Huang, Y. et al. Development and validation of a probability-assessment nomogram for non-suicidal self-injury in hospitalized adolescents and young adults with mental disorders. *Sci Rep* 15, 15142 (2025), doi: 10.1038/s41598-025-00142-y.
- [15] Bhattacharya, M., Roy, S., Chattopadhyay, S., Das, A. K., & Shetty, S. (2023). A comprehensive survey on online social networks security and privacy issues: Threats, machine learning-based solutions, and open challenges. *Security and Privacy*, 6(1), e275, doi: 10.1002/spy2.275.
- [16] Shajari, S., Agarwal, N. Developing a network-centric approach for anomalous behavior detection on youtube. *Soc. Netw. Anal. Min.* 15, 3 (2025), doi:10.1007/s13278-025-01417-y.
- [17] A. Adesokan, S. Madria and L. Nguyen, "TweetACE: A Fine-grained Classification of Disaster Tweets using Transformer Model," 2023 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), St. Louis, MO, USA, 2023, pp. 1-9, doi: 10.1109/AIPR60534.2023.10440656.
- [18] Mostafa, M., Almogren, A. S., Al-Qurishi, M., & Alrubaian, M. (2024). Modality deep-learning frameworks for fake news detection on social networks: A systematic literature review. *ACM Computing Surveys*, 57(3), 1-50, doi: 10.1145/3700748.
- [19] M. Salman, M. Ikram and M. A. Kaafar, "Investigating Evasive Techniques in SMS Spam Filtering: A Comparative Analysis of Machine Learning Models," in *IEEE Access*, vol. 12, pp. 24306-24324, 2024, doi: 10.1109/ACCESS.2024.3364671.
- [20] P. Kumar and A. R. Abhyankar, "A Time Efficient Factorial Hidden Markov Model-Based Approach for Non-Intrusive Load Monitoring," in *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 3627-3639, Sept. 2023, doi: 10.1109/TSG.2023.3245019.
- [21] Mannocci, L., Mazza, M., Monreale, A., Tesconi, M., & Cresci, S. (2024). Detection and characterization of coordinated online behavior: A survey. *arXiv preprint arXiv:2408.01257*, doi: 10.48550/arXiv.2408.01257.
- [22] Alcacer, A., & Epifanio, I. (2024). Outlier detection of clustered functional data with image and signal processing applications by archetype analysis. *Plos one*, 19(11), e0311418, doi: 10.1371/journal.pone.0311418.
- [23] Sarker, I. H. (2024). Cybersecurity Data Science: Toward Advanced Analytics, Knowledge, and Rule Discovery for Explainable AI Modeling. In *AI-Driven Cybersecurity and Threat Intelligence: Cyber Automation, Intelligent Decision-Making and Explainability* (pp. 101-118). Cham: Springer Nature Switzerland, doi: 10.1007/978-3-031-54497-2_6.

- [24] Shah, A., Varshney, S. & Mehrotra, M. Threats on online social network platforms: classification, detection, and prevention techniques. *Multimed Tools Appl* 84, 17083–17115 (2025). <https://doi.org/10.1007/s11042-024-19724-5>.
- [25] S. Terumalasetti and S. R. Reeja, "Enhancing Social Media User's Trust: A Comprehensive Framework for Detecting Malicious Profiles Using Multi-Dimensional Analytics," in *IEEE Access*, vol. 13, pp. 7071-7093, 2025, doi: 10.1109/ACCESS.2024.3521951.