RESEARCH PAPER

Security Risk Analysis and Detection of Suspicious Communications in Social Media Big Data Using Machine Learning

Gaurav Khatri[1*], Vipul M. Dabhi[2], Manoj Patel[3], Hinal Prajapati[4]

[1*] Ph.D. Research Scholar, Assistant Professor, Faculty of Computer Science & Applications, Gokul Global University, Siddhpur, Gujarat, India (Corresponding Author)

[2] Professor, CSE Department, Faculty of Engineering, Gokul Global University, Siddhpur, Gujarat, India

[3] Dean, Faculty of Computer Science & Applications, Gokul Global University, Siddhpur, Gujarat, India

[4] Assistant Professor, Department of Computer & Information Technology, HNGU, Patan, Gujarat, India

Received: 2nd Mar, 2026 | **Revised:** 14th Mar, 2026 | **Accepted:** 4th Apr, 2026 | **Available Online:** 20th Apr, 2026

ABSTRACT

The bridging social media has seen a colossal spread of user generated content and has also spawned a myriad of security threats in the form of spam, rogue messages and threatening communications. The study introduces a machine learning-driven framework of determining the level of security risk and detecting suspicious communications in social media big data. The solution that is proposed makes use of spam detection algorithms that are based on text in order to detect potential harmful traffic. It uses a tedious approach that involves data cleaning, exploratory data analysis, label encoding, and text preprocessing using minimum frequency-inverse document frequency (TF-IDF) technique. The five machine learning models, including the Logistic Regression, random forest, decision tree, K-nearest neighbors and Multinomial Naive Bayes are trained and evaluated by traditional performance measures, which include accuracy, precision, recall, F1-score, and a confusion matrix analysis. It has been experimentally demonstrated that the ensemble and probability models, particularly, the Rand Forest and Multinomial Naive Bayes, have higher detection performance when compared to the other types of classifiers. The findings confirm that the machine learning-based spam detection technologies have potential applications in the security surveillance and reduction of risks in large social media communication networks.

Keywords: Social Media Security, Suspicious Communication Detection, Big Data Analytics, Spam Detection, Machine Learning, Text Classification, TF-IDF, Natural Language Processing, Security Risk Analysis, Random Forest, Naive Bayes, Communication Filtering, Social Media Analytics.

How to cite this article: Khatri G, Dabhi VM, Patel M, Prajapati H. Security Risk Analysis and Detection of Suspicious Communications in Social Media Big Data Using Machine Learning. Int J Drug Deliv Technol. 2026;16(33s):109-115. DOI: 10.25258/ijddt.16.33s.14

Source of support: Nil.

Conflict of interest: The authors declare no conflict of interest.

Author for Correspondence: Gaurav Khatri

1. INTRODUCTION

The rapid development of social media platforms together with online communication systems has generated massive amounts of content that users create. The sites enable users to share information quickly while they connect with others but these platforms introduce serious security threats which include spam attacks and phishing schemes and distribution of false information and fraudulent content. Suspicious digital communication alerts which people send through digital platforms can result in privacy breaches which damage user trust and compromise the security of all systems.

Recent studies show that social media platforms need specialists who possess advanced analytical abilities for both their protection and operational management tasks. The source by Sharma and Jain (2020) highlighted the significance of sentiment analysis as one of the crucial components of the social media security. The system uses sentiment analysis together with text pattern analysis to detect unusual communication behavior which may indicate malicious activities. The system has demonstrated that it can track suspicious communication by using both user sentiment and text pattern analysis. The research shows that analyzing linguistic and emotional elements of messages enables security teams to detect suspicious activities while enhancing their defense systems for extensive communication networks.[1]

Conventional security measures, including filtering on a rule basis or matching keywords, are not always effective to deal with the changing and dynamic malevolent communication. These approaches are inadequate in terms of scalability and flexibility, particularly when they are used with huge volumes of communication data. Therefore, there is a need to have smart data-driven solutions to be able to analyze the content of communication and determine possible security risks.[2]

Detection of suspicious communication has received multiple areas of interest based on machine learning because. it is able to process extensive text data and learn sophisticated patterns by default. By identifying and analyzing the text, sentiment-conscious components, and event classifiers, machine learning models can effectively distinguish between authentic and suspicious messages. Based on this background, the given research is devoted to the analysis of security risks and identification of suspicious messages with the help of machine learning methods. The SMS Spam Collection dataset is used as an exemplary dataset of communication where spam messages are considered as a threat to security. The suggested solution should increase the security of communication and deliver insights that could be used in general social media analytical and security frameworks.[3]

1.1. Security Challenges in Digital and Social Media Communications

The rate of internet users who create content which connects to multiple networks has grown faster than any previous time, which creates security risks for digital and social media platforms. The platforms enable users to share information instantly, but they also create opportunities for dangerous activities such as spamming, phishing, identity theft, and unauthorized data access. The research conducted by Martinez and his colleagues (2024) demonstrates that the present-day network communication systems possess advanced technical features while showing extreme vulnerability to cyberattacks which complicates the maintenance of their protective systems.[4]

The traditional security systems which depend on rules and signature detection methods fail to provide adequate protection against evolving communication threats which develop new patterns of attack. The current methods prove insufficient because they lack adaptable capabilities required to handle substantial quantities of diverse communication information. The detection of suspicious communications needs to be solved because it represents a critical challenge for our systems. Smart security measures, which can expand according to need, will provide advanced protection through digital and social media platforms, because they can analyze communication content to assess potential security threats.[5]

1.2. Machine Learning Approaches for Suspicious Communication Detection

The use of machine learning has been on the increase in the process of detecting suspicious communications due to its ability to automatically develop a pattern of large-scale textual data. Machine learning models in contrast to traditional systems based on rules respond to changes in communication patterns and malicious intent. Sharif et al. (2020) proved that under supervision, machine learning algorithms are able to learn how to identify suspicious texts by examining linguistic features, a structure of a message and contextual patterns with a high degree of reliability in classification within various datasets.[6]

The traditional machine-learning methods for detecting suspicious communication use three types of algorithms which include probabilistic models and linear classifiers and ensemble-based methods. The algorithms use text representation techniques like bag-of-words and TF-IDF to convert unstructured text into useful numerical data. [7]

The machine learning classifiers can accurately identify between authentic messages and suspicious messages by

using the labeled communication data which serves as their training material. The machine learning detection systems provide essential support for improving both security measures and system performance in online communication environments and social media platforms. [8]

2. Literature Review

Prabhu Kavin et al. (2022) The researchers developed a machine learning-based data acquisition system. The system establishes security measures to identify fraudulent accounts within upcoming mobile communication networks. The combination of data protection systems with intelligent classification systems enables organizations to detect users who display suspicious behavior. Machine learning technology increases network communication security through two main benefits which include improved detection accuracy and system performance during different communication conditions.

Taha (2025) examines current progress and problems in big data analytics through an analysis of three areas which include Internet of Things technologies, social media platforms and natural language processing and information security. The article demonstrates how organizations can improve their security capabilities and decision-making processes through the use of scalable data-driven models which help them identify security threats in complex digital communication environments.

Chen and Smys (2020) suggested a hybrid social multi-media security and suspicious activity detention through deep learning in software-defined networking (SDN) systems. Their findings reveal that deep learning models can be used to detect abnormal and malicious activities in multimedia-enhanced social sites with great effectiveness and thus at the network level, they augment network-level security and threat detection.

Terumalasetti and Reeja (2024) investigated the application of predictive analytics to intercept malicious user activities on social networks. The paper identifies the growing amount of user-generated data to be a greater issue of privacy and security concern. Using machine learning and the behavioral patterns of the user, the authors prove that suspicious activities like spamming, phishing, or unauthorized access are possible to identify. Their solution focuses on active threats and real-time tracking, which are added benefits to improved data privacy and social networks experience.

Kumar, Bharati and Prakash (2021) An examination of the machine learning and deep learning algorithms used to safeguard the online social network. The research paper discusses the application of various algorithms. starting with traditional ML classifiers and progressing to more advanced deep learning models for detecting threats like spam, phishing, fraudulent accounts, and harmful content. The authors stress that deep learning methods mainly offer greater accuracy, especially with large and complex datasets, while machine learning models are effective with small or structured datasets. The review also highlights the role of adaptive and automated security to the social network in managing the changing threats.

Saranya et al. (2020). The efficacy of various machine learning-based intrusion detection systems (IDS) was compared Some of the techniques that have been tested in the study include decision trees, SVM, K-NN and ensemble methods in detecting network intrusions and they have been found to be accurate, efficient and scalable. The authors emphasize that selecting the appropriate algorithm relies on factors such as the type of attack, data set sizes, and available computing resources. They also suggest that hybrid and ensemble methods could yield better detection rates for recent and complex network threats.

Gahi and El Alaoui (2020) touched upon the issue of machine learning and deep learning paradigms to solve the issues concerning the analysis of big data, particularly the analysis of cybersecurity applications. The research demonstrates how high-volume and fast-moving and diverse data creates obstacles for data handling processes. The research demonstrates how machine learning and deep learning methods enable efficient processing of extensive data sets to identify suspicious activities and cybersecurity threats and dangerous actions. The authors conclude that deep learning models are suitable for complex and unstructured data, while traditional machine learning is applicable to structured and medium-size data. The researchers established that big data security requires development of solutions which can scale and adjust to different needs.

Alzaabi and Mehmood (2024) conducted a review of recent events on using machine learning methods to detect malicious insider threats. The paper explains the process of detecting insider threats that are hard to detect since they have legitimate access privileges by examining user behavior pattern, access logs and activity abnormalities. The issues that the authors raise are imbalance in data, change in attack strategies, and privacy issues, and hybrid and adaptive ML models are the promising solutions to successful detection.

3. Methodology

The methodological framework is described in this section used to identify and categorize SMS messages as spam or ham through machine learning.

3.1. Dataset Description

The dataset utilized in this paper is the SMS Spam collection, comprising 5,574 messages. Each message is classified as either ham (non-spam) or spam. The data exists as multiple records which contain two distinct columns.

- **Label** The label indicates whether the message qualifies as spam (spam) or not (ham).
- **Message**: The SMS and email message content.

The dataset provides a real world example which allows for the development of a spam detection system that researchers use as a standard testing material to evaluate text classification methods.

3.2. Data Cleaning

All data cleaning procedures began before modeling because researchers needed to establish which dataset elements were suitable for machine learning use :

- **Removal of irrelevant columns**: Only the label and message columns were retained.
- **Handling missing values**: The process of handling missing data involved deletion of all records that contained either null or empty message fields.
- **Label encoding:** The system decoded the categorical labels into numerical values for machine learning purposes which showed ham as 0 and spam as 1.

3.3 Exploratory Data Analysis (EDA)

To investigate the dataset's properties, Exploratory Data Analysis was conducted. This process encompassed several crucial stages:

- **Class distribution analysis**: The number of ham and spam messages was visualized using a bar chart to identify any class imbalance.
- **Message length analysis**: The length of messages was analyzed, and histograms were plotted to compare message lengths for spam vs. ham messages.
- **Word frequency visualization**: Word clouds were generated for spam and ham messages separately to identify common words in each category.

Exploratory data analysis (EDA) allowed researchers to identify patterns within the dataset. These patterns then guided their decisions about data preprocessing and which models to use.

3.4 Text Preprocessing

Preprocessing text is necessary so as to convert raw text into numerical data and utilizable by machine learning:

- **Lowercasing**: To ensure uniformity, all text messages were transformed to lowercase.
- **Stop words removal**: Common words like “the”, “is”, “and” that do not contribute to distinguishing spam from ham were removed.
- **TF-IDF as a method of vectorization:** Text message was translated to numerical feature vectors through Term Frequency-Inverse Document Frequency (TF-IDF) that is a measure of the significance of one word in each message in relation to all the data

3.5 Model Building

Five machine learning algorithms were selected for building spam detection models:

1. **Logistic Regression (LR):** A linear regression that is used in binary classification.
2. **Random Forest (RF):** This is an ensemble model, which builds a number of decision trees and averages their outputs.
3. **Decision Tree (DT)**: A model using a tree structure that divides data according to the values of features.
4. **K-Nearest Neighbors (KNN):**KNN is a distance-based classifier that classifies messages based on the most frequent type of the closest neighbors.
5. **Multinomial Naive Bayes (NB):**A probabilistic model which is highly applicable in text classification issues.

The data was separated into training (80) and testing (20) data. All the models were trained on the training data and evaluated on the test set.

3.6 Model Evaluation

The performances of each model were estimated by:

- **Accuracy:** This is the ratio of correctly predicted messages to the total number of messages.
- **Confusion Matrix:** Visual illustration of the true positives, true negatives, false positives and false negatives.
- **Classification Report:** The report gives accuracy, recall, F1-score, and support of each of the classes.

3.7 Summary

The analysis was conducted through the use of systematic steps, such as the data preparation, data cleaning, EDA, text preprocessing, model construction of five algorithms, and evaluation. The following steps will guarantee a solid and reusable spam detection method, which can be used in academic research and practice.

4. Results and Discussion

The Results section is a systematic research of the intended spam detection system and evaluates the functionality of five machine learning systems, which comprise of Logistic Regression, Random Forest,

Decision Tree, K-Nearest Neighbors and Multinomial Naive Bayes. This section compares performance of the model using the quantitative measures of performance namely accuracy, precision, recall, F1-score and the measures of the confusion matrix. It shows great tendencies that are disclosed based on the results of the experiment and the most efficient model and the credibility of a text-based classification in terms of TF-IDF features.

The work of this part is restricted to reporting and interpreting the empirical results but not details of the implementation, therefore, the objective evidence is given to confirm the effectiveness of the suggested approach.

4.1 Class Distribution

The analysis of the class distribution showed that ham messages are much larger, compared to spam messages, which signifies the class imbalance. This is a reality of the world messaging systems, and spam is a lesser part of the total messages.

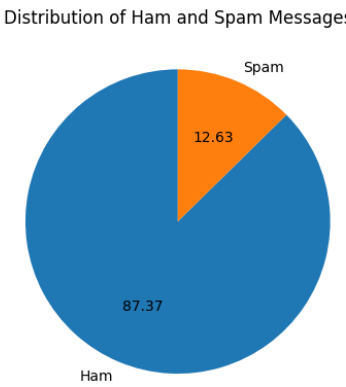

Figure 1. Distribution of Ham and Spam Messages

Despite this imbalance, appropriate text representation and model selection enabled effective spam detection.

4.2 Exploratory Data Analysis Results

Word Distribution Analysis

The word cloud visualizations produced different lexical patterns according to their respective distributions of words.

- The spam messages contained the words free and win and offer and call as their most common terms.
- The Ham messages used informal and conversational words as their main linguistic elements.

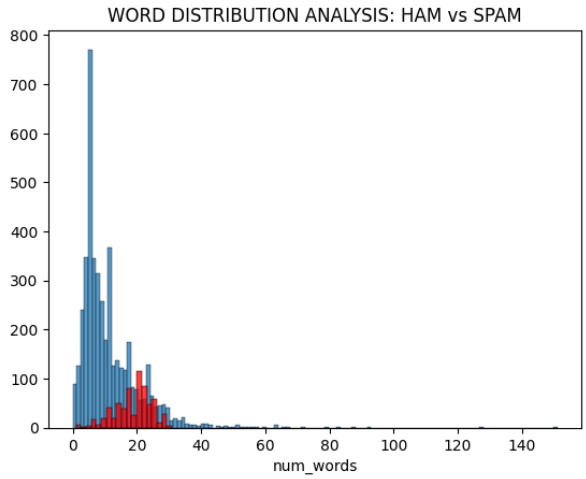

Figure 2. Word Distribution Analysis: Ham Vs Spam

These observations validate the suitability of TF-IDF vectorization for capturing discriminative word importance.

4.3 Model Performance Evaluation

4.3.1 Evaluated Machine Learning Models

The following five supervised machine learning algorithms were implemented and evaluated for spam detection:

- **Logistic Regression (LR)**: This is a linear classification model that is used when performing binary classification.
- **Random Forest (RF):** This is an ensemble learning approach which employs multiple decision trees to improve predictive stability.
- **Decision Tree (DT):** A classifier that organizes the data into a tree and obtains the decision rules depending on features.
- **K-Nearest Neighbors (KNN):** This is a distance-based classifier, whereby instances are labeled by the majority of the nearest neighbors.
- **Multinomial Naive Bayes (NB):** This is a probabilistic classifier that is well suited to a classification problem involving text.

4.3.2 Evaluation Metrics

The performance of the models was evaluated by using the following measures:

- **Accuracy:** Measures the proportion between the number of correctly boiled messages and the total messages.
- **Precision:** Precision measures the proportion of correctly identified spam messages out of all messages the model flagged as spam.
- **Recall:** Recall assesses how well the model correctly identifies genuine spam messages.

- **F1-Score:** The F1-Score, a balanced evaluation metric, is calculated as the harmonic mean of recall and precision.
- **Confusion Matrix:** his is a representation of a matrix that contains the true positives, true negatives, false positives, and false negatives.

- **4.4 Classification Metrics Comparative**

Table 1: Comparative performance of five machine learning models

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.9603	0.96	0.96	0.96
Random Forest	0.9807	0.98	0.98	0.98
Decision Tree	0.9526	0.95	0.95	0.95
K-Nearest Neighbors	0.9043	0.91	0.90	0.88
Multinomial Naive Bayes	0.9758	0.98	0.98	0.97

Table 1 compares the performance of five machine learning models that were used in spam detection. Random Forest obtained the maximum accuracy of 98.06% followed by Multinomial Naive Bayes and Logistic Regression.

Comparison of models Accuracies

According to the analysis of model accuracies, the Random Forest exhibited the highest accuracy, followed by the Multinomial Naive Bayes and Logistic Regression in second and third places, respectively. This suggests their applicability in spam detection. Conversely, the K-Nearest Neighbors algorithm demonstrated the lowest accuracy, highlighting its limitations when applied to high-dimensional text data.

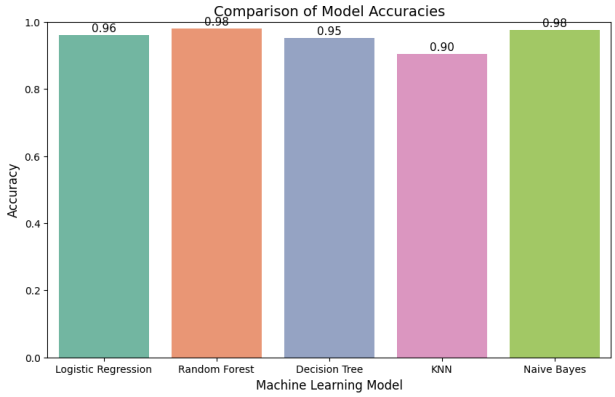

Figure 3: Comparison of Model Accuracies

5. Conclusion

This paper has taken a stringent assessment of five machine learning algorithms namely Logistic Regression, random forest, decision tree, k-nearest neighbors and multinomial naive bayes in addressing the issue at hand which is email/SMS spam detection. It was a systematic search, conducted through the aid of the data cleaning methodology, data exploration methodology, and text-preprocessing with the help of TF-IDF vectorization and massive testing of the model. The results of the given experiment confirm that the approaches, which are based on machine learning, are highly valuable when it comes to distinguishing between spam and regular mail even in the case of the imbalance between the two classes.

The overall accuracy of the models evaluated was greatest using the Random Forest, which used the ensemble learning strategy and was able to detect complex patterns in decision-making. Multinomial Naive Bayes and Logistic Regression also revealed high and regular performance that confirmed their appropriateness in text classification issues because of their effectiveness and strength. Contrarily, K-Nearest Neighbors demonstrated a relatively worse performance, mainly owing to the fact that the distance-based techniques are not effective in high-dimensional and sparse text feature space.

The comparative analysis reveals that in order to have a reliable spam detection model, the choice of model is important. Research findings of the present study can be utilized in building efficient and scalable spam filters systems in the real world setting. Further studies can be conducted in terms of using more sophisticated feature selection techniques, consideration of the changing spam patterns, and the deep learning structures in order to promote higher detection accuracy and versatility.

References:

1. Sharma, S., & Jain, A. (2020). Role of sentiment analysis in social media security and analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(5), e1366.

2. Prabhu Kavin, B., Karki, S., Hemalatha, S., Singh, D., Vijayalakshmi, R., Thangamani, M., ... & Adigo, A. G. (2022). Machine learning-based secure data acquisition for fake accounts detection in future mobile communication networks. *Wireless Communications and Mobile Computing*, *2022*(1), 6356152.
3. Martinez, A., Fitzroy, A., & Hogwart, A. (2024). Network communication security: Challenges and solutions in the digital era. *International Journal of Cyber and IT Service Management*, *4*(1), 47-52.
4. Hussien, A. Y. (2022). Review on social media and digital security. *Qubahan Academic Journal*, *2*(2), 1-4.
5. Sharif, O., Hoque, M. M., Kayes, A. S. M., Nowrozy, R., & Sarker, I. H. (2020). Detecting suspicious texts using machine learning techniques. *Applied Sciences*, *10*(18), 6527.
6. Verma, K. K., Singh, B. M., & Dixit, A. (2022). A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. *International Journal of Information Technology*, *14*(1), 397-410.
7. Bhadoria, R. S., Bhoj, N., Srivastav, M. K., Kumar, R., & Raman, B. (2023). A machine learning framework for security and privacy issues in building trust for social networking. *Cluster Computing*, *26*(6), 3907-3930.
8. Taha, K. (2025). Big Data Analytics in IoT, social media, NLP, and information security: trends, challenges, and applications. *Journal of Big Data*, *12*(1), 150.
9. Chen, J. I. Z., & Smys, S. (2020). Social multimedia security and suspicious activity detection in SDN using hybrid deep learning technique. *Journal of Information Technology*, *2*(02), 108-115.
10. Terumalasetti, S., & Reeja, S. R. (2024). Enhancing data privacy: Predictive analytics for detecting malicious user activity in social networks. In *Computer Science Engineering* (pp. 273-284). CRC Press.
11. Kumar, C., Bharati, T. S., & Prakash, S. (2021). Online social network security: A comparative review using machine learning and deep learning. *Neural Processing Letters*, *53*(1), 843-861.
12. Saranya, T., Sridevi, S., Deisy, C., Chung, T. D., & Khan, M. A. (2020). Performance analysis of machine learning algorithms in intrusion detection system: A review. *Procedia Computer Science*, *171*, 1251-1260.
13. Gahi, Y., & El Alaoui, I. (2020). Machine learning and deep learning models for big data issues. In *Machine Intelligence and Big Data Analytics for Cybersecurity Applications* (pp. 29-49). Cham: Springer International Publishing.
14. Alzaabi, F. R., & Mehmood, A. (2024). A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods. *IEEE Access*, *12*, 30907-30927
15. Bharti, N. S. G., & Gulia, P. (2023). Exploring machine learning techniques for fake profile detection in online social networks. *International Journal of Electrical and Computer Engineering (IJECE)*, *13*(3), 2962-2971.