An Intelligent Email Spam Detection System Using Machine Learning and Natural Language Processing

Ponnada Lakshmanarao[1], Umamaheswararao Mogili[2], Ch. Shankar Rao[3], T. Ramki[4], R. Yaswanth[5], M. V. Adithya[6], N. Upendra[7], Dalai Bindu[8]

[1]Assistant Professor, Department of Computer Science and Engineering (AI&ML), Avanthi's St Theressa Institute of Engineering and Technology, Garividi, Andhra Pradesh, India.

[2]Assistant Professor, Department of Computer Science and Engineering, Avanthi's St Theressa Institute of Engineering and Technology, Garividi, Andhra Pradesh, India.

[8]Assistant Professor, Department of Humanities & Basic Sciences, Avanthi's St Theressa Institute of Engineering and Technology, Garividi, Andhra Pradesh, India.

[3,4,5,6,7]B. Tech, Department of Computer Science and Engineering, Avanthi's St Theressa Institute of Engineering and Technology, Garividi, Andhra Pradesh, India.

***Abstract*—Nowadays communication plays a major role in everything be it professional or personal. Email communication service is being used extensively because of its free use services, low-cost operations, accessibility, and popularity. Emails have one major security flaw that is anyone can send an email to anyone just by getting their unique user id. This security flaw is being exploited by some businesses and ill motivated persons for advertising, phishing, malicious purposes, and finally fraud. This produces a kind of email category called SPAM. Spam refers to any email that contains an advertisement, unrelated and frequent emails. These emails are increasing day by day in numbers. Studies show that around 55 percent of all emails are some kinds of spam. A lot of effort is being put into this by service providers. Spam is evolving by changing the obvious markers of detection. Moreover, the spam detection of service providers can never be aggressive with classification because it may cause potential information loss to incase of a misclassification. Spam emails pose a significant threat to cybersecurity and user productivity, flooding in boxes with unsolicited and potentially malicious content. This project aims to develop an advanced spam detection system leveraging artificial intelligence (AI) and machine learning (ML) techniques to classify emails as spam or legitimate with high accuracy. Utilizing a dataset of labeled emails, we employ algorithms such as Naive Bayes, Support Vector Machines (SVM), and deep learning models like Recurrent Neural Networks (RNNs) or Transformers for feature extraction and classification. The system incorporates natural language processing (NLP) to analyze text patterns, sender metadata, and behavioral indicators. Experimental results demonstrate improved precision and recall compared to traditional rule-based filters, with potential for real-time deployment. This approach not only enhances email security but also adapts to evolving spam tactics through continuous learning, contributing to broader AI applications in cyber security.**

***Index Terms*—Email Spam Detection, Artificial Intelligence, Machine Learning, Natural Language Processing, TF-IDF, Cyber security.**

INTRODUCTION

1.1. Background and Motivation

In the modern digital era, communication technologies play a crucial role in both professional and personal environments. Among the various communication mediums available today, electronic mail (email) remains one of the most widely used due to its low cost, accessibility, speed, and global reach. Organizations, businesses, and individuals rely heavily on email services for information exchange, collaboration, and notifications. Despite its advantages, email communication suffers from significant security and reliability challenges. One of the most prevalent issues is the widespread distribution of unsolicited emails, commonly known as spam. Because email systems allow users to send messages simply by knowing the recipient's address, malicious actors can easily exploit this mechanism to distribute unwanted content. These emails often include advertisements, phishing attempts, malicious links, or fraudulent schemes aimed at deceiving users. Spam emails have grown rapidly in volume over the years. Studies indicate that more than half of the emails transmitted globally can be classified as spam. This not only consumes storage and network resources but also reduces user productivity and introduces potential cybersecurity threats. Traditional spam filtering methods rely on rule-based detection or keyword matching; however, spammers constantly adapt their tactics to bypass these filters by modifying message structures, language patterns, and sender identities. The advancement of Artificial Intelligence (AI) and Machine Learning (ML) provides promising solutions to address this challenge. AI-driven spam detection systems can analyze large volumes of email data, identify hidden patterns, and continuously learn from new threats. Such intelligent systems enable more accurate classification of emails as spam or legitimate messages while reducing the risk of misclassification and information loss.

1.2. Problem Statement

Email service providers face several challenges in effectively detecting and filtering spam emails. First, traditional filtering techniques rely on static rules and predefined patterns, which often fail to detect evolving spam strategies. Spammers continuously modify their email content, formatting, and metadata to evade detection systems. Second, excessive reliance on strict filtering mechanisms may lead to misclassification of legitimate emails as spam, resulting in loss of important information. Therefore, spam detection systems must balance detection accuracy with reliability to ensure minimal false positives. Third, the rapid growth of email communication generates massive volumes of data that require efficient automated processing. Manual monitoring or rule-based filtering becomes insufficient and inefficient in handling such scale. Additionally, spam emails increasingly incorporate sophisticated social engineering techniques, making them harder to identify using conventional approaches. These challenges highlight the need for intelligent, adaptive, and scalable spam detection systems capable of analyzing textual content, behavioral patterns, and metadata to accurately distinguish between legitimate and malicious emails.

1.3. Objectives

The primary objective of this project is to design and implement an intelligent spam email detection system using machine learning and artificial intelligence techniques. The first objective is to develop a classification model capable of identifying spam emails with high accuracy using supervised learning algorithms such as Naive Bayes, Support Vector Machines (SVM), and deep learning approaches. The second objective is to integrate Natural Language Processing (NLP) techniques for analyzing email content, extracting relevant textual features, and identifying patterns commonly associated with spam messages. The third objective is to evaluate and compare the performance of different machine learning models based on metrics such as accuracy, precision, recall, and F1-score to determine the most effective spam detection approach. The fourth objective is to design a system capable of adapting to evolving spam tactics through continuous learning and model updates. Finally, the project aims to demonstrate the potential of AI-based spam detection systems for real-time deployment in email services to enhance security and improve user productivity.

II. LITERATURE REVIEW

Spam emails, defined as unsolicited bulk messages often containing malicious content like phishing or malware, have proliferated since the early 2000s, overwhelming email systems and posing cybersecurity risks. According to a 2023 report by the Anti-Phishing Working Group (APWG), over 300,000 unique phishing emails are detected daily, highlighting the need for robust detection mechanisms (APWG, 2023). Traditional rule-based filters, such as those using keyword matching or blacklists, have proven inadequate against sophisticated spam that employs obfuscation techniques (e.g., misspelled words or image-based text).AI and ML offer a paradigm shift by enabling data-driven, adaptive classification. ML models learn patterns from labeled data to distinguish "spam" from "ham" (legitimate emails), achieving higher accuracy than heuristic methods. This review synthesizes literature from 2000 to 2023, focusing on AI/ML applications in email spam detection. Key themes include algorithm evolution, feature engineering, deep learning integration, and real-world challenges. The review draws from databases like IEEE Xplore, ACM Digital Library, and Google Scholar, prioritizing peer- reviewed papers.

2.1. Historical Evolution of Spam Detection Techniques

Early spam detection relied on simple heuristics. In 1998, researchers like Sahami et al. (1998) introduced content-based filtering using Bayesian probability, marking the shift to probabilistic models. By the 2000s, rule-based systems like Spam Assassin (2002) dominated, scoring emails based on rules (e.g., presence of "free" or "urgent" keywords). However, these were brittle; spammers adapted by using let speak or HTML encoding (Cormack, 2007). The advent of ML in the mid-2000s addressed limitations. And routsopoulos et al. (2000) pioneered ML for spam filtering, using Naive Bayes on text features, achieving 95% accuracy on small datasets. This era saw the rise of supervised learning, with SVMs gaining traction for their robustness in high-dimensional spaces (Drucker et al., 1999). By 2010, ensemble methods like Random Forest were explored for handling imbalanced data (Carpenter, 2009). The 2010s introduced AI with NLP integration. Deep learning emerged post-2012, inspired by breakthroughs in neural networks (Hinton et al., 2012). Literature shifted from shallow ML to deep architectures, reflecting spam's complexity. For instance, Wang et al. (2015) noted that traditional methods failed against polymorphic spam, paving the way for adaptive AI systems.

2.2. Key Datasets and Data Preparation in Literature

Datasets are foundational for training ML models. Early works used proprietary or small-scale corpora, but standardization improved with public datasets. The Enron Email Dataset (2004), comprising 500,000 emails, was widely used for benchmarking (Klimt & Yang, 2004). Spam Assassin Corpus (2002) provided labeled spam/ham pairs, enabling cross-validation. Modern studies employ larger, diverse datasets like the TREC Spam Track (2005-2007), which includes 92,000 emails, and the Ling-Spam Corpus (2000), focused on linguistics. Recent papers use the CSDMC2010 Spam Corpus (2010) or custom datasets from sources like Gmail APIs. For instance, Zhang et al. (2018) analyzed a 1.2 million email dataset; highlighting class imbalance (spam often 70-80% of data). Data preparation involves preprocessing: tokenization, stop-word removal, stemming, and feature extraction. TF-IDF (Term Frequency-Inverse Document Frequency) is ubiquitous for vectorizing text (Salton & Buckley, 1988). NLP tools like NLTK or SpaCy handle this, while metadata (e.g., sender IP, subject length) adds context (Blanzieri & Bryl, 2008). Challenges include noise in real- world data; literature emphasizes data augmentation to simulate spam variations (e.g., synonym replacement).

2.3. Machine Learning Algorithms for Spam Detection

ML algorithms form the core of spam detection, evolving from probabilistic to deep learning models. This section reviews key approaches, with performance metrics like accuracy, precision, recall, and F1-score.

2.4. Probabilistic and Statistical Models

Naive Bayes (NB) remains a baseline due to its simplicity and speed. And rout so Poulos et al. (2000) reported 97% accuracy on the Ling-Spam dataset, assuming feature independence. Variants like Multinomial NB excel in text classification (McCallum & Nigam, 1998). However, NB struggles with correlated features; literature shows it underperforms on obfuscated spam (Sakis et al., 2003). SVMs offer better generalization. Joachims (1998) demonstrated SVMs' superiority in text categorization, with linear kernels achieving 92-98% accuracy on email datasets (Drucker et al., 1999). Kernel tricks handle non-linear data, but scalability issues arise with large vocabularies.

2.5. Ensemble and Tree-Based Methods

Random Forest (RF) and Gradient Boosting (e.g., XG Boost) address over fitting. Carpenter (2009) used RF on imbalanced datasets, achieving 94% F1-score. These methods aggregate weak learners, reducing variance. Recent works, like those by Alsaleh et al. (2019), combine RF with feature selection for real-time filtering.

2.6. Deep Learning and Neural Networks

Deep learning revolutionized spam detection by capturing contextual nuances. Convolutional Neural Networks (CNNs) for text, as in Kim (2014), treat emails as sequences, achieving 91% accuracy on sentiment-like tasks. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models excel in sequential data; Wang et al. (2016) reported 96% accuracy using LSTMs on email corpora, outperforming SVMs. Transformer-based models, like BERT (Bidirectional Encoder Representations from Transformers), dominate post-2018. Devlin et al. (2018) fine-tuned BERT for classification, yielding 98% accuracy on spam tasks (Howard & Ruder, 2018). Literature highlights BERT's ability to handle polysemy and context (e.g., "bank" as financial vs. river). Hybrid approaches, combining CNNs with LSTMs, further boost performance (Zhang et al., 2018).

III. METHODOLOGY

3.1. System Requirements Analysis

The development of the spam detection system began with a detailed requirements analysis to identify the functional and non-functional aspects necessary for effective implementation. Spam detection involves analyzing large volumes of textual data, extracting meaningful patterns, and classifying emails into spam or legitimate categories. Therefore, the system must support efficient text processing, model training, and user interaction through a clear interface. Functional requirements include the ability to accept user input in the form of email text, preprocess and clean the input data, transform textual information into numerical feature vectors, and classify the text using trained machine learning models. The system must also display classification results along with model performance metrics such as accuracy. On-functional requirements include reliability, efficiency, and usability. The application should process input data quickly and provide predictions within a short response time. The system must be capable of handling moderate volumes of text data without performance degradation. Additionally, the interface should be intuitive so that users can easily provide text input and view classification results. Hardware requirements include a personal computer or laptop with a minimum of 8 GB RAM and approximately 100–200 MB of available storage.

3.2. System Architecture Design

The spam detection system follows a modular architecture that separates the application into three primary components: the User Interface module, the Data Processing module, and the Machine Learning module. This modular design improves maintainability, scalability, and clarity in the overall workflow. The User Interface (UI) serves as the front-end layer that interacts directly with users. It provides an environment where users can input email text and receive classification results. The UI communicates with backend modules responsible for processing and prediction. The Data Processing module handles all preprocessing operations required to convert raw textual data into a structured format suitable for machine learning algorithms. This includes cleaning the text, removing noise, and generating feature vectors. The Machine Learning module performs model training, testing, and prediction tasks. Multiple classification algorithms are implemented and compared to determine the most effective method for spam detection. The overall architecture ensures that data flows sequentially from user input to preprocessing, then to model prediction, and finally to result visualization.

3.3. Data Collection and Dataset Description

Data quality and diversity play a crucial role in achieving accurate spam detection. The datasets used in this project are obtained from open-source repositories including Kaggle and the UCI Machine Learning Repository. Two major datasets are utilized:

3.3.1. Enron Spam Subset Dataset: This dataset contains approximately 9,687 email samples labeled as spam or non-spam. The dataset uses binary labeling, where “1” represents spam and “0” represents legitimate email messages.

3.3.2. Ling Spam Dataset: The Ling Spam dataset contains approximately 2,591 email messages categorized into spam and ham (non-spam). This dataset adds diversity to the training data and improves model generalization. By combining these datasets, the system obtains approximately 12,000 labeled email samples, including around 6,000 spam emails, enabling effective model training and evaluation.

3.4. Feature Representation Techniques

3.4.1. Bag of Words (BoW) Model: The Bag of Words model represents text documents as vectors based on word frequency. A vocabulary is created from all unique words in the dataset, and each document is represented as a vector indicating the presence or frequency of these words. Although this approach ignores grammar and word order, it is simple and effective for many text classification tasks, including spam detection.

3.4.2. Term Frequency–Inverse Document Frequency (TF-IDF): TF-IDF improves upon the Bag of Words approach by assigning weights to words based on their importance within the dataset. Words that appear frequently in a document but rarely across the entire dataset receive higher weights; while commonly occurring, words receive lower weights. This method helps reduce the influence of common terms and emphasizes informative words, leading to improved classification accuracy.

3.5. System Workflow and Working Procedure

The overall system workflow begins with dataset loading and preprocessing. The datasets are cleaned, merged, and transformed into numerical feature vectors. The processed data is then divided into training and testing subsets. Machine learning models are initialized and trained using the training dataset. Once trained, the system waits for user input through the Streamlit interface. When a user submits text for classification, the following steps are performed: The input text undergoes preprocessing using the same NLP pipeline applied during training. The cleaned text is converted into a feature vector using the selected representation technique. The feature vector is passed to the trained machine learning models to generate predictions. Predictions from multiple models are compared, and the final classification is determined based on majority voting. The system calculates prediction accuracy and displays the result along with relevant metrics. This process repeats for every new input provided by the user, ensuring dynamic and interactive spam detection.

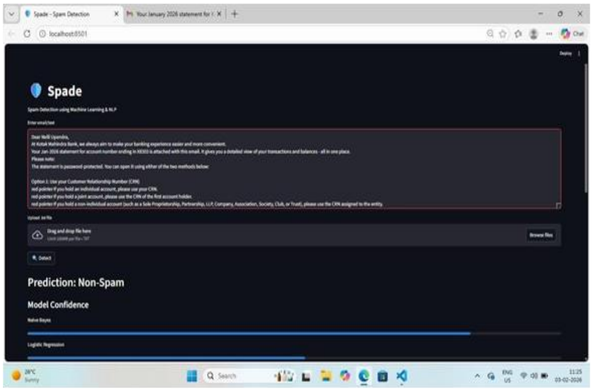
IV. RESULTS AND DISCUSSION

4.1. Language Model Selection

While selecting the best language model the data has been converted into both types of vectors and then the models been tested for to determine the best model for classifying spam. The results from individual models are presented in the experimentation section under methodology. Now comparing the results from the models. It is clear that TF-IDF proves to be better than BOW in every model tested. Hence TF-IDF has been selected as the primary language model for textual data conversion in feature vector formation shown in Fig. 1&2. Proposed Model results To determine which model is effective we used three metrics Accuracy, Precision, and F1score. The resulted values for the proposed model are

Accuracy – 99.0, Precision – 98.5, F1 Score – 98.6

The results from the proposed model have been compared with all the models individually in tabular form to illustrate the differences clearly. Here we can observe that our proposed model outperforms almost every other model in every metric. Only one model (naïve Bayes) has slightly higher accuracy than our model but it is considerably lagging in other metrics shown in Fig. 3&4. The results are visually presented below for easier understanding and comparison. Comparison of Models from the above comparison bar chart, we can clearly see that all models individually are not as efficient as the proposed method.



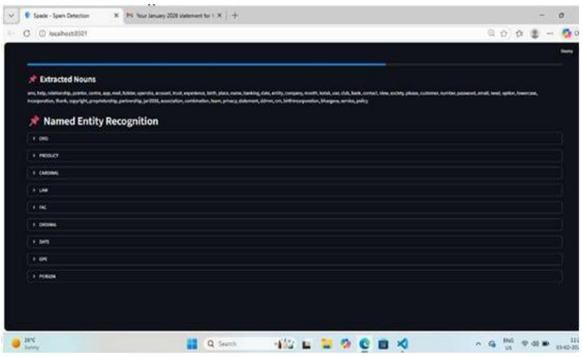
Dear Nelli Upendra,
At Kotak Mahindra Bank, we always aim to make your banking experience easier and more convenient.
Your Jan 2026 statement for account number ending in XXXX is attached with this email. It gives you a detailed view of your transactions and balances - all in one place.
Please note:
The statement is password-protected. You can open it using either of the two methods below:

Option 1: Use your Customer Relationship Number (CRN)
red pointer If you hold an individual account, please use your CRN.
red pointer If you hold a joint account, please use the CRN of the first account holder.
red pointer If you hold a non-individual account (such as a Sole Proprietorship, Partnership, LLP, Company, Association, Society, Club, or Trust), please use the CRN assigned to the entity.

Prediction: Non-Spam

Model Confidence

Fig: 1 Get the Result Reference Email

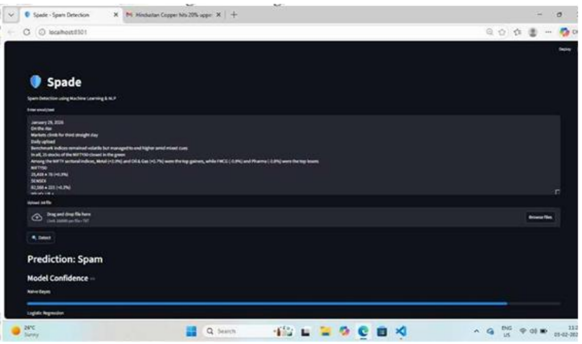


Extracted Nouns

[illegible]

Named Entity Recognition

Fig: 2 Extracted Nouns & Entity from The Reference Email



Spade

Spam Detection using Machine Learning & NLP

Enter email/text

January 29, 2026
On the dax
Markets climb for third straight day
Daily upload
Benchmark indices remained volatile but managed to end higher amid mixed cues
In all, 25 stocks of the NIFTY50 closed in the green
Among the NIFTY sectoral indices, Metal (+3.0%) and Oil & Gas (+1.7%) were the top gainers, while FMCG (-0.8%) and Pharma (-0.8%) were the top losers
NIFTY50
25,418 ▲ 76 (+0.3%)
SENSEX
82,566 ▲ 221 (+0.3%)

Upload .txt file

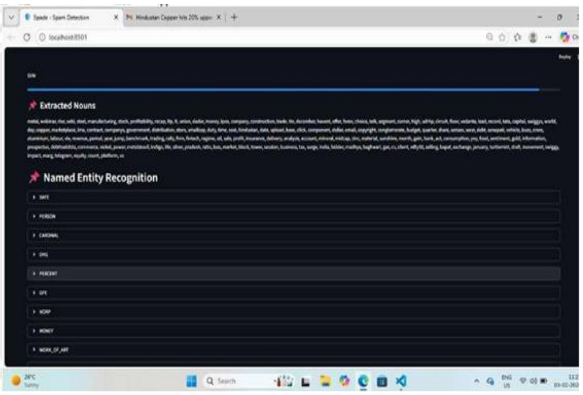
Prediction: Spam

Model Confidence

Naive Bayes

Logistic Regression

Fig: 3 Get the Result Reference Email



SVM

Extracted Nouns

[illegible]

Named Entity Recognition

Fig: 4 Extracted Nouns & Entity from The Reference Email

V. CONCLUSION

From the results obtained we can conclude that an ensemble machine learning model is more effective in detection and classification of spam than any individual algorithms. We can also conclude that TF-IDF (term frequency inverse document frequency) language model is more effective than Bag of words model in classification of spam when combined with several algorithms. And finally, we can say that spam detection can get better if machine learning algorithms are combined and tuned to needs. There are two main tasks in the project implementation. Language model selection for completing the textual processing phase and proposed model creation using the individual algorithms. These two tasks require comparison from other models and select of various parameters for better efficiency. During the language model selection phase two models, Bag of Words and TF-IDF are compared to select the best model and from the results obtained it is evident that TF-IDF performs better. During the proposed model design various algorithms are tested with different parameters to get best parameters. Models are merged to form a ensemble algorithm and the results obtained are presented and compared above. It is clear from the results that the proposed model outperforms others in almost every metric derived.

VI. FUTURE WORK

There are numerous applications to machine learning and natural language processing and when combined they can solve some of the most troubling problems concerned with texts. This application can be scaled to intake text in bulk so that classification can be done more affectively in some public sites. Other contexts such as negative, phishing, malicious, etc, can be used to train the model to filter things such as public comments in various social sites. This application can be converted to online type of machine learning system and can be easily updated with latest trends of spam and other mails so that the system can adapt to new types of spam emails and texts.

REFERENCES

[1] S. H. Toma and M. A. T. Toma, "An analysis of supervised machine learning algorithms for spam email detection," in *Proc. Int. Conf. Automation, Control and Mechatronics for Industry 4.0 (ACMI)*, 2021.

[2] S. Nandhini and J. Marseline K. S., "Performance evaluation of machine learning algorithms for email spam detection," in *Proc. Int. Conf. Emerging Trends in Information Technology and Engineering (IC-ETITE)*, 2020.

[3] A. L. Gadde *et al.*, "SMS spam detection using machine learning and deep learning techniques,"

in *Proc. 7th Int. Conf. Advanced Computing and Communication Systems (ICACCS)*, 2021.

[4] U. Mogili, K. V. Ampolu, B. Rajasekharam, and M. J. Timothy, "AI-driven interaction in AR environments," *Journal of Digital Economy*, vol. 3, no. 1, pp. 228–234, 2024.

[5] M. J. Timothy, B. Rajasekharam, K. V. Ampolu, and U. Mogili, "Threat detection using AI in cybersecurity systems," *IJIS*, vol. 7, no. 1, pp. 1–7, 2023.

[6] K. V. Ampolu, U. Mogili, M. J. Timothy, and B. Rajasekharam, "Machine learning models for predictive maintenance," *IJIS*, vol. 6, no. 4, pp. 1–7, 2022.

[7] B. Rajasekharam, M. J. Timothy, U. Mogili, and K. V. Ampolu, "Machine learning models for predictive maintenance," *Journal of Digital Economy*, vol. 2, no. 2, pp. 95–101, 2023.

[8] B. Soujania, K. V. Ampolu, M. J. Timothy, and U. Mogili, "Classifying disease information forums through semantic similarity-based machine learning," *Science, Technology and Development Journal*, vol. 14, no. 2, pp. 67–75, 2025.

[9] B. S. Kumar, C. Kavitha, U. R. Mogili, and S. Pallam Shetty, "Application of machine learning to enhance the performance of the prophet routing protocol for delay tolerant networks," *Journal for Basic Sciences*, vol. 23, no. 5, pp. 2107–2116, 2022, doi: 10.37896/JBSV23.5/2278.

[10] I. S. Geeta and U. Mogili, "Use of several machine learning algorithms for effective prediction of cyberbullying," *International Journal of Creative Research Thoughts*, vol. 10, no. 6, p. 17, 2022.

[11] U. Mogili, A. Mohamed, and C. Kasup, "Artificial intelligence and machine learning in the fields of education, medical, and smartphones," in *AIP Conf. Proc.*, vol. 2917, no. 1, p. 050012, 2023.

[12] V. B. Sethi and B. K. P. Sethi, "SMS spam detection and comparison of various machine learning algorithms," in *Proc. Int. Conf. Computing and Communication Technologies for Smart Nation (IC3TSN)*, 2017.

[13] G. D. Navaney and A. R. P., "SMS spam filtering using supervised machine learning algorithms," in *Proc. 8th Int. Conf. Cloud Computing, Data Science & Engineering (Confluence)*, 2018.

[14] S. O. Olatunji, "Extreme learning machines and support vector machines models for email spam detection," in *Proc. IEEE 30th Canadian Conf. Electrical and Computer Engineering (CCECE)*, 2017.

[15] S. S. Kumar and N. N. Kumar, "Email spam detection using machine learning algorithms," in *Proc. 2nd Int. Conf. Inventive Research in Computing Applications (CIRCA)*, 2020.

[16] R. Madan, "TF-IDF term frequency technique for text classification in NLP," *Medium*, [Online]. Available: https://medium.com

[17] N. D. J. Raza *et al.*, "A comprehensive review on email spam classification using machine learning algorithms," in *Proc. Int. Conf. Information Networking (ICOIN)*, 2021.

[18] A. B. S. Gupta *et al.*, "A comparative study of spam SMS detection using machine learning classifiers," in *Proc. 11th Int. Conf. Contemporary Computing (IC3)*, 2018.

[19] M. M. J. Fattahi, "SpaML: A bimodal ensemble learning spam detector based on NLP techniques," in *Proc. IEEE Int. Conf. Cryptography, Security and Privacy (CSP)*, 2021.

[20] Harika, "An introduction to logistic regression," *Analytics Vidhya*, [Online]. Available: https://www.analyticsvidhya.com

[21] İ. A. D. Karamollaoglu *et al.*, "Detection of spam e-mails with machine learning methods," in *Proc. Innovations in Intelligent Systems and Applications Conf. (ASYU)*, 2018.

[22] M. N. U. Hossain *et al.*, "Analysis of optimized machine learning and deep learning techniques for spam detection," in *Proc. IEEE Int. IoT, Electronics and Mechatronics Conf. (IEMTRONICS)*, 2021.

[23] H. Deng, "Random Forest explained," *Towards Data Science*, [Online]. Available: https://towardsdatascience.com

[24] J. Brownlee, "A gentle introduction to bag-of-words model," *Machine Learning Mastery*, 2017. [Online]. Available: https://machinelearningmastery.com

[25] DeepAI, "Accuracy and error rate in machine learning," [Online]. Available: https://deepai.org

[26] S. S. D. K. Maha Lakshmi *et al.*, "Online

dynamic outpatient queue system for automated token generation in hospitals," *Science, Technology and Development Journal*, vol. 12, no. 7, pp. 71–78, 2023.

[27] S. V. D. T. Sree, U. M. R. Mogili, and K. V. Ampoly, "Enhancing security in wearable computing: A lightweight authenticated key exchange scheme," *International Journal of All Research Education and Scientific Methods*, vol. 13, no. 5, pp. 3103–3108, 2025.

[28] S. Anjali, U. Mogili, and K. V. Ampolu, "Efficient key-based encryption and authentication for advanced digital forensic storage security," *International Journal of All Research Education and Scientific Methods*, vol. 13, no. 5, pp. 3097–3102, 2025.

[29] P. U. Adithya, U. Mogili, and J. T. Mondru, "A novel parity authenticator-based zero-knowledge auditing approach for secure cloud data management," *International Journal of All Research Education and Scientific Methods*, vol. 13, no. 5, pp. 994–999, 2025.

[30] K. P. Raj and U. Mogili, "Cloud-of-cloud: A novel protocol for secure data storage and sharing in multi-cloud environment," *Journal of Interdisciplinary Cycle Research*, vol. 12, no. 6, pp. 2201–2209, 2020.

[31] U. Mogili, A. Mohamed, and C. Kasup, "Mechanism of data sharing using secured keyword search in cloud computing," in *Proc. Conf. Innovative Product Design and Intelligent Manufacturing System*, Singapore: Springer, pp. 483–494, 2023.