Sparkling Light Publisher

Sparklinglight Transactions on Artificial Intelligence and Quantum Computing

journal homepage: https://sparklinglightpublisher.com/

Email Spam Classifier Using NLP And ML

Sharmila Mudgekar [a], Reesha P R [b], Chaitra Manjunath Naik [c], Vidyarani U A [d], Divya Naveen

[a] *Dept.of Master of Computer Applications, Shree Devi Institute of Technology, Kenjar 574142, India*
[b] *Dept.of Master of Computer Applications, Shree Devi Institute of Technology, Kenjar 574142, India*
[c] *Dept.of Master of Computer Applications, Shree Devi Institute of Technology, Kenjar 574142, India*
[d] *Asst Prof.Dept. of Master of Computer Applications, Shree Devi Institute of Technology, Kenjar 574142, India*

Abstract

Email has become the most largely used official communication mechanism for most of the internet users. In the past few years, the Email usage has increased. Also developed and increased the problems caused by spam Emails. spam or junk email is referred to the act of shared large unrequested messages. Thus, keeping this mind, it is main to build a complete system for spam classification based on semantics based text classification using NLP and ML. The performance of good emerging ML methods and algorithms is reviewed and assessed in this study. Spammers use developed and creative methods to bring their illegal activities using spam emails.

Keywords: Email Communication, Spam Detection, Junk Email, Text Classification, Semantics Web Text Classification.

1. Introduction

Emails are now the primary means of communication for people, companies, and governments in the contemporary, globalized twenty-first century. They make it possible for information to be shared quickly, professionally, and effectively across borders. Recent estimates indicate that the global email industry generated about 246 billion dollars in 2019.Daily email exchanges are expected to reach 320 billion by 2021, comprising 117.7 billion consumer emails and 128.8 billion corporate emails, as the volume of email traffic continues to increase at an exponential rate. These figures gives the necessity of email and the growing reliance on it for exchanging privacy data. Email traffic was predicted to reach 320 billion emails per day by 2021, and the industry was already valued at $246 billion by 2019.

E-mail address of authors: sharmilamudgekar753@gmail.com, reeshapr724@gmail.com, naikchaitra430@gmail.com, vidyas135@gmail.com

Please cite this article as: Sharmila Mudgekar, et al., Email Spam Classifier Using NLP And ML, Sparklight Transactions on Artificial Intelligence and Quantum Computing (2025), 5(2), 57-63. **ISSN (Online):2583-0732.** Received Date: 2025/07/16, Reviewed Date: 2025/07/27, Published Date: 2025/09/05.

A. Existing System

The existing email spam classifiers that use NLP and ML work by examine the content, structure, and metadata of To classify emails as either spam or legitimate (ham). Such systems usually begin with preprocessing steps such as tokenization, stop-word removal, stemming, and lemmatization to clean and prepare the email text. Features like word frequency, TF-IDF scores, presence of suspicious keywords, and metadata (sender address, subject line, links) are then extracted. Machine learning models such as Naive Bayes, Random Forests, or models like RNNs and LSTMs are trained on large labelled datasets of spam and non-spam emails. During classification, the model uses learned patterns to assign probabilities to incoming emails, filtering out spam with high accuracy. Modern spam detection systems often use feedback loops, where user actions—like marking messages as spam or safe—help the model learn and improve over time. By combining natural language processing for analyzing email content with machine learning for predictions, these systems remain efficient, scalable, and increasingly capable of countering new spam strategies.

B. Proposed System

A suggested approach for developing an email spam classifier using NLP and ML consists of several essential stages. The system is organized as a pipeline starting with data collection, where emails—including headers, subjects, and bodies—are gathered from benchmark datasets or provided by users. The collected text undergoes preprocessing to standardize it, which involves converting to lowercase, tokenizing, removing unnecessary characters, and replacing URLs, numbers, or email addresses with placeholders. This cleaning step ensures the data is ready for analysis. Next, numerical features are extracted from the text using methods such as TF–IDF vectorization or embeddings generated by pre-trained models. These features are then fed into a machine learning classifier, such as a finetuned transformer model, which learns to identify patterns that differentiate spam from non-spam messages.

2. Literature Review

To increase email security, researchers regularly use models. Research finds that using machine learning, NLP, and URL analysis is more successful than using these techniques alone[2]. URLs were similarly analyzed based on characteristics like the number of special characters, dots, and total length, which added to their ability to tell them apart when used with text[3] Techniques like NLP and feature extraction have been crucial in spam detection. For the classification of spam and fake reviews, methods such as unigrams, bigrams, and n-grams have been widely employed [4].

Random Tree showed the effectiveness of tree-based techniques for large-scale deployment by obtaining comparable accuracy while consuming less computing time. Evaluations tailored to individual datasets have confirmed findings about algorithm performance[5]. The recent literature review addressed the drawbacks of term-frequency-based models, which result in delayed training and significant computing demands[6]. A recent study improved the accuracy of spam categorization by applying cosine similarity functions to particular speech portions using lemmatization[7].A new study used ensemble techniques to offer an optimal approach for spam email identification and classification. The strategy included several classifiers, including Random Forest, SVM, and Decision Trees, to take advantage of their complementing benefits[8]. A recent study investigated into the classification of spam using email data. Using rule-based approach sender and route patterns were converted into individual values for analysis[9].

Researchers demonstrated that Naive Bayes classifiers are effective in email spam filtering because of their capacity to handle high-dimensional data and provide fast predictions with minimal resources. Early work demonstrated that Bayesian filtering could separate spam and legitimate emails with good precision and recall, making

it among the most practical approaches for real-world applications[10]. Subsequent studies compared different feature engineering methods, including word unigrams, bigrams, and character ngrams, showing that preprocessing choices strongly influence classification accuracy. By combining TF–IDF using algorithms for machine learning such as Support Vector Machines and Logistic Regression, these approaches achieved significant improvements in spam detection performance across benchmark datasets[11].

More recent studies have used deep learning and transformer-based architectures such as BERT and DistilBERT to spam classification. These models capture semantic meaning and context beyond surface-level features, strengthening them against obfuscation and adversarial spam. Although computationally more expensive, they outperform traditional models in detecting sophisticated and disguised spam messages[12]. More recent research has looked into ensemble learning methods and models for deep learning, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), which capture semantic relationships in text more effectively. Additionally, transformer-based architectures such as BERT have shown promising results in handling context-dependent spam detection[13].

A. Problem Description

A safe and user-friendly web platform is also suggested to make the solution feasible, enabling users to sign up, log in, and seamlessly manage their spam, sent mail, and inbox section. The system is designed to provide both technological stability and user convenience in practical applications by integrating efficient detection with an easy-to-use interface.

B. Motivation

Email communication has grown so quickly that is now vital tool for people, companies, and government institutions. however, they also attract a significant volume of spam, which can disrupt regular communication and diminish productivity. Emails make it possible to transmit information quickly, professionally, and effectively, but they also draw a lot of spam messages, this might prevent constant interaction and reduce efficiency. Additionally, spam emails provide significant risks to privacy and security, such as financial fraud, malware dissemination, phishing attempts, and identity theft. Blacklists and rule based filters are examples of spam detecting approaches that are frequently insufficient to manage complex spam schemes that are always evolving to get around these filters The necessity to create a good and flexible spam detecting system using techniques like ML and NLP is what motivates this project. The uses adaptive algorithms and semantic analysis to correctly identify intricate spam patterns, lower false positives, and improve the general security and dependability of email correspondence.

C. Proposed Method

This method utilize a hybrid strategy integrating rule-based filtering and ML to efficiently identify and manage spam emails. The system is developed as a web-based application utilizing Flask, with MySQL functioning as the backend database for secure storage of user data and email records. Users may register by submitting a secure password, a profile picture, and relevant personal information. SHA-256 hashes passwords prior to their storage in the database to ensure confidentiality. In addition to managing their account, registered users can compose emails, browse their mailbox, and log in. Methods of NLP are employed to preprocess both incoming and outgoing emails. The text is subjected to lemmatization, tokenization, removal of stopwords, elimination of special characters, and conversion to lowercase.

This guarantees that preproceesed emails can be taken to further for analysis. The system identifies spam through a hybrid methodology that integrates rule-based filtering, TFIDF vectorization, and a pre-trained machine learning

model. Regular expressions highlight common patterns like dubious keywords, URLs, or phone numbers, ensuring precise and effective spam detection, while the machine learning model predicts spam based on email content. Using TF-IDF vectorization, a pre-trained machine learning model determines whether the email is spam or legitimate based on its content. Common spam patterns, such as dubious keywords, URLs and phone numbers are identified by regular expressions. This guarantees that evident spam is detected right away.

3. Methodology

The proposed email spam classification system used the ML model accurately identifies spam emails. The methodology consisting following steps: A collection of spam and legitimate emails was collected from data sources. To eliminate noise, inconsistent data, and superfluous characteristics while utilizing preprocessing methods like stemming, tokenization, and stop-word removal. To find significant patterns and connections in the data, visualization is done and stop-word removal. To find significant patterns and connections in the data, visualization is done. Emails are cleaned by removing special characters, HTML tags, numbers, and punctuation.

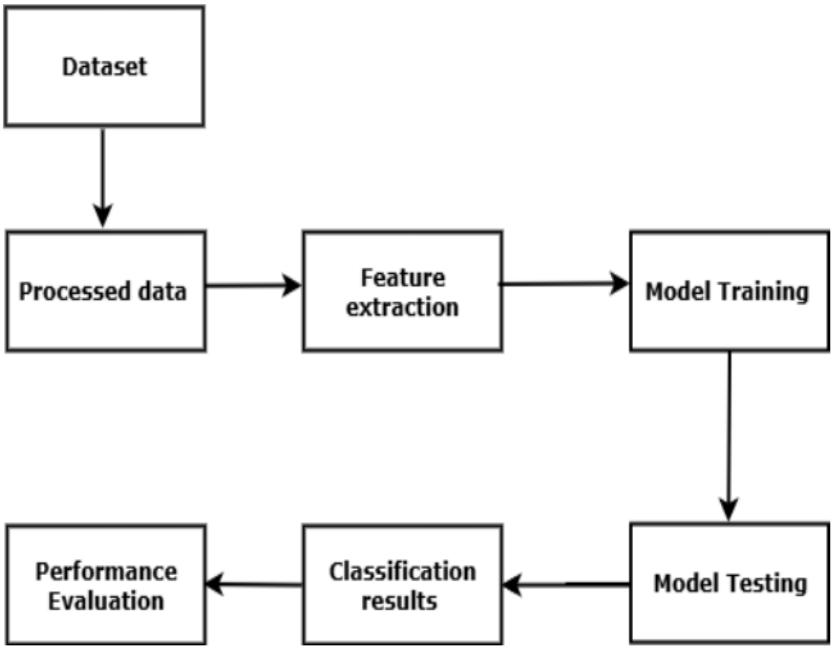


Fig. 1. Functional Block Diagram

Text is then tokenized, converted to lowercase, and stopwords are removed. Stemming or lemmatization is applied to lower words to their root forms, ensuring uniformity in the text data. To assess the model's capacity for generalization, splitting of data takes place, usually at a ratio like 80:20 or 70:30.to enable model training and evaluation on unseen data. Models such as Random Forest algorithms are used. After model selection, each model is trained and learning parameters are adjusted to improve classification performance. Models are trained on the preprocessed and vectorized email data, with hyper-parameters tuned using cross-validation techniques. The models were rated using functionality indicators.

ISSN (Online):2583-0732

Additionally, To show the right and wrong ways to categorize emails like spam as well as non-spam, confusion matrices were employed. In a simulated email environment, the classifier that offers the optimum balance between accuracy and computing efficiency is put into use to instantly identify emails as spam or non-spam(ham). The system can be continuously updated with new email data to adapt to evolving spam patterns. This methodology ensures a robust, adaptive, and reliable approach to spam email classification, balancing accuracy with usability in real-world applications.

4. Results

The Random Forest classifier showed extremely dependable performance with an all-inclusive accuracy of 99%. It attain a precision of 0.99 and a recall of 1.00 for the non-spam class and a precision of 1.00 and a recall of 0.99 for the spam class. The weighted and macro averages remained at 0.99, indicating that each classification experienced an equal distribution of the model's efficiency.

The metrics for evaluation are given below.

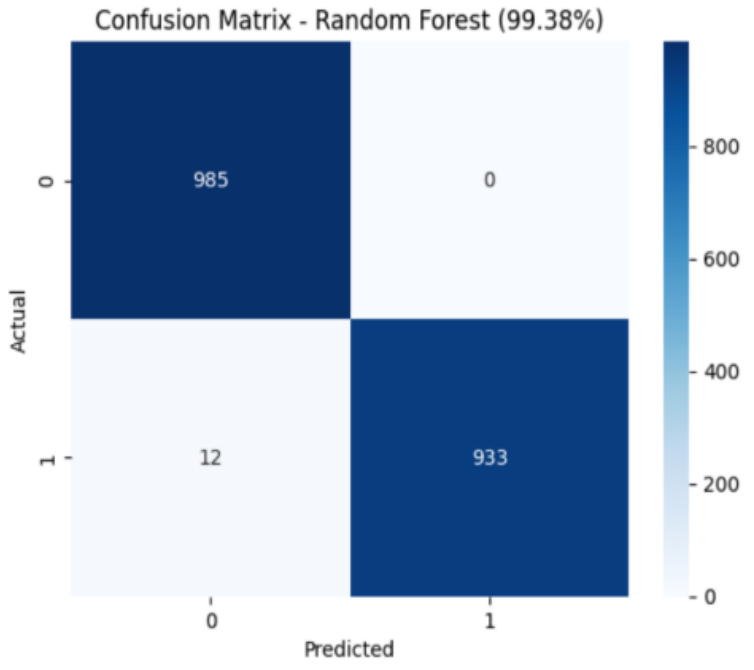
Accuracy - The proportion of all emails are correctly classified.
Precision - The percentage of emails that were truly categorized as spam.
Recall - The real spam emails that the model was able to identify.
F1-score - The ratio of recall to precision. Support - The number of emails used for each class's testing
Confusion Matrix - A table that displays each class's accurate and inaccurate predictions.

A. Random Forest

TABLE I
CLASSIFICATION REPORT

Class	Precision	Recall	F1-score	Support
0	0.99	1.00	0.99	985
1	1.00	0.99	0.99	945
Accuracy		0.99		1930
Macro Avg	0.99	0.99	0.99	1930
Weighted Avg	0.99	0.99	0.99	1930

B. Random Forest Confusion Matrix



In the above figure, Random forest model gained the accuracy of 99.38.It proves to be most reliable and effective model. The Random Forest model's high accuracy is demonstrated by the confusion matrix. The model correctly identified all 985 of the regular (non-spam) emails. It accurately identified 933 spam emails but ignored 12, incorrectly classifying them as non-spam. The model's overall accuracy of 99.38% indicates that it nearly always produces accurate predictions.

5. Discussion

This project is significant because it handles spam emails, which are currently one of the main problems in digital communication. Spam not only clogged inboxes but also presents threats such as malware diffusion, phishing attempts, and scams. The system works by identifying spam and isolating it from real emails with little assistance from humans. The suggested email spam classifier demonstrates how well ML and NLP can be combined to identify spam emails. Unstructured email content is transformed into useful features through tokenization, stopword removal, and stemming. Altogether, NLP and ML offer a strong, scalable, and flexible method of addressing the expanding issue of email spam, which qualifies it for practical implementation in individual and corporate email systems. From a practical perspective, email spam detection systems must be both accurate and lightweight for real-world deployment in mail servers.

False negatives (spam detected as legitimate) pose security threats, while false positives (legitimate emails flagged as spam) reduce usability. Thus, the trade-off between precision and recall must be carefully managed depending on application needs. Furthermore, as spammers continuously evolve their tactics, models require regular retraining with updated datasets. Incorporating deep learning techniques or transformer-based architectures may further enhance detection by capturing semantic and contextual features beyond word frequency.

6. Conclusion

This study combines ML and NLP techniques to give a reliable method for email spam detection. In order to correctly classify spam and legitimate messages, the suggested system efficiently preprocesses and converts email text into meaningful features. Random Forest is the most accurate and robust algorithm among those that were evaluated; in contrast, Decision Tree and Naive Bayes provide dependable performance in certain situations. The system can be implemented in web-based platforms for useful, real-time email management and shows flexibility in responding to changing spam tactics. This method overcomes the drawbacks of conventional rule based filters and provides a scalable response to the expanding problem of email spam by fusing security, usability, and accuracy. Future research might concentrate on dealing with highly obfuscated spam, multilingual emails, and using ensemble or deep learning techniques to improve classification performance even more.

This project demonstrates that NLP combined with ML techniques provides an effective solution for identifying and filtering spam emails. Through preprocessing and techniques for feature extraction like TF-IDF and Bag-of-Words, the classifiers successfully distinguished between spam and legitimate emails with promising accuracy. The results confirm that algorithms like Naive Bayes, Logistic Regression, and Support Vector Machines are well-suited for this task, though their performance depends heavily on dataset quality and balance. Overall, the study highlights the potential of intelligent, automated spam detection systems to enhance email security and user experience, while also emphasizing the need for frequent model upgrades in order to accommodate changing spam strategies.

Reference

[1] Gonzalez-Castro, V., Alaiz-Rodriguez, R., and Alegre, E., ”class distribution estimated based on the Hellinger distance,” *Inf. Sci*., vol. 218, pp. 146-164, 2013. :contentReference[oaicite:1]index=1

[2] Kuchipudi, B., Nannapaneni, R. T., and Liao, Q., ”Adversarial machine learning for spam filters,” in *Proc. 15th Int. Conf. on Availability,:contentReference[oaicite:2]index=2

[3] Labonne, M. and Moran, S., ”Spam-T5: Benchmarking Large Language Models for Few-Show Email Detection,” arXiv preprint arXiv:2304.01238, Apr. 2023.:contentReference[oaicite:4]index=4

[4] shirvani , G. and Ghasemshirazi, S., :Advancing Email Spam Detection:Leveraging Zero-Shot Learning and Large Language Models,” arXiv preprint arXiv:2505.02362, May 2025. :contentReference[oaicite:6]index=6

[5] Bhowmick, A. B. and Hazarika, S. M., ”Machine learning for email spam filtering:Review, techniques and trends,”arXiv preprint arXiv:1606.01042,Jun. 2016. :contentReference[oaicite:7]index=7

[6] Roy, S., and Paul, l.(2020) Email Spam Detection using hybrid machine learning approach.International Journal of Computer Application,176(29),10-14.

[7] Metsis,V.,androutsopoulos,l.,and Paliouras,G(2006.) Spam filtering with Naive Bayes-Which Naive Bayes?.CEAS 2006:Third Conference on Email and Anti-Spam.

****

ISSN (Online):2583-0732