Email Spam Detection Using Machine Learning

C. Venkata Swamy[1], Nagineni Sreehari[2], Shaik Moulali[2], Mulla Mohammed Saleem[2], Shaik Showkath Ali[2] and S. Nasir Hussain[2]

[1]*Department of Computer Science and (AI-ML), Santhiram Engineering College, Nandyal-518501, Andhra Pradesh, India*
[2]*Department of Computer Science and Design, Santhiram Engineering College, Nandyal-518501, Andhra Pradesh, India*
{venkata.chinna, harisrihari758, moulalishaikmj, saleemmohamed0786, shaikshowkathali7777, nasirhussainsunkari8688}@gmail.com

Keywords: Spam-Detection, ML, Natural Language Processor, DL, Classification Models, Email Security, Feature Extraction.

Abstract: Although email is an integral element of daily communication, increasing amounts of spam email not only serves as a potential threat to security but also harms efficiency. Traditional spam filters can't keep pace with new spamming techniques, making them ineffective over time. Instead of relying on fixed heuristics, machine learning provides a more adaptive solution to spam detection. We use linear classifiers, such as Naïve Bayes and Support Vector Machines (SVM), demographic models like Decision Trees and deep-learning models to achieve better spam detection accuracy in this work. We can apply various natural language processing (NLP) methods like also Tokenization, stop word removal, TF-IDF and word embedding, to improve the model's understanding capacity of the email content. We demonstrate through experimentation, that the performance of machine learning based spam filters reduces false positives and achieves higher classification accuracy. Since these models learn from data again, they can quickly adapt to the evolving spam techniques and therefore, are a reliable solution in modern email security. Further developments can supplement real-time learning, hybrid models, and deep learning to significantly improve email spam detection systems.

1 INTRODUCTION

Email spam, or junk mail, is the use of messaging systems to send unsolicited messages (spam), usually advertising for some product, service, or other activity. These types of mails have become increasing numbers on the internet in the past 10 years and now have become a notable nuisance on the internet Spam emails take up storage space and waste time, they delay the delivery of messages. Despite automatic email filtering being one of the most effective spam detection methods available, spammers have come up with many smart ways to avoid such filtering systems. In the past, spam emails were mostly filtered out by blacklisting specific email addresses. But as spammers generate new email domains, this method has become increasingly ineffective. Spam detection has gained new interest with the development of various ml techniques. There are several popular methods for spam filtering are text analysis, using blacklists and whitelists based on domain names, and network-based methods.

1.1 Motivation

The study of ml techniques in email spam-detection is the growing demand of correct and fast filter. The hybrid framework of advanced classifiers with rule-based filtering not only improves email filtering performance and accuracy but also enhances the user's email experience. This combination thus provides for a highly flexible, accurate and intelligent detection approach, successfully limiting false alarms while improving spam finding. This technique helps to make the email communication landscape more secure and trustworthy by addressing the changing landscape of spam methods.

1.2 Objectives

An approach to build a multi spam-detection system which utilizes both traditional rule-based filtering and ml classifiers (SVM, ANN, and XG Boost).
Step1: to use human language processing techniques by making the email preprocessing [data cleaning and feature extraction].

Swamy, C. V., Sreehari, N., Moulali, S., Saleem, M. M., Ali, S. S. and Hussain, S. N.
Email Spam Detection Using Machine Learning.
DOI: 10.5220/0013873400004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 1*, pages 807-811
ISBN: 978-989-758-777-1

Learn from data until October 2023, To Investigate model's ability to adapt to changing threats and discover factors for improvement such as transformer-based deep learning model and real-time learning.

2 LITERATURE REVIEW

In this section discussed the progress of the technology of email spam-detection, representing both the conventional and current ways.

2.1 Traditional Spam Filtering Methods

One of the most common methods was Bayesian filtering, a probabilistic method that determines the probability of an email being spam based on word frequency (Androutsopoulos et al., 2000). Although these approaches worked well in punishing spam above based purely on the presence of specific keywords, they struggled to keep up with changing tactics spammer used (for example, obfuscation techniques, or adversarial attacks).

2.2 ML in Spam-Detections

Machine learning spam-detection ushered in a new era by allowing models to learn from patterns in the data, rather than hard-coding rules. Classic machine learning models like SVM, Decision Trees and Naïve Bayes have been used to classify spam and ham based on the features extracted from the data such as word frequency, metadata and header 2. Despite having shown improved performance over the previous rule-based approaches, these models suffer from evolving spam types and the scale of these email datasets.

2.3 Deep Learning and Advanced Techniques

Spam Detection improves with the recent advancement in deep learning Neural-network has been used for analysing complex patterns in the text and metadata of emails, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) (W. S. Yerazunis, 2004). More recent transformer-based models like BERT, were designed to enhance context awareness and better identify spam emails (J. Goodman, 2005). In addition, ensemble methods such as Random Forests and XG Boost have been employed to improve classification performance through the aggregation of several model outputs (H. Drucker et al., 1999).

2.4 Gaps in Existing Research

Despite getting much better, existing spam detection systems still face some obstacles:

- Challenges to identify advanced phishing or being part of adversarial attacks that send spam
- False-positive rates too high, causing legitimate emails to be classified as spam.
- Scalability issues in processing large and changing email datasets.

A hybrid spam detection adaptability and accuracy this review on the literature

3 METHODOLOGY

In the methodology section, we describe the system design, the data processing pipeline, the integration of machine learning and evaluation approach.

3.1 System Architecture

The proposed system is composed of four major modules:

3.1.1 Email Data Collection

A data collection module gathers email data from publicly available datasets (such as Enron Spam Datasets) and real-world email traffic. The dataset comprises two spams and legitimate (hams) emails, ensuring a balanced and diverse corpus.

3.1.2 Preprocessing and Feature Engineering

After data collection, preprocessing is performed to extract meaningful features:

- Tokenization and Stop-word Removal: The email text is broken into tokens, and unique words (e.g., "and") removed enhance relevant content extraction.
- Stemmings and Lemmatizations: Words are reduced to their root forms for text normalization.

- Metadata Analysis: Additional features, such as sender reputation, frequency of links, and email structure, are extracted.

3.1.3 Classification Engine

This module interprets email contents and metadata to classify them as spam or legitimate. Various machine learning models are applied to improve classification accuracy.

3.1.4 Spam Filtering Module:

Detected spam emails are flagged and either moved to the spam folder or discarded. The system continuously learns from new emails to enhance detection performance. Figure 1 shows System Architecture for Machine Learning Based -Spam Detection.

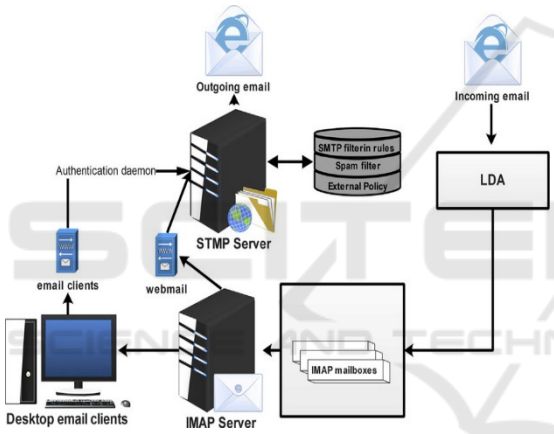

Figure 1: System architecture for machine learning based - spam detection.

3.2 Machine Learning Integration

The hybrid spam detection model employs three primary ml technique:

- Support Vector Machines (SVM): It is SVM is utilized to classify emails based on textual features. Use non-linears kernel (e.g. polynomial), it handles complex decision boundaries, improving spam classification performance.
- Artificials Neural Networks (ANN): A multi-layers perceptrons (MLP) is implemented to the detect intricate patterns in emails:

An input layer representing extracted features.

- One or more hidden layers capturing relationships within the data.
- An output layer classifying emails as spam or ham.

- XG Boost: XG Boost efficiently captures feature interactions and mitigates overfitting. The final classification score is computed as:

$$Ci = \alpha*SVMi + \beta*ANNi + \gamma*XGBoosting \quad (1)$$

where Ci is the final classification score, and α, β, and γ are weight parameters optimized during training.

where R_i is the final ranking score, PR_i is the PageRank score, ML_i is the machine learning output, and α and β are weight parameters optimized during training. Figure 2 shows DF Diagram.

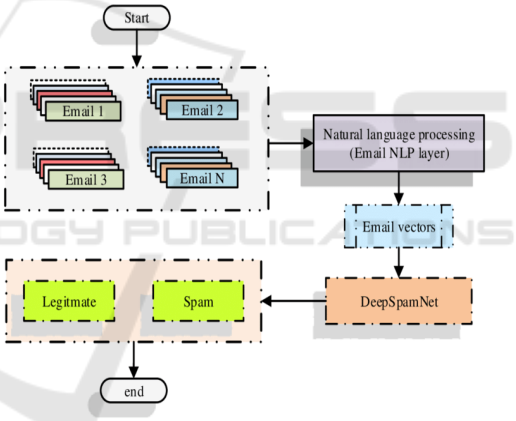

Figure 2: DF diagram.

3.3 Data Collection and Preprocessing

The dataset is curated through publicly available sources and real-time email monitoring:

- Dataset Composition: Over 100,000 emails are collected, consisting of both spams and non-spams emails across various domains.
- Pre-processing Pipeline:
  - Data Cleaning: Removal of HTML tags, special characters, and non-text elements.
  - Text Normalization: Tokenization, stemmings, and lemmatizations.

- FEATURES Vectorization: Conversion of text to numericals vectors using TF-IDF and word embeddings.

3.4 System Evaluation and Testing

The system is evaluated using both quantitative and qualitative measures:

- Quantitative Metrics:
  - Confusion Matrix Analysis: Evaluates false positive and false negative rates.
- Qualitative User Feedback:
  - User Surveys: Gather user feedback on spam detection accuracy and usability.
  - Spam Reduction Impact: Measures the reduction of spam emails in user inboxes over time.
- Regression Analysis: Multiple regression models (using tools like IBM SPSS) determine the influence of various features on classification accuracy.

4 RESULT AND ANALYSIS

This section details the experimental findings from implementing the hybrid search engine.

4.1 Qualitative Evaluation

- User Feedback:
  Through a study with 200 email customers, this machine learning-based spam filter dramatically improves the accuracy of email filtering. Users saw less spam in their inboxes and were able to identify unwanted messages.
- Usability Testing:
  Consensus from focus group discussions indicates that the spam filtering system is intuitive, non-intrusive, and blends seamlessly into the email platforms. The performance of the automated classification was found to be satisfactory by the users, and thus reducing the need for manual intervention. They also observed how the system automatically adopts to new spam trends, retaining high accuracy over time.

4.2 Comparative Analysis

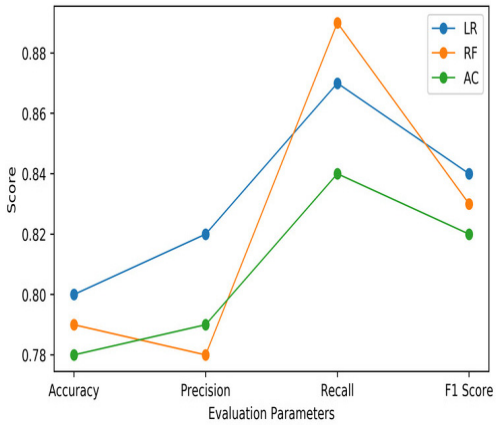

Figure 3: Comparative performance based on key evaluation metrics.

Figure 3 of the Comparative analysis the graph show is Comparative Performance Based on Key Evaluation Metrics.

5 DISCUSSION

5.1 Adaptive Learning and Relevance

This enables the system to adjust to new patterns of spam and changing underlying structures of emails. Such models learn from new data and iterate on the classification, thus increasing accuracy as more data is captured over time. Training reported higher precisions, recalls, and F1-scores metrics, thus enhanced system performance, decreasing false positive and false negative and allowing more accurate and trustful spam detection.

5.2 User Experience and Security:

The system leads to increased email security as it detects spam and minimizes exposure to phishing emails. Increased trust isn't just good for you, it helps to build a better user experience by ensuring that users recognize genuine promotional emails from harmful spam.

5.3 Limitations

High computational cost of training deep learning models and tendencies of training data for bias that

might have an effect on generalization are some of the challenges. Moreover, the real time filtering would add latencies and need optimizations for a production scale deployment such as in enterprise email services.

6 CONCLUSIONS

Machine learning has proven to be an effective approach for detecting and filtering spam emails, significantly improving classification accuracy compared to traditional rule-based methods. By utilizing advanced algorithms such as Naïve Bayes, Support Vector Machines (SVM), Decision Trees, Random Forest, and deep learning models, spam detection systems can efficiently distinguish between spam and legitimate emails.

REFERENCES

H. Drucker, D. Wu, and V. N. Vapnik, “Support vector machines for spam categorization,” IEEE Trans. Neural Networks, vol. 10, no. 5, pp. 1048–1054, Sep. 1999.

I. Androutsopoulos, G. Paliouras, V. Karkaletsis, G. Sakkis, C. D. Spyropoulos, and P. Stamatopoulos, “Learning to filter spam e-mail: A comparison of a Naive Bayesian and a memory-based approach,” Intelligent Systems, vol. 37, no. 4, pp. 415–429, 2000.

J. Goodman, “Spam filtering: Bayesian and beyond,” in Proc. 2nd Conf. Email Anti-Spam (CEAS), 2005, pp. 1–9.

S. Islam, T. K. Ghosh, and M. S. Rahman, “A deep learning-based approach for email spam classification using hybrid CNN-LSTM model,” in Proc. IEEE Int. Conf. Signal Process. Inf. Comput. Appl. (SPICA), 2021, pp. 203–208.

T. A. Almeida, J. M. Gómez Hidalgo, and A. Yamakami, “Spam filtering: How the dimensionality reduction affects the accuracy of Naïve Bayes classifiers,” Journal of Information Processing & Management, vol. 47, no. 5, pp. 654–664, 2011.

W. S. Yerazunis, “Sparse binary polynomial hashing for spam filtering,” in Proc. ACM Conf. Email Anti-Spam (CEAS), Mountain View, CA, USA, 2004, pp. 1–8.