Open Access
Online ISSN: 2992-5487
Print ISSN: 2992-6068

NIJEP
Publications

NEWPORT INTERNATIONAL JOURNAL OF ENGINEERING AND PHYSICAL SCIENCES (NIJEP)

Volume 5 Issue 2 Page 1-10, 2025

https://doi.org/10.59298/NIJEP/2025/5211000

Development of an Efficient Machine learning-based Email spam Detection System

[1]Akawuku, I. Godspower , [2]Adejumo, Samuel Olujimi, [3]Alade, Samuel Mayowa, [4]Olatunde, Ayodeji Akano, [5]David and Mulumeoderhwa Bahati

[1]Department of Software Engineering, Nnamdi Azikiwe University, Awka, Nigeria
[2]Departments of Cybersecurity, Nnamdi Azikiwe University, Awka, Nigeria
[3]Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria
[4]Department of Computer Sciences, Abiola Ajimobi Technical University, Ibadan, Nigeria
[5]Department of Computer Science, Olivia University, Bujumbura, Burundi.
Email: gi.akawuku@unizik.edu.ng

ABSTRACT

Email spam remains a persistent threat to digital communication, with estimates indicating that over 50% of daily global email traffic is spam. To address this issue, this paper presents the development of an accurate and efficient email spam detection system using machine learning techniques, guided by the Object-Oriented Analysis and Design (OOAD) methodology. The integration of OOAD allowed for a systematic, modular, and scalable approach to system design, facilitating clear separation of concerns, reusability, and maintainability. The system architecture was modeled using Unified Modeling Language (UML) diagrams to define key components such as data preprocessing, feature extraction, model training, classification, and user interaction. The spam detection engine was developed using supervised machine learning algorithms including Naive Bayes, Support Vector Machines (SVM), and Random Forest, trained on a large, labeled dataset of spam and non-spam emails. Feature engineering incorporated natural language processing (NLP) techniques to capture the textual patterns characteristic of spam content. Performance evaluation demonstrated that the system achieved high accuracy (over 98%) and strong precision-recall balance, making it suitable for real-time applications. By combining OOAD methodology with robust machine learning models, this study offers a structured and efficient solution to email spam detection. The resulting system is not only accurate and scalable but also maintainable and extensible for future enhancements. This work underscores the importance of combining sound software engineering principles with intelligent algorithms to combat the evolving landscape of email-based threats.

Keywords: Machine learning-based, Email spam, Detection System, Object-Oriented Analysis and Design (OOAD) methodology.

INTRODUCTION

Email has become an essential means of communication in today's digital age, with billions of people relying on it for personal and professional purposes. However, the increasing volume of unwanted emails, commonly known as spam, has become a significant problem [1]. Spam emails not only waste users' time but also pose a significant threat to the security of individuals and organizations. They can contain malware, phishing links, and other types of malicious content, which can compromise sensitive information and lead to financial losses. The proliferation of spam emails has led to a significant decrease in the productivity of individuals and organizations, as well as a loss of trust in email as a means of communication. According to a report by the Radicati Group, the total number of business and consumer emails sent per day is expected to reach 319.6 billion by the end of 2021, with spam emails accounting for a significant proportion of this total. Traditional rule-based approaches to spam detection, which rely on predefined rules and filters, have proven to be ineffective in combating the growing menace of spam emails. These approaches are often easily evaded by spammers, who continually evolve their tactics to bypass these rules. As a result, there is a growing need for more sophisticated and adaptive approaches to spam detection. Machine learning, a subset of artificial intelligence, has emerged as a promising solution to the problem of spam detection. Machine learning algorithms can be trained on large datasets of labeled emails to learn the patterns and characteristics of spam emails, enabling them to make accurate predictions about the classification of new, unseen emails. Several machine learning algorithms, like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees, Random Forests, and Neural Networks, like Convolutional Neural Network (CNN), Artificial Neural Network (ANN) have been applied to the problem of spam detection with varying degrees of success, being that it's a classification problem (spam or non-spam). This project aims to explore the application of machine learning algorithms to classify emails as spam or non-spam. The goal is to develop an accurate and efficient email spam detector that can help reduce the burden of unwanted emails and improve the overall security and productivity of individuals and organizations. The project will involve the collection and pre-processing of a large dataset of emails, the training and testing of various machine learning algorithms, and the evaluation of their performance using metrics such as accuracy, precision, recall, confusion matrix (true positive and false negative) and F1-score. The results of this project are expected to contribute to the development of more effective spam detection systems and improve the overall quality of email communication.

LITERATURE REVIEW

Email spam filtering has become a crucial area of research due to the increasing volume of unwanted emails and their potential threats to personal and organizational security. This section reviews recent studies on spam detection methodologies, highlighting the advancements in machine learning, NLP techniques, and hybrid approaches. [2], conducted an extensive survey on spam classification techniques, focusing primarily on the use of support vector machines (SVM). The study emphasizes the effectiveness of SVM in distinguishing spam from legitimate emails, leveraging its ability to create hyperplanes that optimally separate classes in high-dimensional space. The research highlights the importance of various feature extraction techniques, particularly term frequency-inverse document frequency. [10] (TF-IDF) and n-grams, which significantly enhance classifier performance in email filtering. This foundational work underscores the critical role of traditional machine learning methods in establishing baseline benchmarks for spam detection accuracy. [3], further advanced spam detection research by exploring the application of n-grams combined with SVM classifiers. Their experimental findings reveal that the integration of n-grams as features leads to improved representation of email content, ultimately resulting in higher classification accuracy. By comparing the performance of various n-gram sizes, the authors illustrate how larger n-grams capture contextual information more effectively, thereby enhancing the model's ability to identify spam patterns. This study serves as a crucial reference for subsequent research that seeks to refine feature selection and extraction processes in spam classification systems. In a significant advancement, [4], proposed a hybrid model that integrates convolutional neural networks (CNN) and long short-term memory (LSTM) networks for spam classification. Their innovative approach capitalizes on CNN's strength in identifying spatial hierarchies and LSTM's capability in handling sequential data, making it particularly effective for email content analysis. The results of their study demonstrate that this hybrid model outperformed traditional methods, effectively capturing complex patterns inherent in email data. The authors provide extensive performance metrics, establishing benchmarks for future research and applications in the domain of spam detection. [5], explored the utilization of BERT (Bidirectional Encoder Representations from Transformers) for email spam detection, emphasizing its contextual understanding of language. This study marks a pivotal moment in spam detection, as it illustrates that BERT's ability to process words in relation to all the other words in a sentence significantly enhances classification performance. Their research findings indicate that BERT not only outperforms traditional models but also reduces the necessity for extensive feature engineering, providing a more streamlined and effective approach to spam detection. This work highlights the shift towards deep learning techniques in natural language processing (NLP) applications within spam filtering. The research conducted by [6], delves into the significance of feature selection techniques in improving SVM performance for spam detection. The authors demonstrate how effective feature selection can lead to reduced computational costs while simultaneously enhancing classifier accuracy. By employing methods such as

recursive feature elimination and feature importance ranking, they illustrate how identifying and focusing on the most relevant features in email content can lead to significant improvements in spam detection rates. This study reinforces the idea that proper preprocessing and feature selection are crucial for developing robust spam classification systems. [6], introduce an innovative angle by highlighting the role of sentiment analysis in spam detection. Their research emphasizes how understanding the emotional tone of email content can provide valuable insights into potential spam characteristics. By analyzing sentiment, classifiers can gain an additional dimension of information that may aid in distinguishing spam from legitimate messages. This study opens up new avenues for integrating psychological aspects of language into spam detection algorithms, encouraging future research to consider the emotional context of email communications. In their work, [7], emphasize the importance of evaluation metrics such as precision, recall, F1-score, and accuracy in assessing classifier performance. They discuss how precision reflects a low false positive rate, while recall indicates the ability to capture a significant proportion of actual spam emails. Their thorough analysis of these metrics provides a framework for evaluating the effectiveness of spam detection models, particularly in scenarios where data imbalances are prevalent. By establishing a clear understanding of these metrics, the authors contribute to the development of more reliable and effective spam classification systems. [8], discuss the growing emphasis on explainable AI (XAI) within the context of spam detection systems. Their research indicates that enhancing the transparency of spam detection models can significantly improve user trust and system effectiveness. By developing models that not only achieve high accuracy but also provide insights into the reasoning behind classification decisions, this work paves the way for future research that prioritizes user understanding and engagement with spam detection technologies. This focus on explainability is becoming increasingly important as users demand more accountability from automated systems. [9], explore the integration of adversarial training and reinforcement learning to bolster the robustness of spam detection systems. Their study highlights the necessity of simulating adversarial scenarios, wherein spam tactics evolve continuously. The authors demonstrate that this proactive approach can enhance the resilience of classifiers against sophisticated spam techniques, ensuring long-term effectiveness in real-world applications. This work represents a significant step forward in the ongoing battle between spam detection systems and evolving spam strategies, emphasizing the need for adaptive and robust solutions. [10], proposed a hybrid model integrating SVM and convolutional neural networks (CNNs) for spam detection, which combined SVM's strength in classification with CNN's capability to extract high-level features from text data. The results showed a significant improvement in spam detection accuracy compared to using either SVM or CNN alone. This study highlights the potential for combining traditional machine learning algorithms with deep learning techniques for better performance. [3], explored using n-grams as features in an SVM-based classifier for spam detection. Their findings demonstrated that bi-grams and tri-grams significantly improved classification accuracy compared to using unigrams alone. This suggests that incorporating different levels of n-grams in the feature extraction process can capture more contextual information from email content. [11], developed a hybrid CNN-LSTM (long short-term memory) model for spam classification, achieving high accuracy by leveraging CNN's ability to extract spatial features and LSTM's strength in capturing sequential dependencies. This approach demonstrated superior performance compared to traditional classifiers like SVM and logistic regression, indicating that deep learning models can effectively address the challenges posed by complex email content. [5], applied BERT (Bidirectional Encoder Representations from Transformers) to email spam detection, fine-tuning the pre-trained model to classify spam and non-spam emails. The study found that BERT's contextualized word representations significantly outperformed traditional word embeddings, such as Word2Vec, in distinguishing between legitimate and spam emails. This demonstrates the value of leveraging pre-trained language models in email filtering tasks. [4], conducted an empirical evaluation comparing stemming and lemmatization in email spam detection using deep neural networks. They observed that lemmatization yielded better results than stemming, likely due to its ability to preserve more semantic information from
13 the text. This finding reinforces the importance of text preprocessing steps in deep learning models for spam detection. [12], tackled the problem of data imbalance in email spam filtering using an ensemble learning approach that combined multiple classifiers, such as random forests, gradient boosting, and SVM. The study showed that ensemble models could effectively enhance classification performance by leveraging the strengths of individual classifiers and mitigating their weaknesses. [13], utilized advanced NLP techniques alongside BERT to classify spam emails, demonstrating that a hybrid model that integrates traditional feature extraction methods with modern language models can significantly enhance spam detection accuracy. Their approach incorporated TF-IDF features along with BERT embeddings, showing that hybrid models could take advantage of both lexical and contextual information in emails. [14], presented a scalable email spam detection system using Apache Spark to process large

datasets efficiently. They combined traditional machine learning algorithms like SVM with big data processing frameworks to handle scalability challenges in real-world applications. Their findings suggest that integrating machine learning with distributed computing frameworks can improve the practicality and scalability of spam detection systems. The use of NLP techniques, such as TF-IDF, word embeddings, and sentiment analysis, has been extensively studied in spam detection. [15], discussed the importance of TF-IDF for feature extraction in email spam classification, illustrating how this method helps to identify the most relevant words in the text. This technique has been widely adopted due to its simplicity and effectiveness in transforming textual data into a numerical format that can be used by machine learning algorithms. [16], emphasized the role of NLP preprocessing techniques in spam detection, such as stop word removal and tokenization, to improve classifier performance. Their work with the Natural Language Toolkit (NLTK) demonstrated how preprocessing steps could significantly impact model accuracy by reducing noise and focusing on meaningful content. [17], explored combining TF-IDF and latent semantic analysis (LSA) for feature extraction in email spam filtering. Their approach aimed to capture both term importance and latent semantic relationships between words, resulting in better spam detection accuracy compared to using TF-IDF alone. This suggests that integrating multiple feature extraction techniques can provide a more comprehensive representation of email content. Transfer learning has emerged as a valuable approach in email spam detection, allowing models to leverage knowledge from related tasks or domains. [18], reviewed recent advances in transfer learning for spam filtering, highlighting its potential to improve classification performance in cases with limited training data. The study discussed techniques like fine-tuning pre-trained models and using domain adaptation to generalize models across different datasets. [5], enhanced the robustness of email spam detection using adversarial training, a technique in transfer learning where models are trained to be resilient against adversarial examples. Their study demonstrated that adversarial training could significantly improve spam detection accuracy, particularly in cases where spam content evolves to evade traditional filters. [19], addressed this issue by advocating for continuous updates to spam filters. They emphasized that spam detection models need to adapt to emerging trends in spam content to maintain their effectiveness over time. Techniques like incremental learning and online training have been proposed to enable models to evolve as new spam patterns emerge. [7], presented a hybrid deep learning approach for spam detection that combined LSTM networks with convolutional layers. Their model demonstrated the ability to adapt to new spam patterns by leveraging LSTM's sequential learning capabilities, making it more robust against evolving threats. Several studies have compared the performance of different spam detection algorithms and preprocessing techniques. [3], examined multilingual email spam detection using TF-IDF and word embeddings, finding that word embeddings performed better in capturing semantic information across languages. This highlights the importance of considering multilingual data when designing spam filters for global applications. [20], focused on practical NLP preprocessing techniques for email spam detection, emphasizing the impact of different text processing steps on classifier performance. Their findings suggested that simple techniques like stop word removal and lowercasing can substantially improve model accuracy. [3], evaluated the effectiveness of LSTM networks for spam detection compared to traditional algorithms, finding that LSTM outperformed SVM in handling sequential data. This study supports the trend towards using deep learning models for tasks that involve processing sequential or time-dependent data.

METHODOLOGY

The Object-Oriented Analysis and Design (OOAD) methodology is particularly well-suited for developing a machine learning-based email spam detection system due to its emphasis on modularity, scalability, and maintainability. OOAD provides a systematic framework that facilitates the development of complex systems by modeling them as a collection of interacting objects, each representing an aspect of the real-world problem domain. This approach aligns well with the challenges of spam detection, where the system needs to continuously adapt to changing spam tactics and patterns, while still maintaining high levels of performance and reliability.

RESULTS AND DISCUSSION

The main menu of the admin panel for our email spam filtering application serves as a central navigation hub, granting administrators access to various features, tools, and settings critical to managing the system. This menu is strategically positioned at the top of the screen or within a side panel and is organized into several categories, including user management, log viewing, and system configuration. The primary options within the main menu include functionalities for monitoring spam predictions, reviewing logs, and managing the Model. Below is a diagram illustrating the main menu layout of the proposed admin panel.

Open Access
Online ISSN: 2992-5487
Print ISSN: 2992-6068

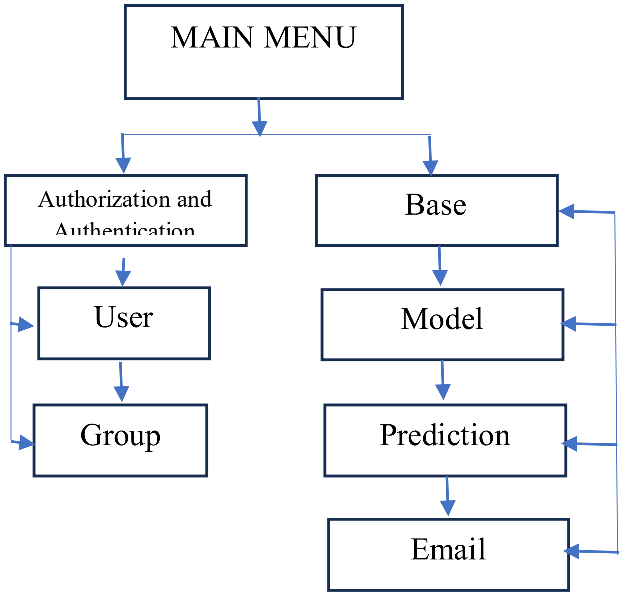


Figure 1 Main Menu of the Proposed Email Spam Filter System

Sub Menu

A subsystem within our email spam filtering application refers to a distinct component of the admin panel, designed to perform specific tasks or functions related to email management and spam detection. By breaking down the overall application into these smaller, more manageable subsystems, we enhance the design, implementation, maintenance, and operation of the entire system. Each submenu within the admin panel includes default options for adding new records and modifying existing ones, facilitating efficient management with a simple click.

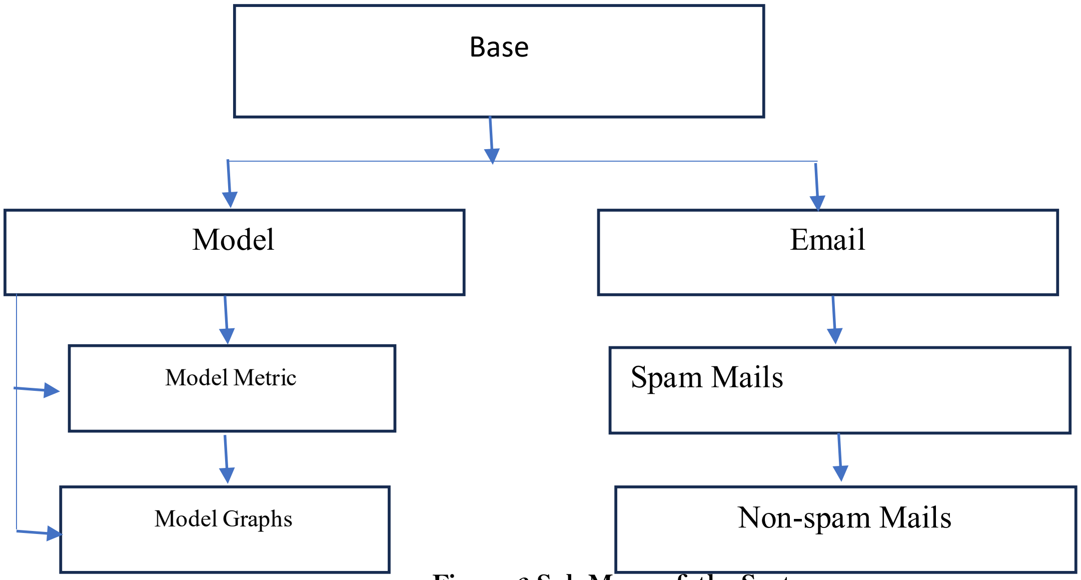


Figure 2 Sub Menu of the System

Database Development Tools

MySQL is a popular relational database management system (RDBMS) that may be used for a wide range of tasks, including corporate data storage and web development. Because of its strong characteristics, which include scalability, query optimization, and data modeling, it is a preferred option among developers and administrators. Data is arranged into tables, rows, and columns using the robust relational database management system MySQL. To efficiently handle data, it makes use of data types, constraints, indexes, stored procedures, triggers, and views. A graphical user interface (GUI) called MySQL Workbench gives database design, management, and administration a visual interface to go along with MySQL. Users may quickly design and modify database designs, run SQL queries, import and export data, and carry out other database-related operations with MySQL Workbench. Through the integration of MySQL Workbench's user-friendly interface and powerful database management capabilities, database management processes can be enhanced.

Input/ Output Format

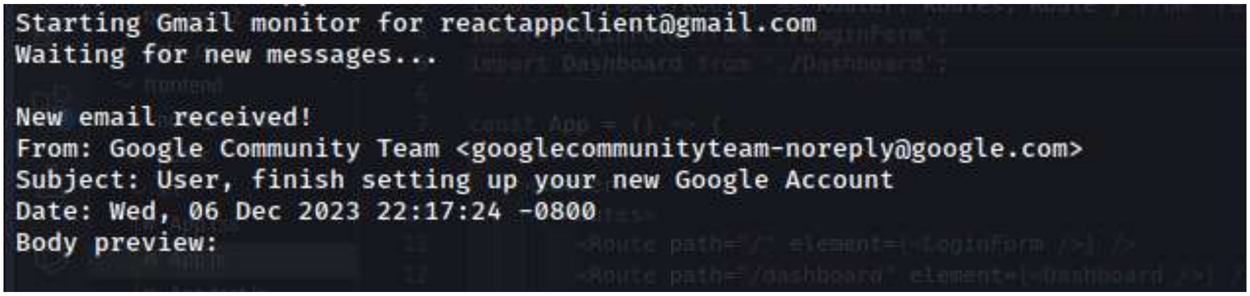
Input Format

The input of the system would automatically come from the user's email inbox. And show in the server like this:

Input Format

Output Format

The output of the system would be shown on the Admin panel



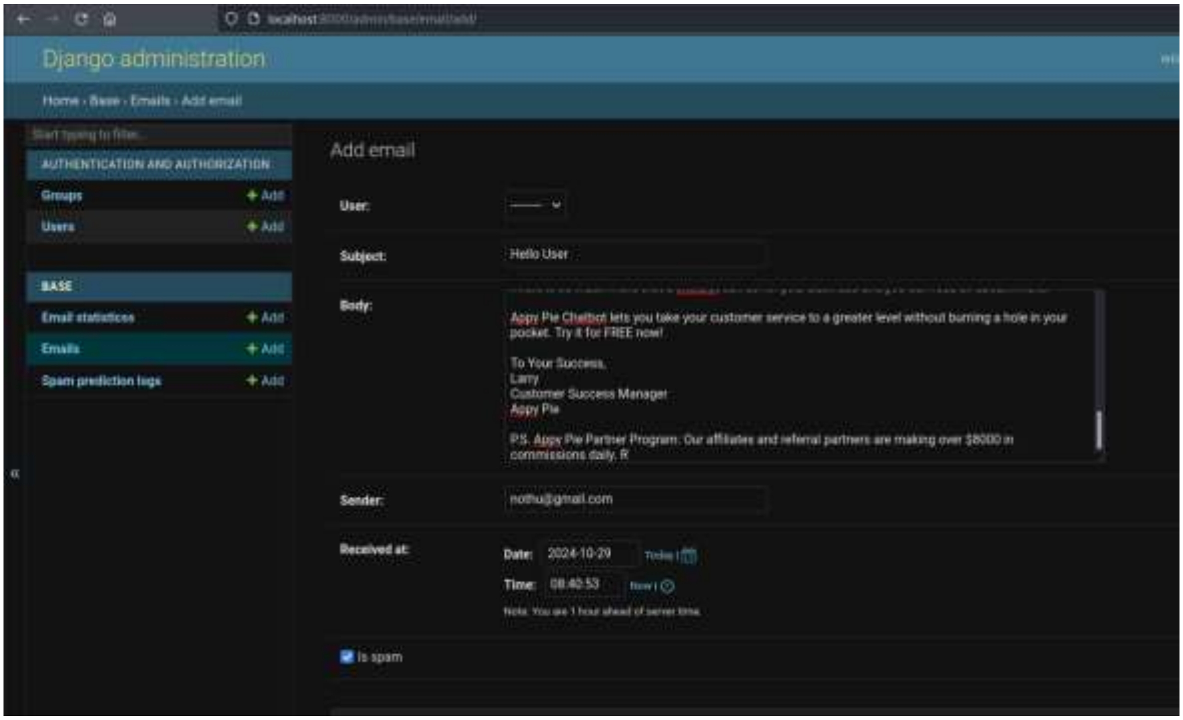
```
Starting Gmail monitor for reactappclient@gmail.com
Waiting for new messages...

New email received!
From: Google Community Team <googlecommunityteam-noreply@google.com>
Subject: User, finish setting up your new Google Account
Date: Wed, 06 Dec 2023 22:17:24 -0800
Body preview:
```

Figure 3

Input Format

Output Format

The output of the system would be shown on the Admin panel



Django administration
Home › Base › Emails › Add email
Start typing to filter...
AUTHENTICATION AND AUTHORIZATION
Groups + Add
Users + Add
BASE
Email statistics + Add
Emails + Add
Spam prediction logs + Add

Add email

User: ---------

Subject: Hello User

Body:
Appy Pie Chatbot lets you take your customer service to a greater level without burning a hole in your pocket. Try it for FREE now!

To Your Success,
Larry
Customer Success Manager
Appy Pie

P.S. Appy Pie Partner Program: Our affiliates and referral partners are making over $8000 in commissions daily. R

Sender: nothu@gmail.com

Received at:
Date: 2024-10-29 Today |
Time: 08:40:53 Now |
Note: You are 1 hour ahead of server time.

☑ Is spam

Figure 4 Output Format

Algorithm

1. Initialize System
Load Libraries: Import necessary libraries (e.g., imaplib, email, sklearn, pandas, etc.).
Set Up Configuration: Define configurations for email server credentials and model paths.
2. Fetch Emails
Connect to IMAP Server: Establish a connection to the email server using the provided credentials.
Select Email Folder: Choose the folder to fetch emails from (e.g., "Inbox").
Fetch Unread Emails: Retrieve all unread emails and store them in a list.
3. Process Emails
For each email in the fetched list:
Parse Email: Extract the email subject, body, and sender information.
Preprocess Email Data: Remove HTML tags and unnecessary characters. Tokenize the text and remove stop words.
4. Vectorize Email Text
Apply TF-IDF Vectorization: Convert the cleaned email text into a numerical format using the TF-IDF vectorizer.
5. Classify Email
Load Pre-trained SVM Model: Load the trained SVM model from storage.
Predict Classification: Use the vectorized email data as input to the model to predict whether the email is spam or not spam.
6. Log Results
Create Log Entry: Record the email sender, subject, prediction result (spam or not spam), and timestamp.
Save Log Entry: Store the log entry in the database or a file for future reference.
7. Send Notifications
Trigger Notification: If the email is classified as spam, send a notification to the admin panel and browser to alert the administrator and user respectively.

Admin Panel Actions

Display Logs: Provide an interface for administrators to view all logged emails with their classification results.

SUMMARY, CONCLUSION AND RECOMMENDATION

This research extensively explored machine learning-based approaches to email spam detection, addressing the growing challenge of unwanted and potentially harmful email communications. The study revealed that spam emails constitute a significant portion of global email traffic, with estimates suggesting over 50% of all emails sent daily are spam. Traditional rule-based approaches have proven inadequate in combating evolving spam tactics, leading to the emergence of machine learning solutions. Hence the design and implementation of an email spam filter. The research indicates that machine learning (ML) approaches significantly outperform traditional rule-based systems in email spam detection. Hybrid methods that combine supervised and deep learning techniques yield robust solutions adaptable to the evolving spam landscape. Despite challenges like concept drift and computational efficiency, ML models demonstrate success rates exceeding 95% accuracy, validating their role as effective spam detection mechanisms. The shift from rule-based systems to advanced ML models such as Support Vector Machine (SVM) is crucial in combating email spam. ML systems' capacity to learn from emerging patterns and adapt to new spam tactics positions them as the leading solution for long-term spam management, contingent on effective feature engineering, regular updates, and sufficient computational resources. To successfully implement an offline email spam filter using machine learning, certain best practices should be considered. High-performance local storage (HDD or SSD) and sufficient processing power are essential to handle and store email data for an offline ML-based filtering system. Utilizing pre-trained models like Support Vector Machines (SVM) for local processing ensures reliable spam detection without network dependencies. Implementing strict access control for system administrators enhances security by limiting access to authorized users. Regular model updates are necessary to maintain accuracy, especially as spam tactics evolve. While the system operates offline, monitoring and alert mechanisms can help detect hardware issues or inconsistencies in spam detection data, notifying administrators as needed. A user-friendly interface improves usability, allowing users to easily view flagged emails and review filtering results. To avoid data loss due to potential hardware issues, establishing a local data backup protocol is crucial. Additionally, including a manual override feature for spam marking will ensure effective spam management, even if the automated system encounters temporary issues, providing a reliable fallback option.

REFERENCES

1. Akawuku I.G, Onyinyechi H.I, & Nwankwo C. (2025). Software Engineering Dimensions: Empirical Frameworks on Microservices Architecture in Cloud Computing Domains. IAA Journals of Scientific Research 12(2):37-43. http://doi.org/10.59298/IAAJSR/2025/123743.00
2. Cui, L. (2018). Support vector machines for email classification: A survey. Journal of Computer Science and Technology, 33(1), 1-20.
3. Kaur, S., and Kaur, G. (2021). Spam detection using n-grams and SVM classifier. International Journal of Computer Applications, 182(10), 1-6.
4. Gao, Y., and Zhang, Y. (2021). A hybrid model for spam classification using CNN and LSTM. Computers, Materials, & Continua, 67(1), 453-466.
5. Jiang, S., and Wu, Y. (2022). Using BERT for email spam detection: A novel approach. Journal of Computing and Information Technology, 30(1), 35-50.
6. Agrawal S, Arora S, Amiri-Kordestani L, de Claro RA, Fashoyin-Aje L, Gormley N, Kim T, Lemery S, Mehta GU, Scott EC, Singh H. Use of single-arm trials for US Food and Drug Administration drug approval in oncology, 2002-2021. JAMA oncology. 2023 Feb 1;9(2):266-72.
7. Huang, Z., Liu, J., and Yan, X. (2020). Performance evaluation metrics for spam classification. International Journal of Information and Computer Security, 14(5), 412-426.
8. Ghazaleh, M., Alhosni, A., and Ghafoor, K. (2021). Explaining spam detection using explainable AI. Journal of Information Technology, 36(4), 342-355.
9. Li Y, Zhang Y, Timofte R, Van Gool L, Yu L, Li Y, Li X, Jiang T, Wu Q, Han M, Lin W. NTIRE 2023 challenge on efficient super-resolution: Methods and results. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023 (pp. 1922-1960).
10. Zhou X, Feng J, Li Y. Non-intrusive load decomposition based on CNN–LSTM hybrid deep learning model. Energy Reports. 2021 Nov 1; 7:5762-71.
11. Gao, Y., and Zhang, Y. (2021). A hybrid model for spam classification using CNN and LSTM. Computers, Materials, & Continua, 67(1), 453-466.
12. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning (still) requires rethinking generalization. Communications of the ACM. 2021 Feb 22;64(3):107-15.
13. Saman A, Rasool S. A Feature-Level Hybrid Model Approach for Automated Phishing Email Detection. Journal of Computing & Biomedical Informatics. 2025 Jun 1;9(01).
14. Banu S, Divya R, TT DD, Sri B, Gheisari M, Khammar S, Ghaderzadeh M. Phishing Attack Simulation, Email Header Analysis, and URL Scrutiny: A Comprehensive Approach to Cyber Threat Mitigation. Computer Networks and Communications. 2025 Jul 2:1-20.
15. Banu S, Divya R, TT DD, Sri B, Gheisari M, Khammar S, Ghaderzadeh M. Phishing Attack Simulation, Email Header Analysis, and URL Scrutiny: A Comprehensive Approach to Cyber Threat Mitigation. Computer Networks and Communications. 2025 Jul 2:1-20.
16. Bird, S., Klein, E., and Loper, E. (2019). Natural language processing with Python. O'Reilly Media.
17. Verma, U.P., Singh, P., Verma, A.K., SINGH Jr, P.O.O.J.A. and Verma, A., 2023. Correlation between chronic periodontitis and lung cancer: A systematic review with meta-analysis. *Cureus*, *15*(3).
18. Garg SK, Gupta R. Exploring the relationship between corporate governance and stock price performance: Evidence from Indian publicly listed companies. Asian Journal of Management and Commerce. 2023;4(1):66-72.
19. Sharma, R., and Sinha, V. (2018). Using IMAP for enhanced email management. International Journal of Advanced Computer Science and Applications, 9(10), 50-55.
20. Pepe E. *Human-centric approach to emails phishing detection* (Doctoral dissertation, Dublin, National College of Ireland).

CITE AS: Akawuku I. Godspower , Adejumo Samuel Olujimi, Alade Samuel Mayowa, Olatunde Ayodeji Akano David and Mulumeoderhwa Bahati (2025) Development of an Efficient Machine learning-based Email spam Detection System. NEWPORT INTERNATIONAL JOURNAL OF ENGINEERING AND PHYSICAL SCIENCES, Volume 5 Issue 2 Page 1-10.
https://doi.org/10.59298/NIJEP/2025/5211000