IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145100

SPAM EMAIL DETECTION USING Machine Learning Algorithms

Gaurav Mani Tripathi[1], Aman Maddheshiya[2], Ankit Verma[3], Ashish Awasthi[4]

Mr. Namita Srivastava[5]

Department Of Computer Science, Goel Institute Of Technology and Management, Lucknow, Uttar Pradesh, India.[1-5]

Abstract: The rise in unsolicited emails, known as spam, has created an urgent need for more trustworthy and powerful antispam filters. Recent advances in machine learning techniques have enabled researchers and developers to effectively identify and filter spam emails. In this paper, we present a thorough analysis of several popular machine learning-based email spam filtering strategies. We provide an overview of key concepts, methods, effectiveness, and current research directions in spam filtering.

We begin by examining how top internet service providers (ISPs), including Gmail, Yahoo, and Outlook, apply machine learning techniques in their email spam filtering processes. We also describe the general process of email spam filtering and highlight the various ways researchers have applied machine learning to combat spam. Our evaluation compares the strengths and limitations of existing machine learning techniques and identifies unresolved challenges in spam filtering research. Based on our analysis, we recommend adopting deep learning and deep adversarial learning approaches to more effectively address the problem of spam emails in the future.

Keywords: Analysis of Algorithms, Machine Learning, Spam Filtering, Deep Learning, Neural Networks, Support Vector Machines (SVM), Naïve Bayes.

I. INTRODUCTION

Background

- In recent years, the internet has become an integral part of everyday life. As a result, the number of people using email has grown significantly. Email is one of the most widely used modes of communication and a powerful tool for personal, academic, and business interactions. However, with the increase in email usage, there has also been a surge in unwanted and unsolicited emails, commonly known as spam.
- Spam emails are often generated in bulk and sent to users without their consent. These emails are frequently used for advertising, phishing, or spreading malware. They are typically received after a user unknowingly shares their email address on untrusted websites or platforms.
- Spam emails are not just a minor annoyance—they pose serious security and operational risks. They flood users' inboxes, making it harder to identify important messages. In organizations, spam can lead to lost productivity as employees spend time sorting through irrelevant or dangerous emails.
- Emails may also contain malicious links or attachments that install spyware, ransomware, or viruses on the user's device. Phishing emails trick users into revealing sensitive information such as passwords or banking credentials.
- The growing complexity of spam tactics has made it difficult for traditional rule-based filtering methods to keep up. Spammers constantly evolve their techniques by using random text, image-based content, or slight variations in wording.
- To tackle this, modern spam detection systems now rely on artificial intelligence and machine learning. These can analyze patterns in email content and sender behavior to distinguish between legitimate and spam messages.
- Natural Language Processing (NLP) plays a key role in this process by helping systems understand language and identify spam-related phrasing. Supervised learning allows models to improve by learning from labeled examples Spam and non-spam emails.
- Though progress has been made, spam detection remains an ongoing challenge. Sophisticated threats like spear phishing continue to evolve, requiring constant updates and improvements in filtering systems.
- A strong spam detection system not only improves security but also enhances the overall user experience by maintaining a clean and organized inbox.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145100

Problem Statement

The increasing volume of spam emails creates serious challenges for users and service providers. These emails:

- Waste users' time and clutter inboxes.
- Decrease internet speed and consume bandwidth.
- Pose risks to security and privacy by stealing sensitive information through phishing links or malware.
- Interfere with the functionality of email systems and applications.

Despite ongoing research, spam detection methods still struggle to accurately and consistently identify spam emails, especially when new tactics are used by attackers. Thus, there is a need for more robust and adaptive filtering techniques.

Objectives

The primary objectives of this research are:

- To analyze the nature and behavior of spam emails.
- To explore and implement machine learning techniques for identifying spam emails.
- To evaluate the performance of various algorithms in detecting spam with high accuracy.
- To reduce the risk of spam-related attacks by improving filtering mechanisms.

Scope

This study focuses on the detection of spam emails using **Natural Language Processing (NLP)** and **machine learning** techniques. It uses a publicly available dataset to train and test various algorithms, such as **Naive Bayes**, and measures their effectiveness in filtering spam. The research does not cover hardware-level or encryption-based spam prevention methods but is confined to content-based filtering using text analysis.

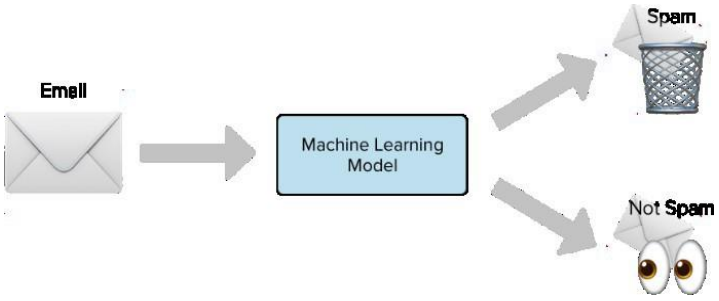

Figure.3 Working Process

II. LITERATURE REVIEW

Literature Review: Email Spam Detection

Email spam detection has been a critical area of research in the fields of machine learning, natural language processing, and cybersecurity for over two decades. The proliferation of unsolicited messages not only clutters inboxes but also poses security threats like phishing and malware distribution. Numerous studies have explored different approaches to detect and filter spam, evolving from simple rule-based systems to advanced machine learning and deep learning models.

1. Traditional Rule-Based and Heuristic Approaches

Early spam detection systems relied on manually crafted rules and heuristic filters. These included keyword- based filters, blacklists, and pattern matching (e.g., Spam Assassin). Although effective initially, these methods quickly became obsolete as spammers adapted their strategies. They also suffered from high false positive rates and required constant manual updating.

2. Machine Learning Approaches

Machine learning (ML) techniques brought a more dynamic and adaptive approach to spam detection. Pioneering work by and routs set all. (2000) applied Naive Bayes classifiers, demonstrating significant improvements over rule-based methods. Since then, several classifiers have been employed:

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025
DOI: 10.17148/IJARCCE.2025.145100

Naive Bayes:
Favored for its simplicity and effectiveness in text classification. It works well with bag-of-words or TF-IDF representations.

Support Vector Machines (SVMs):
Offer better generalization performance and robustness against high-dimensional data, often outperforming Naive Bayes in benchmark tests.

Decision Trees and Random Forests:
Known for interpretability and ensemble power, though sometimes slower with large datasets.

k-Nearest Neighbors (k-NN):
Effective but computationally expensive and sensitive to feature scaling.

3. **Deep Learning Models**

Recent work has explored the use of deep learning for spam detection, which can learn complex patterns without extensive feature engineering:

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks: Effective for capturing sequential dependencies in email text.

Convolutional Neural Networks (CNNs): Applied to text for capturing local features and n-gram patterns.
Offer state-of-the-art performance by leveraging contextual word embeddings.

These models, however, are resource-intensive and often impractical for real-time filtering on edge devices.

4. **Feature Engineering and Representation**

The performance of spam detectors heavily depends on feature extraction methods. Common techniques include:

Bag-of-Words (BOW) and TF-IDF: Widely used in traditional ML pipelines.
Word Embeddings (Word2Vec, Glove): Improve semantic understanding of text.
Meta features: Include header info (sender, subject), link analysis, and HTML structure.

5. **Hybrid and Ensemble Methods**

To improve detection accuracy, researchers have also proposed hybrid approaches combining multiple classifiers or techniques.
For example, combining Naive Bayes with SVM or integrating ML classifiers with heuristic filters.

6. **Evaluation and Benchmarking**

Most studies use publicly available datasets such as the Enron Email Dataset, Spam Assassin Public Corpus, and Ling-Spam.

Common evaluation metrics include precision, recall, F1-score, and accuracy. However, the imbalance in spam vs. ham emails often
necessitates careful metric selection (e.g., ROC-AUC, PR-AUC).

7. **Challenges and Open Issues**

Despite significant progress, challenges remain:

Concept Drift: Spammers continuously change tactics, making models outdated over time. Data Privacy: Email content is sensitive, limiting data sharing and model training.

Adversarial Attacks: Spammers may deliberately craft messages to evade filters.
Real-Time Detection: Balancing accuracy and computational efficiency is crucial for deployment in live systems.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145100

III. MATERIALSAND METHOD

This section outlines the process used to detect spam emails using Natural Language Processing (NLP) and machine learning.

Dataset
The dataset used for this study was obtained from Kaggle:
https://www.kaggle.com/datasets/venky73/spam-mails-dataset

Dataset Overview
The dataset used in this study was sourced from Kaggle and contains a total of **5,171 email messages**. Each message is labeled as either **spam** or **ham (non-spam)**, providing a reliable basis for supervised learning in spam classification tasks.

Data Pre-processing
Pre-processing is an essential step in preparing raw text data for analysis and machine learning. In this project, the following techniques were applied to clean and normalize the dataset:

- **Special characters, digits, and punctuation** were removed, as they do not contribute significantly to the semantic meaning of the messages.
- All text was **converted to lowercase** to ensure uniformity and eliminate case sensitivity issues.**Stop words**, which are common words with little informational value (such as "the", "is", and "and"), were removed to focus on more meaningful terms.
- **Stemming** was performed using the **Porter Stemmer**, and **lemmatization** was applied to reduce words to their root forms, thereby minimizing redundancy.
- The cleaned text was then **tokenized**, meaning it was split into individual words or tokens for further analysis.

These steps helped in converting the raw email messages into structured and analyzable data suitable for feature extraction.

Handling Imbalanced Data

An initial analysis of the dataset revealed a class imbalance: the number of ham (non-spam) emails significantly exceeded the number of spam emails. This imbalance could negatively impact the performance of the machine learning model, particularly in its ability to correctly detect spam.

To resolve this issue, a **down sampling technique** was employed. This involved reducing the number of ham samples to match the count of spam samples, resulting in a more balanced dataset.

- **Figure 1(a)** illustrates the original distribution of spam and ham emails.
- **Figure 1(b)** shows the dataset after applying the down sampling strategy.

This approach helps in mitigating the bias of the classifier toward the majority class and improves the model's ability to generalize.

Feature Extraction

Since machine learning models operate on numerical data, the textual email messages needed to be transformed into numerical feature vectors. The following methods were used for this transformation:

- **Count Vectorizer** and the **Tokenizer API** from **TensorFlow Keras** were used to convert words into integer-based sequences.
- The Tokenizer splits each email into words (tokens) and maps them to numerical values.
- To ensure consistency in input size, **padding** was applied to make all sequences of equal length using the pad sequences () me thod.

These feature vectors serve as the input to the machine learning algorithms used for spam detection.

Data Splitting

To train and evaluate the machine learning model, the dataset was divided into two parts:

- **80%** of the data was used for training the model.
- The remaining **20%** was reserved for testing the model's performance on unseen data.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145100

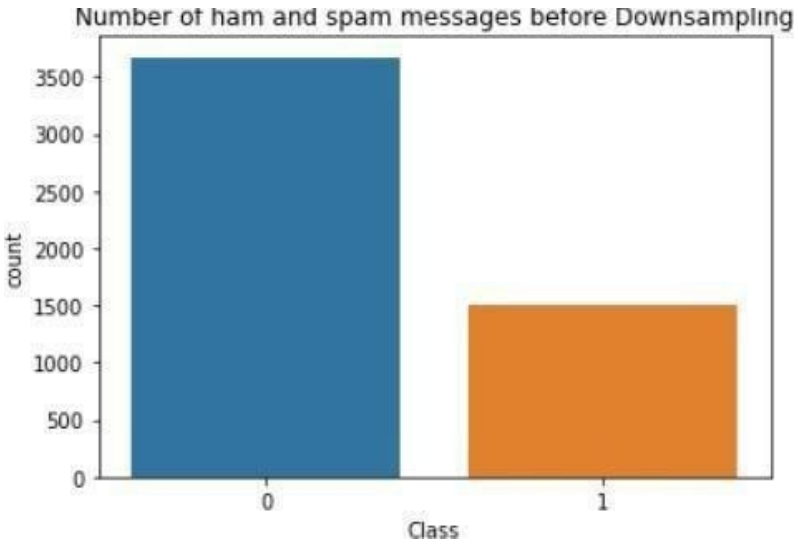

Figure.2 Before Sampling

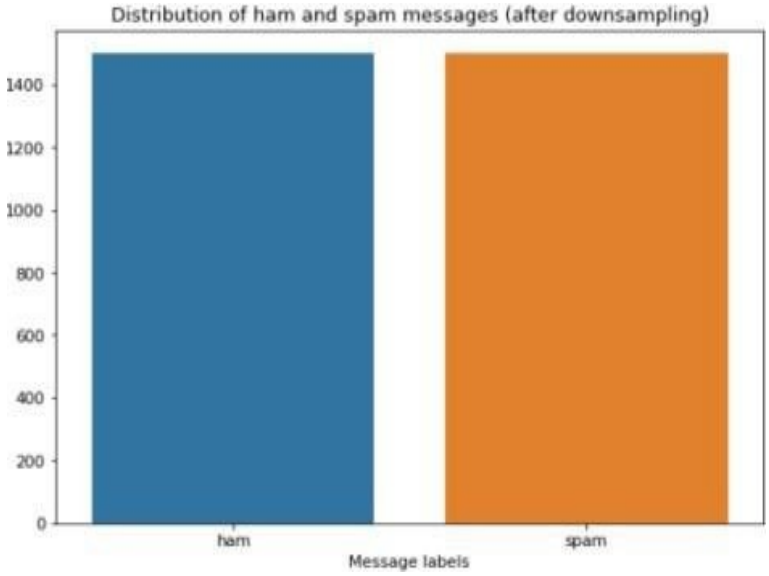

Figure.3 After Sampling

Since it's crucial to extract features, algorithms anticipate vectors. Tensor Flow Keras's Count Vectorizer and Tokenizer API are utilized to extract features. Tokenizer API does integer encoding and breaks up phrases into words.

Each sentence is represented by a numerical sequence using sequencing. Pad sequences () is used to create sequences of the same length. Additionally, we separate the data into training and test sets. Twenty percent were test samples and the remaining eighty percent were training samples in order to regulate how the algorithm operated. Machine learning algorithms are now being used to train the model.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145100

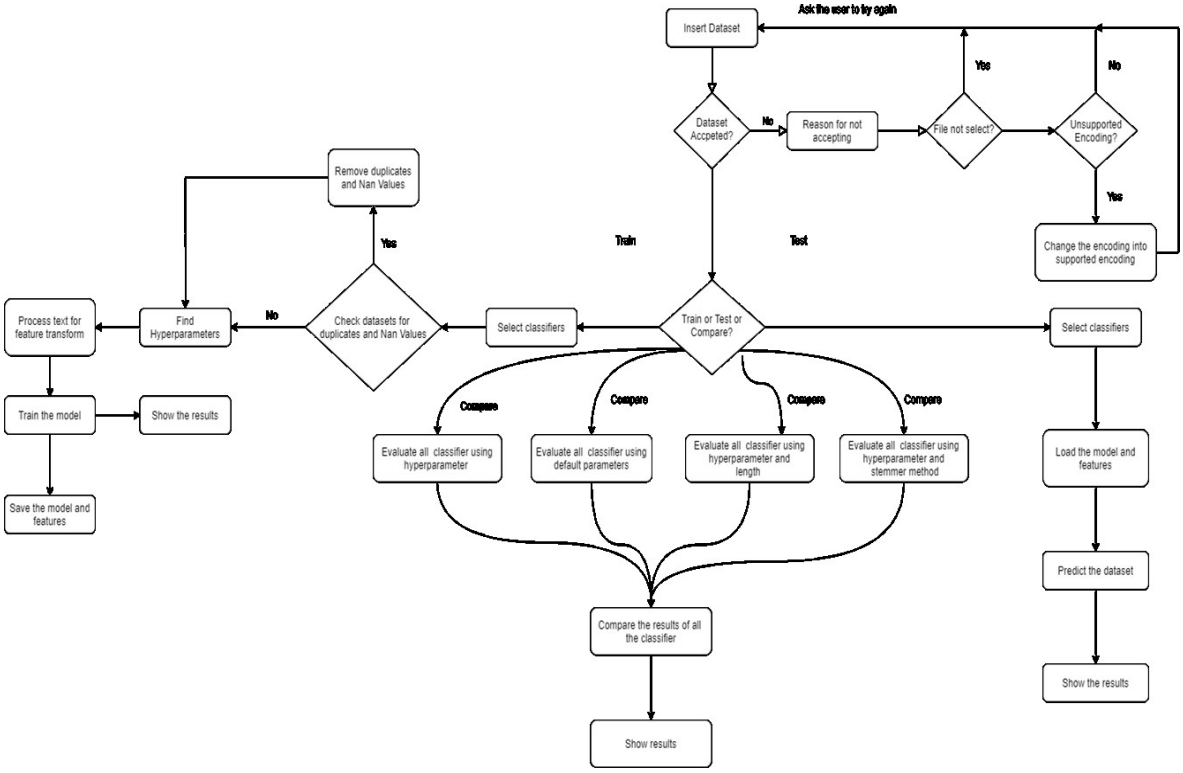

Fig.4. Flow Chart of Model

1. Naive Bayes:

The Bayesian classifier is a frequently used probabilistic method for text classification. A Bayesian classifier's primary function is to identify which terms are present in an email message and which ones are not in order to assess if it is spam or not. As per the literature, the most likely target label is assigned in the Bayesian technique to the new email. A Nave Bayes network is the most basic type of Bayesian network, where all attributes are unaffected by the value of the class variable. One way to think of the categorization problem is as finding the greatest value in the equation below.

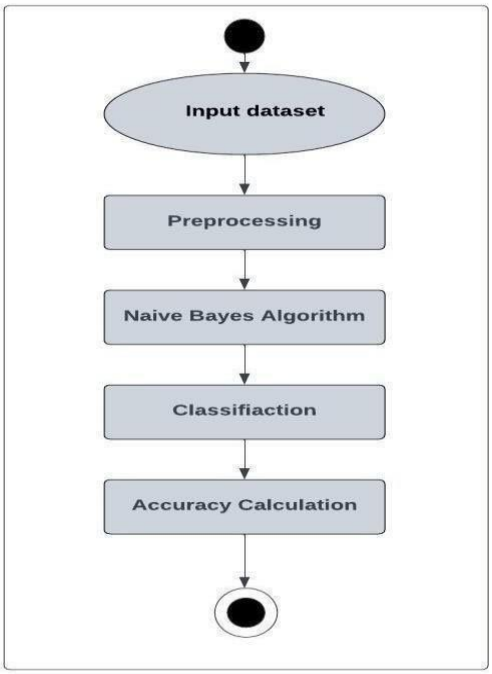

Figure.5 Activity Diagram

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145100

2. SUPPORT VECTOR MACHINE

"The Support Vector Machine (SVM) is a popular Supervised Learning Algorithm, the Support Vector model is used for classification problems in Machine Learning techniques. "The Support Vector Machines totally founded on the idea of Decision points. The Main resolution of Support Vector Machine algorithm is to create the line or decision boundary. The Support Vector Machine algorithm gives hyperplane as a output which classifies new samples. In 2- dimensional space "hyperplane is line dividing a plane into 2 parts where each class is present in one side."

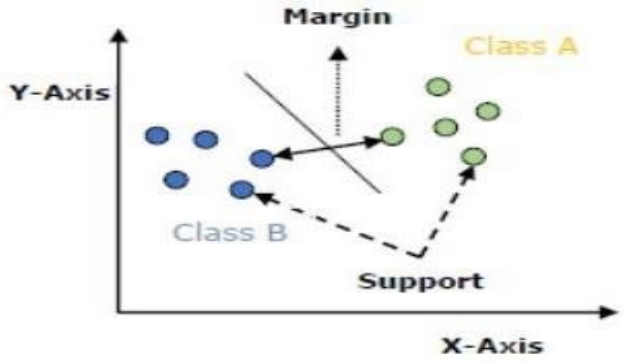

Fig.2 Support Vector Machine

IV. CONCLUSION

Spam email detection is a critical component of modern digital communication security. By employing various techniques such as keyword filtering, blacklisting, machine learning algorithms, and natural language processing (NLP), systems can effectively distinguish between legitimate and spam messages. Over time, spam filters have evolved from rule-based methods to advanced AI-driven models that can adapt to new spam tactics.

The implementation of spam detection improves user experience, protects sensitive information, and enhances productivity by minimizing distractions and potential threats like phishing or malware. However, the challenge remains in maintaining a balance between accurately filtering spam and avoiding false positives (flagging valid emails as spam).

In conclusion, while no spam detection system is perfect, ongoing advancements in artificial intelligence and data analytics continue to significantly improve the accuracy, adaptability, and efficiency of spam filtering technologies.

V. FUTURE SCOPE

- **Advancements in Artificial Intelligence and Machine Learning**

Ongoing progress in artificial intelligence (AI) and machine learning (ML) algorithms is expected to result in more **accurate and adaptive** spam detection systems. These systems can learn from large volumes of data and dynamically adjust to emerging spam patterns, enhancing detection accuracy over time.

- **Behavioral Analysis Integration**

Future spam filters may incorporate **behavioral analysis**, allowing systems to learn from a user's individual email usage patterns, preferences, and interactions. By understanding typical user behavior, spam detectors can more effectively differentiate between **legitimate** and **malicious** messages.

- **Real-Time Threat Intelligence**

The integration of **real-time threat intelligence feeds** can enable spam detection systems to stay updated on the latest spamming techniques and newly discovered threats. This proactive approach can improve the system's ability to identify **zero-day attacks** and other sophisticated spam tactics.

- **User-Centric Customization**

Personalized spam filters can offer **user-level customization**, enabling individuals to fine-tune the sensitivity of spam detection according to their needs. Such flexibility can help users maintain better control over their inbox, reducing false positives and improving user satisfaction.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.145100

REFERENCES

[1]. M. Labonne and S. Moran, "Spam-T5: Benchmarking Large Language Models for Few-Shot Email Spam Detection," vol. 2304.01238, Apr. 2023. Available: https://arxiv.org/abs/2304.01238

[2]. G. Shirvani and S. Ghasemshirazi, "Advancing Email Spam Detection: Leveraging Zero-Shot Learning and Large Language Models," vol. 2505.02362, May 2025. Available: https://arxiv.org/abs/2505.02362

[3]. S. Jamal and H. Wimmer, "An Improved Transformer-based Model for Detecting Phishing, Spam, and Ham," vol. 2311.04913, Nov. 2023. Available: https://arxiv.org/abs/2311.04913

[4]. M. H. Alsuwit, M. A. Haq, and M. A. Aleisa, "Advancing Email Spam Classification Using Machine Learning and Deep Learning Techniques," *Engineering, Technology and Applied Science Research*, vol. 14, no. 4, pp. 14994–15001, Aug. 2024. Available: https://etasr.com/index.php/ETASR/article/view/7631

[5]. K. Thakur, M. L. Ali, M. A. Obaidat, and A. Kamruzzaman, "A Systematic Review on Deep-Learning-Based Phishing Email Detection," *Electronics*, vol. 12, no. 21, p. 4545, Oct. 2023. Available: https://www.mdpi.com/2079-9292/12/21/4545

[6]. S. Das, S. Mandal, and R. Basak, "Spam Email Detection Using a Novel Multilayer Classification-Based Decision Technique," *International Journal of Computers and Applications*, vol. 45, no. 9, pp. 587–599, Sep. 2023. Available: https://www.tandfonline.com/doi/full/10.1080/1206212X.2023.2258328

[7]. K. Taghandiki, "Building an Effective Email Spam Classification Model with spaCy," vol. 2303.08792, Mar. 2023. Available: https://arxiv.org/abs/2303.08792

[8]. N. Al-shanableh, M. S. Alzyoud, and A. M. Alzyoud, "Enhancing Email Spam Detection Through Machine Learning," *California State University, San Bernardino*, vol. 22, no. 1, Article 2, 2023. Available: https://scholarworks.lib.csusb.edu/ciima/vol22/iss1/2/

[9]. K. S. Reddy, V. Harshini, K. Maneesha, and T. Gyaneshwari, "Spammer Detection in Social Networks Using Deep Learning," *IET Conference Proceedings*, vol. 2023, no. 5, Jul. 2023. Available: https://digital-library.theiet.org/doi/abs/10.1049/icp.2023.1481

[10]. G. Borotić, "Effective Spam Detection with Machine Learning," *Croatian Regional Development Journal*, vol. 4, no. 2, pp. 43–64, Dec. 2023. Available: https://sciendo.com/article/10.2478/crdj-2023-0007