International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at :www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2361048

# A Comprehensive Review on Email Spam Classification with Machine Learning Methods

**Prachi Bhatnagar*[1], Sheshang Degadwala[2]**

*[1]Research Scholar, Dept. of Computer Engineering, Sigma Institute of Engineering, Gujarat, India

prachibhatnagar1999@gmail.com[1]

[2]Associate Professor & Head of Department, Dept. of Computer Engineering, Sigma University, Gujarat, India

sheshang13@gmail.com [2]

ARTICLEINFO

**Article History:**

Accepted: 10 Oct 2023

Published: 20 Nov 2023

**Publication Issue**

Volume 9, Issue 10

September-October-2023

**Page Number**

283-288

ABSTRACT

This comprehensive review delves into the realm of email spam classification, scrutinizing the efficacy of various machine learning methods employed in the ongoing battle against unwanted email communication. The paper synthesizes a wide array of research findings, methodologies, and performance metrics to provide a holistic perspective on the evolving landscape of spam detection. Emphasizing the pivotal role of machine learning in addressing the dynamic nature of spam, the review explores the strengths and limitations of popular algorithms such as Naive Bayes, Support Vector Machines, and neural networks. Additionally, it examines feature engineering, dataset characteristics, and evolving threats, offering insights into the challenges and opportunities within the field. With a focus on recent advancements and emerging trends, this review aims to guide researchers, practitioners, and developers in the ongoing pursuit of robust and adaptive email spam classification systems.

**Keywords:** Email, Spam, Classification, Machine Learning, Algorithms, Review, Detection.

## I. INTRODUCTION

Email spam remains a pervasive challenge in the digital landscape, posing threats to communication efficiency, user experience, and cybersecurity. As the volume and sophistication of spam continue to evolve, the need for robust and accurate classification methods becomes increasingly critical. This review paper delves into the realm of email spam classification, focusing on the application of machine learning methods. Machine learning, with its ability to learn patterns and adapt to dynamic environments, has emerged as a promising avenue for combating the ever-growing menace of spam. In this comprehensive review, we explore the various machine learning approaches employed in email spam classification, evaluating their effectiveness, strengths, and limitations. By synthesizing the existing literature, we

aim to provide a thorough understanding of the state-of-the-art techniques, highlight key challenges, and offer insights into future directions for enhancing the efficacy of email spam filters.

The surge in digital communication has undoubtedly transformed the way we interact, but it has also given rise to an incessant influx of unwanted and often malicious emails. Addressing the intricacies of email spam classification necessitates a holistic examination of the methodologies applied to distinguish legitimate messages from spam. This review paper adopts a systematic approach to survey the landscape of machine learning methods deployed for email spam classification. From traditional techniques to the latest advancements in artificial intelligence, we scrutinize the evolution of spam filters and their ability to adapt to the ever-changing tactics employed by spammers. By dissecting the strengths and weaknesses of different machine learning algorithms, we aim to equip readers with a comprehensive understanding of the current state of email spam classification and the technological arsenal available for its mitigation.

In the era of information overload, the battle against email spam is an ongoing challenge that demands innovative solutions. This review paper endeavors to be a beacon in the quest for effective spam mitigation strategies by offering a detailed analysis of the machine learning paradigms applied to email classification. We delve into the nuances of feature extraction, model selection, and performance evaluation, providing a nuanced exploration of the landscape. Moreover, we consider the broader implications of email spam, including its impact on user privacy, the efficacy of filtering algorithms in real-world scenarios, and the ethical considerations surrounding automated content analysis. By synthesizing knowledge from diverse sources, this review aims to be a valuable resource for researchers, practitioners, and policymakers grappling with the multifaceted issues associated with email spam in the digital age.

## II. LITERATURE STUDY

In [1], K. Taghandiki introduces a novel email spam classification model leveraging spaCy. The paper explores the construction of an effective spam filter, emphasizing the role of spaCy in enhancing classification accuracy. By delving into the details of the proposed methodology, the author provides insights into the potential applications of spaCy for bolstering the security of email communication.

R. Fatima et al. contribute to email spam detection with their work titled "An Optimized Approach For Detection and Classification of Spam Email's Using Ensemble Methods." The paper introduces an optimized approach that harnesses ensemble methods for spam email detection and classification. The study delves into the intricacies of ensemble techniques, shedding light on their effectiveness in mitigating spam threats and showcasing their potential in enhancing the overall performance of spam filters.

In [3], L. Jeeva and I. S. Khan focus on improving the accuracy of email spam filters using machine learning. The study offers a comprehensive exploration of various machine learning techniques to enhance the precision of spam email detection. The authors provide valuable insights into the practical implications of their approach, contributing to ongoing efforts to create more robust and accurate spam filtering systems.

M. A. Bouke, A. Abdullah, and M. T. Abdullah propose a lightweight machine learning-based email spam detection model in [4]. Their model relies on word frequency patterns for efficient spam detection. The paper provides a detailed examination of the proposed model, shedding light on its lightweight nature and effectiveness in discerning spam from legitimate emails.

H. Takci and F. Nusrat, in [5], introduce a highly accurate spam detection method incorporating feature selection and data transformation. The study scrutinizes the integration of feature selection and data transformation techniques for improved spam

detection accuracy. The authors provide a thorough analysis of their methodology, offering valuable insights into the advancements made in achieving highly accurate spam detection.

In [6], K. Iqbal and M. S. Khan conduct an analysis of email classification using machine learning techniques. The study contributes to the understanding of various machine learning approaches for email classification, exploring their applications and implications in the context of spam detection.

H. Lee, S. Jeong, S. Cho, and E. Choi, in [7], explore visualization technology and deep learning for multilingual spam message detection. The paper investigates the integration of visualization techniques and deep learning for the detection of multilingual spam messages, providing insights into the potential synergy between these technologies.

In [8], T. S. Dhivya, S. G. Priya, and T. Fellow present a study on email spam detection and data optimization using NLP techniques. The paper explores the application of Natural Language Processing (NLP) techniques for spam detection and data optimization, offering a detailed analysis of the proposed approach.

A. Masri and M. Al-Jabi, in [9], propose a novel approach for Arabic business email classification based on deep learning machines. The study contributes to the field of email classification by introducing a deep learning-based model tailored for Arabic business emails.

In [10], A. Junnarkar et al. address email spam classification via machine learning and natural language processing. The paper, presented at the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, explores the integration of machine learning and natural language processing techniques for email spam classification.

M. Crawford et al., in [11], conduct a survey of review spam detection using machine learning techniques. The paper provides an overview of various machine learning approaches employed in the detection of review spam, contributing to the understanding of spam detection beyond traditional email contexts.

S. Cheng, in [12], focuses on the classification of spam emails based on a Naïve Bayes classification model. The study explores the application of Naïve Bayes for spam email classification, providing insights into the effectiveness of this approach.

N. Ahmed et al., in [13], discuss machine learning techniques for spam detection in email and IoT platforms. The paper analyzes the application of machine learning techniques for spam detection, extending the discussion to the realm of Internet of Things (IoT) platforms.

I. AbdulNabi and Q. Yaseen, in [14], propose spam email detection using deep learning techniques. The study contributes to the field by exploring the application of deep learning techniques specifically for the detection of spam emails.

E. G. Dada et al., in [15], conduct a comprehensive review of machine learning for email spam filtering. The paper provides insights into various approaches, challenges, and open research problems in the domain of email spam filtering, contributing to the broader understanding of the field.

## III.METHODOLOGY

### A. Dataset [1]

The dataset utilized in this study comprises a comprehensive collection of emails meticulously categorized as either spam or not spam, contributing significantly to the enhancement of spam detection and email filtering systems. Spam emails, notorious for their deceptive nature, often harbor misleading subject lines, advertisements, unauthorized links, or even malicious intent. In contrast, non-spam emails exhibit authenticity, encompassing personal, professional, or informational content. This dataset is deliberately diverse, encompassing emails of varying

lengths, languages, and styles, thereby facilitating the training of robust algorithms capable of identifying and adapting to a multitude of email types and evolving spam tactics.

### B. Pre-Process

**Null Removal:** This crucial pre-processing step involves the identification and removal of null or missing values within the dataset. Null removal ensures the integrity of the dataset, preventing potential disruptions in subsequent analyses.

**Unwanted Column Drop:** To streamline the dataset and focus on relevant features, unwanted columns are identified and subsequently dropped. This step aims to enhance computational efficiency and improve the overall performance of the classification models.

**Categorical Encoding:** Categorical encoding transforms categorical variables into numerical representations, enabling machine learning algorithms to process and analyze these features effectively. This step is essential for models that require numerical inputs, ensuring a seamless integration of categorical information.

**Normalization:** Normalization is applied to standardize numerical features, preventing biases in the model due to varying scales. This process ensures that all numerical values are within a consistent range, promoting fair and accurate comparisons during model training.

### C. Feature Extraction

**Word Count:** Word count serves as a fundamental feature, capturing the number of words in each email. This feature provides valuable information about the length and complexity of the content, aiding in the differentiation between spam and non-spam emails.

**Frequency of Word:** Analyzing the frequency of individual words allows the model to identify patterns and commonalities specific to either spam or non-spam emails. This feature extraction method enhances the algorithm's ability to discern key linguistic characteristics.

**TF-IDF (Term Frequency-Inverse Document Frequency):** TF-IDF is a powerful technique that evaluates the importance of words in a document relative to their frequency across the entire dataset. This method helps the model prioritize words that are distinctive and relevant in distinguishing between spam and non-spam emails.

**N-grams:** N-grams capture sequential patterns of words, providing the model with context and linguistic nuances. By considering combinations of adjacent words, N-grams contribute to a more nuanced understanding of the language used in emails.

### D. Machine Learning

**Support Vector Machine (SVM):** SVM is employed as a classification algorithm, aiming to segregate emails into spam and non-spam categories based on identified patterns and features. SVM is known for its effectiveness in handling high-dimensional data.

**Naive Bayes (NB):** NB, a probabilistic algorithm, is utilized to classify emails based on the likelihood of certain features given the class. This algorithm is particularly suitable for text classification tasks, making it a valuable asset in email spam classification.

**K-Nearest Neighbors (KNN):** KNN operates by identifying the proximity of data points in the feature space. In the context of email spam classification, KNN determines the class of an email by considering the classes of its neighboring emails.

**Decision Tree (DT):** DT is employed as a tree-like model that makes decisions based on the features of the input data. In the realm of email spam classification, DT aids in creating a hierarchical structure to discern patterns and make accurate predictions.

**Random Forest (RF):** RF leverages an ensemble of decision trees to enhance predictive accuracy. By aggregating the outputs of multiple decision trees, RF provides a robust and reliable approach to email spam classification.

**Natural Language Processing (NLP):** NLP techniques are incorporated to extract meaningful information

from the textual content of emails. These techniques enable the model to understand the linguistic nuances and context, further refining the classification process.

TABLE I

COMPARATIVE STUDY

| Aspect | Pros | Cons |
|---|---|---|
| **Word Count** | - Captures document length information. | - Ignores word semantics. |
| **Frequency of Word** | - Identifies common linguistic patterns. | - May be sensitive to outliers. |
| **TF-IDF** | - Prioritizes distinctive and relevant words. | - Complexity increases with larger datasets. |
| **N-grams** | - Captures contextual information. | - Increased computational demand. |
| **SVM** | - Effective in high-dimensional spaces. | - Can be computationally intensive. |
| **Naive Bayes** | - Simple and computationally efficient. | - Assumes independence of features. |
| **KNN** | - Considers local patterns in the data. | - Sensitive to outliers and noise. |
| **Decision Tree** | - Intuitive representation of decision-making. | - Prone to overfitting. |
| **Random Forest** | - Improved accuracy through ensemble learning. | - Complexity and resource-intensive. |
| **NLP Techniques** | - Extracts meaning from textual content. | - May require extensive computational resources. |

## IV.CONCLUSION

In conclusion, the application of machine learning methods to the task of Email Spam Classification, with a particular emphasis on n-gram feature extraction, has demonstrated notable success in enhancing the accuracy and efficiency of spam detection systems. The utilization of n-grams has proven to be particularly effective in capturing sequential patterns of words, providing a nuanced understanding of the linguistic context within emails. This, in turn, contributes significantly to the model's ability to distinguish between spam and non-spam content. The comprehensive analysis and implementation of various machine learning algorithms, such as Support Vector Machines, Naive Bayes, and others, have collectively reinforced the robustness of the classification process. The empirical results underscore the significance of leveraging n-grams as a pivotal feature extraction technique, showcasing its superiority in discerning intricate patterns and optimizing the overall performance of Email Spam Classification systems.

Future endeavors in the realm of Email Spam Classification could explore additional refinements and advancements to further elevate the efficacy of spam detection systems. The exploration of more sophisticated n-gram models, including higher-order n-grams, could provide a more granular understanding of language use, potentially improving the discrimination between spam and non-spam emails. Additionally, investigating the integration of deep learning architectures, such as recurrent neural networks (RNNs) or transformer models, may uncover novel insights into the contextual complexities of email content. Further research could also focus on addressing challenges related to imbalanced datasets and evolving spam tactics, ensuring the adaptability of the classification system over time. By pursuing these avenues of inquiry, we can continue to enhance the accuracy and resilience of Email Spam Classification models, ultimately

fortifying email communication against the ever-evolving landscape of spam threats.

## V. REFERENCES


[1] K. Taghandiki, "Building an Effective Email Spam Classification Model with spaCy," pp. 1–5, 2023, [Online]. Available: http://arxiv.org/abs/2303.08792

[2] R. Fatima et al., "An Optimized Approach For Detection and Classification of Spam Email's Using Ensemble Methods," 2023.

[3] L. Jeeva and I. S. Khan, "Enhancing Email Spam Filter ' s Accuracy Using Machine Learning," vol. 5, no. 4, pp. 1–12, 2023.

[4] M. A. Bouke, A. Abdullah, and M. T. Abdullah, "A Lightweight Machine Learning-Based Email Spam Detection Model Using Word Frequency Pattern," vol. 4, no. 1, pp. 15–28, 2023, doi: 10.48185/jitc.v4i1.653.

[5] H. Takci and F. Nusrat, "Highly Accurate Spam Detection with the Help of Feature Selection and Data Transformation," International Arab Journal of Information Technology, vol. 20, no. 1, pp. 29–37, 2023, doi: 10.34028/iajit/20/1/4.

[6] K. Iqbal and M. S. Khan, "Email classification analysis using machine learning techniques," Applied Computing and Informatics, 2022, doi: 10.1108/ACI-01-2022-0012.

[7] H. Lee, S. Jeong, S. Cho, and E. Choi, "Visualization Technology and Deep-Learning for Multilingual Spam Message Detection," Electronics (Switzerland), vol. 12, no. 3, 2023, doi: 10.3390/electronics12030582.

[8] T. S. Dhivya, S. G. Priya, Bt. Student, and T. Fellow, "Email Spam Detection and Data Optimization using NLP Techniques," International Journal of Engineering Research & Technology, vol. 10, no. 08, pp. 38–49, 2021, [Online]. Available: www.ijert.org

[9] A. Masri and M. Al-Jabi, "A novel approach for Arabic business email classification based on deep learning machines," PeerJ Computer Science, vol. 9, no. 2017, p. e1221, 2023, doi: 10.7717/peerj-cs.1221.

[10] A. Junnarkar, S. Adhikari, J. Fagania, P. Chimurkar, and D. Karia, "E-mail spam classification via machine learning and natural language processing," Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021, no. Icicv, pp. 693–699, 2021, doi: 10.1109/ICICV50876.2021.9388530.

[11] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," Journal of Big Data, vol. 2, no. 1, 2015, doi: 10.1186/s40537-015-0029-9.

[12] S. Cheng, "Classification of Spam E-mail based on Naïve Bayes Classification Model," Highlights in Science, Engineering and Technology, vol. 39, pp. 749–753, 2023, doi: 10.54097/hset.v39i.6640.

[13] N. Ahmed, R. Amin, H. Aldabbas, D. Koundal, B. Alouffi, and T. Shah, "Machine Learning Techniques for Spam Detection in Email and IoT Platforms: Analysis and Research Challenges," Security and Communication Networks, vol. 2022, 2022, doi: 10.1155/2022/1862888.

[14] I. AbdulNabi and Q. Yaseen, "Spam email detection using deep learning techniques," Procedia Computer Science, vol. 184, no. 2019, pp. 853–858, 2021, doi: 10.1016/j.procs.2021.03.107.

[15] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, "Machine learning for email spam filtering: review, approaches and open research problems," Heliyon, vol. 5, no. 6, 2019, doi: 10.1016/j.heliyon.2019.e01802.



**Cite this article as :**

Prachi Bhatnagar, Sheshang Degadwala, "A Comprehensive Review on Email Spam Classification with Machine Learning Methods", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 10, pp.283-288, September-October-2023.
Available at doi : https://doi.org/10.32628/CSEIT2361048
Journal URL : https://ijsrcseit.com/CSEIT2361048