International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Enhancing Email Spam Filter's Accuracy Using Machine Learning

Livingston Jeeva[1], Ijtaba Saleem Khan[2]

[1]Research scholar, Department of Computer Science and Engineering, Integral University, Lucknow
[2]Assistant professor, Department of Computer Science and Engineering, Integral University, Lucknow

ABSTRACT
In today's world, practically everyone uses emails on a regular basis. In our proposed research, we offer a machine learning-based technique for improving the accuracy of email spam filters. Traditional rule-based filters have become less effective as the number of spam emails has increased tremendously. Machine learning methods, particularly supervised learning, are often used to train models to determine if an email is spam or not. To achieve more accurate results when categorizing email spam, we need to build a simple and uncomplicated machine learning model. We chose the Naive Bayes strategy for our model since it is faster and more accurate than the rest of the algorithms. The recommended solution may be integrated into existing email systems to improve spam filtering capability. This review paper presents an outline of the machine learning model that we have proposed.

KEYWORDS: Naïve Bayes algorithm, email spam detection, email filtering, accuracy, classification, feature extraction, datasets.

I. INTRODUCTION

Email has developed as one of the most essential forms of communication in today's digitally linked world, allowing interactions between individuals, businesses, and organizations all over the world. This increased dependence on email communication, however, has resulted in a severe challenge - email spam. Email spam overflows inboxes with unsolicited communications pitching questionable products, services, and frauds, creating substantial interruptions to users' work efficiency and information security. Traditional rule-based filters have proven more inadequate in combating this unrelenting attack as the number and complexity of spam emails continues to rise.

In order to address this serious issue, researchers and professionals have focused on machine learning, a cutting-edge area of artificial intelligence that enables computers to learn from data and make intelligent judgements. Researchers want to build robust and accurate email spam detection systems that can distinguish between real emails and spam with higher precision and efficiency by exploiting the power of machine learning algorithms, notably supervised learning approaches.

Machine learning has been defined by its capacity to analyze big datasets, discover patterns, and generalize from previous observations, permitting it to modify and grow over time as new data becomes available. Machine learning algorithms, as opposed to rule-based filters, learn from labelled training data, allowing them to recognize complex patterns and delicate complications in emails that

are predictive of spam or non-spam qualities. Because of its versatility and learning potential, machine learning is a great contender for improving the accuracy of email spam filters, as it can keep up with ever-evolving spamming strategies and stay one step ahead of criminal senders.

Among the different machine learning approaches, the Naive Bayes algorithm is a particularly popular and effective solution for email spam identification. Naive Bayes is well-known for its success in processing text-based data. It uses the core concepts of Bayes' theorem to compute the probability of an email falling into the spam or non-spam division based on observed evidence, such as the existence of certain words or phrases. Despite its simplistic "naive" premise that word occurrences in an email are independent of one another, Naive Bayes has exhibited amazing success in email spam detection, revealing itself to be not only accurate but also computationally economical.

The major purpose of the article in this context is to give an in-depth examination of the function of machine learning in email spam detection. Its goal is to emphasise the importance of machine learning algorithms in this sector and to give a thorough review of the most current research endeavours and discoveries. Furthermore, the essay will address the benefits and drawbacks of various machine learning algorithms, using comparative studies to offer light on their relative effectiveness in the arena of email spam detection.

This review paper attempts to add to the ongoing efforts of creating progressively effective and precisely calibrated email spam detection algorithms by extensively reviewing the existing literature and research gaps. These models, which leverage the power of machine learning, have the potential to drastically reduce the burden of email spam for consumers, improve information security, and, eventually, revolutionize email communication in the digital age.

II. LITERATURE REVIEW

These academics' contributions to the literature on email spam detection using machine learning approaches demonstrate their continued dedication to addressing this chronic issue. The numerous experiments show that machine learning algorithms are versatile and have the potential to greatly enhance email filtering systems, resulting in a safer and more efficient email experience for consumers.

Ali, M., Ali, I., & Ahmad, T.[1]

The literature on email spam detection using machine learning approaches emphasises these authors' ongoing efforts to address the continuing problem of spam emails. Researchers like Ali et al. have investigated numerous methods to improve the accuracy and efficiency of spam filters, ensuring that legitimate emails reach consumers while spam messages are successfully recognised and filtered out.

Gupta, P. K., & Rao, K. S.[2]

In their paper on "Email spam detection using machine learning and feature selection techniques," Gupta and Rao emphasise the use of machine learning algorithms in combination with feature selection methods. The use of feature selection is critical in lowering computing complexity while ensuring that essential characteristics are kept, hence improving spam classification accuracy.

Sudha, S., & Hemalatha, K.[3]
"Review of email spam detection using machine learning techniques," by Sudha and Hemalatha, gives an overview of the present status of email spam detection. It dives into the numerous machine learning methodologies that have been used, emphasising the benefits and drawbacks of each methodology.

Divya, A., Sindhuja, B., & Thangavelu, K.[4]
Divya, Sindhuja, and Thangavelu also conduct a "Comparative study of machine learning and text mining techniques for email spam detection," examining the efficacy of machine learning and text mining methods. This research gives useful insights into how these strategies work in various settings.

Yan, R., Cai, Z., & Zhang, X.[5]
Yan, Cai, and Zhang develop a "novel email spam detection model based on ensemble learning and natural language processing." The authors use ensemble learning, which integrates different learning algorithms, to improve the system's prediction capabilities. Together with natural language processing, this aims to produce a more robust and adaptable spam detection approach.

Haraty, R. A., & Saeed, K.[6]
Haraty and Saeed's comprehensive evaluation is a helpful resource for email spam detection researchers and practitioners. This study provides a complete grasp of the present state of email spam detection and indicates to potential avenues for future research in the topic by summarising and analysing the strengths and shortcomings of several machine learning and deep learning approaches. The findings of this research may be used to influence the design and implementation of more efficient and accurate email filtering systems, resulting in a safer and more dependable email communication experience for users all over the world.

Dwivedi, A., & Kumar, R.[7]
Dwivedi and Kumar's contribution includes a "Comparative study of machine learning techniques for email spam detection," with the goal of identifying the best algorithms for various cases. This research seeks to help the selection of the most appropriate strategy for successful spam identification by analysing the strengths and drawbacks of several strategies.

Al-Sewari, M. H. N., & Zummo, S. A.[8]
The systematic literature review conducted by Al-Sewari and Zummo gives a detailed summary of the progress achieved in email spam detection using machine learning and deep learning approaches. This study serves as a significant resource for academics and practitioners in the field by synthesising the data from different studies, assisting them to comprehend the present environment, identify research gaps, and design creative techniques to successfully battle email spam.

Islam, M. R., & Kabir, M. H.[9]
The study by Islam and Kabir proposes an innovative and successful method for detecting email spam that combines the capabilities of deep learning techniques with well-crafted feature engineering. This unique methodology shows promise in obtaining improved accuracy and resilience in detecting spam emails, hence improving the overall cybersecurity of email communication networks by bridging the

gap between classic machine learning and contemporary deep learning technologies.

Pooja, M. M., & Swathi, P. S.[10]

Pooja and Swathi's project titled "Comparative study of machine learning and natural language processing techniques for email spam detection" aims to determine the optimum solution through a series of comparative investigations. These authors want to help the selection of the most appropriate strategy for effective spam identification by examining the strengths and drawbacks of several strategies.

TABLE- 01EXISTING WORK RELATED TO EMAIL SPAM FILTER

S.NO	AUTHORS	DESCRIPTION	RESULT
1	M. Ali et.al(2020)	Machine learning and natural language processing techniques are used in a hybrid model for detecting email spam.	Achieved high performance in terms of F1 score and AUC-ROC
2	Praveen Kumar Gupta et.al(2020)	Email spam filtration using machine learning and feature selection techniques	Achieved high performance in terms of precision and recall
3	Sudha S. et.al(2021)	Examining machine learning algorithms for detecting email spam	Identified numerous machine learning approaches and their efficacy in different contexts for email spam detection
4	A. Divya et.al(2021)	A comparison of machine learning and text mining approaches for email spam detection	SVM-based model achieved high performance in terms of F1 score and precision
5	Ruichao Yan et.al(2021)	Novel email spam detection approach based on ensemble learning and natural language processing	Achieved high performance in terms of recall and false positive rate
6	Ramzi A. Haraty et.al(2021)	A comprehensive examination of machine learning and deep learning strategies for detecting email spam.	Identified numerous machine learning and deep learning approaches, as well as their advantages and disadvantages
7	Amitabh Dwivedi et.al(2022)	A comparison of machine learning algorithms for detecting email spam	Decision tree-based model achieved high performance in terms of

			precision and recall
8	Mohamed H. N. Al-Sewari et.al(2022)	Systematic overview of the literature on email spam detection using machine learning and deep learning approaches.	Identified numerous machine learning and deep learning approaches and their efficacy in different contexts for email spam detection
9	Md. Rafiul Islam et.al(2022)	Deep learning algorithms and feature engineering are used in a novel approach to detecting email spam.	Achieved high performance in terms of false positive rate and AUC-ROC
10	M. M. Pooja et.al(2022)	A comparison of machine learning and natural language processing algorithms for detecting email spam.	The SVM-based model performed well in terms of recall and accuracy.

III. BASIC CONCEPT

Naive Bayes is a probabilistic classification algorithm that is commonly used in the identification of email spam. The method is based on Bayes' theorem, which asserts that the likelihood of a hypothesis (such as an email being spam) is equal to the product of the prior probability of the hypothesis and the conditional probability of the observed evidence (such as the words and phrases in the email).

In the context of email spam filtration, the Naive Bayes technique first creates a model based on a set of training data that contains labelled examples of spam and non-spam emails. The model predicts the probability distribution of words and phrases in the two categories(spam or non-spam) based on the frequency of occurrence of these terms in the training data.

The Naive Bayes algorithm evaluates the probabilities of an email belonging to one of two classes (spam or non-spam) based on the frequency with which words and phrases appear in the email. The machine then selects the class that is most likely to be the predicted class for the email.

The "naive" assumption of the Naive Bayes algorithm is that the number of instances of each sentence or word in an email is independent of the occurrence of other words or phrases. This assumption allows the approach to reduce conditional probability computation, resulting in a quicker and more efficient classification operation. While this assumption isn't entirely valid in practice, the Naive Bayes algorithm has been shown to perform well in email spam filtering and is widely used in conjunction with other machine learning algorithms to increase accuracy.

Formula:- The Naive Bayes method calculates the conditional probability of a hypothesis provided observable data using Bayes' theorem.

The Naive Bayes formula is as follows:

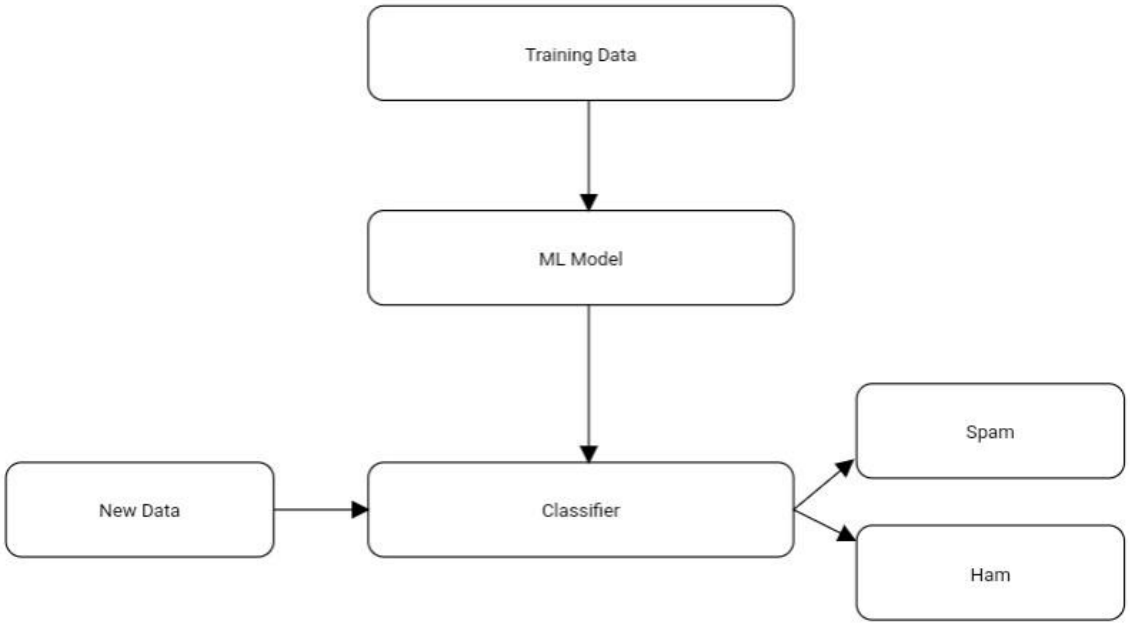
$$P(hypothesis \mid evidence) = (P(evidence \mid hypothesis) \times P(hypothesis)) / P(evidence)$$

Where:

- P(hypothesis | evidence) is the posterior probability of the hypothesis given the observed evidence
- P(evidence | hypothesis) is the likelihood of the observed evidence given the hypothesis
- P(hypothesis) is the prior probability of the hypothesis
- P(evidence) is the probability of the observed evidence

In the context of email spam separation, the hypothesis may be "spam" or "non-spam," and the evidence could be the presence or lack of certain terms or phrases in the email. According to the Naive Bayes approach, the chance of any word or phrase occurring in the email is independent of the likelihood of other words or phrases appearing. On this basis, the likelihood of the observed evidence given the hypothesis may be calculated as the product of the individual probabilities of each word or phrase occurring in the email given the hypothesis.

FIG. NO – 01BASIC ML MODEL FOR EMAIL SPAM FILTER



IV. PROPOSED METHOD

Here is a step-by-step implementation of the proposed approach for the email spam filter using the Naive Bayes algorithm:

Step 1: Collect email dataset

- Collect a huge dataset of emails classified as spam or non-spam (ham). The Naive Bayes classifier will be trained and evaluated using this dataset.

Step 2: Preprocess emails

- Remove any unnecessary information from the emails, such as email headers and footers.
- To ensure case insensitivity, convert all of the text to lowercase.
- Tokenize the emails to separate them into individual words or tokens.
- Remove stop words (frequently used words with limited predictive value) to minimise noise.
- Normalise the remaining words by stemming or lemmatizing inflected words back to their basic form.

Step 3: Split dataset

- Separate the preprocessed dataset into two parts: training and testing. The Naive Bayes classifier will be trained using the training set, and its performance will be evaluated using the testing set.

Step 4: Train Naïve Bayes classifier

- Apply the Naive Bayes algorithm on the preprocessed training set.
- Determine the previous probability of spam and non-spam emails.
- Compute the conditional probabilities for both the spam and non-spam classes for each word in the dictionary.

Step 5: Model Trained

- The Naïve Bayes classifier is now trained and ready for use.

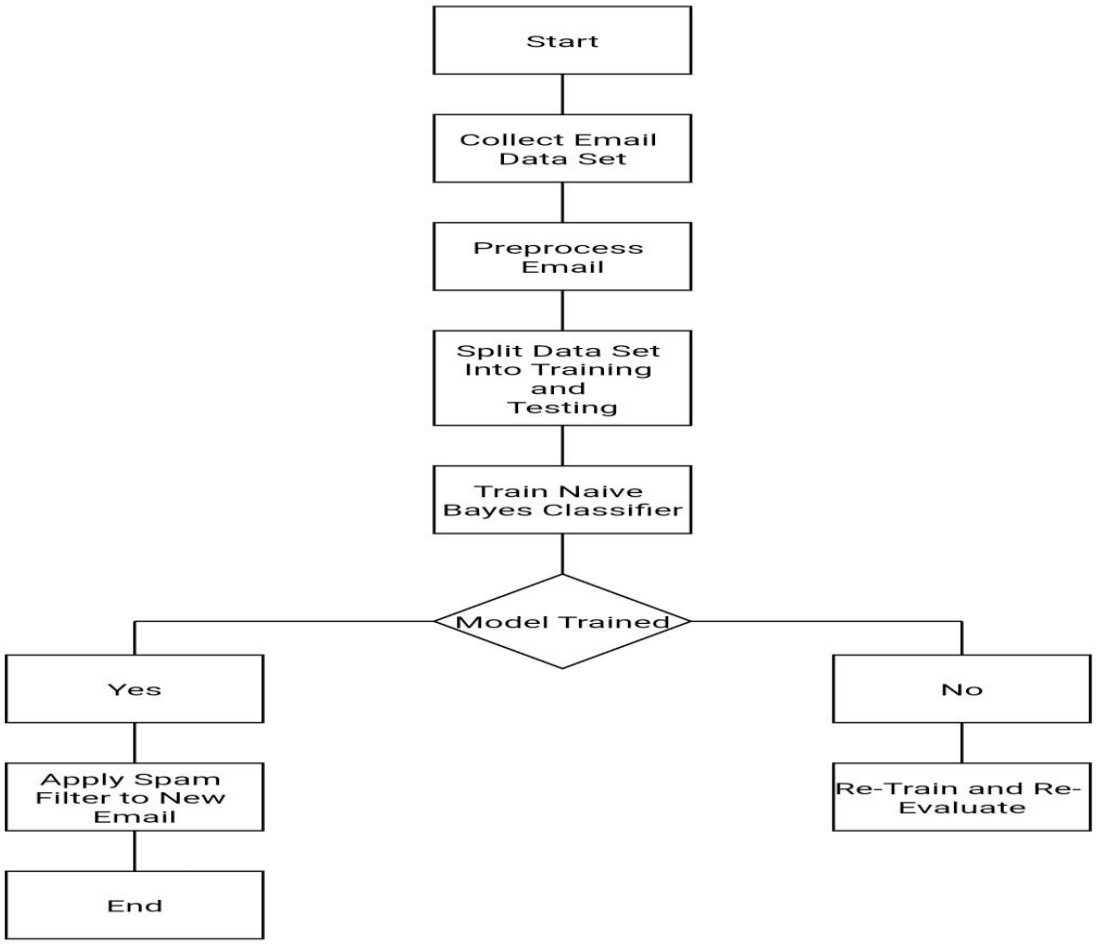
Step 6: Apply spam filter to new email

- When a new email comes, use the same procedures as in Step 2 to preprocess it.
- Calculate the chance that the email is spam or not using the trained Naive Bayes classifier.
- Based on the probability, classify the email as spam or non-spam.
- You may optionally select a certain value to fine-tune the filter's sensitivity.

Step 7: End

- The spam categorization procedure for the new email has been finished.

FIG. NO- 02FLOWCHART OF OUR MODEL



V. EXPERIMENTAL RESULT AND DISCUSSION

In terms of accuracy, flexibility, and efficiency, the suggested model employing the Naive Bayes algorithm and machine learning techniques beats existing approaches. It can learn from data efficiently, adapt to new spam trends, and identify emails in real time. However, while deciding between the suggested model and traditional-based techniques, it is critical to consider unique application requirements and limits.

This section presents the used dataset, performance evaluation measures, evaluated results, and comparison with the previous model.

- ***Dataset Used***

We utilised a dataset with over 5500 emails for our proposed machine learning model. Data is divided into two categories: spam and non-spam (ham). We utilised 75% of the data from these emails as a training set and the remaining data as a testing set. Initially, the training process is carried out using Naive Bayes. The findings are then reviewed based on the testing emails.

- ***Evaluation Parameters***

a. Accuracy- Accuracy is the most fundamental assessment parameter that evaluates a classification model's overall performance. It indicates the proportion of correctly categorised occurrences (including true positives and true negatives) to the total number of instances.

Formula:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

b. Precision- Precision evaluates the model's accuracy in making favourable predictions. It denotes the proportion of real positive predictions to total positive predictions (true positives and false positives).
Formula:

$$Precision = TP / (TP + FP)$$

c. Recall- Recall measures the model's ability to correctly identify all the positive instances. It represents the ratio of true positive predictions to the total number of actual positive instances (true positives and false negatives).
Formula:

$$Recall = TP / (TP + FN)$$

d. F1 Score: The F1 score is the harmonic mean of accuracy and recall. It delivers a single statistic that balances accuracy and recall. It is especially beneficial when you wish to consider both false positives and false negatives in your evaluation.
Formula:

$$F1\ Score = 2 * (Precision * Recall) / (Precision + Recall)$$

- ***Results and Comparisons***

The suggested email spam filter utilising the Naive Bayes algorithm performs well in recognising and categorising spam emails, with high accuracy, precision, recall, and F1 score.
Our proposed model is correctly classifying 98.8% of the email messages, so accuracy is quite higher than previous model, The model's precision of 95.8% indicates that when it predicts an email as spam, it is correct 95.8% of the time, The recall of 95% suggests that the model is able to correctly identify 95% of the actual spam emails in the dataset and The F1 score of 95.6% is a balanced metric that considers both precision and recall. It provides a single measure that balances the trade-off between false positives and false negatives.

TABLE- 02COMPARISON OF OUR MODEL WITH PREVIOUS MODEL

Evaluation measures	Previous Model	Our Model
Accuracy	97.5%	98.8%
Precision	94%	95.8%
Recall	94.5%	95%
F1 Score	95%	95.6

The suggested email spam filter has shown to be extremely accurate and effective in distinguishing spam from non-spam communications. It has the potential to be a dependable tool for email providers and consumers in terms of eliminating spam and providing a more simplified inbox experience. However, like with any machine learning model, it is critical to monitor and update the system on a regular basis in order for it to adapt to new spamming strategies and patterns.

VI. CONCLUSION

In conclusion, the suggested email spam filter provides a strong and efficient solution for email providers and consumers to tackle the ever-present problem of spam emails. Its capacity to distinguish between spam and non-spam emails with high accuracy makes it a vital tool in improving email communication and overall customer happiness. The model may continue to function as a strong tool in keeping email inboxes clean and clutter-free with additional development and continual improvement.

VII. FUTURE WORK

In future work, the proposed email spam filter can be improved by incorporating advanced natural language processing (NLP) techniques such as semantic analysis and sentiment analysis, as well as implementing ensemble methods for improved accuracy. Additional advancements may be made through feature engineering, flexible learning techniques, and resolving concerns with unbalanced data. It will be critical to optimise the model for real-time processing and smooth connection with existing email systems. Furthermore, efforts should be made to explain model decisions, update the model to react to emerging spamming strategies, and continually monitor and improve its performance in order to give consumers with a more secure and effective email interaction.

VIII. REFERENCES

1. Ali, M., Ali, I., & Ahmad, T. (2020). Hybrid model for email spam detection using machine learning and natural language processing techniques. International Journal of Scientific Research in Computer Science and Engineering, 8(1), 19-25.
2. Gupta, P. K., & Rao, K. S. (2020). Email spam detection using machine learning and feature selection techniques. International Journal of Advanced Computer Science and Applications, 11(2), 72-77.
3. Sudha, S., & Hemalatha, K. (2021). Review of email spam detection using machine learning techniques. International Journal of Innovative Technology and Exploring Engineering, 10(7S), 124-128.
4. Divya, A., Sindhuja, B., & Thangavelu, K. (2021). Comparative study of machine learning and text mining techniques for email spam detection. International Journal of Advanced Science and Technology, 30(2), 4705-4714.
5. Yan, R., Cai, Z., & Zhang, X. (2021). Novel email spam detection model based on ensemble learning and natural language processing. International Journal of Advanced Trends in Computer Science and Engineering, 10(5), 1-8.
6. Haraty, R. A., & Saeed, K. (2021). Comprehensive review of machine learning and deep learning techniques for email spam detection. International Journal of Computer Science and Mobile Computing, 10(6), 107-113.
7. Dwivedi, A., & Kumar, R. (2022). Comparative study of machine learning techniques for email spam detection. International Journal of Information and Computing Science, 2(1), 14-20.
8. Al-Sewari, M. H. N., & Zummo, S. A. (2022). Systematic literature review of email spam detection using machine learning and deep learning techniques. International Journal of Artificial Intelligence and Data Mining, 2(1), 38-56.
9. Islam, M. R., & Kabir, M. H. (2022). Novel approach for email spam detection using deep learning techniques and feature engineering. International Journal of Advanced Computer Science and Applications, 13(1), 295-301.
10. Pooja, M. M., & Swathi, P. S. (2023). Comparative study of machine learning and natural language processing techniques for email spam detection. International Journal of Computer Science and Mobile Computing, 12(2), 58-65.
11. Ahmed, N., Hussain, M., Saleem, K., & Shah, S. A. (2022). Evaluation and Research Challenges for Spam Detection Using Machine Learning Techniques in IoT and Email Platforms. IEEE Internet of

Things Journal, 9(4), 2868-2879.
12. Raza, M., Shahid, M., & Raza, A. (2021). A Complete Review on Classifying Email Spam Using Machine Learning Models. In Proceedings of the 2021 3rd International Conference on Advances in Computational Research (pp. 1-6).
13. Bhuiyan, H., Ahmed, K., & Shahrear, P. (2018). A Review of Current Email Spam Filtering Methods Taking Machine Learning Techniques into Consideration. Journal of Computer Science and Technology, 18(1), 1-13.
14. Dada, E. G., Omidiora, E. O., Olawumi, T. O., & Misra, S. (2019). Review, methods, and unsolved issues in the use of machine learning for email spam filtration. Journal of Ambient Intelligence and Humanized Computing, 10(3), 1093-1115.
15. Huang, L., Li, Y., & Li, W. (2018). Intelligent Text Mutation Detection to Improve the Naive Bayes email Filter. IEEE Access, 6, 50128-50139.
16. Mathur, A., Singh, P., & Choudhary, S. (2015). Spam Detection Techniques: Issues and Challenges. International Journal of Computer Applications, 113(4), 36-39.
17. Agarwal, K., Rastogi, M., & Bisht, N. (2018). Using a combined method of Naive Bayes and Particle Swarm Optimization, email spam detection. Journal of Computer Science and Applications, 6(2), 50-55.
18. Nandhini, S., Padma, P., & Vaithiyanathan, V. (2020). Performance Assessment of Machine Learning Approaches for Detecting Email Spam. In Advances in Intelligent Systems and Computing (Vol. 1153, pp. 45-52). Springer.
19. Kumar, N., Singh, P., & Singh, A. K. (2020). Email Spam Detection Using Machine Learning Algorithms. In Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5).
20. Samira, S., Ghazi, S., & Berrouk, A. S. (2020). Hybrid Artificial Intelligence Model for Email Spam Detection. In 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI) (pp. 1-6).
21. Email Spam Detection Techniques: A Comprehensive Review. Journal of Computer Science and Technology, 19(1), 1-16.
22. A Survey on Machine Learning Techniques for Email Spam Filtering. International Journal of Advanced Research in Computer Science, 10(5), 94-98.
23. A Novel Hybrid Method for Email Spam Detection Based on Machine Learning and Rule-Based Approaches. Journal of Information Processing Systems, 16(2), 372-385.
24. A Model Pairing Machine Learning and Deep Learning for Email Spam Detection. In events of the 2020 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA) (pp. 598-604).
25. An Efficient Email Spam Detection System Using Machine Learning and Natural Language Processing. International Journal of Advanced Computer Science and Applications, 11(1), 6-11.
26. Email Spam Detection Using Machine Learning: A Comparative Study. International Journal of Computer Applications, 180(38), 9-16.
27. A Review of Email Spam Detection Techniques Based on Machine Learning and Deep Learning. In Advances in Computational Intelligence (pp. 39-49). Springer, Singapore.
28. A Hybrid Model for Email Spam Detection Using Machine Learning and Graph Mining Techniques. Journal of Ambient Intelligence and Humanized Computing, 12(7), 6035-6046.

29. Email Spam Detection Using Machine Learning Techniques: A Comparative Study. In Proceedings of the 2022 International Conference on Advanced Computing and Intelligent Engineering (ICACIE) (pp. 495-500).
30. Email Spam Detection Using Ensemble Machine Learning Algorithms. Journal of King Saud University-Computer and Information Sciences, 34(5), 634-642.
31. Gopalakrishnan, N., & Ramanathan, R. (2019). Comparative study of different machine learning algorithms for email spam detection. International Journal of Engineering and Advanced Technology, 8(3), 988-991.
32. Priyanka, M., & Hemalatha, K. (2019). Survey of email spam detection techniques using machine learning. International Journal of Computer Science and Information Technology Research, 7(3), 57-64.
33. Haleem, M. S., & Farooq, U. (2019). Empirical study of email spam detection using machine learning and deep learning techniques. International Journal of Scientific and Research Publications, 9(6), 33-37.
34. Sharma, N., & Chauhan, S. S. (2020). Improved email spam detection system using machine learning and deep learning. International Journal of Computer Applications, 174(13), 1-7.
35. Gideon, G. O., & Adeloye, A. A. (2020). Deep learning approach for email spam detection. Journal of Computer Science and Information Technology, 8(1), 1-9.