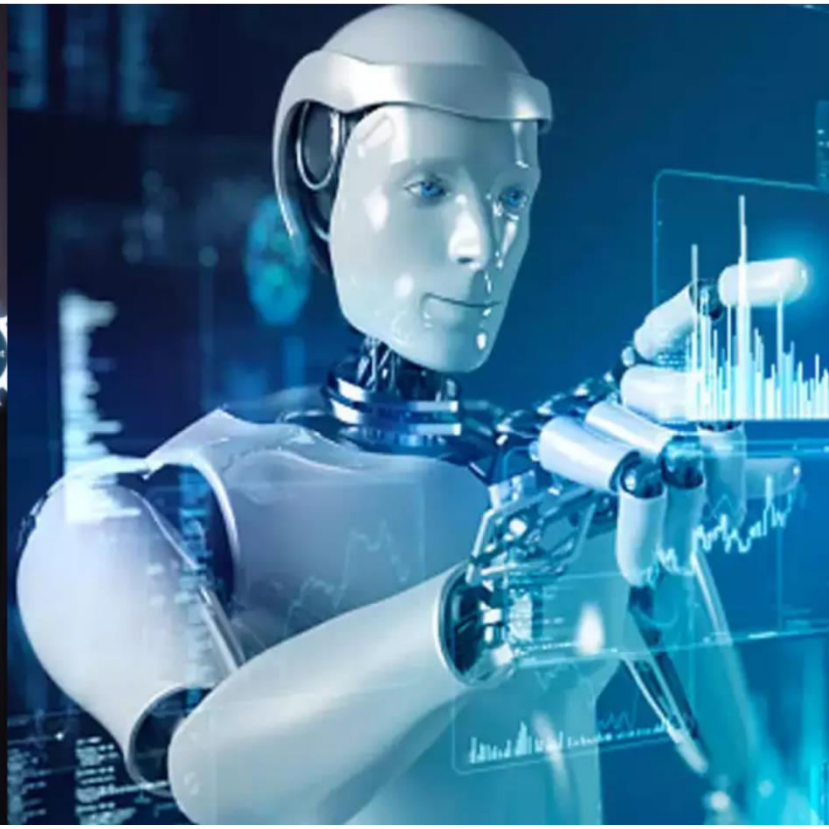
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 11, November 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411120|

Design and Implementation of a Spam Email Detection

Vishwas Vijay khade[1], Aniruddha Manoj Pardeshi[2], Kamlesh Siddharam Hirapure[3], Aditya Babasaheb Sarak[4], S.N. Shendge[5]

U.G. Student, Department of Computer Science and Engineering, VVPIET, Solapur, Maharashtra, India[1]

U.G. Student, Department of Computer Science and Engineering, VVPIET, Solapur, Maharashtra, India[2]

U.G. Student, Department of Computer Science and Engineering, VVPIET, Solapur, Maharashtra, India[3]

U.G. Student, Department of Computer Science and Engineering, VVPIET, Solapur, Maharashtra, India[4]

Assistant Professor, Department of Computer Science and Engineering, VVPIET, Solapur, Maharashtra, India[5]

ABSTRACT: Spam email detection is an essential task in modern communication systems to protect users from unwanted, harmful, or fraudulent messages. As the volume of email traffic continues to grow, distinguishing legitimate emails from spam has become increasingly challenging. This project focuses on developing an efficient spam detection model using machine learning techniques. The system preprocesses email data through tokenization, stop-word removal, and feature extraction methods such as TF-IDF. Various classification algorithms—including Naïve Bayes, Support Vector Machine (SVM), and Decision Trees—are evaluated to determine their effectiveness in identifying spam. Experimental results demonstrate that machine learning-based approaches significantly improve detection accuracy compared to traditional rule-based methods. The study concludes that automated spam email detection enhances user security, reduces inbox clutter, and strengthens overall email communication reliability.

KEYWORDS: Spam Detection, Email Classification Machine Learning, Naïve Bayes, Support Vector Machine (SVM), Text Mining, Natural Language Processing (NLP), Feature Extraction, TF-IDF, Data Preprocessing, Classification Algorithms, Cybersecurity, Phishing Detection, Supervised Learning, Phishin, Pattern Recognition

I. INTRODUCTION

Email is one of the most widely used communication platforms in personal, academic, and commercial environments. However, the increasing dependence on email services has also led to a rapid rise in spam—unwanted and often harmful messages that may include advertisements, phishing attempts, malware, or fraudulent content. Spam emails not only clutter inboxes but also pose significant security threats, including identity theft, financial loss, and unauthorized access to sensitive information. Due to the sheer volume and evolving nature of spam messages, traditional rule-based filtering techniques are no longer sufficient.

To address these challenges, spam email detection has emerged as a critical application of machine learning and natural language processing (NLP). By analyzing patterns, word frequencies, and contextual features within email content, machine learning models can automatically classify emails as spam or legitimate (ham). The use of algorithms such as Naïve Bayes, Support Vector Machine, Decision Trees, and deep learning methods has shown promising results in improving accuracy and reducing false positives.

This project focuses on developing an automated spam detection system using modern machine learning approaches. It emphasizes data preprocessing, feature extraction, model training, and performance evaluation. By implementing a reliable classifier, the system aims to enhance security, reduce user workload, and improve the overall efficiency of email communication.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411120|

II. LITERATURE SURVEY

A literature survey of spam email detection reveals that the field primarily uses machine learning (ML) techniques to analyze various features, including sender information, subject lines, and message content. Common ML algorithms include Naïve Bayes, Support Vector Machines, and Decision Trees. Advanced methods combine ML with other techniques like deep learning and neural networks for improved accuracy, though spammers continually adapt their methods to bypass filters.

Common features used in detection:
1) Sender details: Spam often comes from dubious or unknown email addresses or domains, unlike legitimate emails.
2) Subject line: Deceptive or attention-grabbing subject lines are a common characteristic of spam.
3) Message content: Spam emails frequently contain words and phrases like "free," "urgent," or "click here".
4) Email headers: Information like routing data, IP addresses, and mail transfer agents can provide clues about the email's origin and path.

III. METHODOLOGY

The methodology for spam email detection primarily involves using machine learning, particularly supervised learning techniques like Naive Bayes and Support Vector Machines, to classify emails as either spam or legitimate ("ham").
Front-End:- We are using language Python for Programming.
Back-End:-We are using Database Management system for programming (SQL, Oracle).

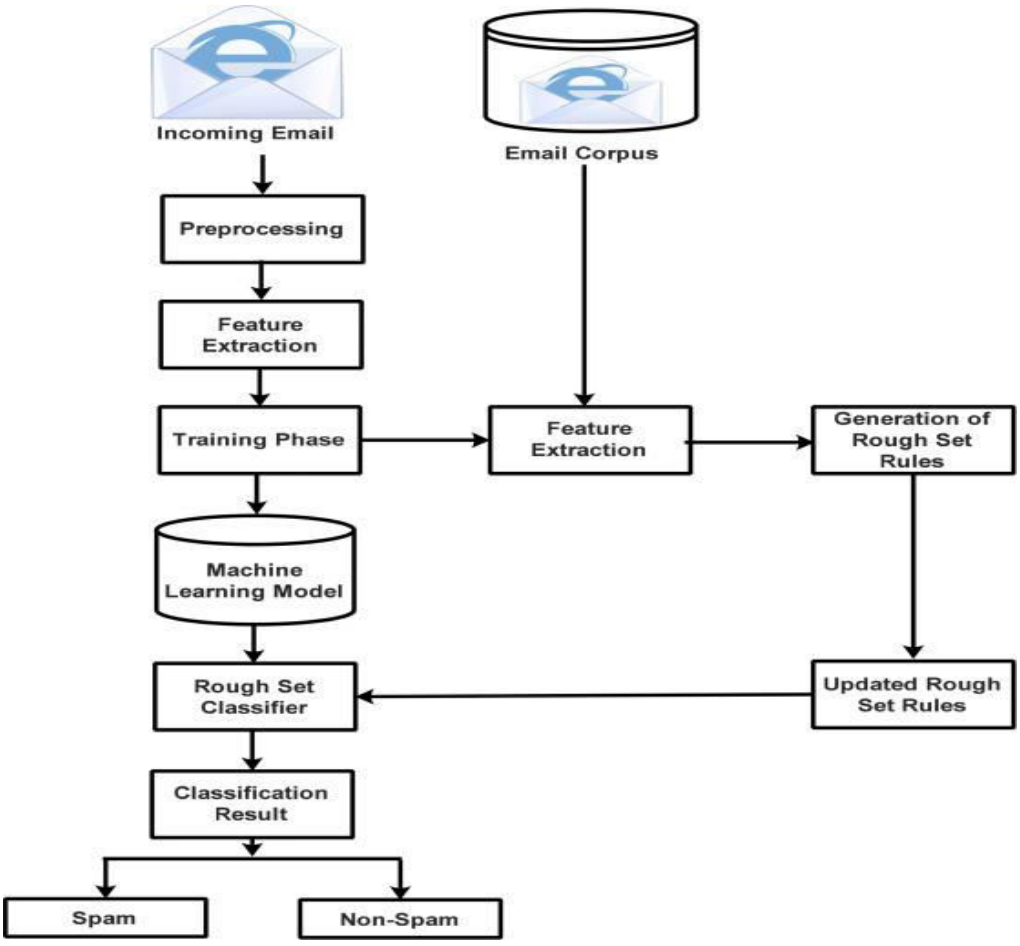


Fig (A)

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411120|

IV. EXPERIMENTAL OUTCOME

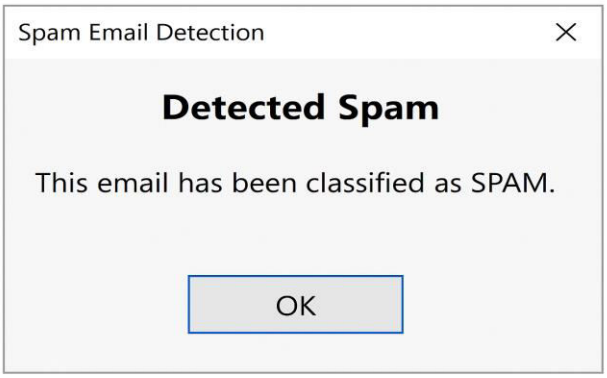


FIG (B)

V. CONCLUSION

The conclusion is that machine learning, particularly through techniques like Naïve Bayes and SVM, provides a highly effective and adaptive solution for spam email detection, surpassing traditional rule-based methods. This approach improves security, increases user productivity, and enhances user experience by accurately filtering unwanted emails and mitigating risks from malicious content. Despite ongoing challenges like evolving spam tactics and data imbalance, continuous improvement in AI models ensures spam detection remains a dynamic and crucial defense.

Key takeaways:
1) Machine learning is superior to traditional methods: Machine learning models can automatically learn and adapt to new spam tactics, unlike traditional rule-based systems that are time-consuming to update manually.
2) Enhanced security and productivity: Effective spam filters protect users from malicious content, such as phishing and malware, while allowing individuals and organizations to focus on legitimate emails.
3) Ongoing challenges: The dynamic nature of spam requires continuous research and improvement. Challenges include the constant evolution of spam tactics, the need for vast datasets for training, and issues like class imbalance.
4) Future directions: Future efforts focus on creating more robust AI models, incorporating user feedback, and enhancing datasets to improve accuracy and efficiency in the face of evolving threats.

REFERENCES

Academic journals and articles:
1) ScienceDirect: Provides access to numerous articles on the topic, such as "Spam email detection using hierarchical attention hybrid deep ..." and "A large-scale empirical analysis of email spam detection through network characteristics in a stand-alone enterprise".
2) IEEE Xplore: Publishes many papers on spam detection, including comparative studies and works on specific algorithms like Naive Bayes and Particle Swarm Optimization.
3) Journal of Survey in Fisheries Sciences (SFS): Features research on using machine learning techniques for spam detection.
4) ITM Web of Conferences: Publishes research on various methods, including Naive Bayes and other machine learning approaches.
5) IJRASET: Features research on spam filtering, including the optimization of neural networks.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details