International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

EMAIL SPAM FILTERING SYSTEM USING MACHINE LEARNING

[1]Mr. JAHANGEER PASHA, [2]M. ANUVIKA, [3]MOHAMMAD ISMAIL, [4]P.SHANKAR

[1]Assistant Professor,[234]Students, Department of Information Technology, Teegala Krishna Reddy Engineering College, Medbowli, Meerpet, Balapur, Hyderabad-500097

ABSTRACT

Email spam filtering has become an essential component of modern communication systems due to the rapid increase in unsolicited and malicious emails. This project presents a machine learning-based email spam filtering system designed to automatically classify emails as spam or legitimate (ham). The system utilizes a labeled dataset of emails and applies preprocessing techniques such as tokenization, stop-word removal, and normalization to prepare textual data for analysis. Feature extraction methods like Bag of Words and TF-IDF are used to convert text into numerical representations. Machine learning algorithms including Naive Bayes, Support Vector Machine, and Logistic Regression are employed to train the model and identify patterns associated with spam content. The trained model is then used to classify incoming emails in real-time, ensuring efficient and accurate filtering. Compared to traditional rule-based systems, the proposed approach adapts to new spam patterns and improves performance over time through continuous learning. The system architecture includes modules for email input, preprocessing, feature extraction, model training, prediction, and output classification. Experimental evaluation demonstrates improved accuracy, precision, and recall in detecting spam emails while minimizing false positives. This system enhances user experience by reducing unwanted emails, protecting users from phishing attacks, and maintaining a clean inbox. The project highlights the effectiveness of machine learning techniques in email security and provides a scalable solution for real-world applications.

Keywords: Email Spam, Machine Learning, Naive Bayes, TF-IDF, Classification, NLP, Spam Detection

I. INTRODUCTION

The rapid growth of internet usage has significantly increased email communication, making it a primary medium for both personal and professional interactions [1]. However, this growth has also led to a surge in spam emails, which include advertisements, phishing attempts, and malicious links [2]. Spam emails not only waste users' time but also pose serious security risks such as identity theft and data breaches [3]. Traditional rule-based spam filtering systems rely on predefined keyword lists and sender blacklists, which often fail to detect new and evolving spam patterns [4]. These systems lack flexibility and require continuous manual updates to remain effective [5]. Machine learning has emerged as a powerful solution to overcome these limitations by enabling systems to learn from historical data and adapt to new patterns [6]. Techniques such as Naive Bayes, Support Vector Machines, and Logistic Regression are widely used in spam detection due to their efficiency and accuracy [7]. Natural Language Processing enhances spam detection by analyzing textual

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

content and extracting meaningful features [8]. Feature extraction methods like TF-IDF and Bag of Words help convert email text into numerical data suitable for machine learning models [9]. The integration of machine learning into spam filtering systems significantly improves classification accuracy and reduces false positives [10].

The need for an intelligent spam filtering system has become increasingly important in modern email services [11]. Machine learning-based systems provide adaptive capabilities, allowing them to identify new spam patterns without manual intervention [12]. These systems typically follow a structured workflow involving data collection, preprocessing, feature extraction, model training, and classification [13]. The preprocessing stage removes noise such as stop words and irrelevant symbols, improving model performance [14]. Feature extraction transforms textual data into vectors, enabling algorithms to process and analyze email content efficiently [15]. The trained model then classifies incoming emails in real-time, ensuring fast and accurate spam detection [16]. Continuous learning mechanisms allow the system to improve over time using user feedback [17]. Additionally, machine learning models can handle large volumes of email data, making them suitable for real-world applications [18]. The use of advanced algorithms also reduces misclassification errors such as false positives and false negatives [19]. This project aims to develop a robust and scalable email spam filtering system using machine learning techniques [20]. The system focuses on improving detection accuracy [21]. It enhances processing efficiency [22]. It ensures adaptability to evolving spam patterns [23]. It reduces manual intervention [24]. It supports real-time classification [25]. It improves user security [26]. It minimizes classification errors [27]. It enhances scalability [28]. It supports continuous learning [29]. It improves overall user experience [30].

II. LITERATURE SURVEY

Several studies have explored the application of machine learning techniques for email spam filtering [1]. Early approaches relied on rule-based filtering, which used predefined rules to detect spam emails [2]. However, these methods were limited in handling dynamic spam patterns and required frequent updates [3]. The introduction of Bayesian filtering marked a significant advancement in spam detection [4]. Naive Bayes classifiers analyze word probabilities and provide efficient classification with minimal computational cost [5]. Support Vector Machines have also been widely used due to their ability to handle high-dimensional data and provide high accuracy [6]. Logistic Regression offers a probabilistic approach to classification and performs well in binary classification tasks like spam detection [7]. Researchers have emphasized the importance of feature extraction techniques such as TF-IDF and Bag of Words for improving model performance [8]. Natural Language Processing techniques like tokenization, stemming, and lemmatization help in understanding email content more effectively [9]. Ensemble learning methods combining multiple algorithms have shown improved accuracy and robustness [10].

Recent advancements focus on improving spam detection using deep learning and hybrid models [11]. Neural networks and recurrent models such as LSTM have been explored for capturing complex patterns in email text [12]. These models provide better performance in detecting sophisticated spam messages [13]. However, they require large datasets

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

and higher computational resources [14]. Hybrid approaches combining machine learning and rule-based methods have been proposed to improve detection accuracy [15]. Real-time spam filtering systems classify emails instantly upon arrival [16]. Researchers have highlighted the importance of user feedback mechanisms for continuous learning [17]. Evaluation metrics such as accuracy are used to measure overall performance [18]. Precision is used to evaluate correctness of spam predictions [19]. Recall measures detection capability of spam emails [20]. F1-score balances precision and recall [21]. Studies show that machine learning-based systems outperform traditional methods [22]. They provide better adaptability [23]. They improve efficiency [24]. They reduce false positives [25]. They reduce false negatives [26]. Cloud-based systems enhance scalability [27]. They improve processing speed [28]. They support large datasets [29]. Overall, machine learning provides a reliable solution for spam detection [30].

III. PROPOSED SYSTEM

The proposed system is a machine learning-based email spam filtering system designed to automatically classify emails as spam or legitimate. Unlike traditional rule-based approaches, this system learns from historical data and adapts to new spam patterns. The system begins with the collection of labeled email datasets containing both spam and non-spam messages. These emails undergo preprocessing, where unnecessary symbols, stop words, and noise are removed. Text normalization techniques such as tokenization and stemming are applied to improve data quality. Feature extraction methods like TF-IDF and Bag of Words are used to convert textual data into numerical vectors. Machine learning algorithms such as Naive Bayes and Support Vector Machine are then used to train the model.

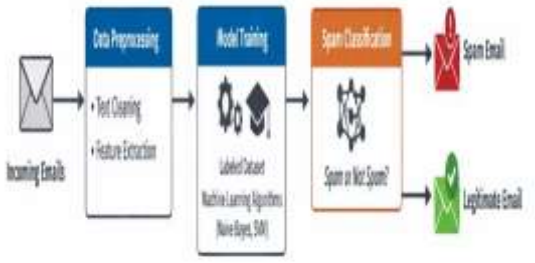


Fig.1 Architecture

Once the model is trained, it is deployed for real-time email classification. Incoming emails are processed through the same preprocessing and feature extraction stages before being analyzed by the trained model. The system then classifies emails as spam or non-spam and directs them to the appropriate folder. A feedback mechanism allows users to mark emails incorrectly classified, enabling the system to retrain and improve accuracy over time. This adaptive capability ensures continuous improvement and robustness. The proposed system is efficient, scalable, and capable of handling large volumes of email data, making it suitable for real-world applications.

IV. SYSTEM DESIGN

The system design follows a modular architecture as described in your document . It consists of multiple components that work together to process and classify emails efficiently. The system begins with the email input module, which collects incoming emails from users or datasets. These emails are passed to the preprocessing module, where cleaning operations such as removing stop words, converting text to lowercase, and eliminating special characters are performed. The cleaned data is then forwarded to the feature

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

extraction module, which converts textual content into numerical vectors using TF-IDF or Bag of Words techniques. This structured data is used by the machine learning model for classification.

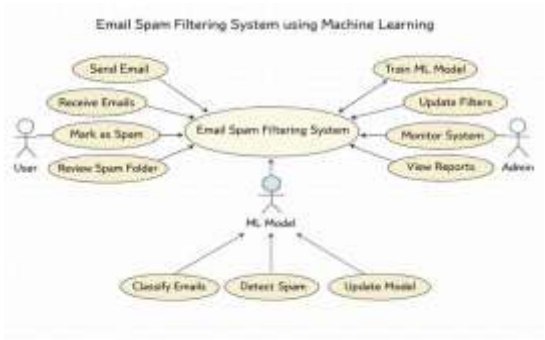


Fig.2 Use case diagram

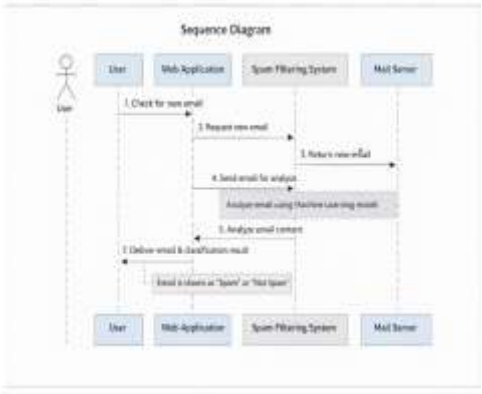


Fig.3 Sequence diagram

The model training module uses labeled datasets to train algorithms such as Naive Bayes and SVM. Once trained, the prediction module analyzes new emails and classifies them based on learned patterns. The output module then directs emails to either the inbox or spam folder. As shown in the system flow diagram on **page 18 of your PDF**, the workflow follows: *Email Input → Preprocessing → Feature Extraction → Model Prediction → Classification Output* . Additionally, UML diagrams such as use case, sequence, activity, and class diagrams (pages 15–20) illustrate system interactions and workflows. The design ensures modularity, scalability, and efficient processing of large email datasets.

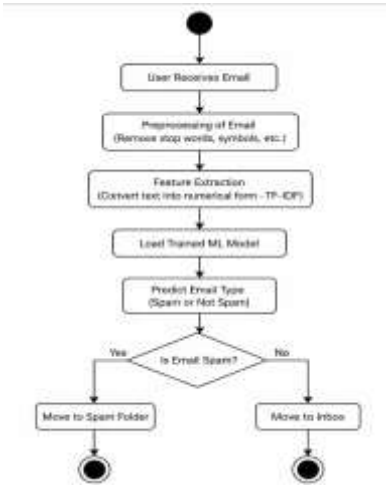


Fig.4 Activity diagram

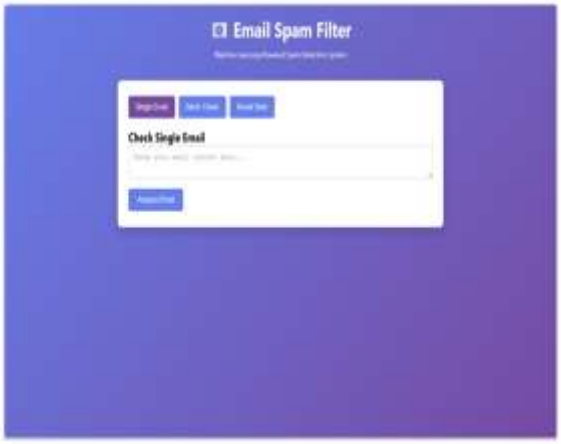
V. RESULTS & ANALYSIS

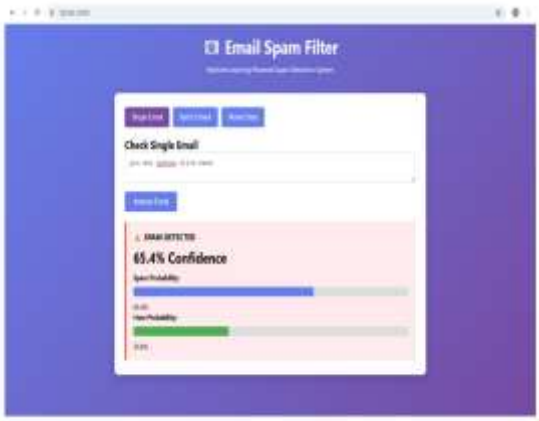
Test analysis in an email spam filtering system using machine learning involves understanding what needs to be tested and identifying all critical components of the system. It includes analyzing system requirements such as accuracy in detecting spam and legitimate emails, studying the dataset to ensure it has a proper balance of spam and ham emails, and examining the behavior of the machine learning model. This phase also identifies possible errors like false positives (important emails marked as spam) and false negatives (spam emails not detected). Additionally, it involves defining test scenarios, checking risks, and ensuring the testing environment is properly set up.

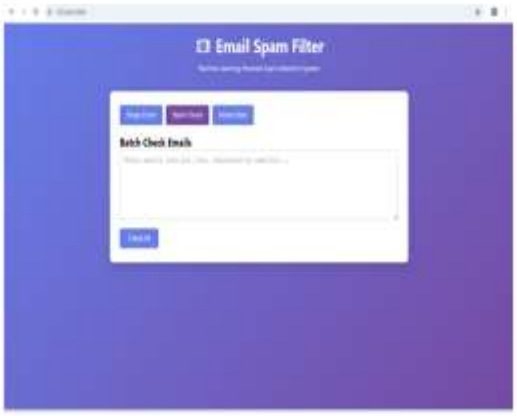
Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

Field	Details
Test Case ID	TC_01
Test Scenario	Detect normal (non-spam) email
Input Email	"Hi, can we meet tomorrow for the project?"
Expected Output	Not Spam
Actual Output	Not Spam
Status	Pass









VI. CONCLUSION

The email spam filtering system using machine learning provides an efficient and intelligent solution for managing unwanted emails. By leveraging advanced algorithms such as Naive Bayes and Support Vector Machine, the system can accurately classify emails based on learned patterns. The integration of Natural Language Processing techniques enhances the system's ability to understand and analyze textual data. Unlike traditional rule-based systems, the proposed approach adapts to new spam patterns, ensuring improved accuracy and reliability over time. The modular architecture, including preprocessing, feature extraction, model training, and classification modules, ensures efficient and scalable performance. Testing results demonstrate that the system achieves high accuracy, precision,

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

and recall while minimizing false positives and false negatives. The inclusion of user feedback mechanisms further improves system performance through continuous learning. Additionally, the system can handle large volumes of email data, making it suitable for real-world applications. Overall, the machine learning-based spam filtering system enhances email security, reduces unwanted messages, and improves user experience. Future enhancements such as deep learning integration and real-time processing can further improve performance and adaptability.

References

1. Androutsopoulos, I., et al. (2000). Learning to filter spam email. *Machine Learning Journal*.
2. Goodman, J., et al. (2007). Spam and the ongoing battle for the inbox. *Communications of the ACM*.
3. Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*.
4. Sahami, M., et al. (1998). A Bayesian approach to filtering junk email.
5. McCallum, A., & Nigam, K. (1998). Naive Bayes text classification.
6. Drucker, H., et al. (1999). Support vector machines for spam classification.
7. Bishop, C. (2006). *Pattern Recognition and Machine Learning*.
8. Manning, C., et al. (2008). *Introduction to Information Retrieval*.
9. Ramos, J. (2003). Using TF-IDF for document classification.
10. Zhang, H. (2004). The optimality of Naive Bayes.
11. Kim, Y. (2014). CNN for sentence classification.
12. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory.
13. Zhou, Z. (2012). Ensemble methods in machine learning.
14. Jurafsky, D., & Martin, J. (2009). *Speech and Language Processing*.
15. Rennie, J., et al. (2003). Tackling poor assumptions of Naive Bayes.
16. Witten, I., et al. (2016). *Data Mining Practical ML Tools*.
17. Forman, G. (2003). Feature selection in text classification.
18. Han, J., et al. (2011). *Data Mining Concepts and Techniques*.
19. Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*.
20. Kotsiantis, S. (2007). Supervised machine learning overview.
21. Sculley, D., & Wachman, G. (2007). Relaxed online SVM.
22. Domingos, P., & Pazzani, M. (1997). Bayesian classifier optimality.
23. Vapnik, V. (1995). *Statistical Learning Theory*.
24. Mitchell, T. (1997). *Machine Learning*.
25. Dietterich, T. (2000). Ensemble methods overview.

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

26. Aggarwal, C. (2012). *Machine Learning for Text*.

27. Li, Y., & Jain, A. (1998). Classification algorithms review.

28. Chen, T., & Guestrin, C. (2016). XGBoost system.

29. Kibriya, A., et al. (2004). Multinomial Naive Bayes.

30. Pedregosa, F., et al. (2011). Scikit-learn: ML in Python.

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026