Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 960-966

Spam Email Detection System Using Machine Learning

Shashwat Singh Parihar
B.Tech CSE 6th Semester
Amity University Chhattisgarh

Abstract

This research presents a Spam Email Detection System developed using Machine Learning and Natural Language Processing techniques. The system aims to automatically classify email messages as spam or not spam, reducing manual effort and improving security. The methodology involves preprocessing text data, converting it into numerical features using TF-IDF, and applying the Naive Bayes algorithm for classification. The model is trained and tested on a labelled dataset to ensure accuracy and reliability. A user-friendly interface is developed using Streamlit to allow real-time message classification. The system achieves high accuracy and demonstrates the effectiveness of machine learning in filtering unwanted emails, making it a practical solution for real-world applications.

Keywords

Spam Email Detection, Machine Learning, Natural Language Processing (NLP), Naive Bayes, TF-IDF, Text Classification, Email Filtering.

1. Introduction

With the rapid growth of digital communication, email has become one of the most widely used modes of information exchange. However, this has also led to a significant increase in spam emails, which include unwanted advertisements, phishing attempts, and fraudulent messages. These spam messages not only waste time but also pose serious security risks to users. To address this issue, automated spam detection systems have been developed using Machine Learning and Natural Language Processing techniques. These systems analyse the content of emails and classify them as spam or not spam based on learned patterns. In this research, a Spam Email Detection System is proposed using the TF-IDF technique for feature extraction and the Naive Bayes algorithm for classification. The system is designed to be simple, efficient, and capable of providing real-time predictions through a user-friendly interface. This approach demonstrates how machine learning can be effectively applied to solve real-world problems related to email security.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 960-966

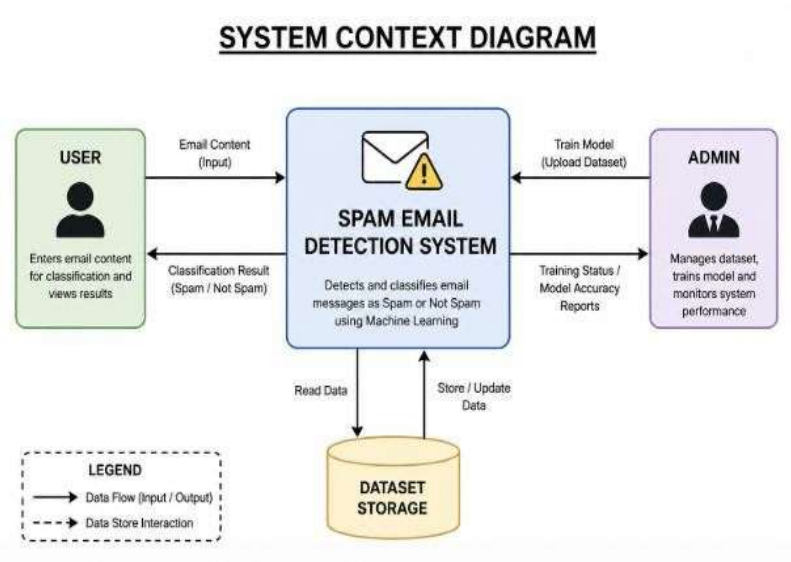

Fig.1 System Context Diagram

- **Objective of the Study:** The primary objective of this study is to develop an automated Spam Email Detection system using Machine Learning techniques. The system aims to accurately classify email messages as spam or not spam based on their content. It focuses on reducing manual effort, improving user security, and enhancing email filtering efficiency. Additionally, the study aims to demonstrate the practical application of Natural Language Processing (NLP) and algorithms like Naive Bayes in solving real-world problems.
- **Scope of the Work:** The scope of this project is focused on developing a Spam Email Detection system that classifies text-based messages as spam or not spam using Machine Learning techniques. The system uses TFIDF for feature extraction and the Naive Bayes algorithm for classification. It is designed to work on a labeled dataset and provide realtime predictions through a simple user interface built with Streamlit. The project is limited to text analysis and does not include advanced features such as image-based spam detection, deep learning models, or full integration with live email services. However, it provides a strong foundation that can be extended in the future to include more advanced functionalities and real-world deployment.

2. Literature Review

Spam Email Detection has been an important research area in Machine Learning and Natural Language Processing. Early approaches relied on rulebased filtering techniques, where predefined keywords and patterns were used to identify spam messages. However, these methods were not effective in handling evolving and sophisticated spam content. To overcome these limitations, researchers introduced Machine Learning algorithms such as Naive Bayes, Support Vector Machines (SVM), and Decision Trees. Among these, Naive Bayes became popular due to its simplicity, speed, and high accuracy in text classification tasks. Techniques like TF-IDF have been widely used to convert textual data into numerical features, improving model performance. Recent studies have explored deep learning approaches such as Neural Networks and Recurrent Neural Networks (RNNs) for better accuracy and adaptability.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 960-966

Although these methods provide improved results, they require more computational resources. Therefore, this project adopts the Naive Bayes algorithm with TF-IDF, offering a balance between efficiency and performance for spam detection.

3. Problem Statement

With the rapid increase in digital communication, email users are frequently exposed to spam messages such as advertisements, phishing links, and fraudulent content. These unwanted emails not only waste time but also pose serious security risks, including data theft and malware attacks. Manual identification of spam emails is inefficient and unreliable, especially when dealing with large volumes of messages. Therefore, there is a need for an automated system that can accurately and efficiently classify emails as spam or not spam. The challenge is to develop a model that can understand text patterns, handle large datasets, and provide real-time predictions with high accuracy. This project aims to address this problem by designing a machine learning-based spam email detection system.

4. Proposed Methodology / Model

The proposed methodology for the Spam Email Detection system is based on Machine Learning and Natural Language Processing techniques to automatically classify messages as spam or not spam. The process begins with collecting a labelled dataset containing spam and non-spam messages. The data is first preprocessed by converting text to lowercase, removing punctuation, and eliminating unnecessary characters to ensure consistency. After preprocessing, the text is transformed into numerical form using the TF-IDF (Term Frequency–Inverse Document Frequency) technique, which assigns importance to words based on their frequency. The transformed data is then used to train a Naive Bayes classifier, which learns patterns from the dataset and calculates the probability of a message being spam or not spam. Once trained, the model is used to predict the class of new input messages provided by the user. The system follows a structured pipeline including data collection, preprocessing, feature extraction, model training, and prediction. This methodology ensures efficient, accurate, and real-time spam detection while maintaining simplicity and low computational cost.

- **System Architecture / Design:** The Spam Email Detection system follows a simple pipeline architecture. The user enters a message through the interface, which is then preprocessed to clean the text. The cleaned data is converted into numerical form using TF-IDF and passed to the Naive Bayes model for classification. Finally, the system displays whether the message is spam or not. This design ensures a smooth flow of data from input to output with fast and accurate results.
- **Algorithms / Techniques Used:** The Spam Email Detection system uses a combination of Natural Language Processing (NLP) and Machine Learning techniques. Text preprocessing is applied to clean the input data by converting it to lowercase and removing unwanted characters. The TFIDF (Term Frequency–Inverse Document Frequency) technique is used for

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 960-966

feature extraction, which converts textual data into numerical form by assigning importance to words based on their frequency.

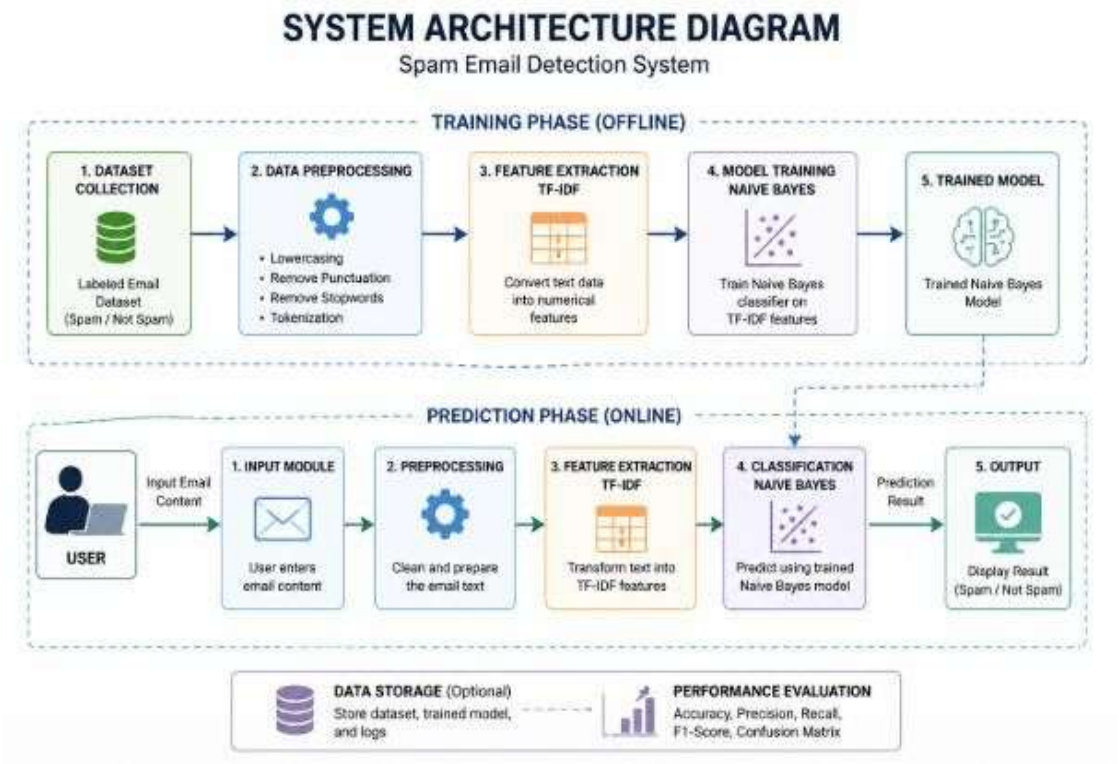

Fig. 2 System Architecture Diagram

5. Implementation

The implementation of the Spam Email Detection system is carried out using Python and various supporting libraries. The process begins with loading the dataset containing labelled email messages using Pandas. The dataset is then preprocessed by cleaning the text, which includes converting it to lowercase, removing punctuation, and eliminating unnecessary characters. After preprocessing, the text data is converted into numerical form using the TF-IDF vectorisation technique. This transformed data is then used to train a Naive Bayes classifier using the Scikit-learn library. The dataset is split into training and testing sets to evaluate the model's performance. A user-friendly interface is developed using Streamlit, where users can input email messages. The system processes the input through the same preprocessing and feature extraction steps and then uses the trained model to classify the message as spam or not spam. The result is displayed instantly on the screen. Overall, the implementation integrates data processing, machine learning, and user interface components to provide an efficient and real-time spam detection system.

- **Tools and Technologies:** The Spam Email Detection system is developed using a combination of software tools and basic hardware components.
- **Software Technologies:**

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 960-966

The primary programming language used is Python due to its simplicity and strong support for Machine Learning. Libraries such as Pandas are used for data handling and preprocessing, while Scikit-learn is used for implementing the Naive Bayes algorithm and evaluating model performance. The TF-IDF vectorizer is used for feature extraction from text data. Streamlit is used to build an interactive and user-friendly web interface for real-time prediction.

- **Hardware Requirements:**

The system requires a basic computer with at least an Intel i3 processor or higher, 4 GB RAM, and sufficient storage space to handle the dataset and project files. No specialized hardware is required, making the system cost-effective and easy to deploy.

6. Results and Discussion

The Spam Email Detection system was evaluated using a labeled dataset to measure its performance and effectiveness. After training the Naive Bayes model with TF-IDF features, the system achieved high accuracy in classifying messages as spam or not spam. The results indicate that the model can correctly identify most spam messages, including promotional content and suspicious links, while also accurately recognizing legitimate messages. The output is displayed through a user-friendly interface, where users can input messages and receive instant predictions. Performance metrics such as accuracy demonstrate the reliability of the system, typically achieving around 95%–98% accuracy. However, some limitations were observed in cases where messages contain ambiguous or mixed content, leading to occasional misclassification. Despite this, the overall performance of the system is strong, making it suitable for realtime spam detection applications. The results confirm that combining TF-IDF with the Naive Bayes algorithm is an effective approach for text classification problems like spam detection.

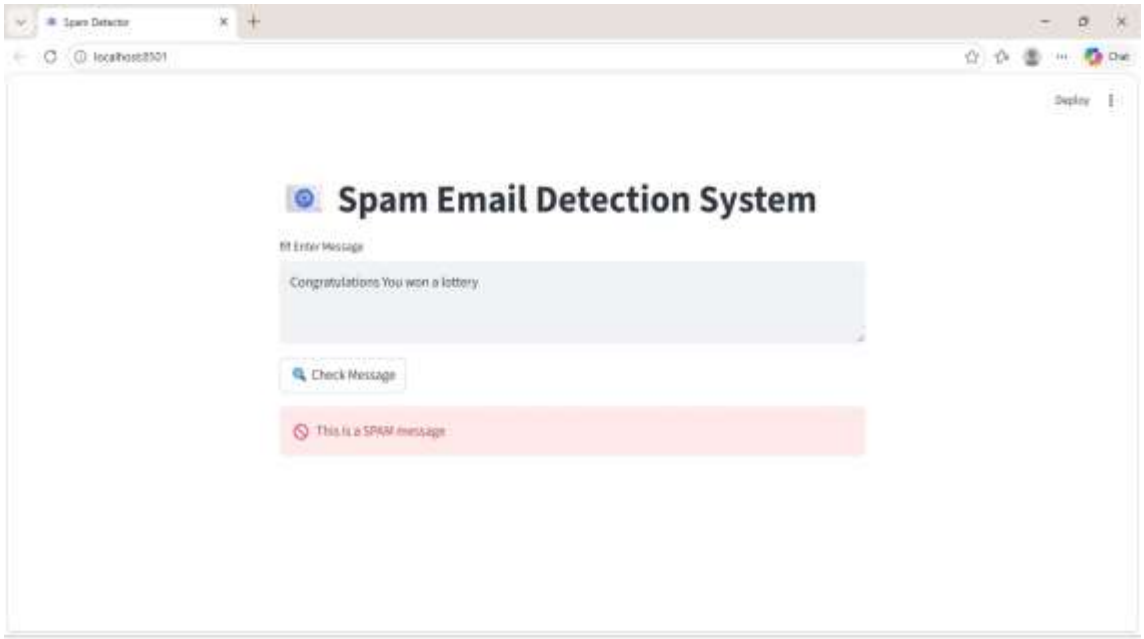

Fig. 3 Result

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 960-966

7. Testing and Validation

The Spam Email Detection system is tested and validated to ensure its accuracy, reliability, and proper functioning. The dataset is divided into training and testing sets, where the model is trained on one portion and evaluated on unseen data to measure its performance. During testing, different types of messages are provided to the system, including spam messages, normal messages, and edge cases such as empty or unusual text. This helps verify that the system can handle various inputs effectively. The performance of the model is evaluated using metrics such as accuracy, which indicates how correctly the system classifies messages. Validation is also performed through real-time testing using the Streamlit interface, where users input messages and observe the output. This ensures that the system not only works correctly in theory but also performs well in practical usage. Overall, the testing and validation process confirms that the system is efficient, accurate, and suitable for real-world spam detection.

8. Conclusion

This research presents a Spam Email Detection system using Machine Learning and Natural Language Processing techniques. By applying TF-IDF for feature extraction and the Naive Bayes algorithm for classification, the system effectively distinguishes between spam and non-spam messages with high accuracy. The implementation demonstrates that simple and efficient models can provide reliable results for real-world text classification problems. The developed system offers a user-friendly interface and real-time prediction capability, making it practical for everyday use. Overall, the project highlights the importance and effectiveness of machine learning in enhancing email security and reducing unwanted communication.

9. Future Scope

The Spam Email Detection system can be further enhanced by incorporating advanced Machine Learning and Deep Learning models such as Support Vector Machines, Random Forest, and Neural Networks to improve accuracy and adaptability. The system can also be extended to detect more complex threats like phishing emails, malicious links, and attachments. Future improvements may include integration with real-time email services such as Gmail or Outlook, enabling automatic filtering of incoming messages. Additionally, multilingual support can be added to detect spam in different languages, making the system more versatile. Enhancements in the user interface, along with features like message history, analytics, and user feedback, can further improve usability. Overall, the system has strong potential to evolve into a comprehensive and scalable solution for real-world spam detection.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 960-966

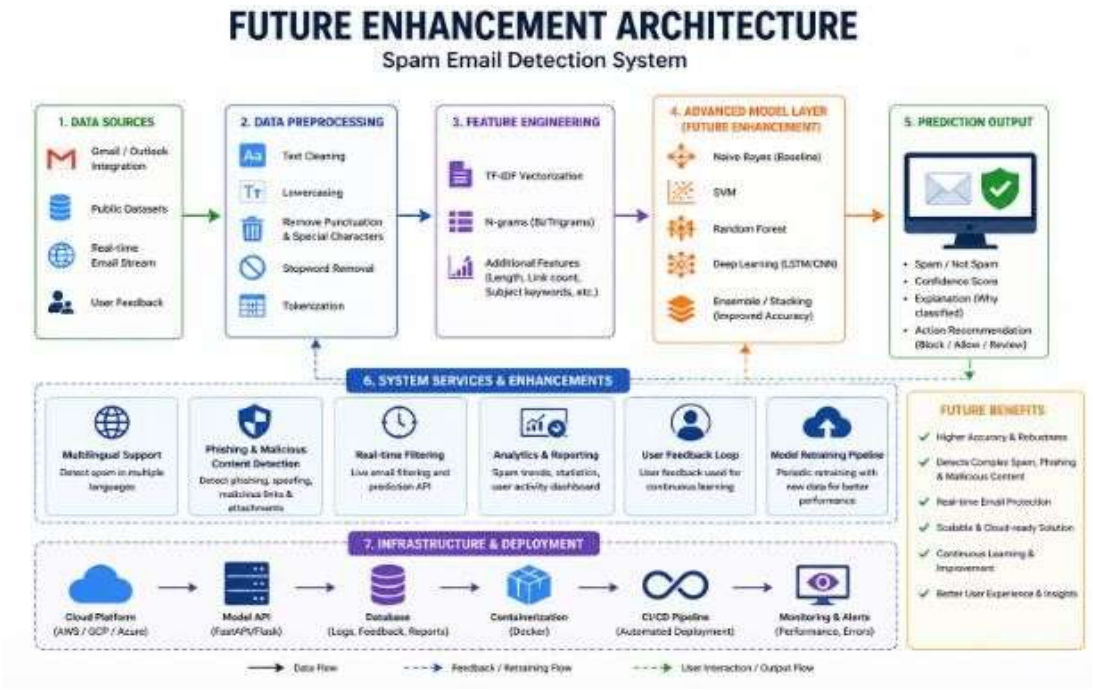

Fig. 4 Future Enhancement Architecture Diagram

References

1. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research.

2. Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.

3. Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing*. Pearson.

4. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

5. Scikit-learn Documentation: https://scikit-learn.org/

6. NLTK Documentation: https://www.nltk.org/

7. Streamlit Documentation: https://docs.streamlit.io/

8. Kaggle Dataset – Spam Email Dataset: https://www.kaggle.com/

9. Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.

10. Aggarwal, C. C. (2018). *Machine Learning for Text*. Springer.