

Spam Detection System Using Machine Learning

Shaikh Shahid Mahemood¹, Prof. N. M. Kolhe², Prof. D. S. Kshirsagar³, Dr. H. B. Jadhav⁴

Student, Computer Department¹

Assistant Professor, Computer Department^{2,3}

Head of Department, Computer Department⁴

Ansul's Technical Campus, Ahilyanagar, Maharashtra, India

Abstract: Email and messaging services have become one of the most important methods of communication in today's digital world. However, along with the growth of communication technologies, the number of unwanted messages known as spam has also increased significantly. Spam messages usually contain advertisements, fraudulent offers, phishing links, or malicious content that can cause security threats to users and organizations. Detecting these spam messages manually is difficult due to the large volume of messages received every day. Therefore, an intelligent system is required to automatically identify and filter spam messages. The objective of this project is to develop a Spam Email Detection System using Machine Learning techniques that can automatically classify messages as Spam or Safe (Ham). The system analyzes the content of messages using Natural Language Processing (NLP) techniques and converts the text data into numerical format using TF-IDF vectorization. A machine learning algorithm such as Multinomial Naive Bayes is used to train the model and make predictions based on previously learned data. To make the system user-friendly, a web-based interface is developed using Streamlit. The application provides features such as user login, signup, message analysis, dashboard visualization, and feedback system. Users can enter a message into the system, and the trained model will instantly predict whether the message is spam or safe. Additionally, the dashboard displays graphical representations such as bar charts showing the number of spam and safe messages, which helps users understand the system's performance and statistics.

Keywords: Machine Learning, spam, Data, Authentication

I. INTRODUCTION

In the modern digital world, electronic communication such as email and messaging services plays a very important role in both personal and professional environments. Millions of messages are exchanged every day through email platforms, social media applications, and messaging services. However, along with useful communication, users also receive a large number of unwanted or irrelevant messages commonly known as spam messages. Spam messages usually contain advertisements, fake offers, phishing links, or malicious content intended to mislead users. These messages not only waste time but may also create serious security threats such as identity theft, financial fraud, and malware attacks. As the number of internet users continues to grow, the volume of spam messages is also increasing rapidly. Manually identifying and filtering spam messages is a difficult and time consuming task. Traditional filtering methods that rely on simple keyword matching or rule based systems are not always effective because spammers frequently change the structure and content of their messages. Therefore, more intelligent techniques are required to automatically detect and filter spam messages in an efficient manner. Machine Learning provides an effective solution to this problem. Machine learning algorithms can analyze large amounts of data, learn patterns from previously labeled messages, and classify new messages as spam or legitimate (safe) automatically. By using Natural Language Processing techniques, the system can understand the textual content of messages and convert it into a format that machine learning models can process. These techniques allow the system to detect spam messages with higher accuracy compared to traditional methods. The main goal of this project is to develop a Spam Email Detection System using



Machine Learning that can automatically identify whether a message is spam or safe. The system uses machine learning algorithms such as Multinomial Naive Bayes along with text preprocessing and TF-IDF vectorization to analyze message content and make predictions. In addition to the machine learning model, the project also includes a user-friendly web interface built using Streamlit.

II. PROBLEM STATEMENT

With the rapid growth of internet usage and digital communication, email and messaging services have become an essential part of daily life. However, along with legitimate communication, users also receive a large number of unwanted messages known as spam. Spam messages often contain advertisements, fake promotional offers, phishing links, or malicious content that may lead to security risks such as fraud, data theft, and malware attacks. These unwanted messages create inconvenience for users and reduce the efficiency of communication systems. Traditional spam filtering techniques mainly rely on rule-based or keyword-based approaches. These methods are not always reliable because spammers continuously modify message content to bypass filters. As a result, many spam messages still reach users' inboxes, while sometimes legitimate messages may also be incorrectly classified as spam. Therefore, there is a need for an intelligent and automated system that can accurately detect and filter spam messages. Machine learning techniques provide a powerful solution by learning patterns from previously labeled data and automatically classifying new messages as spam or safe. The problem addressed in this project is to design and develop a Spam Email Detection System using Machine Learning that can analyze message content, classify messages efficiently, and provide a user-friendly interface for message testing and monitoring through a web-based dashboard.

III. OBJECTIVES

The main objectives of the Spam Email Detection System project are as follows:

- To develop an intelligent system that can automatically detect and classify messages as spam or safe using machine learning techniques.
- To build a model that can learn patterns from previously labeled data and accurately predict whether a message is spam or ham.
- To design a user-friendly web interface using Streamlit that allows users to easily enter messages and check spam detection results.
- To implement user authentication features such as login and signup to provide secure access to the system.
- To create a dashboard for that displays statistics of spam and safe messages using charts.
- To include a feedback system where users can provide feedback about the prediction results to improve the system in the future.
- To demonstrate the practical use of AI in solving real-world problems related to email and message filtering.

IV. SCOPE

The scope of the Spam Email Detection System project focuses on developing an intelligent system that can automatically identify and classify messages as spam or safe using machine learning techniques. The system analyzes the textual content of messages and applies Natural Language Processing methods to process the data before making predictions. By using machine learning algorithms, the system learns patterns from previously labeled messages and improves its ability to detect spam effectively. This project also includes the development of a web-based application using Streamlit, which provides an interactive and user-friendly interface for users. Through this interface, users can sign up, log in, and test messages to check whether they are spam or legitimate. The system also includes a dashboard that visually represents spam and safe messages using bar charts, helping users understand message statistics more clearly. The project can be useful in various communication platforms such as email services, messaging applications, and online communication systems where spam messages are a common problem. It demonstrates how machine learning and data science techniques can be applied to improve message filtering and enhance communication security.



Although the system works effectively for basic spam detection, it can be further expanded in the future by integrating it with real-time email systems, improving the machine learning models, and deploying it on cloud platforms for large-scale applications.

V. SYSTEM ARCHITECTURE

1. Software Requirements

- Operating System :- Windows 11
- Programming Language :- Python 3
- Development Environment :- Jupyter Lab
- Libraries and Frameworks :- a. Pandas b. NumPy c. Scikit-learn d. Streamlit
- Dataset : Spam SMS Dataset (CSV format)

2. Hardware Requirement

- Processor :- Intel Core i5
- RAM :- 8 GB RAM
- Storage :- 4 GB free disk space
- System Type :- 64-bit computer system
- Internet Connection :- Required for installing libraries and accessing datasets

VI. ARCHITECTURE

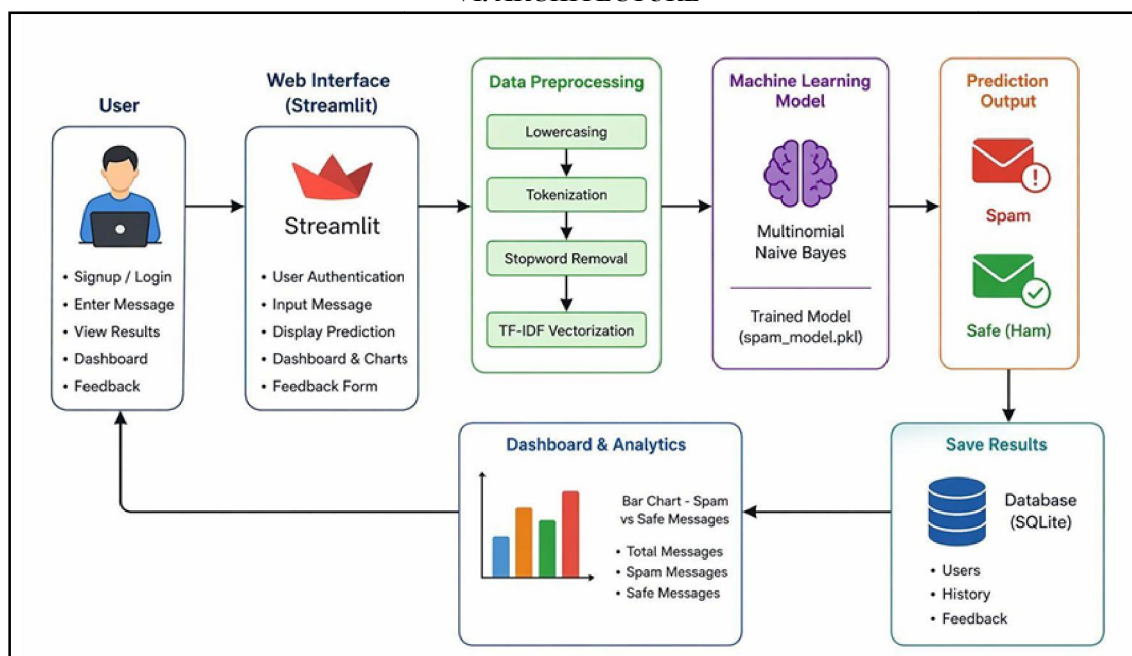


Fig 1. Architecture

1. User Module

Allows users to sign up or log in to the system.
 Enables users to enter an email or message for spam detection.
 Displays the prediction results to users.
 Provides access to the dashboard, message history, and feedback options.



2. Web Interface (Streamlit)

Provides a simple and user-friendly interface using the Streamlit framework.
Handles user authentication and message input.
Displays the spam detection result.
Provides dashboard visualizations, charts, and a feedback form.
Connects the user interface with the preprocessing and machine learning modules.

3. Data Preprocessing Module

This module converts raw text into a suitable format for machine learning.
Lowercasing: Converts all text into lowercase to maintain consistency.
Tokenization: Splits the message into individual words or tokens.
Stopword Removal: Removes common words that may contribute little to classification.
TF-IDF Vectorization: Converts the processed text into numerical features based on the importance of words in the message.

4. Machine Learning Model

Uses the Multinomial Naive Bayes algorithm to classify messages.
Receives the numerical features generated by the TF-IDF vectorizer.
Uses a trained model stored in the spam_model.pkl file.
Calculates the predicted class based on learned patterns in the training data.
Produces one of two classifications: Spam or Safe (Ham).

5. Prediction Output Module

Receives the classification generated by the machine learning model.
Displays a spam warning when the message is classified as spam.
Displays a safe confirmation when the message is classified as ham.
Presents the prediction in an understandable format using visual indicators.

6. Save Results Module

Stores relevant system information in an SQLite database.
Maintains user details, message prediction history, and feedback, as indicated in the diagram.
Supports the retrieval of previously saved results.
Provides stored information for dashboard analysis.

7. Dashboard and Analytics Module

Displays statistics about messages processed by the system.
Shows a bar chart comparing spam and safe messages.
Presents the total number of messages, spam messages, and safe messages.
Helps users understand the system's classification results through visual summaries.
Uses saved results to support historical analysis.

8. Feedback Module

Allows users to provide feedback about the prediction.
Helps identify messages that may have been classified incorrectly.
Stores feedback for future review and possible model improvement.



VII. WORKING FLOW

The working flow of the Spam Email Detection System describes the sequence of steps through which the system processes the input message and generates the prediction result.

The flow of the system is explained below:

- 1. User Accesses the System:** The user opens the web application through the browser and accesses the main interface of the Spam Detection System.
- 2. User Authentication:** The user logs into the system using valid login credentials. If the user is new, they can create an account using the signup option.
- 3. Message Input:** After logging in, the user enters a message (email or SMS) into the text input field provided on the detection page. 5
- 4. Text Preprocessing:** The system processes the entered message by performing text cleaning operations such as converting text to lowercase, removing punctuation, and eliminating unnecessary words.
- 5. Feature Extraction:** The cleaned message is converted into numerical format using the TF-IDF vectorization technique, which helps the machine learning model understand the importance of words in the message.
- 6. Spam Classification:** The numerical data is passed to the trained Machine Learning model (Naive Bayes classifier). The model analyzes the message and predicts whether it is Spam or Safe (Ham).
- 7. Display of Result:** The prediction result is displayed on the user interface, informing the user whether the message is spam or not.
- 8. Data Storage:** The message and its prediction result are stored in the system database for future reference and analysis.
- 9. Dashboard Visualization:** The dashboard updates the statistics of spam and safe messages and displays them using charts such as bar graphs.
- 10. User Feedback:** Users can submit feedback about the prediction results, which can be used to improve the system performance in future updates.

VIII. ADVANTAGES & LIMITATIONS

Advantages

The Spam Email Detection System provides several advantages that help improve communication efficiency and security. The main advantages are listed below:

Automatic Spam Detection

The system automatically identifies and filters spam messages without requiring manual checking.

Improves Communication Security

It helps protect users from fraudulent messages, phishing links, and malicious content.

Saves Time and Effort

Users do not need to manually sort or delete unwanted messages, which saves time and effort.

High Accuracy

By using machine learning algorithms, the system can classify messages with good accuracy.

User-Friendly Interface

The system provides an easy-to-use web interface where users can quickly check messages.

Real-Time Message Analysis

The system can instantly analyze messages and display whether they are spam or safe.

Data Visualization

The dashboard provides graphical representation of spam and safe messages using charts.

Learning Capability

Machine learning models improve performance as more data becomes available.



Useful for Multiple Platforms

The system can be applied to email services, messaging applications, and online communication platforms.

Scalable System

The system can be further expanded with advanced models and integrated with larger communication systems.

Limitations

Although the Spam Email Detection System provides an effective solution for detecting unwanted messages, it also has some limitations. These limitations are described below:

Dependence on Training Data

The accuracy of the system depends on the quality and size of the dataset used to train the machine learning model.

Difficulty in Detecting New Spam Patterns

If spammers use completely new words or patterns that are not present in the training data, the system may fail to detect them.

Possibility of Misclassification

Sometimes legitimate messages may be classified as spam or spam messages may be classified as safe.

Limited Dataset

If the dataset used for training is small, the model may not perform well in real-world scenarios.

Language Dependency

The system mainly works with the language used in the dataset, so it may not perform well with messages written in other languages.

Requires Regular Updates

The model needs to be updated regularly with new data to maintain accuracy.

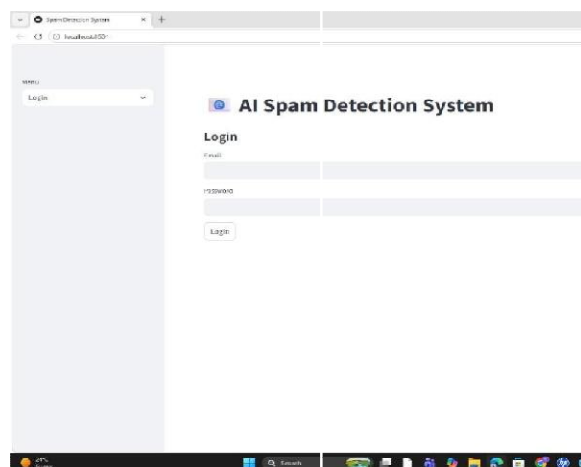
Basic Machine Learning Model

Using simple algorithms like Naive Bayes may not capture very complex patterns in spam messages compared to advanced deep learning models.

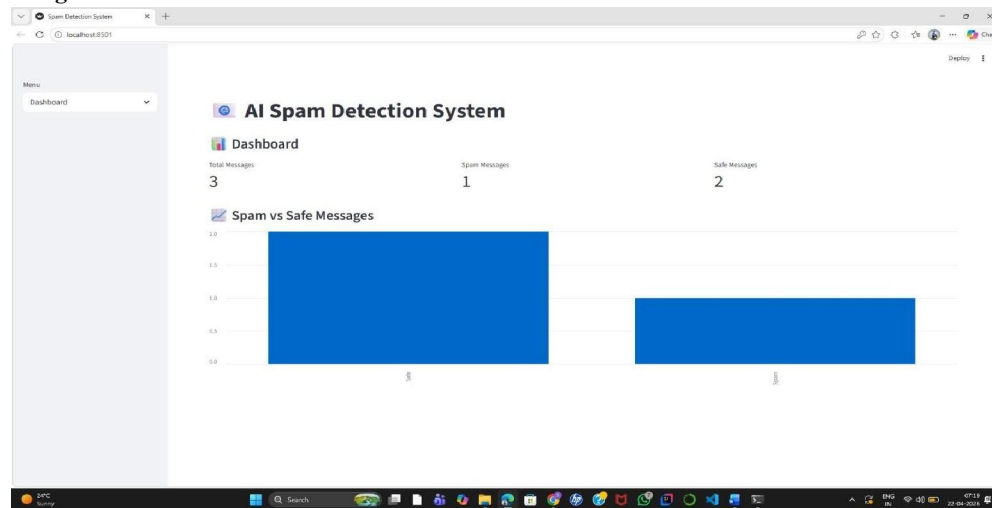
No Direct Integration with Email Servers

IX. RESULTS

Login Page



Dashboard Page



X. FUTURE SCOPE

The Spam Email Detection System can be further improved and expanded in several ways to make it more powerful and useful. The possible future enhancements of the system are listed below:

Use of Advanced Machine Learning Models.

In the future, advanced algorithms such as Deep Learning, LSTM, or BERT can be used to improve the accuracy of spam detection.

Integration with Real Email Services.

The system can be integrated with real email platforms such as Gmail or Outlook to automatically filter spam messages from user inboxes.

Multi-language Support

The system can be extended to detect spam messages written in different languages.

Cloud Deployment

The application can be deployed on cloud platforms so that it can be accessed by a large number of users from anywhere.

Mobile Application Development

A mobile application version can be developed to allow users to detect spam messages directly from their smartphones.

Continuous Model Training

The system can be designed to continuously learn from new messages and user feedback to improve prediction accuracy.

Real-time Spam Monitoring Dashboard

A more advanced dashboard can be developed to display real-time analytics and trends of spam messages.

Integration with Messaging Platforms

The system can be integrated with messaging applications such as WhatsApp, Telegram, or SMS services to detect spam messages automatically.

XI. CONCLUSION

The Spam Email Detection System using Machine Learning is designed to automatically identify and classify messages as spam or safe. With the rapid increase in digital communication, spam messages have become a common problem that can cause inconvenience, waste time, and sometimes lead to serious security threats such as phishing attacks and



fraud. Therefore, developing an intelligent system to detect and filter spam messages is very important. In this project, a machine learning model is developed to analyze message content and classify it into spam or legitimate categories. Natural Language Processing techniques are used to preprocess the text data and convert it into a numerical format using vectorization. The Naive Bayes classification algorithm is used to train the model and make predictions based on learned patterns from the dataset.

A user-friendly web interface using Streamlit is implemented to allow users to interact with the system easily. The application provides features such as login, signup, message analysis, dashboard visualization, and feedback. The dashboard displays graphical statistics such as bar charts showing the number of spam and safe messages detected by the system. The results of the project show that machine learning techniques can effectively classify spam messages with good accuracy. The system helps reduce unwanted messages and improves communication efficiency. Although there are some limitations, the system can be further improved by using advanced algorithms, larger datasets, and integration with real email platforms.

REFERENCES

1. Ioannis Moutafis, Antonios S. Andreatos, Petros Stefaneas: "*Spam Email Detection. Using Machine Learning Techniques.*" European Conference on Cyber Warfare and Security, 2023
2. Stanley Munga, Richard Mathenge, Josphat Karani: "*Spam Detection in Emails Using Machine Learning Techniques: A Review.*" International Journal of Computer and Information Technology, 2024.
3. Shaik Mohammed Fateen, Shaik Mohammad Asif, Shaik Afroz, Mohammed Sami "*Email Spam Detection Using Machine Learning.*" International Journal of Scientific Research in Computer Science Engineering and Information Technology, 2025.
4. Pooja Malhotra, Sanjay Malik : "*Spam Email Detection Using Machine Learning and Deep Learning Techniques.*" International Conference on Innovative Computing & Communication (ICICC), 2022.

