International Journal of Novel Trends and Innovation
IJNTI.ORG | ISSN : 2984-908X
An International Open Access, Peer-reviewed, Refereed Journal

Spam Detection Using Machine Learning

[1]Ms. S. T. Sawale, [2]Ms. Sakshi Pravin Bhusari, [3]Ms. Chanchal Ganesh Mali, [4]Ms. Sakshi Sunil Deshmukh,
[5]Ms. Vaishnavi Vijay Jasudkar, [6]Ms. Priyanka Arun Bhendale
Anuradha College of Engineering and Technology, Chikhli, Maharashtra, India

Abstract : The rapid growth of digital communication has led to a significant increase in spam messages, which pose major challenges for users and organizations. This paper presents an efficient machine learning approach for detecting spam messages. The system employs supervised learning algorithms to classify messages as either spam or non-spam (ham). Among the various algorithms tested, Support Vector Machine (SVM) was chosen for its higher accuracy. The proposed model is trained on a labelled dataset and evaluated using several performance metrics, including accuracy, precision, recall, and F1-score. The results demonstrate that the model is effective in identifying spam messages and can be implemented in real-world applications.

IndexTerms - Spam Detection, Machine Learning, SVM, Text Classification, TF-IDF.

I. INTRODUCTION

In today's digital world, the use of emails and messaging platforms has increased significantly. However, this rise in communication has also led to an increase in spam messages, which are unwanted and often contain harmful content. Spam can include advertisements, phishing attempts, or malicious content. Consequently, it is essential to develop systems that can automatically detect and filter such messages [1].

Machine learning plays a crucial role in addressing this issue by enabling systems to learn patterns from data and make predictions. This project focuses on building a spam detection system using machine learning techniques to classify messages accurately.

II. LITERATURE REVIEW

Various methods have been proposed for spam detection. Traditional approaches, such as rule-based filtering and keyword matching, tend to be less effective against changing spam patterns. In contrast, machine learning algorithms like Naive Bayes, Logistic Regression, and Support Vector Machines (SVM) have demonstrated significant improvements in classification accuracy. Notably, SVM is recognized for its high accuracy and its ability to manage high-dimensional data.

III. METHODOLOGY

The proposed system follows these steps:

3.1 Data Collection

A dataset containing labelled messages (spam and ham) is used. Each message is pre-classified to help the model learn patterns [2].

3.2 Data Preprocessing

The text data is cleaned and processed through several steps: removing punctuation and special characters, converting to lowercase, tokenization, and eliminating stop words.

3.3 Feature Extraction

Text data is converted into numerical form using techniques like Term Frequency-Inverse Document Frequency (TF-IDF) [3].

3.4 Model Training

The Support Vector Machine (SVM) algorithm is used to train the model. It finds the optimal hyperplane that separates spam and non-spam messages.

3.5 Web Implementation (Django Framework)

The spam detection model is integrated into a web application built using the Django framework. This application enables users to enter messages and receive real-time predictions.

Key components of the application include:

1. A form module (MessageForm) that allows users to input text messages.
2. URL routing to connect the front-end and back-end logic.
3. View functions that process the user input and return the prediction results.
4. This implementation ensures that the system is user-friendly and accessible through a web interface [5].

3.6 Model Evaluation

The model is evaluated using:

1. Accuracy: Accuracy is one factor to consider when rating categorization models.
   Accuracy = (TP+TN) / (TP+TN+FP+FN)
2. Precision: Precision can be used to judge how well an identifying system works.
   Precision = TP / (TP+FP)
3. Recall: Recall is a quantitative measure that indicates the proportion of instances correctly identified by the method among all possible positive labels.
   Recall = TP / (TP + FN)
4. F1-score: The accuracy metric quantifies the frequency at which the model accurately predicted the entirety of the dataset.
   F1-Score = 2 × (Precision × Recall) / (Precision + Recall)

These metrics collectively help evaluate the effectiveness of a classification model.

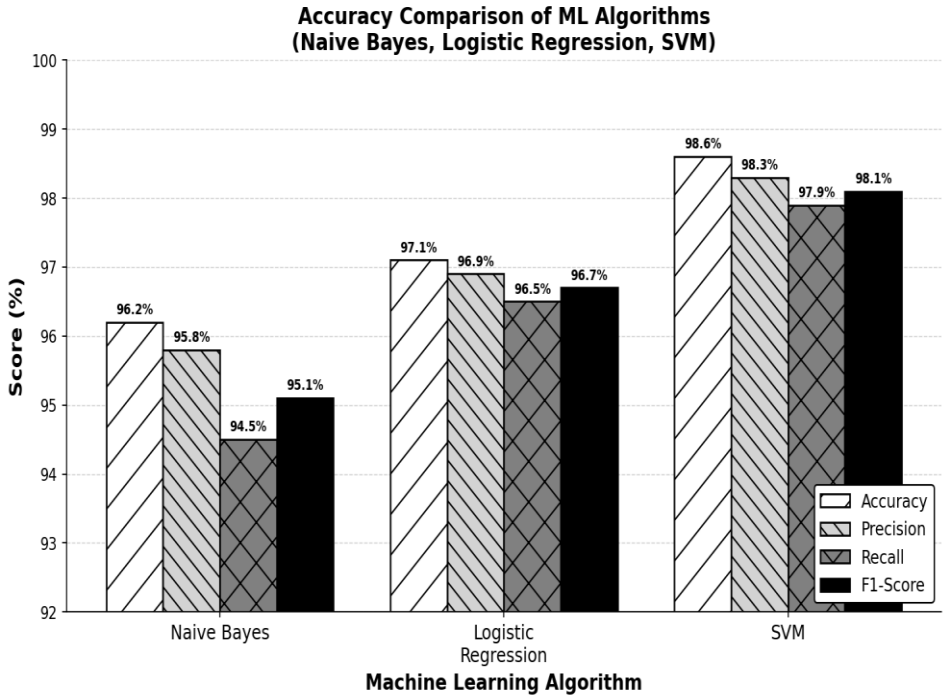


Fig.1: Accuracy Comparison of ML Algorithms (Naive Bayes, Logistic Regression, SVM)

IV. SYSTEM ARCHITECTURE AND DIAGRAMS

4.1 System Architecture Diagram:

The overall architecture of the proposed system is shown below:

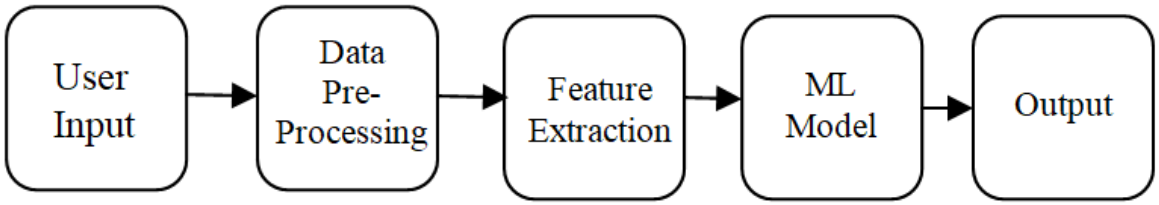


Fig.2: System Architecture

This architecture demonstrates how user input flows through different components before generating the final prediction.

4.2 Flowchart of Spam Detection Process:

Start → User Inputs Message → Preprocessing → Feature Extraction (TF-IDF) → Model Prediction (SVM) → Result (Spam / Not Spam) → End

This flowchart represents the step-by-step working of the system.

4.3 Confusion Matrix (Conceptual Representation):

Where:

1. TP (True Positive): Correctly detected spam
2. TN (True Negative): Correctly detected non-spam
3. FP (False Positive): Non-spam marked as spam
4. FN (False Negative): Spam marked as non-spam.

V. FUTURE SCOPE

- Implementation of deep learning models like LSTM
- Development of a real-time spam detection system
- Integration with email and messaging platforms

VI. RESULTS AND DISCUSSION

The trained model demonstrated a high level of accuracy in detecting spam messages. The confusion matrix indicates that the majority of spam messages were correctly classified, resulting in very few false positives and false negatives. The application of Support Vector Machines (SVM) significantly enhanced performance compared to other algorithms, due to their effectiveness in handling text data.

VII. CONCLUSION

This paper presents a spam detection system based on machine learning using SVM. The results demonstrate that the model is both effective and reliable. Future work may involve implementing deep learning techniques and enhancing the dataset to improve accuracy.

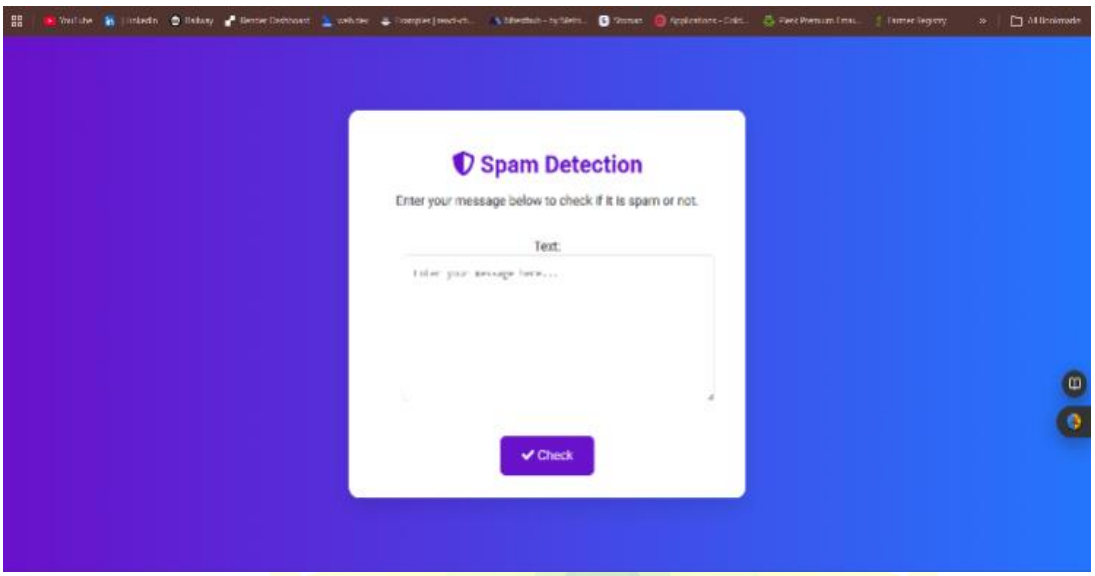
VIII. PROJECT DEPLOYMENT

The proposed spam detection system has been successfully deployed as a web application using the Django framework. The live project can be accessed through the following link:
https://spamdetection-u7y7.onrender.com

The deployment demonstrates the practical applicability of the model in a real-world environment. Users can input messages through the interface and instantly receive predictions, showcasing the efficiency and usability of the system.

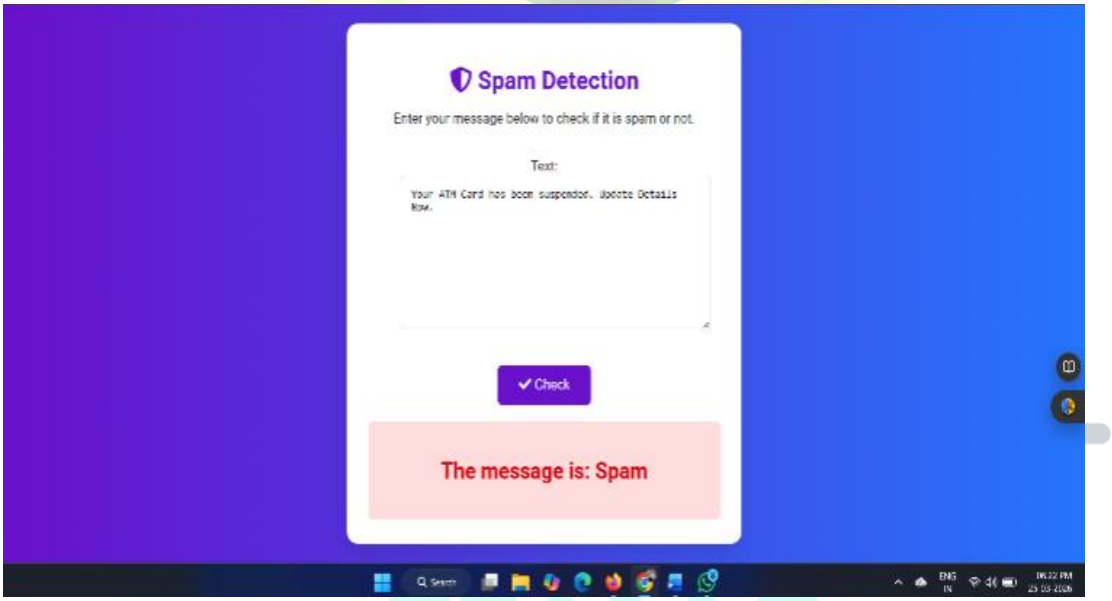
IX. SCREENSHOTS OF WEB APPLICATION

9.1 Home Page Interface:



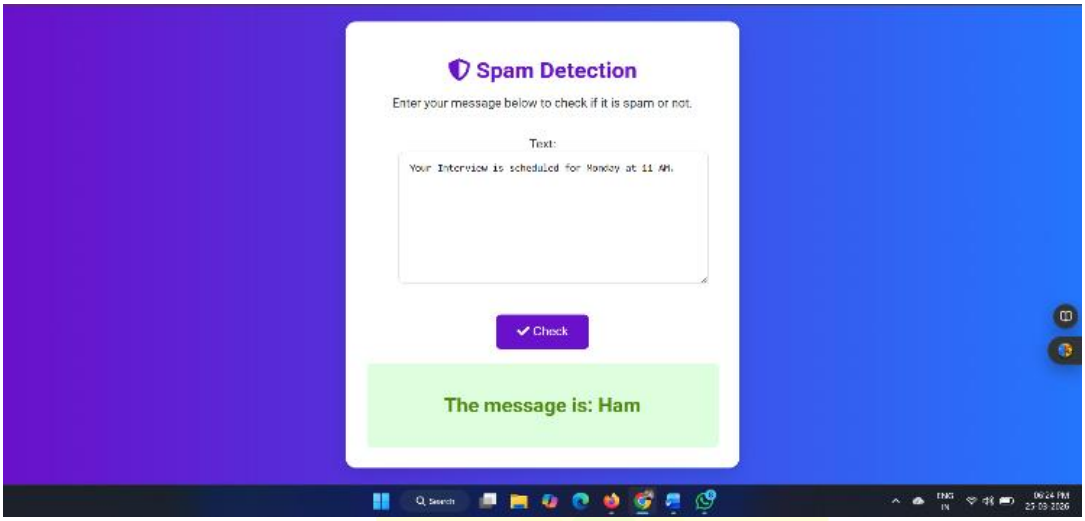
Description: This screen shows the main interface where users can enter a message in the text box.

9.2 Prediction Result (Spam):



Description: The system classifies the message as spam and displays the result.

9.3 Prediction Result (Not Spam):



Description: The system identifies a message as non-spam (ham).
These screenshots provide visual proof of the working system and enhance the credibility

ACKNOWLEDGMENT

We feel great pleasure in expressing our deepest sense of gratitude and sincere thanks to our guide **Ms. S. T. Sawale** for her valuable guidance during the Research paper work, without which it would have been very difficult task. We have no words to express my sincere thanks for valuable guidance, extreme assistance and co-operation extended to all the **Staff Members** of our Department.
This acknowledgement would be incomplete without expressing our special thanks to **Mr. P. T. Talole, Head of Department** (Information Technology) for his support during the work.

REFERENCES

[1] S. N. Ilyasa and A. O. Khadidos, "Optimized SMS spam detection using SVM-DistilBERT and voting classifier: A comparative study on the impact of lemmatization," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 11, pp. 1323-1333, 2024
[2] M. Salman, M. Ikram, and M. A. Kaafar, "Investigating evasive techniques in SMS spam filtering: A comparative analysis of machine learning models," IEEE Access, vol. 12, pp. 24306-24333, 2024.
[3] H. H. Mansoor and S. H. Shaker, "Using classification techniques to SMS spam filter," Int. J. Innov. Technol. Exploring Eng., vol. 8, no. 12, pp. 1734-1739, 2019.
[4] W. Etaiwi and G. Naymat, "The impact of applying different preprocessing steps on review spam detection," Procedia Comput. Sci., vol. 113, pp. 273-279, 2017.
[5] E. H. Tusher et al., "Email spam: A comprehensive review of optimized detection methods, challenges, and esearch problems," IEEE Access, vol. 12, pp. 143627-143653, 2024.