IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

AI-Powered Spam Email Detection System

Mallikarjun S M[*1] and Dr. Kruthi R[*2]
[*1]Student, Department of Computer Application
[*2]Assistant Professor, Department of Computer Application
Maharaja Institute of Technology, Mysore, Karnataka, India

Abstract: *Electronic mail remains one of the most widely used communication platforms for personal, educational, and business purposes. However, the rapid growth of spam emails has introduced significant security risks, including phishing attacks, malware distribution, fraudulent advertisements, and identity theft. Conventional spam filtering techniques primarily rely on manually defined rules or blacklists, which often fail to detect newly emerging spam patterns. This paper presents an AI-Based Spam Email Detection System using Machine Learning and Natural Language Processing (NLP) to automatically classify emails as spam or legitimate. The proposed system performs text preprocessing through tokenization, stop-word removal, stemming, and TF-IDF feature extraction before training a machine learning classifier for accurate prediction. In addition to email classification, the system provides blacklist management, subject-based filtering, real-time prediction, and a user-friendly web dashboard for monitoring classification results. Experimental evaluation demonstrates that the proposed model achieves high detection accuracy with low prediction time, making it suitable for practical email security applications. The lightweight architecture, ease of deployment, and low implementation cost make the system beneficial for students, educational institutions, and small organizations seeking an efficient spam filtering solution*

Keywords: *Spam Email Detection, Machine Learning, Natural Language Processing, TF-IDF, Email Classification, Artificial Intelligence, Cyber Security*

I. INTRODUCTION

Electronic mail has become one of the most widely used communication platforms for personal, educational, and business communication because of its speed, accessibility, and low operational cost. Millions of emails are exchanged every day for information sharing, online transactions, customer support, and official communication. However, the continuous growth of email usage has also resulted in a rapid increase in unsolicited spam emails. These spam messages include advertisements, phishing attempts, malware attachments, fake promotions, and fraudulent links that threaten user privacy and organizational security. As cybercriminals continuously modify their attack strategies, identifying spam emails has become a significant challenge for traditional filtering systems.

Conventional spam detection techniques mainly depend on manually created rules, keyword matching, and blacklist databases. Although these approaches can identify known spam patterns, they often fail to detect newly generated spam messages that use modified text, obfuscated words, or social engineering techniques. Frequent updates to rule-based systems are required to maintain detection accuracy, making them less efficient in dynamic email environments. Consequently, intelligent spam filtering techniques capable of learning from previously observed data are becoming increasingly important.

A. Problem Statement

Despite significant improvements in email communication technologies, spam emails continue to pose serious security and privacy risks. Existing rule-based filtering techniques frequently fail to detect newly emerging spam messages because they depend heavily on predefined keywords and manually updated blacklist databases. Modern spam campaigns continuously evolve using sophisticated language patterns, making traditional filtering approaches less effective.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

Consequently, there is a need for an intelligent, automated spam detection system capable of accurately identifying both known and previously unseen spam emails using Machine Learning and Natural Language Processing techniques.

B. Motivation

The motivation behind this research is the increasing number of phishing attacks, fraudulent advertisements, and malicious email campaigns targeting individuals and organizations worldwide. Many existing commercial email security solutions are expensive and difficult to customize for educational or small-scale environments. Therefore, developing a lightweight, intelligent, and cost-effective spam email detection system that automatically classifies emails with high accuracy can significantly improve email security while remaining accessible to students, educational institutions, and small organizations.

C. Contribution of this Work

Development of an AI-based spam email detection system using Machine Learning and Natural Language Processing.
Implementation of comprehensive text preprocessing techniques including tokenization, stop-word removal, stemming, and TF-IDF feature extraction.
Integration of supervised machine learning algorithms for accurate spam email classification.
Development of blacklist management and subject-based filtering to enhance spam detection capabilities.
Design of a user-friendly web interface that provides real-time spam prediction, prediction history, and classification reports.
Creation of a lightweight, scalable, and cost-effective solution suitable for academic research and practical email security applications.

II. RELATED WORK

Research on spam email detection has advanced significantly with the adoption of Machine Learning and Natural Language Processing techniques. Early spam filtering systems mainly relied on keyword matching, manually maintained blacklists, and rule-based filtering methods. Although these approaches were effective for identifying known spam emails, they struggled to detect newly emerging spam campaigns that use modified vocabulary, obfuscated text, and social engineering techniques. Recent studies have demonstrated that Machine Learning models trained on large email datasets provide better classification accuracy by automatically learning hidden textual patterns from email content.

A. Existing Systems and Their Limitations

Traditional spam filtering systems primarily depend on keyword-based filtering, manually maintained blacklists, and predefined rules. These approaches require frequent updates to remain effective and often fail to detect newly emerging spam messages that use modified text or previously unseen attack patterns. Many existing systems also generate false positives, causing legitimate emails to be incorrectly classified as spam. Commercial spam filtering solutions generally provide higher accuracy but require expensive subscriptions and limited customization options. Furthermore, several existing systems focus only on spam classification without providing additional features such as blacklist management, prediction history, or user-friendly visualization dashboards.

B. Research Gap

Although numerous Machine Learning models have been proposed for spam email detection, many existing studies primarily focus on improving classification accuracy while providing limited attention to practical usability and system integration. Few lightweight solutions combine advanced Natural Language Processing, supervised Machine Learning, blacklist management, subject-based filtering, real-time prediction, prediction history, and an interactive web interface within a single application. The proposed system addresses this gap by integrating these functionalities into a unified platform that provides accurate spam detection, improved user interaction, and efficient performance suitable for academic research and real-world deployment.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

III. OBJECTIVES AND SCOPE

A. Objectives

Develop an intelligent spam email classification system using Machine Learning algorithms.
Perform efficient text preprocessing through tokenization, stop-word removal, stemming, and text normalization.
Apply TF-IDF feature extraction to convert textual email data into numerical feature vectors suitable for machine learning models.
Train and evaluate supervised Machine Learning classifiers to improve spam detection accuracy while minimizing false positives.
Implement blacklist management to automatically identify and block emails from suspicious senders.
Develop subject-based filtering for additional spam identification based on email subject content.
Design an interactive web dashboard that enables users to classify emails, monitor prediction history, and visualize spam detection results.
Generate prediction reports and maintain historical records for future analysis.

B. Scope

The proposed Spam Email Detection System is designed primarily for educational and research purposes while also supporting practical email filtering applications. The system focuses on detecting spam emails using textual analysis rather than analyzing attachments or external URLs.
The system performs email preprocessing using Natural Language Processing techniques before applying Machine Learning algorithms for classification. Users can manually enter email content or upload email data for prediction through a web-based interface. The application supports blacklist management, subject-based filtering, prediction history, and report generation to improve usability and monitoring.

IV. PROPOSED SYSTEM ARCHITECTURE

The proposed AI-Based Spam Email Detection System follows a modular architecture that integrates Natural Language Processing (NLP), Machine Learning (ML), and a web-based application to provide accurate and real-time spam email classification. The architecture is designed to process incoming email content, perform text preprocessing, extract meaningful features, classify emails using a trained machine learning model, and present the prediction results through an interactive user interface. This modular design improves system scalability, maintainability, and ease of deployment while ensuring efficient spam detection.

A. User Interface Layer

The User Interface Layer provides an interactive web application through which users can enter email content or upload email data for spam classification. The interface allows users to submit email messages, view prediction results, monitor classification history, manage blacklist entries, and generate reports. The dashboard is developed using HTML, CSS, JavaScript, and Flask, providing a simple and user-friendly environment for interacting with the system.

B. Text Preprocessing Layer

The preprocessing layer prepares raw email text for machine learning analysis by removing unnecessary information and standardizing the textual content. The preprocessing operations include:

- Conversion of text into lowercase.
- Removal of punctuation marks and special characters.
- Tokenization of email text into individual words.
- Removal of stop words that do not contribute to classification.
- Stemming to reduce words to their root forms.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

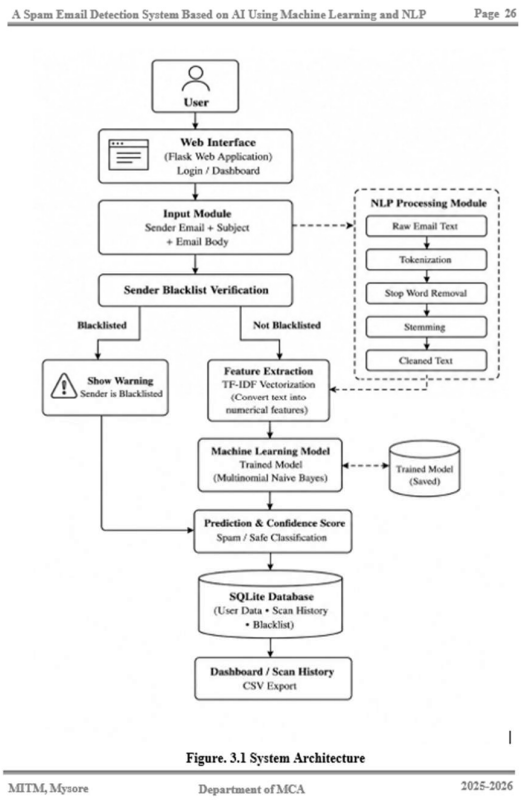

Fig. 1. Layered system architecture

C. Machine Learning Classification Layer

The Machine Learning layer is responsible for classifying emails as **Spam** or **Ham (Legitimate)**. The extracted TF-IDF features are provided as input to a trained supervised Machine Learning model. During prediction, the classifier analyzes the textual patterns learned during training and determines the appropriate email category with high accuracy.

D. Database and Report Management Layer

The final layer stores prediction results, blacklist information, user activities, and historical classification records using an SQLite database. This layer enables report generation, prediction history visualization, and efficient management of classified email records. The stored information can also be used for future analysis and model improvement.

E. Feature Extraction Layer

After preprocessing, the cleaned email text is converted into numerical feature vectors using the **Term Frequency–Inverse Document Frequency (TF-IDF)** technique. TF-IDF assigns importance scores to words based on their occurrence within individual emails and across the entire dataset.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

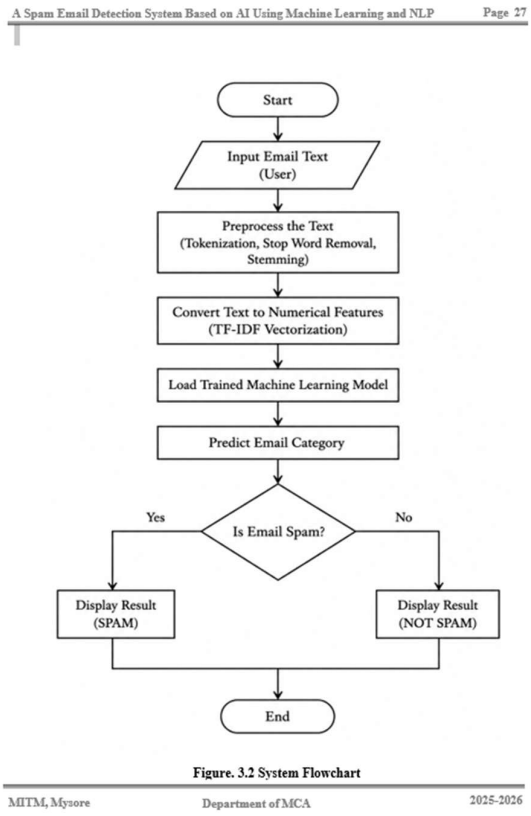

Fig. 2. System flowchart

The detailed design of the proposed AI-Based Spam Email Detection System describes the internal organization of the system modules and the interaction between various software components. The design emphasizes modularity, scalability, and maintainability by separating the system into independent functional units. The proposed system consists of five major modules: User Interface, Text Preprocessing, Feature Extraction, Machine Learning Classification, and Database Management. These modules work together to provide accurate and efficient spam email classification while maintaining historical records and blacklist information.

The Class Diagram represents the static structure of the proposed AI-Based Spam Email Detection System and illustrates the relationships among the major software components. The system is designed using an object-oriented approach, where each class performs a specific functionality while interacting with other classes to complete the spam detection process efficiently.

DOI: 10.48175/IJARSCT-37893

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

3.11.4 Class Diagram

A Class Diagram kind represents the static layout of the whole system, by showing the main classes, their attributes, methods and the relationships. Basically it gives a pretty good view of how different parts of the application are put together and how they sort of work together.

The suggested Spam Email Detection System Using Machine Learning and NLP includes classes like User, Spam Detection System, Text Preprocessor, TF IDF Vectorizer, ML Model, and Prediction Result. Each class does a distinct job, like handling user interaction, text preprocessing, converting the text into features, running the prediction, and producing the final output. The way these classes are connected with each other, makes the spam detection flow run more smoothly.

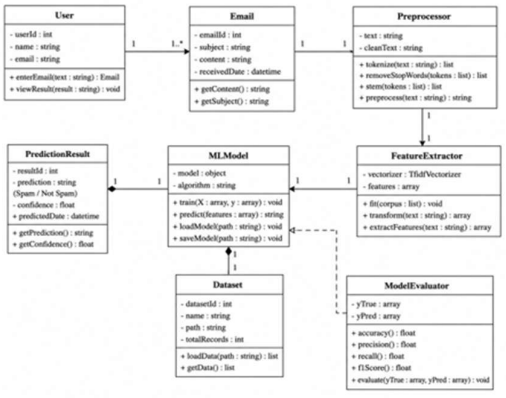

Figure. 3.4 Class Diagram

Fig. 3. Class diagram showing the static structure of the detection, response, and data layers.

DOI: 10.48175/IJARSCT-37893

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

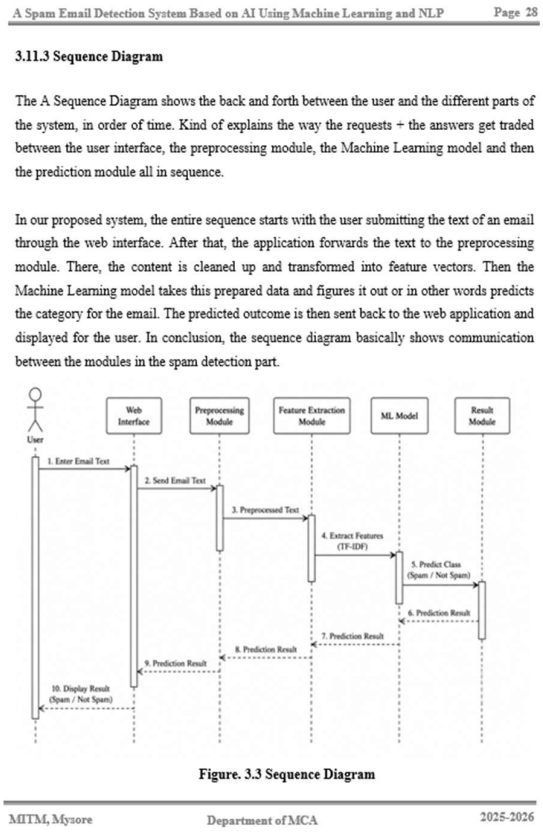

Fig. 4. Sequence diagram .

VI. IMPLEMENTATION

The proposed AI-Based Spam Email Detection System is implemented using Python and integrates Machine Learning, Natural Language Processing, and web technologies to provide accurate and efficient spam email classification. The system is developed with a modular architecture, allowing each functional component to operate independently while communicating seamlessly with other modules. This implementation ensures flexibility, maintainability, and scalability for future enhancements.

A. Software Tools and Technologies

- **Programming Language:** Python 3.x
- **Framework:** Flask
- **Machine Learning Library:** Scikit-learn
- **Natural Language Processing:** NLTK
- **Feature Extraction:** TF-IDF Vectorizer
- **Database:** SQLite
- **Frontend:** HTML, CSS, JavaScript
- **IDE:** Visual Studio Code

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

B. Spam Detection Algorithm

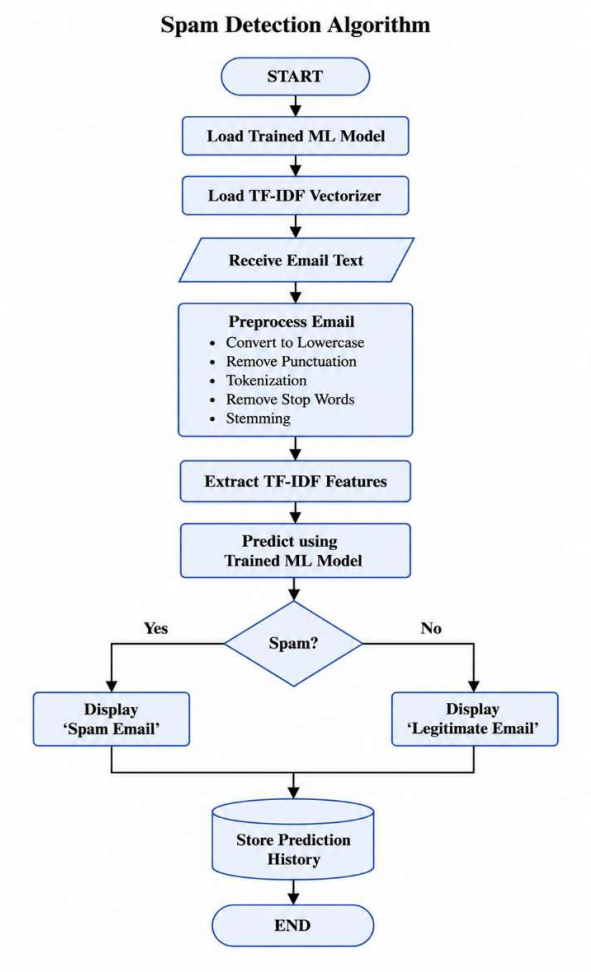

Fig. 5. Spam Detection Algorithm

C. Implementation Workflow

The implementation workflow of the proposed system consists of the following sequential steps:
User enters email content through the web application.
Email text undergoes preprocessing using NLP techniques.
TF-IDF converts textual data into numerical feature vectors.
The trained Machine Learning model predicts Spam or Ham.
Prediction result is displayed on the dashboard.
Prediction details are stored in the SQLite database.
Historical records and reports are generated for future analysis.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

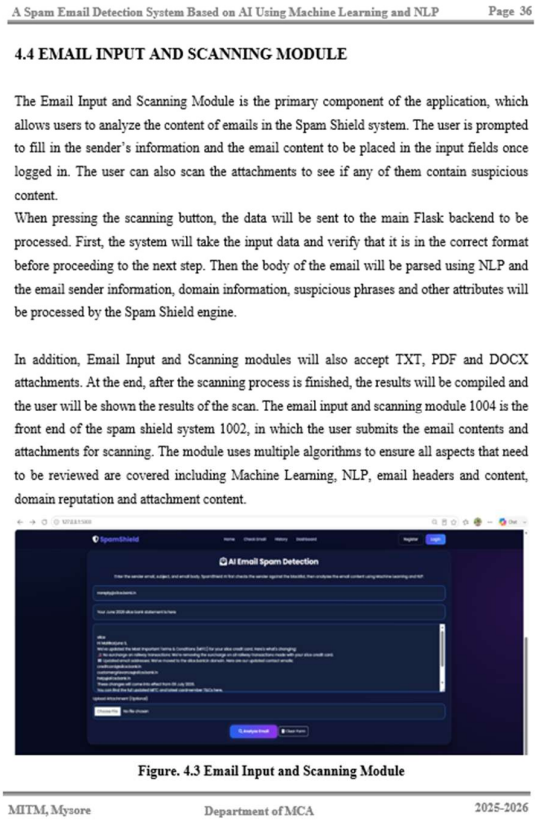

Fig. 6. Email Input and Scanning Module

VII. TESTING AND VALIDATION

Testing and validation play a significant role in evaluating the performance, reliability, and accuracy of the proposed AI-Based Spam Email Detection System. The developed application was tested under different input conditions to verify that each module performs its intended functionality correctly. The testing process included validation of the user interface, text preprocessing module, feature extraction process, machine learning prediction module, blacklist management, database operations, and report generation.

A. Functional Testing

Functional testing was performed to verify whether each module of the system operates according to the specified requirements. Different email samples containing both spam and legitimate messages were used during testing. The system successfully classified incoming emails, updated prediction history, managed blacklist records, and generated prediction reports without any functional errors.

Test Case	*Input*	*Expected Result*	*Status*
User Login	*Valid Credentials*	*Login Successful*	*Pass*
Email Prediction	*Spam Email*	*Classified as Spam*	*Pass*
Email Prediction	*Legitimate Email*	*Classified as Ham*	*Pass*
Blacklist Check	*Blacklisted Email*	*Sender Blocked*	*Pass*
Prediction History	*View Records*	*History Displayed*	*Pass*

DOI: 10.48175/IJARSCT-37893

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

Report Generation	*Generate Report*	*Report Created Successfully*	*Pass*

C. Black Box Testing

Black Box Testing evaluates the external behavior of the application without examining its internal implementation. Various valid and invalid email inputs were provided to verify the correctness of spam prediction results.

Input Scenario	Expected Output	Actual Output	Result
Spam Email	Spam	Spam	Pass
Legitimate Email	Ham	Ham	Pass
Empty Email	Validation Message	Validation Message	Pass
Invalid Input	Error Notification	Error Notification	Pass

VIII. APPLICATIONS

The proposed **AI-Based Spam Email Detection System Using Machine Learning and Natural Language Processing** has a wide range of applications in environments where secure and efficient email communication is essential. By automatically identifying spam emails and filtering unwanted messages, the system enhances productivity, improves cybersecurity, and minimizes the risk of phishing attacks. The major applications of the proposed system are described below.

Cloud-based email service providers can integrate the proposed spam detection model to automatically classify incoming emails in real time. Furthermore, cybersecurity researchers and software developers can use the system for evaluating spam detection techniques, analyzing email patterns, and developing more advanced intelligent email filtering solutions. Overall, the proposed system provides a scalable, reliable, and cost-effective solution that can be adapted for various real-world applications requiring secure and intelligent email communication.

IX. CONCLUSION AND FUTURE SCOPE

The proposed **AI-Based Spam Email Detection System Using Machine Learning and Natural Language Processing** provides an efficient and intelligent solution for identifying spam emails with high accuracy. The system combines Natural Language Processing techniques with supervised Machine Learning algorithms to automatically classify incoming emails as spam or legitimate. Text preprocessing and TF-IDF feature extraction improve the quality of textual data, enabling the classifier to achieve reliable prediction performance.

Although the proposed system achieves accurate spam detection using Machine Learning and Natural Language Processing techniques, several enhancements can further improve its performance and functionality.

Future work may include the integration of **Deep Learning** models such as Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), and Transformer-based architectures like **BERT** to improve the understanding of complex email content and contextual relationships. These advanced models can further increase classification accuracy, especially for sophisticated phishing and social engineering attacks.

REFERENCES

[1] A. Almomani, B. B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani, "A Survey of Phishing Email Detection Techniques," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2070–2090, 2013.

[2] J. S. Bhat, M. M. Ameen, and S. Kumar, "Spam Email Detection Using Machine Learning Techniques," *International Journal of Computer Applications*, vol. 179, no. 21, pp. 15–20, 2018.

[3] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *Proceedings of the European Conference on Machine Learning (ECML)*, pp. 137–142, 1998.

[4] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," *Proceedings of the First Instructional Conference on Machine Learning*, pp. 133–142, 2003.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

[5] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[6] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[7] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O'Reilly Media, 2009.

[8] J. Brownlee, *Machine Learning Mastery with Python*. Machine Learning Mastery, 2016.

[9] K. S. Jones, "A Statistical Interpretation of Term Specificity and Its Application in Retrieval," *Journal of Documentation*, vol. 28, no. 1, pp. 11–21, 1972.

[10] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and Their Compositionality," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 26, 2013.