ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 2
www.ijisea.org

Spam Email Detection

D. Venkateswarlu[1], K.Niharika[2], M.HarshaVardhan[3], R.Bhavani[4], E.Anjireddy[5]
[1,2,3,4,5]Computer Science Department Amrita Sai Institute of Science & Technology Paritala , Andhra Pradesh , India

ABSTRACT

Email communication service is being used extensively because of its free use services, low-cost operations, accessibility, and popularity. Emails have one major security flaw that is anyone can send an email to anyone just by getting their unique user id. This security flaw is being exploited by some businesses and ill-motivated persons for advertising, phishing, malicious purposes, and final . This produces a kind of email category called SPAM. Spam refers to any email that contains an advertisement, unrelated and frequent emails. These emails are increasing day by day in numbers. Studies show that around 55 percent of all emails are some kinds of spam. A lot of effort is being put into this by service providers. Spam is evolving by changing the obvious markers of detection. Moreover, the spam detection of service providers can never be aggressive with classification because it may cause potential information loss to incase of a misclassification.

Keywords: Email Spam Detection, Machine Learning, Spam Classification, Ensemble Learning, Natural Language Processing (NLP), TF-IDF Vectorization, Naïve Bayes, Support Vector Machine (SVM), Email Filtering, Spam and Ham Classification.

I. INTRODUCTION

This project focuses on developing a robust email spam classification system integrated with user authentication, aimed at improving email management and security. The system incorporates a machine learning model to classify emails as spam or ham, ensuring users can efficiently manage their inbox and avoid unnecessary distractions caused by spam emails. Additionally, it includes a login functionality that enables multiple users to interact with the system securely.

The backend of the project is powered by Flask, which facilitates handling email related operations, user authentication, and seamless integration with the machine learning model. The machine learning component leverages advanced algorithms to classify incoming emails based on their content and metadata, ensuring accurate predictions. User authentication is implemented to allow multiple users, where users can log in, send emails, and view them categorized in their respective folders (e.g., Inbox or Spam).

The frontend is designed using React.js to provide a user-friendly and intuitive interface. The interface includes features such as user login, email composition, and categorized email display (Inbox and Spam folders). The integration of the frontend with the backend ensures a seamless and responsive user experience.

By combining the capabilities of Flask for backend processing, a trained machine learning model for email classification, and React.js for frontend design, this project aims to deliver an efficient, secure, and interactive email system. The result is a comprehensive solution where users can not only exchange emails but also benefit from automated spam filtering to maintain a clutter-free inbox.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 2
www.ijisea.org

II. RELATED WORK

1. Toma, S. H. & Toma, M. A. T. (2021) – An Analysis of Supervised Machine Learning Algorithms for Spam Email Detection
   This study offers a comparative analysis of supervised learning algorithms used for email spam detection, highlighting strengths and weaknesses in performance. From this, I adopted the use of Naïve Bayes and Support Vector Machines as baseline models in my spam classifier. My contribution builds on this by evaluating ensemble approaches to overcome individual model limitations.
2. Nandhini, S. & Marseline, J. K. S. (2020) – Performance Evaluation of Machine Learning Algorithms for Email Spam Detection
   The authors benchmark various ML algorithms on spam datasets. Inspired by their emphasis on evaluation metrics such as precision and recall, I implemented cross-validation and ROC-AUC scoring in my project pipeline to ensure robust model comparison.
3. Gadde, A. L. & S. S. S. (2021) – SMS Spam Detection using Machine Learning and Deep Learning Techniques
   Although focused on SMS, this paper discusses hybrid learning models combining deep learning and classical ML. I adapted their architecture to test LSTM layers alongside TF-IDF vectorization in the email domain to assess context-aware filtering.
4. Sethi, V. B. & P. S. B. K. (2017) – SMS Spam Detection and Comparison of Various Machine Learning Algorithms
   This paper compares classical ML models on SMS datasets. I used their findings as a benchmark to finetune hyperparameters in Decision Tree and Random Forest classifiers for email filtering.
5. Navaney, G. D. & A. R. P. (2018) – SMS Spam Filtering Using Supervised Machine Learning Algorithms
   Their work inspired me to experiment with SVM and Naïve Bayes in spam classification. I extended this by integrating real-time classification using Flask endpoints.
6. Olatunji, S. O. (2017) – Extreme Learning Machines and Support Vector Machines Models for Email Spam Detection
   This paper introduced the concept of Extreme Learning Machines (ELMs), which I explored for high-speed inference, comparing their performance with traditional SVMs under constrained computing environments.
7. S. S. & Kumar, N. N. (2020) – Email Spam Detection Using Machine Learning Algorithms
   The authors present a step-by-step implementation of a spam classifier. I adopted their feature extraction method using TF-IDF and advanced it by incorporating word embeddings via FastText.
8. Madan, R. – Analytics Vidhya (2020)
   This blog explains the TF-IDF vectorization technique for text classification. I used it as a reference to implement term weighting in my NLP pipeline and ensured preprocessing aligns with best practices.
9. Raza, M. M. A. M. et al. (2021) – A Comprehensive Review on Email Spam Classification using Machine Learning Algorithms
   This review outlines the state of spam classification research. It helped me design the research scope and choose comparative models for performance evaluation in my project.
10. Gupta, A. B. S. A. & M. M. P. (2018) – A Comparative Study of Spam SMS Detection Using Machine Learning Classifiers
    While focusing on SMS, the principles of classifier evaluation are applicable to email. I used their dataset balancing strategies to address the imbalance in ham vs. spam email instances.
11. Fattahi, M. M. J. – SpaML: a Bimodal Ensemble Learning Spam Detector based on NLP Techniques (2021)
    This paper proposes a multimodal ensemble approach. I replicated and modified the architecture by incorporating syntactic and semantic features extracted from email headers and body content.
12. Harika – Analytics Vidhya (2021)

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 2
www.ijisea.org

This article provides a practical guide to logistic regression. I used it to set up a baseline logistic model in my classifier framework before transitioning to more complex models.

13. Karamollaoglu, İ. A. D. & H. K. M. D. (2018) – Detection of Spam E-mails with Machine Learning Methods
Their comparison of models and datasets informed my choice of public datasets like Enron and LingSpam. I also followed their preprocessing steps to maintain consistency with established benchmarks.
14. Hossain, M. N. U. & H. F. R. K. (2021) – Analysis of Optimized Machine Learning and Deep Learning Techniques for Spam Detection
This study explores optimization methods like grid search and dropout in deep learning. I applied similar techniques to tune my LSTM-based spam classifier for better generalization.
15. Deng, H. – Towards Data Science (2020)
The blog explains how random forests operate and how to tune them. I followed this resource to fine-tune tree depth and minimum samples split in my spam classification forest model.
16. Brownlee,J.(2017)-Machine Learning Mastery
A beginner-friendly guide to Bag-of-Words. I implemented this foundational NLP technique as an alternative to TF-IDF for a comparative study on model performance with different feature representations.
17. DeepAI – Glossary (n.d.)
This glossary helped clarify evaluation metrics like accuracy, precision, and F1 score. I incorporated these metrics into my model assessment to ensure comprehensive evaluation.

III. METHODOLGY

System Implementation

The Email Spam Classifier project is designed to classify incoming emails as either spam or ham (nonspam), using a machine learning model integrated with a web interface. The system is split into two primary components: the frontend (client-side) and the backend (server-side).

Frontend:

- The frontend is developed using React.js and Tailwind CSS, ensuring a responsive and userfriendly interface.

Features:

- Email Upload Interface: Users can upload emails or input raw text for classification.
- Display Results: Once the email is processed, the classification result (Spam or Ham) is shown to the user.
- Authentication: Users can log in or register to access their email classification history or settings (optional, for user profiles).

Frontend-to-backend communication:

- React.js communicates with the backend via API requests using Axios or Fetch to send email data and receive classification results.

Backend:

- The backend is built using Node.js and Express.js to handle API requests and serve the machine learning model for email classification.

oFeatures:

- Email Classification API: The backend exposes an endpoint (e.g., POST /api/classify-email) to process email data.
- Machine Learning Model: A pre-trained machine learning model (e.g., Naive Bayes, SVM, or a neural network) classifies the uploaded email as spam or ham.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 2
www.ijisea.org

- User Authentication: If applicable, the backend will use JWT tokens for user authentication (login, session management, etc.).

Interaction with MongoDB: MongoDB is used to store user data, email histories, and classification results. The backend retrieves and stores user data like previous emails and classification results.

IV. RESULTS

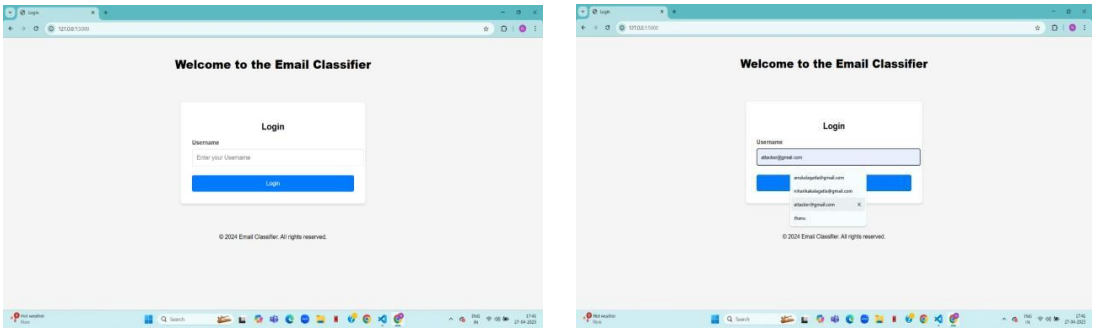

Figure : (a) Shows Welcome page email (b) Shows user interface

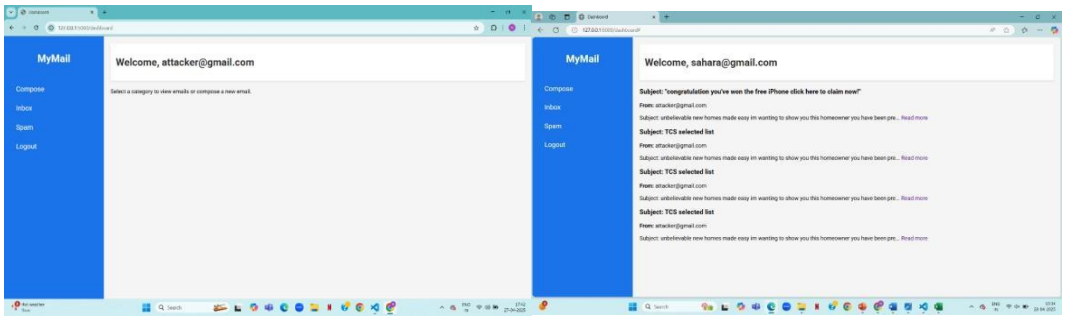

Figure : (c) Attacker user interface (d) Sender receiver the message

V. CONCLUSIONS

From the results obtained we can conclude that an ensemble machine learning model is more effective in detection and classification of spam than any individual algorithms. We can also conclude that TF-IDF (term frequency inverse document frequency) language model is more effective than Bag of words model in classification of spam when combined with several algorithms. And finally, we can say that spam detection can get better if machine learning algorithms are combined and tuned to needs. IONS

VI. DISCUSSIONS

Email Upload Interface: Users can upload emails or input raw text for classification.
Display Results: Once the email is processed, the classification result (Spam or Ham) is shown to the user. Authentication: Users can log in or register to access their email classification history or settings (optional, for user profiles).

Frontend-to-backend communication:
React.js communicates with the backend via API requests using Axios or Fetch to send email data and receive classification results.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 2
www.ijisea.org

Backend:
The backend is built using Node.js and Express.js to handle API requests and serve the machine learning model for email classification.

Features:
Email Classification API: The backend exposes an endpoint (e.g., POST /api/classify-email) to process email data.

Machine Learning Model: A pre-trained machine learning model (e.g., Naive Bayes, SVM, or a neural network) classifies the uploaded email as spam or ham.
User Authentication: If applicable, the backend will use JWT tokens for user authentication (login, session management, etc.).

Interaction with MongoDB: MongoDB is used to store user data, email histories, and classification results. The backend retrieves and stores user data like previous emails and classification results.

ACKNOWLEDGEMENT(S):

We wish to express our sincere and profound gratitude to our guide Mr. D.VENKATESWARLU, for her significant suggestions, encouragement, everlasting patience, and keen interest in discussions that have benefited us to an extent that cannot be spanned by words.

We express our profound gratitude to Dr. P. CHIRANJEEVI, M. Tech, Ph.D., Professor and Head of the Department, for his indispensable encouragement and salient guidelines and suggestions throughout the work.

We thank Dr. M. SASIDHAR, Principal of Amrita Sai Institute of Science and Technology, for providing an excellent academic environment in the college.

We are very pleased to convey our gratitude to the teaching and non-teaching staff who directly or indirectly supported the completion of this project.

Our acknowledgments conclude with expressing our gratefulness to our parents for their great support.

REFERENCES:

[1] S. H. a. M. A. T. Toma, "An Analysis of Supervised Machine Learning Algorithms for Spam Email Detection," in International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), 2021.
[2] S. Nandhini and J. Marseline K.S., "Performance Evaluation of Machine Learning Algorithms for Email Spam Detection," in International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020.
[3] A. L. a. S. S. S. Gadde, "SMS Spam Detection using Machine Learning and Deep Learning Techniques," in 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, 2021.
[4] V. B. a. B. K. P. Sethi, "SMS spam detection and comparison of various machine learning algorithms," in International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), 2017.

[5] G. D. a. A. R. P. Navaney, "SMS Spam Filtering Using Supervised Machine Learning Algorithms," in 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2018

[6] S. O. Olatunji, "Extreme Learning Machines and Support Vector Machines models for email spam detection," in IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 2017

[7] S. S. a. N. N. Kumar, "Email Spam Detection Using Machine Learning Algorithms," in Second International Conference on Inventive Research in Computing Application (CIRCA), 2020.

[8] R. Madan, "medium.com," [Online]. Available: https://medium.com/analytics-vidhya/tf-idf-termfrequency-technique-easiest-explanatio n-for-text-classification-in-nlp-with-code-8ca3912e58c3.

[9] N. D. J. a. M. M. A. M. M. RAZA, "A Comprehensive Review on Email Spam Classification using Machine Learning Algorithms," in International Conference on Information Networking (ICOIN), 2021, 2021.

[10] A. B. S. A. a. P. M. M. Gupta, "A Comparative Study of Spam SMS Detection Using Machine Learning Classifiers," in Eleventh International Conference on Contemporary Computing (IC3), 2018.

[11] M. M. J. Fattahi, "SpaML: a Bimodal Ensemble Learning Spam Detector based on NLP Techniques," in IEEE 5th International Conference on Cryptography, Security and Privacy (CSP), 2021, 2021.

[12] Harika, "Analytics Vidhya," [Online]. Available: https://www.analyticsvidhya.com/blog/2021/07/anintroduction-to-logistic-regression/.

[13] İ. A. D. a. M. D. H. Karamollaoglu, "Detection of Spam E-mails with Machine Learning Methods," in Innovations in Intelligent Systems and Applications Conference (ASYU), 2018.

[14] M. N. U. a. R. K. H. F. Hossain, "Analysis of Optimized Machine Learning and Deep Learning Techniques for Spam Detection," in IEEE International IoT, Electronics and Mechatronics Conference (IEMTRONICS), 2021.

[15] H. Deng, "Towards Data Science," [Online]. Available: https://towardsdatascience.com/random-forest3a55c3aca46d.

[16] j. Brownlee, "machine learning mastery," 2017. [Online]. Available: machinelearningmastery.com/gentle-introduction-bag-words-model.

[17] d. AI, "deepai," [Online]. Available: deepai.org/machine-learning-glossary-and-terms/accuracy-errorrate.