Secure Guard: Robust Spam Detection With SVM Algorithm For Enhanced Email Filtering

L.Yogeshwaran[1], Dr.Bhuvaneswari.M[2]

[1]Dept of Computer Applications

[2]AssociateProfessor, Dept of Computer Applications

[1, 2] Dr. M.G.R. Educational and Research Institute, Chennai – 95

Abstract- *"Secure Guard: Robust Spam Detection with SVM Algorithm for Enhanced Email Filtering" appears to be a title describing a system or method for improving email filtering through the use of a Support Vector Machine (SVM) algorithm. Secure Guard: This could be the name given to the spam detection system. It suggests a focus on security, possibly emphasizing the protective nature of the email filtering solution. Robust Spam Detection: This implies that the system is designed to be strong, resilient, and effective in identifying spam emails. A robust spam detection system is capable of handling various types of spam and adapting to new spamming techniques.*

Keywords- SecureGuard, Spam detection, SVM algorithm, Email filtering, Machine learning, Robustness

I. INTRODUCTION

E-mail or e-mail spam refers to "the use of e-mail to send unwanted e-mail or promotional messages to a group of recipients. Unsolicited e-mail means that the recipient has not given permission to receive "The use of spam e-mail has become more popular since the last decade . Spam has not become a major disaster on the Internet. Spam is a waste of storage space, time and message speed [1].

Automated email filtering can be the most effective way to detect spam, but now spammers can easily bypass all spam filters. A few years ago, most spam could be manually blocked from certain email addresses. Machine learning is used to detect spam.The main approaches to spam filtering include text analysis [2].

Text evaluation of message content is a widely used method of spamming. Many server-deployable responses and aspects of lifter are available. Naive Bayes is one of the most famous algorithms used in these procedures. However, rejecting essentially content-dependent submissions from inspection can be a difficult problem in case of false positive results [3].

The spam domain name classification of this technology works much longer. Whitelisting is a method that accepts emails from publicly allowed domains addresses and places others in a much lower priority queue that is most efficiently delivered after the sender responds to a verification request sent through a spam filtering system [4].

Machine learning approaches are more efficient, training data is used, these samples are emailed and pre-classified. There are many algorithms in machine learning techniques that can be used to filter emails. These algorithms include “Naive Bayes, Support Vector Machines, Neural Networks, K-Nearest Neighbor, Random Forests, etc [5].

II. LITERATURE SURVEY

According to**Sylwia Rapacz**.et al.,2021The meta-algorithm applies a cross-validation approach to supervised learning between different datasets. To illustrate how the meta-algorithm works, we compared three machine learning methods that allow the user to classify emails as desirable (ham) or potentially harmful (spam). The methods used are simple, but as the results have shown, they are quite effective. We use the following classifiers: k-nearest neighbors (k-NN), support vector machines (SVM), and naive Bayes (NB) [6].

According to**Nebojsa Bacanin**.et al.,2022 Spam is a real nuisance for email users because it often disturbs them at work or in their free time. Machine learning methods are often used as the engine of spam detection solutions because they are efficient and usually have high classification accuracy [7].
According to **Nikhil Kumar**.et al.,2020 Email spam has become a big problem nowadays, with the rapid increase in the number of internet users, email spam is also increasing. People use them for illegal and unethical activities, phishing and fraud. Sending malicious link through spam which can damage our system and also get into your system[8].

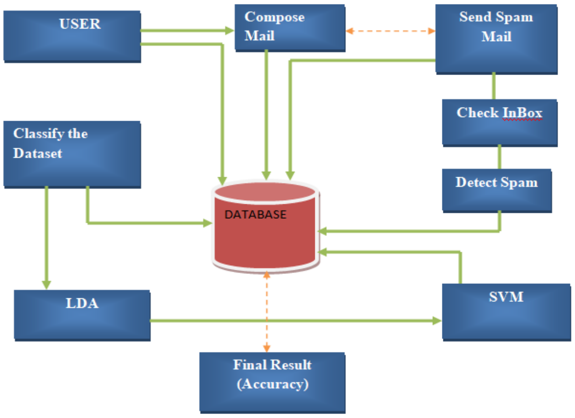
According to **Krishnan Kannoorpatti**.et al.,2019 Email spam has become a big problem nowadays, with the rapid increase in the number of internet users, email spam is also increasing. People use them for illegal and unethical activities, phishing and fraud. Sending malicious link through spam which can damage our system and also get into your system [9].

According to **Romany F Mansour**.et al.,2022Email communication is considered the most important professional communication channel that allows business people and business and non-business organizations to communicate with each other or share globally important official documents and reports, some attractive contents and post them as email to global user [10].

III.PROPOSED SYSTEM

Spam Filtering is to find out whether the emails are spam or not. In our experiment, the class of personal letter was the only class that was defined as ham. But for a specific user, the recruitment massages could also be important and useful; as a consequence, the emails in the class of recruitment should be identified as ham for this user. So, with the multiple classifications the personalized spam filtering system was much easier to develop. In the Proposed System its have implemented SVM Algorithm.If the Email is having following attributes means, that email is a spam Attributes are, Invoicing, training, recruiting, eroticism, website, selling, letter, defrauding, Etc. Its have created the web applications like Gmail. By using the LDA and SVM Algorithm, spam mails are filtered based on the category.

ARCHITECTURE DIAGRAM:



Explanation:

1. Data Collection and Preprocessing:

Gather a large dataset of labeled emails, where each email is tagged as either spam or non-spam (ham).Preprocess the data by cleaning and transforming the text, which includes removing stop words, stemming, and converting text into numerical representations (e.g., TF-IDF vectors).

2. Feature Extraction:

Extract relevant features from the preprocessed emails. Features could include word frequency, presence of specific keywords, email metadata (e.g., sender, subject), and structural features (e.g., length of email, number of links).

3. SVM Show Training:

SVM show utilizing the preparing dataset and the extricated highlights. SVM could be a administered learning calculation that learns to classify information by finding the hyperplane that best isolates the classes (spam vs. non-spam) in a high-dimensional highlight space.

4. Model Evaluation:

Evaluatethe prepared SVM demonstrate utilizing thet estingdataset. Common assessment measurements incorporate exactness, exactness, review, and F1-score.

5. Integration with E-mail System:

Integrate the prepared SVM demonstrate into the email sifting framework. Approaching emails will be passed through the show, which can classify them as either spam or non-spam based on their features.

6. TopicModelingwithLDA:

Apply Inactive Assignment (LDA) to the pre processedemails. LDA couldbea generative factual show usedt odiscover theoretical points inside acollectionofdocuments.LD A speaks to each record as a blend of subjects, where each point may be a conveyance over words. It accept that each word in a report is inferable to one of the document's topics.

IV.RESULTS AND DISCUSSION

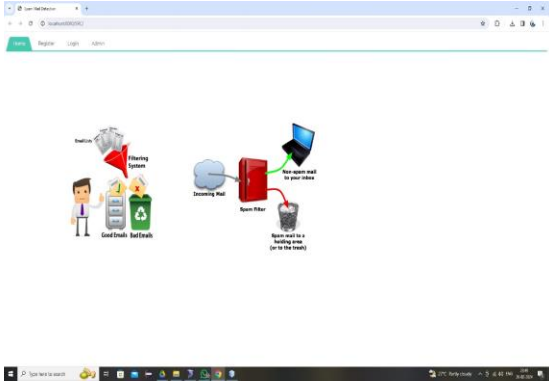
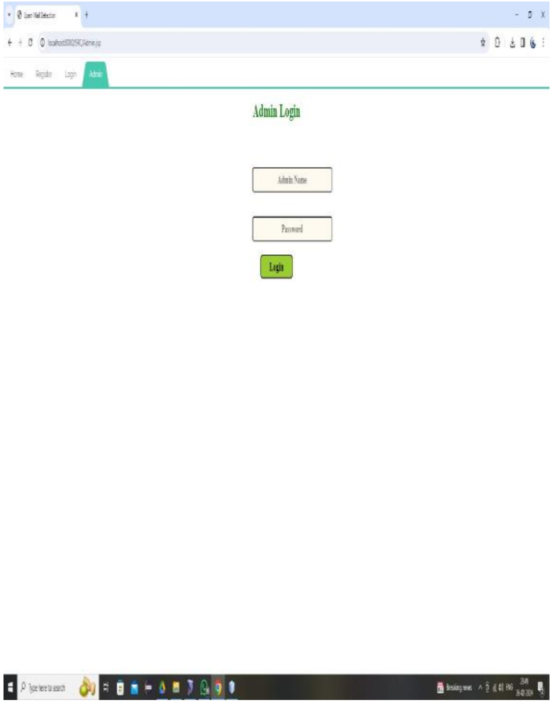


Fig1. Home Page

a)HomePage: The home page of SecureGuard offers a user-friendly interface for easy navigation, providing essential features like search, navigation menu, and announcements for system updates. Iterative improvements based on user feedback ensure optimal user experience.



Admin Login

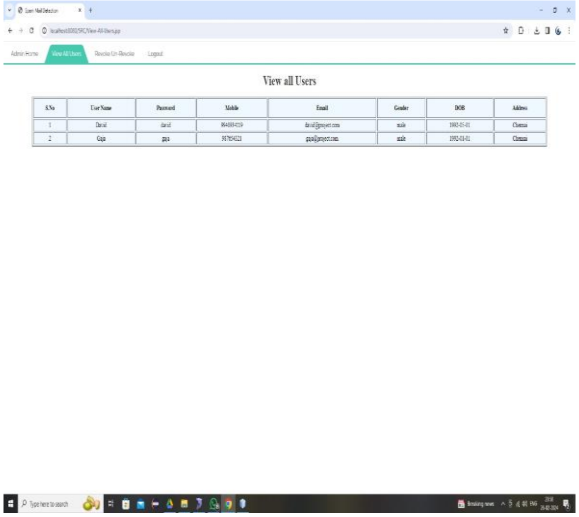
Admin Name

Password

Login

Login Page

b)Admin Login Page: The admin login page prioritizes security with robust authentication measures like username/password and possibly multi-factor authentication. Additional features include password recovery options, CAPTCHA verification, and session management for access control.

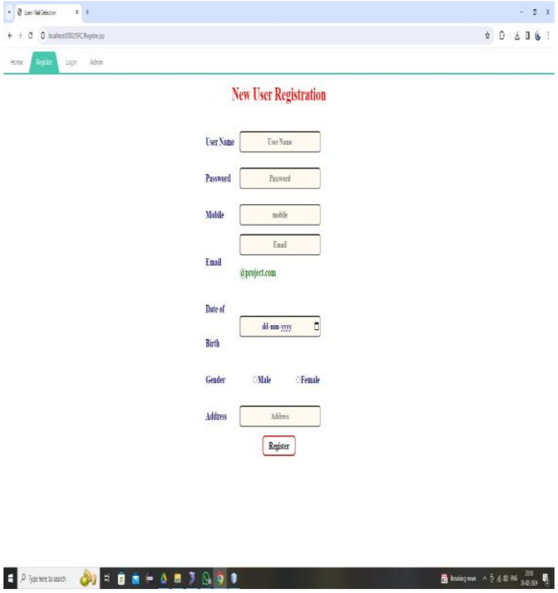


View all Users

S.No	User Name	Password	Mobile	Email	Gender	DOB	Address
1	David	david	9640834119	david@project.com	male	1992-05-01	Chennai
2	Gaja	gaja	987654321	gaja@project.com	male	1992-01-01	Chennai

View All User

c)View All User: The "View All User" feature enables administrators to manage user accounts efficiently, with search, filter, and pagination options for usability. Security measures include restricted access, user privacy protection, and regular audits to ensure compliance and prevent unauthorized access or data breaches.



New User Registration

User Name: User Name

Password: Password

Mobile: mobile

Email: Email @project.com

Date of Birth: dd-mm-yyyy

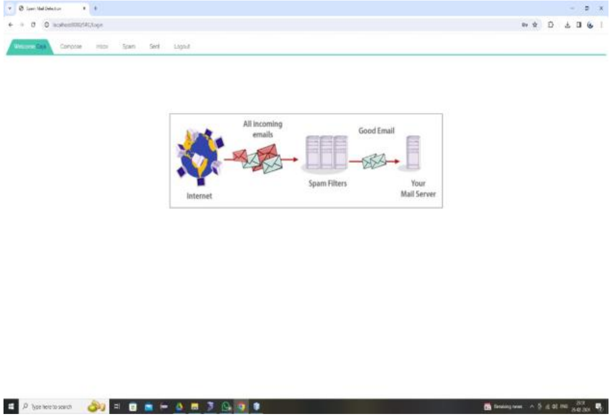
Gender: ○Male ○Female

Address: Address

Register

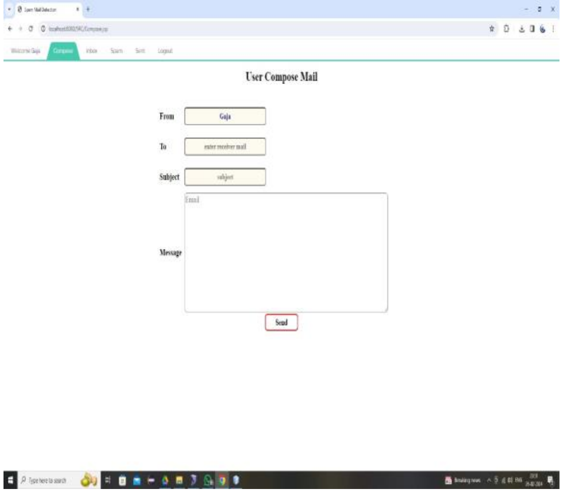
Fig2.Register Page

a)Register Page: The Register Page offers a user-friendly interface with stringent validation checks and security measures, ensuring a seamless and secure registration process.



Home Page

b)User Home Page: The User Home Page features a personalized dashboard, intuitive navigation, and customization options, promoting efficiency and ease of use for users.



User Compose Mail

From: Gaja

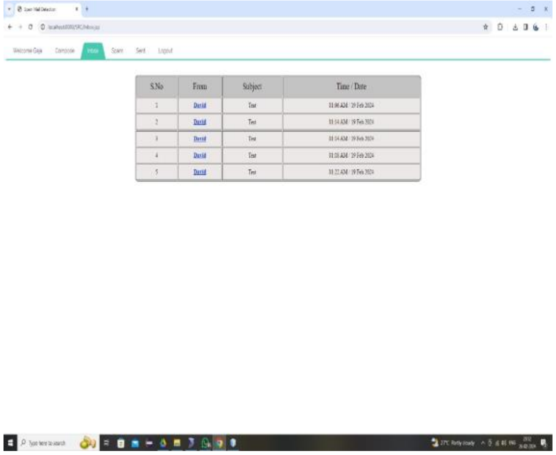
To: enter receiver mail

Subject: subject

Message: Email

Mail Page

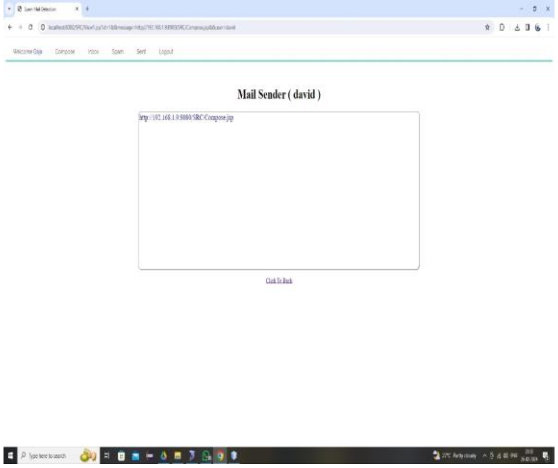
c)Mail Compose Page: The Mail Compose Page includes a WYSIWYG editor, auto-save feature for drafts, and intelligent recipient suggestions, streamlining the email composition process for users.



S.No	From	Subject	Time / Date
1	David	Test	01:06 AM / 19 Feb 2024
2	David	Test	01:14 AM / 19 Feb 2024
3	David	Test	01:14 AM / 19 Feb 2024
4	David	Test	01:18 AM / 19 Feb 2024
5	David	Test	01:22 AM / 19 Feb 2024

Fig3. Inbox Page

a)Inbox Page: Efficiently organizes incoming emails with sorting options and quick actions for improved user productivity.

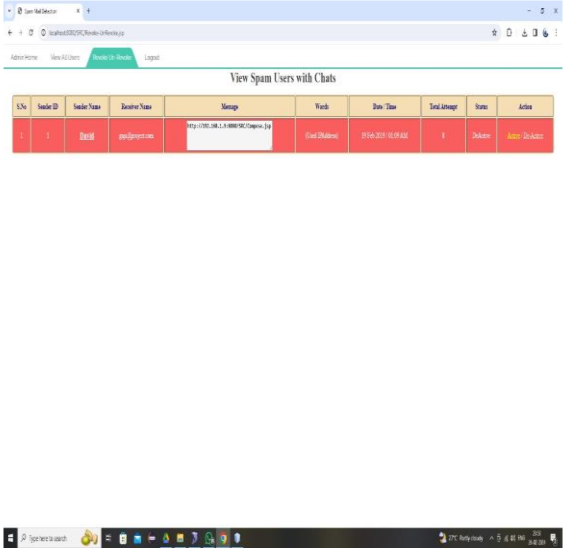


Mail Sender (david)

http://192.168.1.9:8080/SRC/Compose.jsp

Spam Mail

b)Spam Mail: Automatically filters out spam emails using advanced algorithms and quarantines them to protect users from potential security threats.



View Spam Users with Chats

S.No	Sender ID	Sender Name	Receiver Name	Message	Words	Date / Time	Total Attempt	Status	Action
1	1	David	gaja@project.com	http://192.168.1.9:8080/SRC/Compose.jsp	(Used IPAddress)	19 Feb 2019 / 01:09 AM	8	DeActive	Active / De-Active

View Spam Users with Chats

c)View Spam Users with Chats: Provides administrators insights into spam-related user activities, facilitating proactive management and maintenance of a secure email environment.

V.CONCLUSION

In conclusion, SecureGuard represents a robust and efficient approach to spam detection within email systems, utilizing the Support Vector Machine (SVM) algorithm. By leveraging a comprehensive dataset for training and employing advanced feature extraction techniques, SecureGuard is able to accurately classify incoming emails as either spam or legitimate (ham). The integration of the SVM model into SecureGuard's email filtering system ensures real-time detection and mitigation of spam, thereby enhancing the overall security and user experience of email communication. Moreover, the system's ability to adapt and evolve over time through continuous learning and periodic retraining ensures its effectiveness in combating evolving spam tactics. With SecureGuard, organizations and users can confidently manage their email communication with reduced risk of falling victim to spam, enabling them to focus on more critical tasks while maintaining a secure and efficient email environment.

REFERENCES

[1] Suryawanshi, Shubhangi&Goswami, Anurag & Patil, Pramod. (2019). Email Spam Detection: An Empirical Comparative Study of Different ML and Ensemble Classifiers.6974.10.1109/IACC48062.2019.8971582.

[2] Karim,A.,Azam,S.,Shanmugam,B.,Krishnan,K.,&Alazab, M.(2019).AComprehensiveSurveyforIntelligentSpamEmailDetection.IEEEAccess,7,168261168295.[08907831].https://doi.org/10.1109/ACCESS.2019.2954791.

[3] K.AgarwalandT.Kumar,"EmailSpamDetectionUsingIntegratedApproachofNaïveBayesandParticleSwamOptimization,"2018SecondInternationalConferenceonIntelligentComputingandControlSystems(ICICCS),Madurai,India,2018,pp.685-690.

[4] Shradhanjali,Prof. ToranVerma "E-Mail SpamDetectionandClassificationUsingSVMandFeatureExtraction"inInternationalJouranlOfAdvanceReasearch,IdeasandInnovationIn Technology,2017 ISSN: 2454-132X Impact factor: 4.295.

[5] Diren,D.D.,Boran,S.,Selvi,I.H.,&Hatipoglu,T.(2019).RootCauseDetectionwithanEnsembleMachineLearningApproachinthe Multivariate Manufacturing Process.

[6] SylwiaRapacz, Piotr Chołda, Marek Natkaniec Electronics 10 (17) 2021,"A method for fast selection of machine-learning classifiers for spam filtering"2083,.

[7] NebojsaBacanin, MiodragZivkovic, CatalinStoean, Milos Antonijevic, StefanaJanicijevic, Marko Sarac, Ivana Strumberger 2022,"Application of natural language processing and machine learning boosted with swarm intelligence for spam email filtering Mathematics 10 (22), 4173.

[8] Nikhil Kumar, SanketSonowal 2020 ,"Email spam detection using machine learning algorithms",Second International Conference on Inventive Research in Computing Applications (ICIRCA), 108-113.

[9] Asif Karim, Sami Azam, BharanidharanShanmugam, Krishnan Kannoorpatti, MamounAlazab 2019, "A comprehensive survey for intelligent spam email detection" Ieee Access 7, 168261-168295.

[10] Safaa SI Ismail, Romany F Mansour, Rasha M Abd El-Aziz, Ahmed I Taloba 2022,"Efficient E-mail spam detection strategy using genetic decision tree processing with NLP features" Computational Intelligence and Neuroscience2022