IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

Detecting Spam Email With Machine Learning Optimized With Bio-Inspired Metaheuristic Algorithms

Vatan Koshti[1], Aditi Gaherwar[2], Twinkle Ramteke[3] ,Yogeshwari Durgam[4], Prof. Madhavi Sadu[5]
Students, Department of Information Technology[1,2,3,4]
Professor, Department of Information Technology[5]
Rajiv Gandhi College of Engineering Research and Technology, Chandrapur, Maharashtra, India

Abstract: *Electronic mail has eased communication methods for many organizations as well as individuals. Spammers use this strategy to make fraudulent gains by sending unsolicited emails. This project aims to present a method for detection of spam emails with machine learning algorithms that are optimized with bio-inspired methods. A literature review is carried to explore the efficient methods applied on different datasets to achieve good results. The bio-inspired algorithms like Particle Swarm Optimization and Genetic Algorithm were implemented to optimize the performance of classifiers. Multinomial Naïve Bayes with Genetic Algorithm performed the best overall. The comparison of our results with other machine learning and bio-inspired models to show the best suitable model is also discussed.*

Keywords: Bio inspired algorithm, Particle Swarm Optimization algorithm.

I. INTRODUCTION

Nowadays, Emails are used in almost every field, from business to education. Emails have two subcategories, i.e., ham and spam. Email spam, also called junk emails or unwanted emails, is a type of email that can be used to harm any user by wasting his/her time, computing resources, and stealing valuable information. The ratio of spam emails is increasing rapidly day by day.

People are using them for illegal and unethical conducts, phishing and fraud. Sending malicious link through spam emails which can harm our system and can also seek in into your system. Creating a fake profile and email account is much easy for the spammers, they pretend like a genuine person in their spam emails, these spammers target those peoples who are not aware about these frauds. Machine learning models have been utilized for multiple purposes in the field of computer science from resolving a network traffic issue to detecting a malware.Spam e-mail are message randomly sent to multiple addresses by all sorts of groups, but mostly lazy advertisers and criminals who wish to lead you to phishing sites.Recently unsolicited commercial/bulk e-mail also known as spam, become a big trouble over the internet. Spam is waste of time, storage space and communication bandwidth.So, it is needed to Identify those spam mails which are fraud, this project will identify those spam by using techniques of machine learning Optimized with Bio-Inspired Metaheuristic Algorithms", this project will discuss the machine learning algorithms and apply all these algorithm on our data sets and best algorithm is selected for the email spam detection having best precision and accuracy.

The Scope of this project aims to achieve the following:

1) To explore machine learning algorithms for the spam detection problem.
2) To investigate the workings of the algorithms with the acquired datasets.
3) To implement the bio-inspired algorithms.
4) To test and compare the accuracy of base models with bio-inspired implementation.

1.1 Spam

Unsolicited usually commercial messages (such as emails, text messages, or Internet postings) sent to a large number of recipients or posted in a large number of places.

1.2 What is Spam how it is Harmful

Most spam is irritating and time-consuming, but some spam is positively dangerous to handle. Usually email scams are

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

trying to get you to give up your bank details so that the fraudsters can either withdraw money or steal your identity. Such messages include phishing scams and advanced fee fraud.

II. RELATED STUDY

Naive Bayes classifiers are widely used to filter spam emails, however, the strong independence assumptions between features limit their performance in accurately identifying spams. To address this issue, we proposed a support machine vector based naive Bayes - SVM-NB - filtering system[4]. The SVM-NB first constructs an optimal separating hyperplane that divides samples in the training set into two categories. For samples located nearby the hyperplane, if they are in different categories, one of them will be eliminated from the training set. In this way, the dependence between samples is reduced and the entire training sample space is simplified. With the trimmed training set, the naive Bayes algorithm is applied to classify emails in the test set. The SVM-NB system is evaluated with the dataset obtained from DATAMALL. Experiment results demonstrate that SVM-NB can achieve a higher spam-detection accuracy and a faster classification speed. upsurge in the volume of unwanted emails called spam has created an intense need for the development of more dependable and robust antispam filters. Machine learning methods of recent are being used to successfully detect and filter spam emails. We present a systematic review of some of the popular machine learning based email spam filtering approaches. Our review covers survey of the important concepts, attempts, efficiency, and the research trend in spam filtering. The preliminary discussion in the study background examines the applications of machine learning techniques to the email spam filtering process of the leading internet service providers (ISPs) like Gmail, Yahoo and Outlook emails spam filters. Discussion on general email spam filtering process, and the various efforts by different researchers in combating spam through the use machine learning techniques was done. Our review compares the strengths and drawbacks of existing machine learning approaches and the open research problems in spam filtering. We recommended deep leaning and deep adversarial learning as the future techniques that can effectively handle the menace of spam emails.

The increasing volume of unsolicited bulk e-mail (also known as spam) has generated a need for reliableanti-spam filters. Machine learning techniques now days used to automatically filter the spam e-mail in avery successful rate. In this paper we review some of the most popular machine learning methods(Bayesian classification, k-NN, ANNs, SVMs, Artificial immune system and Rough sets) and of theirapplicability to the problem of spam Email classification. Descriptions of the algorithms are presented,and the comparison of their performance on the SpamAssassin spam corpus is presented.[4]

Email has become one of the fastest and most economical forms of communication. However, the increase of email users has resulted in the dramatic increase of spam emails during the past few years. As spammers always try to find a way to evade existing filters, new filters need to be developed to catch spam. Generally, the main tool for email filtering is based on text classification. A classifier then is a system that classifies incoming messages as spam or legitimate (ham) using classification methods. The most important methods of classification utilize machine learning techniques. There are a plethora of options when it comes to deciding how to add a machine learning component to a python email classification. This article describes an approach for spam filtering using Python where the interesting spam or ham words (spam-ham lexicon) are filtered first from the training dataset and then this lexicon is used to generate the training and testing tables that are used by variety of data mining algorithms. Our experimentation using one dataset reveals the affectivity of the Naïve Bayes and the SVM classifiers for spam filtering[12].

III. LITERATURE REVIEW

1) Emmanuel Gbenga Dada et al “Machine learning for email spam filtering: review, approaches and open research problems”, Heliyon 5 (2019) e01802 Received 3 September 2018; Received in revised form 25 February 2019; Accepted 20 May 2019There is a rapid increase in the interest being shown by the global research community on email spam filtering. In this section, we present similar reviews that have been presented in the literature in this domain. This method is followed so as to articulate the issues that are yet to be addressed and to highlight the differences with our current review. Lueg presented a brief survey to explore the gaps in whether information filtering and information retrieval technology can be applied to postulate Email spam detection in a logical, theoretically grounded manner, in order to facilitate the introduction of spam filtering technique that could be operational in an efficient way. However,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

the survey did not present the details of the Machine learning algorithms, the simulation tools, the publicly available datasets, and the architecture of the email spam environment. It also fails short of presenting the parameters used by previous research in evaluating other proposed techniques. Wang reviewed the different techniques used to filter out unsolicited spam emails. The paper also to categorized email spams into different hierarchical folders, and automatically regulate the tasks needed to response to an email message. However, some of the limitations of the review article are that; machine learning techniques, email spam architecture, comparative analysis of previous algorithms and the simulation environment were all not covered.

2) Jai Batra et al "A comprehensive study of spam detection in e-mails using bio-inspired optimization techniques", International Journal of Information Management Data Insights 1 (2021) 100006 Received 20 October 2020; Received in revised form 30 November 2020; Accepted 19 December 2020 Spam messages sent by marketers to promote and advertise their products are considered as a nuisance by most people whose limited server storage is filled-up by these unwanted e-mails CRaad et al., 2010). The amount of time it takes to remove all the spam e-mails hinders the productivity of individuals on a daily basis. Some innocent users may also become prey to attacks like e-mail spoofing CHu and Wang, 2018) and phishing schemes CBanu and Banu, 2013; Halaseh and Alqatawna, 2016), that include mass-forwarded e-mails by scammers who try to steal money and bank account details from users. Spam e-mails are also used by attackers and hackers for the distribution of viruses and other such dangerous software hidden behind attractive links and exciting of- fers CKarim et al., 2019). The problem of spam e-mails, therefore, needs to be addressed promptly and effective measures need to be taken to curb the prob- lem. Attempts have been made in reducing spam e-mails acquired on a daily basis by e-mail users. These range from creation of advanced spam-filtering tools to introduction of anti-spamming laws in the United States which would prevent spammers from sending unwanted mes- sages. Several approaches have been proposed to detect and filter spam e-mails whose overview can be found in Wang and Cloete (2005) . To reduce the number of spam messages, IP address filtering CBajaj et al., 2017) has also been introduced, which is a heuristic approach that de- pends on the source's IP address in an incoming e-mail to determine its legitimacy as a non-spam e-mail. Malicious URLs detection systems are proposed which remove spam content and malicious URLs in email (Ranganayakulu and C., 2013; Rathod and Pattewar, 2015). A complete spam detection system named Cosdes has also been created which pos- sesses an efficient near-duplicate matching scheme and a progressive update scheme CTseng et al., 2011). The Blacklist method is another ap- proach, which rejects the acceptance of an e-mail with an address that can be found on the list of sources that cannot be trusted. Spam detection methods that employ text categorization using hybrid algorithms like Neuro-SVM CDhawan and Simran, 2018) and NLP pre-processing meth- ods CEtaiwi and Naymat, 2017) have also been developed. Cryptography has been explored which is a tool that requires a digital signature on an e-mail as proof of authorization, otherwise, it is discarded by the filter. Despite all these efforts, sooner or later, spammers succeed in finding ways to evade these approaches. Thus, there is a need for more reli- able and sophisticated approaches that are capable of minimizing, if not eliminating, all spam e-mail. For this purpose, various machine learning approaches have been adopted in which usually a group of spam and legitimate e-mail messages are used for training a learning algorithm so that future incoming e-mail can be automatically categorized. Examples include Artificial Neural Network CÖzgür et al., 2004), Deep Neural Net- work CBarushka and Hájek, 2018), Granular Support Vector Machine –Boundary Alignment algorithm CTang et al., 2006), Naive Bayesian algorithm (Deshpande et al., 2007), Random Weight Network CFaris et al., 2019), and a quadratic-neuron-based neural tree (QUANT) CSu et al., 2010) to name a few. An excessive number of features while detect- ing spam e-mails may negatively affect the performance of a learning classifier, therefore methods have also been proposed for feature selection CDutta et al., 2018) and extraction CDiale et al., 2019; Inuwa-Dutse et al., 2018) like Common Vector Approach (Gunal et al., 2006) and Local Concentration CZhu and Tan, 2010). In this study, this categorization between spam and legitimate e- mails will be done with the help of bio-inspired algorithms. Some re- search has already been conducted in this domain using algorithms like Antlion Optimization algorithm CNaem et al., 2018), Krill Herd Opti- mization algorithm (Faris et al., 2015), Octopods CSadek et al., 2019), and Modified K-Means integrated Levy flight Firefly Algorithm with chaotic maps (Aswani et al., 2018). This study seeks to understand the utility of other bio-inspired techniques in solving the spam detection problem.

Dr. V. Malsoru, et al "Detecting Spam Email With Machine Learning Optimized With Bio-Inspired MetaheuristicAlgorithms", JAC : A Journal Of Composition Theory, ISSN : 0731-6755, Page No: 40-47

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

A specific algorithm is used to learn the classification rules from these messages. Those algorithms are used for classification of objects of different classes. The algorithms are provided with input and output data and have a self-learning program to solve the given task. Searching for the best algorithm and model can be time consuming. The two-class classifier is best used to classify the type of message either spam or ham. This algorithm is used to predict the probability and classification of data outcome.

4) P. VANAJA, et al "Machine Learning based Optimization for Efficient Detection of Email Spam", Positif Journal, Issn No : 0048-4911, Page No : 310-319. This section reviews literature on various methods of email spam detection. Gibson et al. [1] used ML models along with bio-inspired heuristics in order to improve prediction performance. Dada et al. [2] investigated on various spam filtering methods that are useful for preventing spam emails. Those methods are based on machine learning models. Krithigaet al. [3] studied different methods found in the literature for spam profile detection in social media. Shuaibet al. [4] explored whale optimization method along with feature selection and ML models for spam detection and filtering. Truong et al. [5] used artificial intelligence methods for developing defence against spam contents. Imam et al. [6] proposed a supervised learning approach to detect spam drift in Twitter which is one of the social media. Rathoreet al. [7] proposed a spam filtering method based on SMO and Bayesian approaches. Karim et al. [8] explored different intelligent spam detection models available. Abdolahnezhadet al. [9] proposed a hybrid method based genetic algorithm and negative selection approach for email spam detection. Krithigaet al. [10] proposed a modified whale optimization method along with feature selection to detect spam profiles in Twitter.

5) K.Varun Kumar, et al "Machine Learning-based spam detection using Naïve Bayes Classifier in comparison with Logistic Regression for improving accuracy", Journal of Pharmaceutical Negative Results ¦ Volume 13 ¦ Special Issue 4 ¦ 2022, Page No. 548-554

6) SIMRAN GIBSON, et al "Detecting Spam Email With Machine Learning Optimized With Bio-Inspired Metaheuristic Algorithms, IEEE Access · January 2020, Page No. 187914-187932

This research will experiment Bio-inspired algorithms alongwith Machine learning models. This will be conducted on In this experiment WEKA acted as a black box and provided the better performing algorithms which were Support Vector, Random Forest, Naïve Bayes and Decision Tree.Since spam email detection falls into classification category, supervised learning method will be used. Supervised learning is a concept where the dataset is split into two parts: 1) Training data and 2) Testing data. The main aim of this learning method is to train a classier with a given data and parameters and then predict the outcome with the testing dataset which will not be known to the program or classier [12].The models will be trained with a training dataset of 60%,70%, 75% and 80%. Once the model is trained, model will be provided with the testing dataset which is distributed as 40%,30%, 25% and 20%respectively with training dataset. This will provide a better knowledge of what percentage split is best suited and thus be more efficient to work with majority of the datasets. This will provide results on classifiers working best with more or less training data.

7) N. Sutta, et al " A Study of Machine Learning Algorithms on Email Spam Classification", EPiC Series in Computing Volume69, 2020, Page No.170-179 According to [7], e-mail spam and ham classification can be done using either a machine learning approach or a non-machine learning approach. Bayesian, SVM, Neural Network, Markova model, memory-based pattern discovery, etc. come under machine learning methods and Blacklist/White list, signatures, hash base, grant listing come under non-machine learning methods. To distinguish spam and ham emails, several algorithms like Naïve Bayesian (NB), Multi-layered Perceptron (MLP), J48, and Linear Discriminant Analysis (LDA) have been suggested and their performance was compared by N. Radha and R Lakshmi [8].

Mujtaba et al. reviewed Bayesian classification, k-NN, ANNs, SVMs, Artificial Immune System, and Rough sets, and compared their performance on the Spam Assassin public mail corpus [9]. Banday et al. [3] examined spam filter design procedures by integrating Naïve Bayes, KNN, SVM, and Bayes Additive Regression Tree and evaluated them in terms of accuracy, recall, precision, etc. Chhabra et al. [4] used the Support Vector Machine to develop spam filtering by considering nonlinear SVM classifiers with specific kernel functions over Enron Dataset. Rusland et al. [5] conducted the study using the Naïve Bayes spam filtering algorithm on two datasets that are evaluated based on accuracy, recall, precision, and F-measure. Wang [6] categorized email spams into various hierarchical folders and regulated the tasks needed to respond to an email message automatically.

8) Chode Abhinav, et al "Spam Mail Detection using Machine Learning", International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538, Page

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

no.2327-2329 On email spam detection, as well as social media and Twitter signaling spam detection, a lot of research and literature studies have been done. Because this is a relatively new area of research, there is no thorough systematic literature review on SMS spam detection. Although SMS communication first became popular in 2000, it gained traction in 2006 and acquired much greater traction after the introduction of Android phones. SMS spam is growing more popular with spammers as the number of individuals utilizing SMS as a way of communication grows. As a result, SMS spam detection research evolved out of need, and it primarily began around 2007. Our goal with this review is to gain proper knowledge in the field of spam mail detection, gain knowledge about the algorithms currently used for spam mail detection, their benefits and drawbacks, compare the accuracy of algorithms and identifying any gaps in current research so that need to be investigated further.

Spam and ham mails are classified using a variety of algorithms. On a spam base dataset, feature selection has a vital role in identifying the optimal classification method in terms of computational time, accuracy, misclassification rate, and precision, followed by algorithm selection.

9) Thashina Sultana, et al "Email based Spam Detection", International Journal of Engineering Research & Technology (IJERT) IJERTV9IS060087 (Vol. 9 Issue 06, June-2020) Page No. 135-139

In the paper[1], authors have highlighted several features contained in the email header which will be used to identify and classify spam messages efficiently .Those features are selected based on their performance in detecting spam messages. This paper also communalizes each feature contains in Yahoo mail, Gmail, and Hotmail so a generic spam messages detection mechanism could be proposed for all major email providers. In the paper[2], a new approach based on the strategy that how frequently words are repeated was used. The key sentences, those with the keywords, of the incoming emails must be tagged and thereafter the grammatical roles of the entire words in the sentence need to be determined, finally they will be put together in a vector to take the similarity between received emails. K-Mean algorithm is used to classify the received e-mail. Vector determination is the method used to determine to which category the e-mail belongs to. In the paper[3],authors described about cyber-attacks. Phishers and malicious attackers are frequently using email services to send false kinds of messages by which target user can lose their money and social reputations. These results into gaining personal credentials such as credit card number, passwords and some confidential data .In This paper ,authors have used Bayesian Classifiers .Consider every single word in the mail. Constantly adapts to new forms of spam. In the paper[4],proposed system attempts to use machine learning techniques to detect a pattern of repetitive keywords which are classified as spam. The system also proposes the classification of emails based on other various parameters contained in their structure such as Cc/Bcc, domain and header. Each parameter would be considered as a feature when applying it to the machine learning algorithm. The machine learning model will be a pre-trained model with a feedback mechanism to distinguish between a proper output and an ambiguous output. This method provides an alternative architecture by which a spam filter can be implemented. This paper also takes into consideration the email body with commonly used keywords and punctuations.

IV. ALGORITHM USED

4.1 Bio-Inspired Optimization Algorithms

There are two bio-inspired optimization approaches that are discussed here which helped to improve the results of the experiments, i.e., Particle Swarm Optimization and Genetic Algorithm.

A. Particle Swarm Optimization Algorithm

The PSO is based on the swarming methods observed in fish or birds. The particles are evaluated based on their best position and overall global position. Particles within a search space are scattered to find the global best position.

The Py swarm's library offers different calculations and techniques for PSO to be used with an ML model such as feature subset selection or parameter tuning optimization. As researched in the previous sections, the feature selection can reduce feature space but can also discard some features that can be useful during the classification. Therefore, PSO will be used to tune and find the hyper-parameter for a given ML/NN model.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

B. Genetic Algorithm

The GA algorithm is an evolutionary algorithm based on Darwinian natural selection that selects the fittest individual from the given population. This involves the principle of variation, inheritance and selection. The algorithm maintains a population size and the individuals have a unique number. (Chromosomes) that are binary represented Implementation of the GA was conducted with the help of TPOT library. The program selects the best parameters from a given dictionary of parameters. The TPOT classifier is then trained with cross validation. The parameters given to the TPOT are as follows:

- Generation: Number of times the pipeline will conduct the optimization processes. The default value is 100. The program has set this parameter as '10'.
- Population size: Number of individuals participating for Genetic programming within each generation. Default is100. The program has set this parameter as '40'
- Offspring size: Offspring to be produced in each generation. Default is 100. The program has set this parameteras '20'

V. CONCLUSION

Spam detection and filtration gained the attention of a sizeable research community. The reason for a lot of research in this area is its costly and massive effect in many situations like consumer behavior and fake reviews.

The survey covers various machine learning techniques and models that the various researchers have proposed to detect and filter spam in emails and IoT platforms. The study categorized them as supervised, unsupervised, reinforcement learning, etc.

The study compares these approaches and provides a summary of learned lessons from each category. This study concludes that most of the proposed email and IoT spam detection methods are based on supervised machine learning techniques. A labeled dataset for the supervised model training is a crucial and time-consuming task.

Supervised learning algorithms SVM and Naïve Bayes outperform other models in spam detection. The study provides comprehensive insights of these algorithms and some future research directions for email spam detection and filtering.

REFERENCES

[1] Emmanue Gbenga Dada et al "Machine learning for email spam filtering: review, approaches and open research problems", Heliyon 5 (2019) e01802 Received 3 September 2018; Received in revised form 25 February 2019; Accepted 20 May 2019

[2] Jai Batra et al "A comprehensive study of spam detection in e-mails using bio-inspired optimization techniques", International Journal of Information Management Data Insights 1 (2021) 100006 Received 20 October 2020; Received in revised form 30 November 2020; Accepted 19 December 2020

[3] Dr. V. Malsoru, et al "Detecting Spam Email With Machine LearningOptimized With Bio-Inspired MetaheuristicAlgorithms", JAC : A Journal Of Composition Theory, ISSN : 0731-6755, Page No: 40-47

[4] P. VANAJA, et al "Machine Learning based Optimization for Efficient Detection of Email Spam", Positif Journal, Issn No : 0048-4911, Page No : 310-319

[5] K.Varun Kumar, et al "Machine Learning-based spam detection usingNaïve Bayes Classifier in comparison withLogistic Regression for improving accuracy", Journal of Pharmaceutical Negative Results ¦ Volume 13 ¦ Special Issue 4 ¦ 2022, Page No. 548-554

[6] SIMRAN GIBSON, et al "Detecting Spam Email With Machine Learning Optimized With Bio-Inspired Metaheuristic Algorithms, IEEE Access • January 2020, Page No. 187914-187932

[7] N. Sutta, et al " A Study of Machine Learning Algorithms on Email Spam Classification", EPiC Series in Computing Volume69, 2020, Page No.170-179

[8] Chode Abhinav, et al "Spam Mail Detection using Machine Learning", International Journal for Research in Applied Science & Engineering Technology (IJRASET)ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538, Page no.2327-2329

[9] Thashina Sultana, et al "Email based Spam Detection", International Journal of Engineering Research & Technology (IJERT) IJERTV9IS060087 (Vol. 9 Issue 06, June-2020) Page No. 135-139

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

[10] Rajesh Kumar J, et al "Email Spam Detection using Machine Learning Techniques", International Advanced Research Journal in Science, Engineering and Technology Vol. 8, Issue 6, June 2021, DOI: 10.17148/IARJSET.2021.8632 Page No. 189-193

[11] Naresh Vinod Wankhade, et al "Paper on Spam Email Detection with Classification Using Machine Learning", INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY , IJIRT 156181, Page No. 1055-1059

[12] Ms.A. Sowshna, et al "Detecting Spam Email With Machine Learning Optimized With Bio Inspired Metaheuristic Algorithms", International Journal of Scientific Development and Research (IJSDR), Page No. 160-163

[13] Neha Karadkar, et al "Spam Mail Classification Using SVM and Genetic Algorithm", Journal of Emerging Technologies and Innovative Research (JETIR), Page No. e513-e517

[14] Miss. Pratiksha Mantri, et al "A Proposed Paper on Spam Email Detection using Machine Learning", Journal of Emerging Technologies and Innovative Research (JETIR), Page No. a573-577

[15] Nebojsa Bacanin, et al "Application of Natural Language Processing and Machine Learning Boosted with Swarm Intelligence for Spam Email Filtering", Mathematics 2022, 10, 4173. https://doi.org/10.3390/math10224173

[16] Pooja Malhotra, et al "Spam Email Detection using Machine Learning and Deep Learning Techniques", https://ssrn.com/abstract=4145123

[17] Omar Almomani, et al "A Hybrid Model Using Bio-Inspired Metaheuristic Algorithms for Network Intrusion Detection System", Tech Science Press, Page No. 410-429