IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

A novel approach for spam detection using natural language processing with AMALS models

Agarwal R.[1], Dhoot A.[2], Surya Kant[3], Vimal Singh Bisht[4], Hasmat Malik[5], Mohd. Fahim Ansari[6], Asyraf Afthanorhan[7], Mohammad Asef Hossaini[8*]

[1]JDepartment of Computer Applications, JIMS Engineering Management Technical Campus, Greater Noida, 201303, India
[2]Department Phystech School of Radio Engineering and Computer Technology, Moscow Institute of Physics & Technology, Moscow, 141701, Russia
[3]Department of Electronics and Communication Engineering, Graphic Era Hill University, Bhimtal, India
[4]Department of Electronics and Communication Engineering, Graphic Era Hill University, Bhimtal, India
[5]Department of Electrical Power Engineering, Faculty of Electrical Engineering, University Technology Malaysia (UTM), Johor Bahru 81310, Malaysia, Skudai 81310, Johor, Malaysia (Email: hasmat.malik@gmail.com)
[6]Department of Electrical Engineering, Graphic Era Deemed to be University, Dehradun 248002, India, (Email: drfahim.ee@geu.ac.in)
[7]Universiti Sultan Zainal Abidin (UniSZA), Gong Badak, Kuala Terengganu 21300, Terengganu, Malaysia (Email: asyrafafthanorhan@unisza.edu.my)
[8]Department of Physics, Badghis University, Afghanistan (Email: asef.hossaini_edu@basu.edu.af)

Corresponding author[*]: Mohammad Asef Hossaini (asef.hossaini_edu@basu.edu.af).

ABSTRACT To enhance their company operations, organizations within the industry leverage the ecosystem of big data to manage vast volumes of information effectively. To achieve this objective, it is imperative to analyze textual data while prioritizing the safeguarding of data integrity and implementing robust measures for organizing and validating data through the utilization of spam filters. Various methodologies can be employed, including Word2Vec, bag-of-words, BERT, as well as term frequency & reciprocal document frequency (TF-IDF). Nevertheless, none of these solutions effectively address the problem of data scarcity, which might lead to the existence of missing information in the collected documents. To properly address this problem, it is necessary to employ a strategy that categorizes each document based on the topic matter and uses statistical approaches for approximation. This research paper presents a novel approach for spam detection using natural language processing. The proposed strategy utilizes a least-squares model to modify themes and incorporates gradient descent and altering least-squares (i.e., AMALS) models for estimating missing data. TF-IDF and uniform-distribution methods perform the estimation. The performance evaluation reveals that the suggested technique exhibits a superior performance of 98% compared to the existing industry TF-IDF model in accurately predicting spam within big data ecosystems. By this model, the environment of an organization or a company can be saved from spamming or other attacks, which can lead to extracting their data for unauthorized users to protect the details.

INDEX TERMS Artificial Intelligence, Big Data, Machine Learning, Spam Detection

I. INTRODUCTION

In the contemporary age of information technology, the process of transferring information has become significantly streamlined and expedited. Numerous platforms exist that enable users to disseminate knowledge globally. Email is often regarded as one of the most straightforward, cost-effective, and expeditious means of disseminating information around the globe. However, owing to their inherent simplicity, electronic mail (email) systems are susceptible to several forms of malicious activities, with the most prevalent and perilous being unsolicited bulk messages, commonly referred to as spam [1]. The receipt of unsolicited emails that are unrelated to one's interests is generally undesirable since it results in the wastage of recipients' time and resources. In addition, it is important to note that emails might potentially contain harmful content that is concealed within attachments or URLs, posing a risk to the security of the host system [2]. Spam refers to the transmission of unsolicited and irrelevant messages or emails by an individual or entity to a large number of recipients using various means of information dissemination, such as email or other communication channels [3]. Therefore, there is a significant need for robust security measures in place for the email system. Spam emails have the potential to contain malicious software such as viruses, rats, as well as Trojans. This approach is predominantly employed by attackers to entice consumers into internet services. The individuals in question can transmit unsolicited emails that include attachments including a variety of file extensions. These attachments may contain URLs that have been manipulated to direct users to websites that engage in hazardous activities, such as spamming and fraudulent behaviour. As a result, users may experience detrimental consequences, including data or financial fraud, as well as identity theft [4, 5]. Numerous email service providers offer their users the capability to establish rule-based filters that automatically categorize incoming emails based on keywords. However, this methodology seems to be of limited utility as it presents challenges in terms of complexity, and users exhibit a reluctance to personalize their emails, rendering their email accounts vulnerable to spam attacks.

Over the past few decades, the Internet of Things (i.e. IoT) has emerged as an integral aspect of contemporary society, seeing significant and rapid expansion. The Internet of Things (IoT) has emerged as a crucial element within the context of smart cities. There exists a multitude of social media applications and platforms that are based on the Internet of Things (IoT) technology. The proliferation of the Internet of Things (IoT) has led to a significant escalation in the prevalence of spamming issues. The researchers put forth a range of spam detection techniques to identify and eliminate spam content and individuals engaging in spamming activities. The current methods for spam identification can be broadly classified into two categories: behaviour pattern-based approaches as well as semantic pattern-based approaches. These methodologies possess inherent restrictions and disadvantages. The proliferation of spam emails has experienced a notable expansion in tandem with the emergence and widespread adoption of the Internet & global communication [6]. Spam messages are produced globally through the utilization of the Internet, employing techniques to conceal the identity of the attacker. Numerous antispam methods and approaches have been developed; yet, the prevalence of spam remains significantly elevated. The most perilous forms of unsolicited electronic communications are malicious emails that include hyperlinks directing recipients to websites designed to inflict harm upon the victim's data. The presence of spam emails has the potential to impede server response times due to the occupation of server memory or capacity. To effectively identify and prevent the proliferation of spam emails, organizations undertake a meticulous assessment of the various instruments at their disposal to address this escalating problem. Several well-known techniques for identifying and analyzing incoming emails to detect spam include whitelisting/blacklisting [7], email header analysis, and keyword verification, among others.

According to estimates provided by social networking professionals, over 40% of accounts on social networks are utilized for spam [8]. Spammers employ widely used social networking technologies to selectively target distinct segments, and review the pages, or fan pages to discreetly embed hyperlinks that direct users to pornographic and other commercial websites. These websites are typically associated with false accounts and aim to promote the sale of illicit products. The poisonous emails that are disseminated to persons or organizations of a similar nature have recurring characteristics. Through a thorough examination of these key points, one can enhance the efficacy of identifying and detecting such forms of electronic correspondence. The classification of emails into spam & non-spam categories can be achieved by the application of artificial intelligence (AI) [9]. One alternative approach to solving this problem involves extracting features from the headers, subject, & body of the messages. Once the data has been extracted and categorized according to their characteristics, they can be classified into two groups: spam or ham. Currently, spam detection is frequently accomplished by the utilization of learning-based classifiers [10]. In the context of learning-based classification, the approach to detection operates under the assumption that spam emails possess distinct properties that can be used to identify them from valid emails [11]. Several aspects contribute to the heightened complexity of the spam identification process in learning-based models. The elements encompassed in this context are spam subjectivity, concept drift, linguistic difficulties, overhead processing, as well as text latency.

Prominent multinational firms like Amazon have established an extensive infrastructure comprising numerous servers and databases. These resources are utilized not only for the storage of literary works but also to accommodate a substantial volume of product-related data. The aforementioned data facilities have been intentionally created to attain optimal productivity and have the potential to be offered as services to other organizations [1]. Various forms of structured data are grouped inside big data ecosystems. However, text data often lacks structure and necessitates analysis to offer additional services utilizing consumer big data. The capturing of the features of company and customer actions in the online environment may be achieved through the use of textual communication [2]. The utilization of Natural Language Processing (i.e. NLP) methodologies for the analysis of unstructured textual data encompasses approaches such as Word2Vec and bag-of-words.

Bag-of-words (BOW), Bidirectional Encoder Representations from Transformers (BERT), & term frequency–inverse document frequency (i.e. TF-IDF) are three commonly used techniques in natural language processing (NLP). Nevertheless, the task of analyzing surface-level textual information obtained through Natural Language Processing (NLP) poses challenges, particularly about the scarcity and omission of textual data. To address this issue, traditional models employ a range of methodologies in conjunction with machine learning and statistical methods. Furthermore, several models have conducted comparisons and experiments on documentary clustering matrices by transforming the document-word matrix into a document-factor scoring matrix (Jun et al., 2018). Nevertheless, the issue of sparsity continues to have an impact on the performance of document clustering. This paper introduces a novel approach for spam identification using natural language processing (NLP). The proposed technique combines the ratios of topic-altering least squares (i.e. TALS), approximations gradient descent (i.e. AMGD), & approximations alternating least squares (i.e. AMALS) models:

- The TALS framework categorizes feature-related concerns by putting them into the process of addressing sparsity issues and approximating them through the utilization of a probability distribution. This approach aims to enhance the predictability and suitability of the features.

- The AMGD algorithm employs a gradient descent (i.e. GD) function as well as a uniform distribution to address the issue of missing information by approximating the model.

- The remaining scarcity issue is addressed by AMALS by the implementation of alternating least squares (i.e. ALS), L2 normalization, & uniform distribution.

This research presents a unique machine learning methodology to address the challenges of shortage and missing information in large-scale data documents.

- This research successfully reduces the performance gap between the testing & training sets of documents.
- This study provides a novel natural language processing (NLP)--based spam detection model that exhibits enhanced performance in comparison to the conventional term frequency-inverse document frequency (TF-IDF) approach.
- This study presents a new finding that supports the advantages of utilizing the ALS function in conjunction with the GD algorithm for effectively classifying spam text inside a large-scale data environment.

The subsequent sections of this work are organized in the following manner. Section II provides an overview of the backdrop. Section III provides an elucidation of the underlying factors that drive the research endeavour. Section IV introduces the recommended methodology. Section V of the paper provides an analysis and assessment of the subject matter, while Section VI serves as the concluding section, summarizing the main findings and implications of the study.

An instance of learning-based models can be observed in the form of an extreme learning machine (i.e. ELM). The present study introduces a contemporary machine-learning approach designed for feedforward neural networks, specifically focusing on architectures with a solitary hidden layer [12]. When compared to standard neural networks, it effectively addresses issues related to sluggish training speed and overfitting. In the ELM framework, a single iteration cycle is sufficient. Due to its enhanced capacity for generalization, robustness, as well as controllability, this method has gained widespread adoption across various domains. This study examines various machine-learning techniques utilized in the context of spam identification. The contributions made by our team are categorized as follows:

- The present paper examines a range of machine learning-based filters for spam, exploring their architectural design and evaluating their respective advantages and disadvantages. In addition, we engaged in a discussion regarding the fundamental characteristics of unsolicited email communications, commonly referred to as spam.
- A complete examination of the proposed strategies and the nature of spam revealed some intriguing research gaps in the field of spam detection and filtering.
- This section presents a discussion on open research topics and future research objectives aimed at enhancing email security and spam email filtration through the utilization of machine learning algorithms.
- In this paper, the authors examine the existing obstacles encountered by spam filtering algorithms and analyze the impact of these challenges on the efficiency of the models.
- This paper presents a thorough examination of several machine learning techniques & concepts, with a specific focus on their application in the field of spam identification.
- The paper classifies several machine learning techniques-based spam detection approaches to gain a comprehensive understanding of their underlying principles.
- This section presents a range of potential avenues for future research in the field of spam detection and filtration. These areas aim to enhance the detection capabilities and bolster the security of email platforms.

II. LITERATURE REVIEW

Email spam refers to the dissemination of fraudulent or unsolicited bulk messages through various accounts or automated systems. The proliferation of unsolicited emails, commonly referred to as spam, has exhibited a steady upward trend, emerging as a prevalent issue during the past ten years. Spam emails are commonly obtained through the utilization of spambots, which are automated programs designed to scour the Internet for email addresses. The utilization of machine learning techniques has significantly contributed to the identification and detection of unsolicited and unwanted emails commonly referred to as spam. Researchers are employing a range of models and strategies to advance the development of innovative spam detection & filtering models [13]. In their study, Kaur and Verma [14] conducted a comprehensive survey on the topic of email spam detection. They focused on employing a supervised approach that incorporates feature selection techniques. The authors engage in a discourse regarding the knowledge discovery process employed in the context of spam detection systems. In addition, the authors provide detailed explanations of numerous strategies and technologies that have been presented for the detection of spam. This survey also discusses the selection of features using N-gram analysis. The N-Gram [15, 16] algorithm is a predictive-based method employed for estimating the likelihood of the subsequent word appearing after identifying $N - 1$ phrases inside a sentence as well as text corpus. The N-Gram model employs probabilistic methods to anticipate the subsequent word. The study conducts a comparative analysis of different ways for email spam detection, including both machine learning techniques such as multilayer perceptron neural network, support vector machine, and Naïve Bayes, as well as non-machine learning methods such as Signatures, Blacklist as well as Whitelist, including mail header verification.

In their publication, (Saleh et al., 2020) provide an extensive examination of the topic of smart spam email detection through the use of a survey. The authors engage in a comprehensive examination of security vulnerabilities associated with electronic mail, with a particular emphasis on spam emails. The discourse encompasses an exploration of the breadth of spam analysis, as well as an examination of several methodologies employed in both machine learning and non-machine learning approaches for spam identification and filtration. The researchers concluded that there is a significant prevalence of supervised learning algorithms, as evidenced by the adoption rates, in the context of email spam detection [18]. The authors assert that the primary reason for the widespread adoption of supervised learning is due to the high level of accuracy and consistency exhibited by supervised techniques. The researchers also engaged in a discussion on multialgorithm frameworks and determined that such frameworks exhibit more efficiency compared to their single-algorithm counterparts. It has been observed that the majority of research endeavours involving the use of email content to identify spam, namely phishing emails, mostly rely on word-based classification as well as clustering techniques.

(Sun et al., 2017) provide a comprehensive overview of learning-based methodologies employed in the domain of email spam filtering. This study discusses the issue of spam and presents a comprehensive analysis of learning-based spam filtering techniques. The authors elucidate diverse characteristics of unsolicited electronic communications commonly referred to as spam emails. This study examines the impact of spam emails on various domains. This study also examines the diverse economic and ethical concerns associated with spam. The prevalent antispam strategy involves the utilization of learning-based filtering, which has undergone significant advancements. The filters that are frequently employed rely on diverse classification approaches that are applied to the different elements of email communications. This paper posits whether the Naïve Bayes classifier occupies a distinct place among various learning algorithms employed in the context of spam filtering. The tool exhibits remarkable efficiency and clarity, yielding outcomes of great accuracy.

In their study, Bhuiyan et al. (2020) provide a comprehensive analysis of contemporary methodologies employed in email spam filtering. The authors provide an overview of several spam filtering methodologies and evaluate the performance of several suggested systems by examining multiple metrics through a comprehensive analysis. The authors engage in a discussion regarding the efficacy of various approaches employed to filter unsolicited and unwanted emails commonly referred to as spam. Certain individuals have achieved favourable outcomes, while others are endeavouring to integrate alternative methods to enhance their level of accuracy. Despite their overall success, experts remain concerned about the various challenges encountered in spam filtering technologies. The researchers are endeavouring to develop an advanced spam filtering mechanism capable of comprehending vast quantities of multimedia data to effectively filter out spam emails. The authors conclude that the predominant approach for email spam filtering involves the utilization of the Naïve Bayes and Support Vector Machine (SVM) algorithms. To evaluate the efficacy of spam filtration models, it is possible to train these models using many datasets, such as the "ECML" and UCI datasets [21].

In their study, [Ferrag et al., 2020] conducted a comprehensive examination of deep learning techniques utilized in intrusion detection systems as well as spam detection datasets. The authors engaged in a comprehensive examination of detection systems that rely on deep learning models, subsequently assessing the efficacy of those models. The researchers analyzed a total of 35 widely recognized cyber datasets, which were then classified into seven distinct groups. The aforementioned categories encompass datasets that are classified as Internet traffic-based, networking traffic-based, Intranet traffic-based, electric network-based, virtualized private network-based, Android apps-based, IoT traffic-based, & Internet linked device-based datasets. The researchers concluded that deep learning models exhibit superior performance compared to classical machine learning and lexical models in the context of intrusion as well as spam detection.

In their study, (Vyas et al. 2016) provide a comprehensive analysis of supervised machine-learning techniques employed in the context of spam email filtering. The researchers concluded that the Naive Bayes method exhibits superior speed and satisfactory precision compared to the other methods reviewed, except SVM and

ID3. Support Vector Machines (SVM) and Iterative Dichotomiser 3 (ID3) algorithms provide higher precision compared to the Naïve Bayes algorithm, albeit at the cost of significantly increased system construction time. A trade-off exists between the factors of timing and precision. The authors conclude that the choice of learning algorithm is contingent upon the specific circumstances and the desired levels of accuracy and efficiency. It is asserted that to develop a more resilient spam filtering architecture, careful consideration should be given to all components of the email.

This survey study examines three primary categories of machine-learning techniques that can be employed for spam filtering. In this study, we undertake a comprehensive examination of multiple scholarly articles, analyzing the suggested methodologies and deliberating on the obstacles encountered in spam detection & filtration systems. This paper also examines the merits and drawbacks of the proposed methodologies for spam identification and filtration that have not been previously evaluated.

III. SPAM DETECTION

The origin of the term "spam" can be traced back to a Monty Python episode [23], whereby the Hormel canned beef product is humorously exaggerated and repetitively emphasized. The term "spam" was reportedly first employed in 1978 to refer to unsolicited email. However, its prevalence grew significantly in the mid-1990s, extending beyond academic and research communities [24]. One such type is the development expense deception, wherein a recipient is sent an electronic communication with a proposition that purportedly leads to a reward. During the contemporary technological era, the dodger or spammer presents a narrative wherein an unlucky individual requires immediate financial assistance, enabling the fraudster to amass a significantly larger sum of money, which they would subsequently distribute amongst themselves. The individual engaging in fraudulent activities may choose to either generate financial gains or cease all forms of communication once the unsuspecting victim fulfils the agreed-upon instalment.

A. METHODOLOGY FOR SPAN FILTERING FOR IoT PLATFORMS & EMAIL

The prevalence of unsolicited emails, sometimes referred to as spam, is experiencing a notable surge across several domains including marketing, chain communication, stock market tips, politics, as well as education [24]. At present, multiple organizations are engaged in the development of diverse approaches and algorithms aimed at enhancing the effectiveness of spam detection & filtering processes. In this section, we examine several filtering procedures to have a comprehensive understanding of the filtering process.

- METHOD OF SPAM FILTERING

The standard spam filtering mechanism is the filtering system that employs a predefined set of rules and operates as a classifier based on these protocols. Figure 1 depicts a conventional approach to the filtration of unsolicited electronic communications, commonly referred to as spam. The initial phase involves the implementation of content filters, which employ artificial intelligence methodologies to discern and identify spam [25]. The implementation of the email headers filter, which involves the extraction of header data from the email, occurs during the second stage. Subsequently, a series of backlist filters are implemented to effectively identify and intercept emails originating from the backlist file, thereby mitigating the influx of spam emails. Following this phase, rule-based filters are employed to identify the sender by utilizing the subject line and parameters specified by the user. The utilization of allowance & task filters is achieved through the implementation of a technique that enables the account holder to initiate the transmission of mail [26].

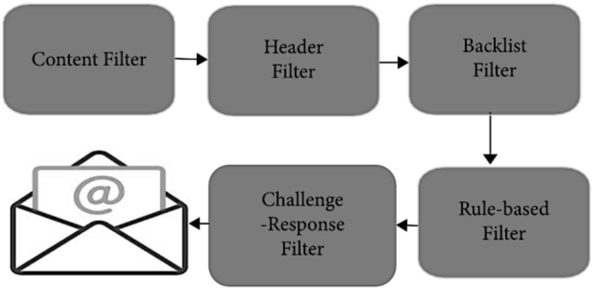

Figure 1: **Email filtering process**

- SPAM FILTERING ON THE CLIENT SIDE

A client refers to an individual who can utilize the Internet as well as an email network to transmit or receive electronic mail [27]. Client-side spam detection provides various rules and techniques to ensure the secure delivery of communications between individuals and organizations. To facilitate the transfer of data, a client should implement several pre-existing frameworks on their system. These systems establish connections with client mail agents as well as carry out the task of filtering the client's mailbox by composing, accepting, and managing incoming emails [28, 29].

- SPAM FILTERING AT ENTERPRISE LEVEL

Email spam detection at the enterprise level involves the implementation of diverse filtering frameworks on the server. These frameworks are responsible for managing the mail transfer agent as well as categorizing the received emails as either spam or legitimate (ham) [30]. The system client continuously and successfully utilizes the enterprise filtering methodology to filter emails on a network. Current approaches to spam detection employ a scoring system to evaluate emails. This principle outlines the specification of a rating function, which generates a score for each post. The categorization of messages as either junk mail, as well as

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3391023

ham, is determined through the assignment of certain scores as well as ranks [31]. Due to the varying tactics employed by spammers, a list-based technique is frequently employed to automatically block their communications, necessitating continual modifications to all associated activities. The reproduction of Figure 2 has been sourced from the work of (Bhuiyan et al., 2018). The architecture of both the client & enterprise-level filtering of spam process is depicted in Figure 2.

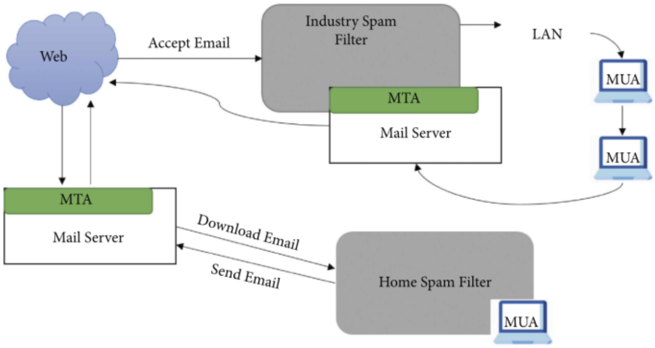

Figure 2: **Enterprises level spam process**

- SPAM FILTERING BASED ON ANY CASE

The case-based as well as sample-based filtering of spam systems is a widely recognized and traditional machine-learning approach for detecting spam [32]. Figure 3 depicts a standard case-based filtering framework. The filtering process in question involves multiple stages, facilitated by the collecting method. In the initial phase, data (namely, emails) is gathered. Subsequently, the primary transition persists using the preprocessing procedures executed via the graphical user interface of the client. These steps involve delineating abstraction and selecting the method for classifying email data. The overall process is then tested using vector expression, resulting in the classification of the data into two distinct categories: spam as well as legitimate email.

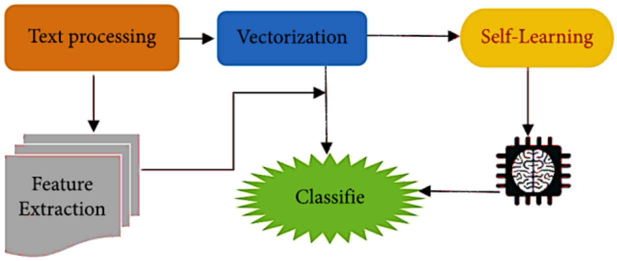

Figure 3: **Standard case of filtering framework**

IV. UNITS

The Internet of Things (i.e. IoT) refers to a network of interconnected objects that are connected to the Internet and capable of collecting and transmitting data wirelessly, without requiring human involvement. The Internet of Things (IoT) facilitates the seamless integration and deployment of physical items in various geographical locations. In the given context, the effective management and monitoring of network performance pose significant challenges and necessitate the implementation of robust privacy and security solutions. To address security concerns in IoT applications, it is imperative to prioritize the protection of privacy against various threats, including but not limited to intrusions, phishing attempts, DoS attacks, spamming, as well as malware. The iOS operating system, encompassing both its objects and networks, exhibits susceptibility to network & physical threats as well as privacy breaches. Figure 4 provides a visual representation of the primary categories of attacks targeting the Internet of Things (IoT).

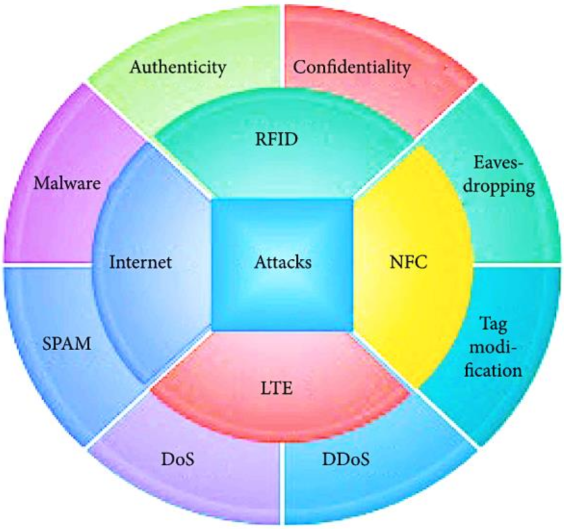

Figure 4: **Primary categories of IoT**

The enumerated instances of attacks targeting Internet of Things (IoT) systems are presented as follows:

- A self-promotion attack: This attack involves a hacked node attempting to gain superiority over all other nodes within the Internet of Things (IoT) environment for a specific recommendation.
- Criticism or derogatory remarks directed towards someone or something. In the context of this attack, a compromised node erroneously accepted an incorrect recommendation, potentially undermining the trustworthiness of the trustworthy node. The services provided by the trusted node experienced a decline.
- The topic under discussion is the ballot stuffing attack. Within the context of the Internet of Things (IoT) ecosystem, it is observed that a compromised node can amplify the functionality and effectiveness of other compromised nodes. The compromised node has an opportunity to provide its services. This phenomenon is commonly referred to as the collision advice assault.
- The topic of discussion is an opportunistic service attack. In this particular form of attack, a compromised node actively cooperates with other malicious nodes in order to execute the mouthing & ballot stuffing attack.

- The topic of discussion is the On-Off Attack. In this particular style of attack, the infiltrated node exhibits substandard service provision, as it engages in the random execution of detrimental services.
- The concept of node tampering. The perpetrator manipulates the malevolent node and obtains targeted data, including a security key.
- The topic of discussion is the malicious node attack. The perpetrator physically inserts the malicious node into the group of nodes.
- The topic of discussion is the Man in the Middle Attack. In this particular form of attack, the assailant covertly intercepts the conversation between two nodes across the Internet. The perpetrator acquires crucial information through the act of surreptitiously listening in on private conversations.
- The Sybil Attack is a type of security threat that involves an adversary creating many fake identities in a network to gain control or manipulate the system. The compromised node illicitly appropriates the reputation of legitimate nodes and assumes the role of a trustworthy node.

A study conducted by Nozomi Networks reveals a notable rise in attacks and threats targeting Operational Technology (i.e. OT) & Internet of Things (IoT) networks over the initial six months of 2020. Figure 5 illustrates the frequency of cyber assaults on Internet of Things (IoT) devices throughout different years.

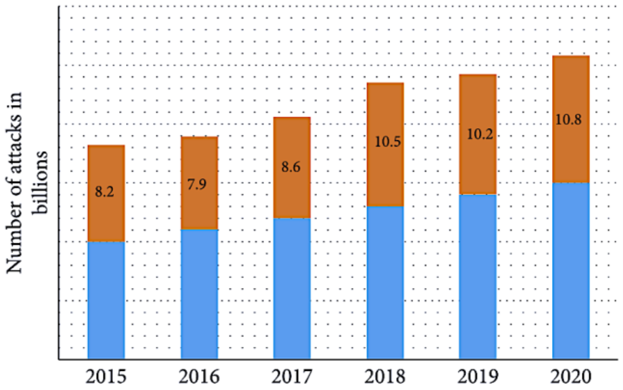

Figure 5: **Frequency of getting cyber assaults**

Machine learning methodologies have demonstrated considerable efficacy in the realm of preventing and detecting such attacks, exhibiting superior performance. Numerous research projects have been conducted to identify and mitigate the aforementioned difficulties outlined in Section 5.

V. MACHINE LEARNING

Machine learning is widely recognized as a significant and valuable implementation of artificial intelligence (i.e. AI), enabling computer systems to autonomously acquire knowledge and improve their performance without the need for explicit programming [34]. The basic objective of machine learning algorithms is to construct automated systems that enable the retrieval and utilization of data for training. The initial stage of the learning process involves acquiring labelled data, which is commonly referred to as the training dataset. The user's input can encompass several forms such as real-life experiences, reviews, examples, or feedback. These forms serve the purpose of identifying patterns within the data, hence enabling improved decision-making in the future. The primary goal of machine learning methods is to acquire knowledge autonomously, without requiring human interaction. Machine learning encompasses three primary categories that are employed for a wide range of activities.

Over the past decade, scholars have endeavoured to enhance the efficacy of email communication beyond its current state. The implementation of spam filtering techniques for email systems is widely recognized as a crucial measure in safeguarding email networks [35]. Numerous scholarly publications have been dedicated to employing diverse machine-learning methodologies to detect and manage spam emails. However, certain areas within this research domain remain unexplored or inadequately addressed. The study of junk mail is a prominent and compelling area of research that addresses existing knowledge gaps [36]. Numerous studies have been conducted to enhance the reliability and use of email communication by employing various techniques in spam classification. This study aims to provide a concise overview of several machine-learning techniques and approaches currently employed in the field of email spam detection. This research additionally assesses the prevailing machine learning methodologies, namely K-Nearest Neighbors (KNN), Support Vector Machines (SVM), random forest, as well as Naïve Bayes.

A. SPAM FILTERING BASED ON ML

Machine learning plays a crucial role in enabling the efficient processing of large volumes of data. While the use of this technology generally yields expedited and precise outcomes in identifying undesirable content, it may necessitate additional investments of time as well as finances to adequately train the models for optimal performance. The combination of machine learning, artificial intelligence (AI), and cognitive computing has the potential to enhance the processing capabilities of large datasets. Figure 6 illustrates a range of machine learning methodologies.

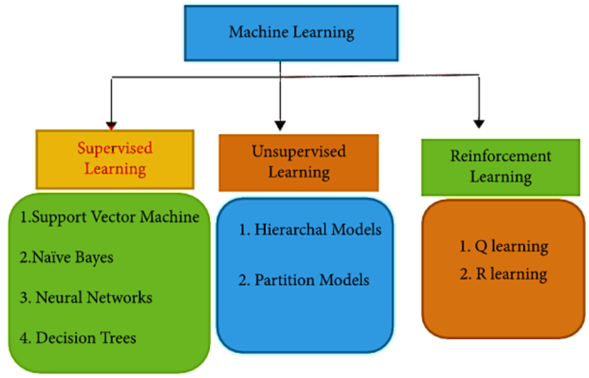

Figure 6: Methodologies of ML

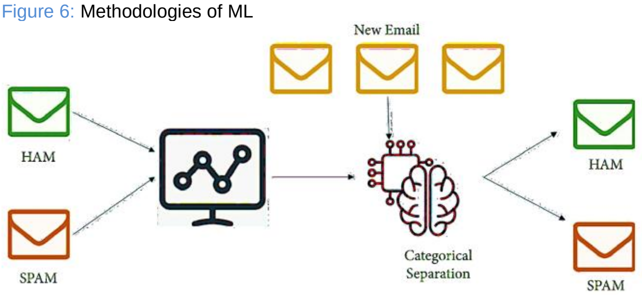

Figure 7: Supervised Learning Method of ML

1. *Supervised ML*

Supervised ML algorithms refer to machine learning models that require annotated data to learn and make predictions. The models are initially trained using labelled training data and subsequently used to make predictions about future events. To clarify, these models initiate the process by examining a pre-existing training dataset, from which they derive a methodology for predicting success ratings. After undergoing appropriate training, the system can generate predictions for any new data that is relevant to the data provided by the user during the training phase [38]. In addition, the learning algorithm effectively evaluates the generated output against the desired output and detects mistakes to refine the model.

Supervised learning is a machine learning approach that relies on the utilization of labelled data during the training phase, enabling the model to make predictions on unseen data. This form of learning has applications in diverse problem domains, including assessing the attractiveness of advertisements, classifying spam emails, recognizing faces, and categorizing objects. The method of supervised learning is depicted in Figure 7.

VI. DT Classifier

The decision tree classifier is an approach to machine learning that has gained significant popularity in the field of classification during the past decade [39]. The present technique utilizes a straightforward approach for resolving categorization problems. A classifier based on decision trees refers to a set of precisely defined inquiries about the properties of test records. With each response obtained, a subsequent inquiry arises, leading to a continuous cycle of questioning until a definitive conclusion is reached and documented [40]. Tree-based decision algorithms are a class of models that are generated by an iterative or recursive process, utilizing the available data. The objective of decision tree-based algorithms is to forecast the value of a target variable based on a given set of input values. The technique described in this study utilizes a hierarchical tree-based structure to effectively address classification and regression difficulties [41]. The basic structure underlying the decision tree is depicted in Figure 8.

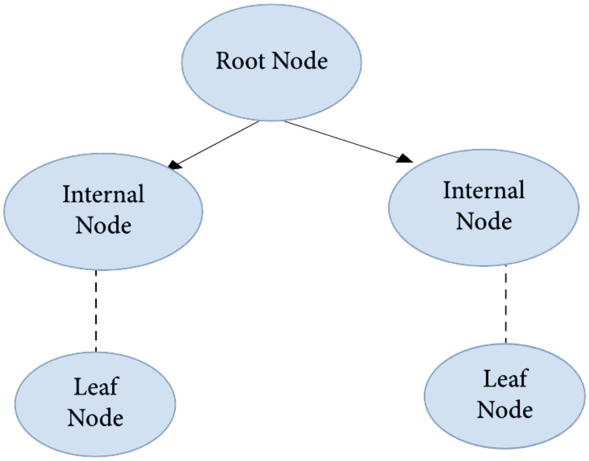

Figure 8: Decision Tree Structure

Several decision tree algorithms include the following:

- The random forest algorithm is a popular machine-learning technique that combines many decision trees to make predictions.
- Classification and regression trees (CART) are a type of decision tree algorithm used for both classification and regression tasks.
- C4.5 and C5.0 are specific versions of decision tree algorithms that use information gain and gain ratio measures to construct the trees.
- The chi-square test is a statistical method used to determine the independence of two categorical variables.

This section presents an examination of various decision tree algorithms that have been developed to detect and prevent email spam:

Larson et al., (2018) examine a spam filtering methodology that uses random forest algorithms to categorize spam emails, while also incorporating active learning techniques to enhance the accuracy of the classification (43). The researchers utilized the dataset comprising email messages sourced from RFC 822 (Internet) [44] and subsequently partitioned each email into two distinct portions. Next, the researchers calculate the term frequency as well as inverse document frequency for all features present in each email, commonly referred to as

TF/IDF. To construct the training dataset, a clustering technique is employed to label a collection of emails. Following an evaluation of the cluster prototype emails for training purposes, the researchers proceeded to conduct experiments utilizing supervised machine learning methods, namely random forests, Naïve Bayes, support vector machine, as well as KNN [45]. The findings of the study indicate that the "random forest" method demonstrates superior efficiency in data classification, with an accuracy rate of 95.2%.

Takhmiri and Haroonabadi [43] propose an alternative approach for spam detection, utilizing a fuzzy decision tree in conjunction with the Naïve Bayes method. The bake voting algorithm is employed to extract patterns of spam behavior. This behaviour is shown due to the absence of overt qualities in the tangible realm. The degree of cross-linking utilized to explicate or depict personalities is both sensible and impartial. Decision trees employ fuzzy Mamdani rules to classify spam and ham emails. Subsequently, the authors employ the Naïve Bayes classifier [45] to analyze the dataset. Ultimately, the electoral process employs the technique of partitioning votes into more manageable segments. This solution provides an optimum weight that may be applied to derived percentages to reach a higher level of accuracy. The dataset utilized in this research included a total of 1000 electronic mail messages, out of which 350 (35%) were identified as spam, while the remaining 650 (65%) were classified as legitimate messages (ham).

Verma and Sofat (2018) employed the supervised machine learning technique ID3 (Quinlan, 1986) to construct decision trees for the given task [46]. Additionally, they utilized the hidden Markov model (Fine, 1998) to estimate the probabilities of various occurrences, which were then combined to categorize emails as either junk mail or ham [47]. The suggested model employs a method of initially classifying emails as either spam or valid by assessing the overall probability of each email based on the later classification of email phrases. Subsequently, the system proceeds to construct decision trees for individual emails. This analysis utilizes the Enron dataset [48], which has a total of 5172 emails. Out of the total 5172 emails analyzed, 2086 were identified as spam, while an equal number of 2086 were classified as legitimate emails. The model can classify emails as either spam or ham by utilizing the feature set derived from the Enron dataset. An 11% inaccuracy was obtained when utilizing the fitness function from the sk-learn library in the suggested model. The model achieved an accuracy rate of 89% on the provided dataset.

The email classification methodology for IoT systems presented by Li et al. [49] is founded on the principles of supervised machine learning. The employed methodology involves the utilization of a multiview methodology that prioritizes the acquisition of more comprehensive data for classification. A dataset with two distinct feature sets, namely internal and exterior, is generated. The suggested methodology has the potential to be applied to both labelled as well as unlabeled data, and its effectiveness was assessed using two datasets inside an authentic network setting. The findings of this study suggest that the implementation of the multiview model yields higher levels of accuracy compared to the straightforward approach of email classification. Ultimately, the multiview model is juxtaposed with other extant models.

Subasi et al. (2019) proposed a spam filtering methodology that utilizes various decision tree algorithms [50]. The objective of their study was to assess the accuracy of these algorithms and determine the most effective one for their specific dataset. The researchers applied various algorithms, including regression, classification, and tree (CART), NBT, C4.5, LAD tree, REP Tree, random forest, & rotation forest, to the dataset to perform email classification. The findings of the study indicate that the customized random forest model outperformed other decision tree models in terms of accuracy when applied to publically available datasets.

- SVM

The support vector machine (i.e. SVM) is a crucial and highly esteemed machine learning model [51]. The Support Vector Machine (SVM) is a prejudiced supervised learning classifier that is technically defined. It operates by utilizing labelled examples during the training phase and produces a hyperplane as its output, which is used to categorize fresh data [52]. Objects in a given set are segregated based on their respective class memberships using decision planes. The classification principle of linear SVMs is depicted in Figure 9. The depicted diagram includes many circular and star-shaped entities, which are referred to as objects. These objects have the potential to be classified into one of two categories, specifically the category of stars as well as dots. The selection of items between those that are green and those that are brown is determined by the isolated lines. The objects located on the bottom half of the plane exhibit a brown star shape, while the objects situated on the top edge of this plane are represented by green dots. This distinction indicates that two distinct objects have been categorized into separate classes. When presented with a new object, specifically a black circle, the model will utilize the training instances provided during the training phase to categorize the circle into a single of the available classes.

In their study, Banday and Jan [53] provide a comprehensive analysis of the statistical spam filter methodology. The filters are designed with Naïve Bayes, support vector machines (i.e. SVM), KNN, as well as regression trees [54]. Various supervised machine learning methods are employed, and the obtained results are assessed by metrics such as precision, recall, as well as accuracy. Based on the application of these machine learning techniques, it was shown that the dataset yielded optimal results when utilizing classification & regression trees (CART) [55] as well as Naïve Bayes classifiers. According to this method, the computational cost of evaluating false positive instances is higher than that of false negative instances in the context of spam filtering.

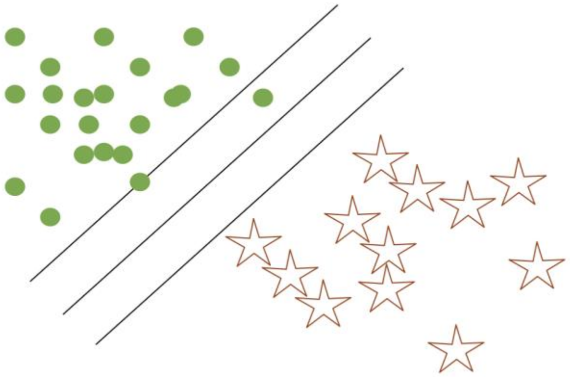

Figure 9: **Linear SVMs**

The approach proposed by Zeng et al. [56] aims to identify and classify spammers as well as spam communications within a given social network. In contemporary society, the use of social media has become ubiquitous, with a substantial portion of individuals devoting a significant portion of their time to engaging in interpersonal communication with their close acquaintances. Spammers exploit diverse social media networks as well as the content posted by users to disseminate malicious content, ads, information, and other undesirable materials within the accounts of social media users. This study examines the methods for identifying and detecting posts or information with malicious intent on social media sites. The researchers in this study employ the Sina Weibo social network and utilize a machine learning method known as support vector machine (SVM) to identify and classify spammers. The dataset employed in this study comprised 16 million messages obtained from many individuals. A set of 18 features was employed as a component of the vector set. The network's clientele can be classified into two distinct groups: legitimate users as well as spammers. The model's training phase utilized 80% of the available data, with the remaining 20% allocated for testing purposes. To enhance the precision of the results, a ratio of 1:2 was employed between spammers and non-spammers in the training dataset. The suggested model achieves a classification accuracy of 99.5% for distinguishing between spammers and non-spammers, as reported in reference [57].

Jamil et al. [58] describe a fitness framework that utilizes Internet of Things (IoT)-enabled blockchain technology & machine learning approaches. The model they have suggested consists of two distinct parts. The first system is a network that utilizes blockchain technology to ensure the security of devices that sense. It also incorporates intelligent contracts to facilitate relationships as well as an inference engine the fact that reveals concealed insights and actionable information from data collected from Internet of Things (IoT) sensors and user devices. The enhanced smart contract provides customers with a valuable application that enables real-time monitoring, enhanced control, and expedited access to multiple devices dispersed across diverse domains. The primary objective of the inference engine's module is to analyze the data collected from the IoT environment to identify hidden patterns and extract valuable information. This process aids in facilitating efficient decision-making and offering easy services. According to the findings of the researchers, the model they have proposed has the potential to enhance system throughput and optimize resource utilization. The technology suggested in this paper has potential applications in many domains, such as healthcare and intelligent enterprises.

The spam filtering program was developed by Olatunji [59], employing support vector machine as well as extreme learning machine techniques. The researcher utilized a commonly employed dataset to construct the spam detection model. The support vector machine (SVM) earned an accuracy rate of 94.06% in the study, while an extreme learning machine (ELM) model acquired an accuracy rate of 93.04%. This indicates that the SVM outperformed the ELM by a marginal improvement of 1.1% in terms of performance. The individual suggested that the improvement in accuracy of Support Vector Machines (SVM) compared to Extreme Learning Machines (ELM) is minimal. This suggests that in scenarios where the timeliness of detection is of utmost importance, including in real-time systems, it is advisable to prioritize the utilization of the ELM spam detector over the SVM spam detection method. While doing his research, it was observed that the Support Vector Machine (SVM) exhibited a greater level of accuracy. However, it was also noted that the training process of the SVM system required more time compared to the Extreme Learning Machine (ELM) system. Tretyakov (2012) provided an extensive analysis of different machine-learning methodologies employed in the context of email spam filtering [60]. This study conducted a comparative analysis of precision outcomes between instances of false positives against precision outcomes after the removal of false positives. The findings demonstrate the outcomes after the removal of false positives, which exhibited enhanced accuracy and reliability compared to previous iterations.

VII. NB Classifier

The Naïve Bayes classifier is derived from the Bayes theorem. The assumption is made that the predictors exhibit independence, implying that the knowledge of one characteristic does not influence the value of any other attribute. Naïve Bayes classifiers are characterized by their ease of construction, as they do not necessitate an iterative procedure. Furthermore, they exhibit notable efficiency when applied to extensive datasets, while maintaining a commendable degree of accuracy. Despite its straightforwardness, Naïve Bayes has been well recognized for its superior performance compared to other classification approaches across a range of issues.

In their study, Rusland et al. (61) investigate the topic of email spam filtering as well as employ the Naïve Bayes machine learning method to conduct their investigation. Two datasets were utilized and analyzed based on the metrics of accuracy, F-measure, precision, & recall. Naïve Bayes is a classification algorithm that uses probability theory to assign class labels to instances. Specifically, it calculates the likelihood by examining the frequency and mix of values present in a given dataset. This study employs a three-step approach for email filtration, namely preprocessing, feature selection, & implementation of features through the Naïve Bayes classifier. The initial stage of preprocessing is the elimination of conjunction words, articles, as well as stop words from the content of the email. Subsequently, the researchers employed the WEKA program [64] to generate two distinct datasets, namely the spam data as well as the spam base dataset. The mean accuracy achieved across the two datasets was 89.59%, with the spam dataset exhibiting a higher accuracy of 91.13%. The accuracy achieved by the spam-based dataset was 82.54%. The precision findings for the spam data set were found to be 83% on average, whereas for the spam base data set, the precision results were 88%. It has been asserted that the Naïve Bayes classifier exhibits superior performance when used to spam base data in comparison to spam data.

Sharma and Sahni (2011) published a scholarly study discussing the utilization of machine learning algorithms to detect spam in Internet of Things (IoT) devices [62]. The researchers employed a total of five machine-learning models and analyzed their outcomes utilizing a range of performance indicators. A substantial quantity of characteristics of the input were employed in the training of the proposed models. The spam score of each model is computed by considering the input attributes. The aforementioned score serves as an indicator of the reliability and credibility of an Internet of Things (IoT) device, taking into account a range of pertinent aspects. The proposed methodology is verified by employing the REFIT home automation dataset [63]. The authors assert that their suggested system exhibits superior spam detection capabilities compared to existing systems in use. The application of their work extends to smart homes as well as additional environments where intelligent gadgets are employed.

In their study, Kumar et al. (2020) examined the application of multiple machine-learning techniques for email spam identification [64]. The essay delves into the examination of machine learning methodologies and their practical application on various datasets. The identification of the most optimal method for email spam detection, which exhibits the best precision and accuracy, is achieved through the evaluation of multiple machine learning techniques. The researchers concluded that the use of the Multinomial Naïve Bayes algorithm yields the most favourable outcomes. However, it is important to acknowledge that this approach has certain drawbacks stemming from its reliance on class-conditional independence. Consequently, there are instances where the machine misclassifies certain inputs. In this study, it was observed that ensemble models yielded superior and dependable outcomes compared to Multinomial Naïve Bayes. The approach described in this study is limited to the detection of spam solely from the content within the body of email.

Singh and Batra (2019) introduced a semi-supervised machine learning approach for spam identification in social Internet of Things (IoT) platforms [65]. An ensemble-based framework including four classifiers was employed. The architectural design relies on the utilization of probabilistic data structures (i.e. PDS) that include a Quotient Filter (QF) for querying the database containing URLs, spam users, and databases of spam keywords. Additionally, Locality Sensitive Hashing (LSH) is employed for doing similarity searches. The suggested model employs the adaptive weighted voting strategy to minimize its decision-making process, taking into account the output of each classifier. The hybrid sampling technique reduces computational efforts by selectively collecting data based on each classifier. The findings of this study suggest that the methodology described in this research holds the potential for effectively detecting spam in extensive datasets. The efficacy of the suggested model was assessed by conducting a comparison between PDS and conventional data models, using commonly employed assessment criteria such as accuracy, recall, as well as F-score.

- ANNs

The artificial neural network (i.e. ANN) is a computer model that is derived from the functional characteristics of biological neural networks, commonly referred to as the neural network (NN) [66]. A neural network consists of many sets of interconnected neurons, wherein information is processed through computational connections. In the majority of scenarios, an artificial neural network (ANN) exhibits adaptability as a system, wherein its structure undergoes modifications based on the influx of either internal or external information throughout the learning phase. Contemporary neural networks represent non-linear methodologies for the analysis of statistical data. These are frequently employed in situations where there exist intricate connections between inputs and outcomes or atypical performance patterns. The diagram presented in Figure 10 illustrates the fundamental architecture within a neural network.

This section provides an elaboration on various proposed strategies for detecting and preventing email spam through the utilization of neural networks.

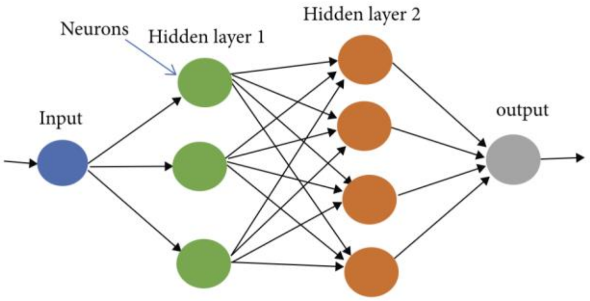

Figure 10: **Architecture of NN**

The approach proposed by Xu et al. [67] aims to detect spam within online social networks. The research conducted by the authors centres on the amalgamation of unsolicited messages across different social networking platforms. The researchers collected a total of 1937 tweets classified as spam and 10943 tweets classified as ham for further analysis, utilizing the Twitter platform. In addition, a total of 1338 spam posts as well as 9285 ham posts were utilized in the analysis. In the context of Twitter Spam Detection (TSD), it was observed that 75.6% of tweets analyzed had URL links, which were identified as spam tweets. On the other hand, 24.4% of the tweets consisted of distinct phrases, indicating a different type of content. Among a total of 10,942 tweets categorized as ham, it was observed that 62.9% of these tweets had both URL links as well as words, while the remaining 37.1% consisted only of words. According to the findings, it has been observed that approximately 32.8% of the spam posts generated by FSD are comprised of various web links, while the remaining 67.2% of these spam posts solely consist of textual content [68]. Out of a total of 9285 postings classified as ham, 95.1% of them contain web links, while the remaining 4.9% solely consist of textual content. The researchers employed the most frequently occurring twenty feature words extracted from datasets comprising Facebook spam and Twitter spam. The TSD and FSD are partitioned into two distinct sets, namely the training dataset and the testing dataset. The aforementioned datasets were employed in the training of diverse machine learning classifiers, including Naïve Bayes, logistic regression random tree, random forest, as well as Bayes Net. Upon analyzing the precision of various classifiers, the researchers integrated the spam dataset from Facebook with the learning dataset of Twitter, and likewise, incorporated the spam dataset from Twitter with the training datasets of Facebook. Subsequently, the researchers utilized the merged dataset to train and evaluate the performance of the classifiers. Ultimately, the researchers conduct a comparative analysis of the classifiers' outcomes on the aforementioned social networks, after an assessment of precision, accuracy, recall, as well as the F-1 measure. It was discovered that the precision of aggregated datasets surpassed that of alternative datasets [68, 69].

The spammer detection technique developed by Guo et al. [70] involves the utilization of a collaborating neural network in the context of Internet of Things (IoT) applications. The authors introduce an innovative spam detection mechanism named Cospam, specifically designed for Internet of Things (IoT) applications. Initially, the individual and the speech content at various time intervals are seen as sequences of features. The subsequent phase involves the utilization of a cooperative neural network model. The collaborative model comprises three distinct models, namely the Bi-AE model, the GCN model, and the LSTM model. These models are employed to determine the characteristics or attributes of the user. Ultimately, a sequence of tests was carried out to assess the efficacy of the proposed methodology. The model under consideration demonstrated a 5% increase in accuracy compared to currently employed methods for detecting spammers. The time required for Cospam is greater compared to existing techniques due to the presence of numerous parameters.

In the realm of the Internet of Things (IoT), Makkar & Kumar [71] introduced a deep learning framework aimed at identifying and mitigating web spam. The method in question improves the cognitive capabilities of search engines to identify instances of web spam effectively. The efficacy of this strategy lies in its ability to eliminate spam pages through the utilization of a website's rank score, which is derived from calculations performed by a search engine. The framework employed in their study leverages the comprehensive capabilities of deep learning. The first application of the LSTM model for spam detection has since been extended to several domains, including weather forecasting. This study involves a comparison between the suggested model and ten distinct machine learning models. This study utilizes the WEBSPAM-UK 2007 standardized dataset. The dataset undergoes preprocessing using a unique technique referred to as "Split by Oversampling as well as Train by Underfitting." The proposed model demonstrated a level of accuracy of 95.25%. Following the use of system optimization techniques, the suggested model achieved a high level of accuracy, specifically 96.96%.

In their publication, Zavvar et al. (72) discuss the topic of spam detection. They propose a methodology that involves the integration of particle swarm optimization techniques and neural networks for feature selection. In addition, support vector machines (SVM) were employed for spam classification and segregation. The researchers conducted a comparative analysis of the proposed methodology and alternative methodologies, namely a self-organizing map along with k-means data grouping, utilizing region under curve characteristics. This study employs the UCI base dataset to assess the effectiveness of spam categorization and proposes a spam detection methodology based on the Particle Swarm Optimization-Artificial Neural Network (PSO-ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) algorithms. The training dataset consisted of 70% of the data, while the remaining 30% was allocated to evaluate the models. The principles of Root Mean Square

Error (RMSE), Normalized Root Mean Square Error (NRMSE), and Standard Deviation (STD) were examined, yielding findings of 0.08733, 0.0185, and 0.08742, respectively, during the testing phase. The findings indicate that the proposed approach exhibits favourable levels of accuracy and performance in the detection of spam emails. Table 2 provides a summary of the supervised machine-learning algorithms that have been described for the purpose of spam identification.

This paper will examine several significant issues encountered by spam filters:

- The proliferation of data on the Internet, characterized by its diverse range of properties, presents a significant obstacle for spam detection systems.
- Evaluating the features of spam filters poses challenges in various dimensions, including temporal, style of writing, semantic, as well as statistical aspects.
- (iii) The majority of models are trained using datasets that are balanced in nature, whereas self-learning models aren't feasible.
- There exists a significant challenge in the realm of spam detection models, as they are susceptible to adversarial machine-learning approaches that can significantly undermine their efficacy. During the testing and training stages of machine learning models, adversaries can launch a diverse range of attacks. Adversarial actors possess the capability to manipulate training data to induce misclassification by a classifier, a technique known as a poisoning attack. Additionally, they can generate unfavourable samples during the testing phase to avoid detection, referred to as an evasion assault. Furthermore, these adversaries can acquire sensitive training data by exploiting a learning model, constituting a privacy attack.
- The user's text is already academic and does not need to be rewritten. The emergence of deep fake technology poses a significant problem for spam detection systems. Neural network models, such as GPT-2 and GPT-3, are utilized to generate, modify, and stylizing images and videos. Additionally, image generation models, namely BigGAN, StyleGAN, and CycleGAN, are also employed for this purpose. The utilization of deep fakes has the potential to propagate inaccurate information.

VIII. RESEARCH GAP AND PROBLEM

This section examines the areas of research that have not yet been addressed and the unresolved issues within the field of spam detection and filtration. In forthcoming research endeavours, it is advisable to employ real-life data for training experiments and models, as opposed to relying on manually generated datasets. This recommendation is based on the observation that models trained on fake datasets exhibit notably inadequate performance when applied to real-life data, as highlighted in multiple scholarly articles. Presently, the field of spam detection employs reinforcement learning, supervised learning, and unsupervised learning algorithms. However, the potential for enhanced accuracy and efficiency in spam detection can be realized through the utilization of hybrid algorithms in forthcoming research endeavours. In the future, the enhancement of the extraction of features can be achieved by the utilization of deep learning techniques for feature extraction. The utilization of clustering techniques in the context of spam filtering, specifically for relevance feedback with dynamic updating, has the potential to enhance the clustering of spam and ham messages. In addition to machine learning, the utilization of blockchain concepts and models holds potential for future applications in email spam detection.

In the future, there is potential for collaboration between linguistics and psycholinguistics experts in the manual annotation of datasets. This collaboration might lead to the creation of spam datasets that are both successful and adhere to standardized practices, characterized by high dimensionality. In the future, it is possible to enhance the performance of spam filters by leveraging Graphics Processing Units (GPUs) as well as Field Programmable Gate Arrays (FPGAs). These technologies provide advantages such as improved processing speed, higher classification accuracy, reduced energy consumption, enhanced flexibility, and the ability to analyze data in real-time. Furthermore, it is recommended that future research focuses on the provision of standardized labelled datasets that can be utilized by researchers for training classifiers. Additionally, enhancing the accuracy and dependability of spam detection algorithms can be achieved by incorporating supplementary features into the dataset, such as the IP address and geographical location of the spammer. The subsequent sections outline other avenues for further research and highlight unresolved issues within the field of spam identification.

Over the past two decades, there has been a significant focus from the scientific community on the subject of spam identification and filtration. The rationale behind extensive research in this domain stems from its significant and far-reaching implications, particularly about customer behaviour and the prevalence of counterfeit reviews. The survey encompasses a range of machine learning methods and models that have been presented by researchers to identify and mitigate spam in emails as well as IoT systems. The study classified the many types of learning approaches such as supervised, unsupervised, and reinforcement learning. This study does a comparative analysis of several methodologies and presents a comprehensive overview of the key insights gained from each category. The present study draws the conclusion that a majority of the suggested solutions for detecting spam in email and Internet of Things (IoT) systems rely on supervised machine learning approaches. The process of creating a labelled dataset to use in training a supervised model is essential and requires a significant amount of time. In the domain of spam detection, it has been observed that supervised learning algorithms, specifically Support Vector Machines (SVM) and Naïve Bayes, exhibit superior performance compared to alternative models. This paper offers a thorough examination of various algorithms utilized in the identification and filtering of email spam, along with an exploration of potential avenues for future research in this field.

IX. METHODOLOGY

The Multinomial Naive Bayes algorithm constitutes a probabilistic learning technique commonly employed in the field of Natural Language Processing (i.e. NLP). The algorithm utilized in this study is grounded on the principles of Bayes' theorem, enabling it to make predictions regarding the classification of various textual forms, including but not limited to emails and newspaper articles. The algorithm computes the likelihood of each tag given a given sample & afterwards outputs the tag containing the greatest likelihood.

The Naive Bayes technique is widely recognized for its efficacy in analyzing text input and addressing classification problems involving several classes. To comprehend the functioning of the Naive Bayes theorem, it is imperative first to grasp the concept of the Bayes theorem, as the former is built upon the latter.

Bayes' theorem, originally proposed by Thomas Bayes, is a mathematical formula that enables the calculation of the likelihood of an event's occurrence by incorporating prior knowledge regarding conditions associated with the event. The formula upon which it is based is as follows:

$$\boldsymbol{P(A|B) = P(A) * P(B|A)/P(B)} \tag{1}$$

In this, the computation of the likelihood of class A given the presence of predictor B. The symbol P(B) represents the prior probability of event B. The symbol P(A) represents the prior probability of the class A. The conditional probability P(B|A) represents the likelihood of the incidence of predictor B given class A. This model provides comparatively best results for spam detection and gives an accuracy of up to 98%.

Table 1: Comparative table of different methods of ML

Model	**SC% (Spams Caught)**	**BH% (Blocked Hams)**	**Accuracy (%)**
Multinomial NB	94.47	0.50	98
SVM	92.99	0.30	96.37
k-Nearest Neighbor	82.60	0.41	97
Random Forest	90.62	0.30	96
Adaboost+DT	92.17	0.50	97.50

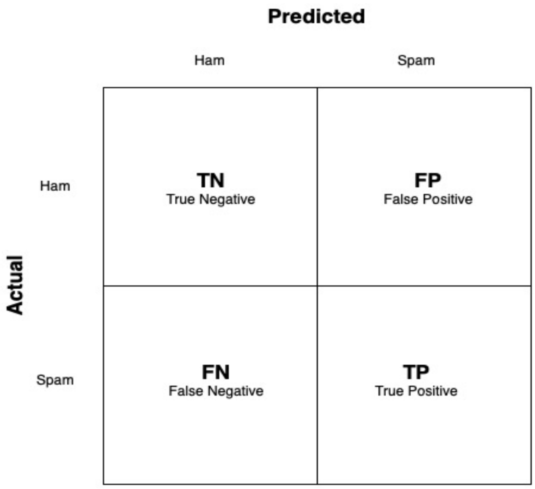

Figure 10: Actual and predicted sets of Ham and Spam

This table 1 shows the comparative table to find out different methods of ML in which Multinomial NB shows 98% accuracy, SVM shows 96.37% accuracy, K-NN gives an output of 97% of accuracy, the random forest gives 96% accuracy, and Adaboost+DT shows 97.50% result.

X. RESULTS AND DISCUSSION

In this model, the prediction designing architecture has been used to gather the best possible data information which has 5171 entries and four columns to show the integer (int64) type data, with two objects and it has used 161.78+ KB. This MultinomialNB model shows an accuracy of 0.9777458722182341, which shows the model precision value is high enough to process the facet distribution that shows a better result than the existing model for spam detection with this dataset. This model contains a pipeline which includes CountVectorizer and MultinomialNB

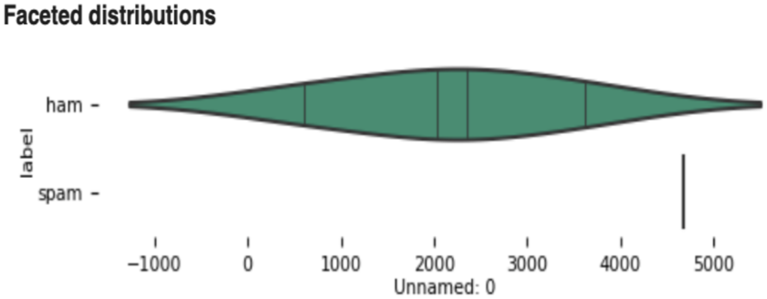

Figure 10: **Faceted Distribution of Label and Unnamed (Category) available in the dataset**

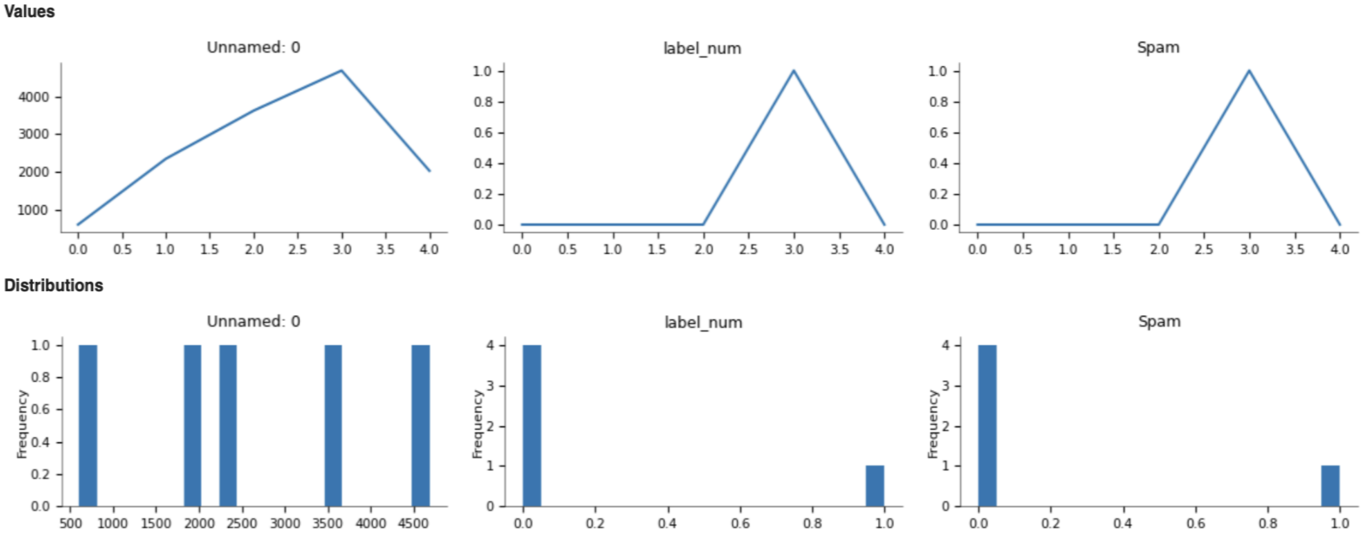

Figure 11: **Graphical representations of values and distributions of spam emails**

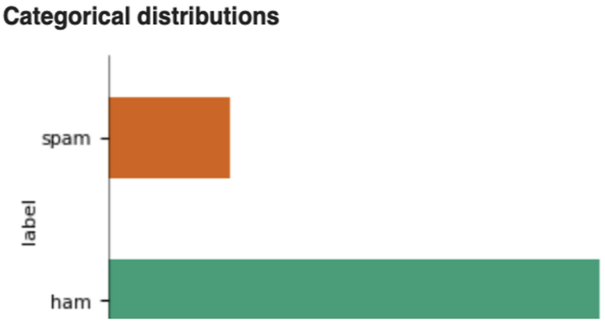

Figure 12: **Graph shows the different categorical distribution of dataset "ham" and "spam"**

XI. CONCLUSION

This research paper presents a novel approach for spam detection using natural language processing, utilizing a least-squares model to modify themes and incorporate gradient descent and AMALS models for estimating missing data. The technique exhibits a superior performance of 98% compared to existing industry TF-IDF models in accurate spam prediction within big data ecosystems. The paper also discusses various spam detection techniques, including Co-spam, a collaborative neural network model for IoT applications. The paper also discusses the challenges faced by spam filters, such as the proliferation of data, evaluating spam filters' features, training models using balanced datasets, and the vulnerability of models to adversarial machine learning approaches.

The emergence of deep fake technology poses a significant problem for spam detection systems, as it can propagate inaccurate information. Future research should focus on real-life data for training experiments and models, rather than manually generated datasets. Hybrid algorithms, deep learning techniques, clustering techniques, blockchain concepts, and collaboration between linguistics and psycholinguistics experts can enhance accuracy and efficiency in spam detection. Graphics Processing Units (GPUs) and Field Programmable Gate Arrays (FPGAs) can improve the performance of spam filters. Standardized labelled datasets and incorporating supplementary features like IP addresses and geographical locations can enhance the accuracy and dependability of spam detection algorithms. The Multinomial Naive Bayes algorithm, a probabilistic learning technique, is widely used in Natural Language Processing (NLP) for spam detection. It provides comparatively the best results for spam detection, with an accuracy of up to 98%. Future research should explore potential avenues for further research in this field. For future work, the security network will be required to improve the consistency of the result and maintain more accuracy than this model.

REFERENCES

[1] Reaves, B., Blue, L., Tian, D., Traynor, P., & Butler, K. R. (2016, July). Detecting SMS spam in the age of legitimate bulk messaging. In *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks* (pp. 165-170).

[2] Khan, W. Z., Khan, M. K., Muhaya, F. T. B., Aalsalem, M. Y., & Chao, H. C. (2015). A comprehensive study of email spam botnet detection. *IEEE Communications Surveys & Tutorials*, *17*(4), 2271-2295.

[3] Tuteja, S. K., & Bogiri, N. (2016, September). Email Spam filtering using BPNN classification algorithm. In *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)* (pp. 915-919). IEEE.

[4] Burnes, D., DeLiema, M., & Langton, L. (2020). Risk and protective factors of identity theft victimization in the United States. *Preventive medicine reports*, *17*, 101058.

[5] Cassim, F. (2015). Protecting personal information in the era of identity theft: Just how safe is our personal information from identity thieves?. *Potchefstroom Electronic Law Journal/Potchefstroomse Elektroniese Regsblad*, *18*(2), 68-110.

[6] Madakam, S., Lake, V., Lake, V., & Lake, V. (2015). Internet of Things (IoT): A literature review. *Journal of Computer and Communications*, *3*(05), 164.

[7] Ibrahim, D. S. (2018). Hybrid Approach to Detect Spam Emails using Preventive and Curing Techniques. *Journal of Al-Qadisiyah for computer science and mathematics*, *10*(3), 16-24.

[8] Salb, M., Jovanovic, L., Zivkovic, M., Tuba, E., Elsadai, A., & Bacanin, N. (2022). Training logistic regression model by enhanced moth flame optimizer for spam email classification. In *Computer Networks and Inventive Communication Technologies: Proceedings of Fifth ICCNCT 2022* (pp. 753-768). Singapore: Springer Nature Singapore.

[9] Roy, S. S., & Viswanatham, V. M. (2016). Classifying spam emails using artificial intelligent techniques. *International Journal of Engineering Research in Africa*, *22*, 152-161.

[10] Pathak, Y., Shukla, P. K., Tiwari, A., Stalin, S., & Singh, S. (2022). Deep transfer learning based classification model for COVID-19 disease. *Irbm*, *43*(2), 87-92.

[11] Karim, A., Azam, S., Shanmugam, B., Kannoorpatti, K., & Alazab, M. (2019). A comprehensive survey for intelligent spam email detection. *IEEE Access*, *7*, 168261-168295.

[12] Jun, I., Han, H. S., Edwards, J. R., & Jeon, H. (2018). Electrospun fibrous scaffolds for tissue engineering: Viewpoints on architecture and fabrication. *International journal of molecular sciences*, *19*(3), 745.

[13] Osborne, S. P. (2018). From public service-dominant logic to public service logic: are public service organizations capable of co-production and value co-creation?. *Public Management Review*, *20*(2), 225-231.

[14] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273-1282). PMLR.

[15] Dada, E. G., Bassi, J. S., Chiroma, H., Adetunmbi, A. O., & Ajibuwa, O. E. (2019). Machine learning for email spam filtering: review, approaches and open research problems. *Heliyon*, *5*(6).

[16] Das, G., Biswal, B. P., Kandambeth, S., Venkatesh, V., Kaur, G., Addicoat, M., ... & Banerjee, R. (2015). Chemical sensing in two dimensional porous covalent organic nanosheets. *Chemical science*, *6*(7), 3931-3939.

[17] Habib, M., Faris, M., Qaddoura, R., Alomari, A., & Faris, H. (2021). A predictive text system for medical recommendations in telemedicine: a deep learning approach in the Arabic context. *IEEE Access*, *9*, 85690-85708.

[18] Mallick, P. K., Mishra, S., & Chae, G. S. (2020). Digital media news categorization using Bernoulli document model for web content convergence. *Personal and Ubiquitous Computing*, 1-16.

[19] Saleh, S. A., Adly, H. M., Abdelkhaliq, A. A., & Nassir, A. M. (2020). Serum levels of selenium, zinc, copper, manganese, and iron in prostate cancer patients. *Current urology*, *14*(1), 44-49.

[20] Douzi, S., AlShahwan, F. A., Lemoudden, M., & El Ouahidi, B. (2020). Hybrid email spam detection model using artificial intelligence. *International Journal of Machine Learning and Computing*, *10*(2).

[21] Sun, G., Li, S., Chen, T., Li, X., & Zhu, S. (2017). Active learning method for chinese spam filtering. *International Journal of Performability Engineering*, *13*(4), 511.

[22] Bhuiyan, M. S. H., Miah, M. Y., Paul, S. C., Aka, T. D., Saha, O., Rahaman, M. M., ... & Ashaduzzaman, M. (2020). Green synthesis of iron oxide nanoparticle using Carica papaya leaf extract: application for photocatalytic degradation of remazol yellow RR dye and antibacterial activity. *Heliyon*, *6*(8).

[23] Torabi, Z. S., Nadimi-Shahraki, M. H., & Nabiollahi, A. (2015). Efficient support vector machines for spam detection: a survey. *International Journal of Computer Science and Information Security*, *13*(1), 11.

[24] Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, *50*, 102419.

[25] Vyas, S., Zaganjor, E., & Haigis, M. C. (2016). Mitochondria and cancer. *Cell*, *166*(3), 555-566.

[26] Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation through neural population dynamics. *Annual review of neuroscience*, *43*, 249-275.

[27] Leggott, J. (2020). 4. THE ROYAL PHILHARMONIC GOES TO THE BATHROOM: THE MUSIC OF MONTY PYTHON. *And Now for Something Completely Different: Critical Approaches to Monty Python*, 75.

[28] Cunningham, K., Headey, D., Singh, A., Karmacharya, C., & Rana, P. P. (2017). Maternal and child nutrition in Nepal: examining drivers of progress from the mid-1990s to 2010s. *Global food security*, *13*, 30-37.

[29] Siegel, J. J. (2021). *Stocks for the long run: The definitive guide to financial market returns & long-term investment strategies*. McGraw-Hill Education.

[30] Yadav, S., Saini, A., Dhamija, A., & Narnauli, Y. (2016). Discerning spam in social networking sites. *Adv Vis Comput Int J*, *3*(2), 2-9.

[31] Duman, S., Kalkan-Cakmakci, K., Egele, M., Robertson, W., & Kirda, E. (2016, June). Emailprofiler: Spearphishing filtering with header and stylometric features of emails. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 1, pp. 408-416). IEEE.

[32] Elhoseny, M., Ramírez-González, G., Abu-Elnasr, O. M., Shawkat, S. A., Arunkumar, N., & Farouk, A. (2018). Secure medical data transmission model for IoT-based healthcare systems. *Ieee Access*, *6*, 20596-20608.

[33] Park, S., Zhang, A. X., Murray, L. S., & Karger, D. R. (2019, May). Opportunities for automating email processing: A need-finding study. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

[34] Bhuiyan, H., Ashiquzzaman, A., Juthi, T. I., Biswas, S., & Ara, J. (2018). A survey of existing e-mail spam filtering methods considering machine learning techniques. *Global Journal of Computer Science and Technology*, *18*(2), 20-29.

[35] Sipahi, D., Dalkı lı ç, G., & Özcanhan, M. H. (2015). Detecting spam through their Sender Policy Framework records. *Security and Communication Networks*, *8*(18), 3555-3563.

[36] Bassiouni, M., Ali, M., & El-Dahshan, E. A. (2018). Ham and spam e-mails classification using machine learning techniques. *Journal of Applied Security Research*, *13*(3), 315-331.

[37] Ahsan, M. I., Nahian, T., Kafi, A. A., Hossain, M. I., & Shah, F. M. (2016, December). An ensemble approach to detect review spam using a hybrid machine learning technique. In *2016 19th International Conference on Computer and Information Technology (ICCIT)* (pp. 388-394). IEEE.

[38] Saidani, N., Adi, K., & Allili, M. S. (2020). A semantic-based classification approach for an enhanced spam detection. *Computers & Security*, *94*, 101716.

[39] Mohammad, R. M. A. (2020). A lifelong spam emails classification model. *Applied Computing and Informatics*, (ahead-of-print).

[40] Johnson, J., Hariharan, B., Van Der Maaten, L., Hoffman, J., Fei-Fei, L., Lawrence Zitnick, C., & Girshick, R. (2017). Inferring and executing programs for visual reasoning. In *Proceedings of the IEEE International Conference on computer vision* (pp. 2989-2998).

[41] Foster, I. D., Larson, J., Masich, M., Snoeren, A. C., Savage, S., & Levchenko, K. (2015, October). Security by any other name: On the effectiveness of provider-based email security. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security* (pp. 450-464).

[42] Larsson, D. J., Andremont, A., Bengtsson-Palme, J., Brandt, K. K., de Roda Husman, A. M., Fagerstedt, P., ... & Wernersson, A. S. (2018). Critical knowledge gaps and research needs related to the environmental dimensions of antibiotic resistance. *Environment International*, *117*, 132-138.

[43] Takhmiri, H., & Haroonabadi, A. (2016). Identifying valid email spam emails using decision tree. *International Journal of Computer Applications Technology and Research*, *5*(2), 61-65.

[44] Kille, S. E. (1986). *Mapping between X. 400 and RFC 822*(No. rfc987).

[45] Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46).

[46] Verma, M., & Sofat, S. (2014). Techniques to detect spammers in twitter-a survey. *International Journal of Computer Applications*, *85*(10).

[47] Fine, S., Singer, Y., & Tishby, N. (1998). The hierarchical hidden Markov model: Analysis and applications. *Machine learning*, *32*, 41-62.

[48] Keila, P. S., & Skillicorn, D. B. (2005). Structure in the Enron email dataset. *Computational & Mathematical Organization Theory*, *11*, 183-199.

[49] Li, W., Meng, W., Tan, Z., & Xiang, Y. (2019). Design of multi-view based email classification for IoT systems via semi-supervised learning. *Journal of Network and Computer Applications*, *128*, 56-63.

[50] Subasi, A., Kevric, J., & Abdullah Canbaz, M. (2019). Epileptic seizure detection using hybrid machine learning methods. *Neural Computing and Applications*, *31*, 317-325.

[51] Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification.

[52] Wang, Q., Guan, Y., & Wang, X. (2006). SVM-Based Spam Filter with Active and Online Learning. In *TREC*.

[53] Banday, M. T., & Jan, T. R. (2009). Effectiveness and limitations of statistical spam filters. *arXiv preprint arXiv:0910.2540*.

[54] Peng, W., Huang, L., Jia, J., & Ingram, E. (2018, August). Enhancing the naive bayes spam filter through intelligent text modification detection. In *2018 17th IEEE International Conference on trust, security and privacy in computing and communications/12th IEEE International Conference on big data science and engineering (TrustCom/BigDataSE)* (pp. 849-854). IEEE.

[55] Steinberg, D., & Colla, P. (2009). CART: classification and regression trees. *The top ten algorithms in data mining*, *9*, 179.

[56] Zeng, Z., Zheng, X., Chen, G., & Yu, Y. (2014, December). Spammer detection on Weibo social network. In *2014 IEEE 6th International Conference on Cloud Computing Technology and Science* (pp. 881-886). IEEE.

[57] Lin, C., He, J., Zhou, Y., Yang, X., Chen, K., & Song, L. (2013, August). Analysis and identification of spamming behaviors in sina weibo microblog. In *proceedings of the 7th workshop on social network mining and analysis* (pp. 1-9).

[58] Jamil, F., Kahng, H. K., Kim, S., & Kim, D. H. (2021). Towards secure fitness framework based on IoT-enabled blockchain network integrated with machine learning algorithms. *Sensors*, *21*(5), 1640.

[59] Olatunji, S. O. (2019). Improved email spam detection model based on support vector machines. *Neural Computing and Applications*, *31*, 691-699.

[60] Tretyakov, K. (2004, May). Machine learning techniques in spam filtering. In *Data Mining Problem-oriented Seminar, MTAT* (Vol. 3, No. 177, pp. 60-79). Citeseer.

[61] Rusland, N. F., Wahid, N., Kasim, S., & Hafit, H. (2017, August). Analysis of Naïve Bayes algorithm for email spam filtering across multiple datasets. In *IOP conference series: materials science and engineering* (Vol. 226, No. 1, p. 012091). IOP Publishing.

[62] Sharma, A. K., & Sahni, S. (2011). A comparative study of classification algorithms for spam email data analysis. *International Journal on Computer Science and Engineering*, *3*(5), 1890-1895.

[63] Kumar, N., & Sonowal, S. (2020, July). Email spam detection using machine learning algorithms. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 108-113). IEEE.

[64] Singh, A., & Batra, S. (2018). Ensemble based spam detection in social IoT using probabilistic data structures. *Future Generation Computer Systems*, *81*, 359-371.

[65] Sattu, N. (2020). *A study of machine learning algorithms on email spam classification* (Doctoral dissertation, Southeast Missouri State University).

[66] Xu, H., Sun, W., & Javaid, A. (2016, March). Efficient spam detection across online social networks. In *2016 IEEE International Conference on Big Data Analysis (ICBDA)* (pp. 1-6). IEEE.

[67] Faris, H., Aljarah, I., & Alqatawna, J. F. (2015, November). Optimizing feedforward neural networks using krill herd algorithm for e-mail spam detection. In *2015 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT)* (pp. 1-5). IEEE.

[68] Wang, A. H. (2010, June). Detecting spam bots in online social networking sites: a machine learning approach. In *IFIP Annual Conference on Data and Applications Security and Privacy*(pp. 335-342). Berlin, Heidelberg: Springer Berlin Heidelberg.

[69] Guo, Z., Shen, Y., Bashir, A. K., Imran, M., Kumar, N., Zhang, D., & Yu, K. (2020). Robust spammer detection using collaborative neural network in Internet-of-Things applications. *IEEE Internet of Things Journal*, *8*(12), 9549-9558.

[70] Makkar, A., & Kumar, N. (2020). An efficient deep learning-based scheme for web spam detection in IoT environment. *Future Generation Computer Systems*, *108*, 467-487.

[71] Zavvar, M., Rezaei, M., & Garavand, S. (2016). Email spam detection using combination of particle swarm optimization and artificial neural network and support vector machine. *International Journal of Modern Education and Computer Science*, *8*(7), 68.