SURYA GROUP OF INSTITUITIONS

NAME:SRIRAM V

Exam no:422221104702

College code :42222

1. Introduction

In the era of information technology, information sharing has become very easy and fast. Many platforms are available for users to share information anywhere across the world. Among all information sharing mediums, email is the simplest, cheapest, and the most rapid method of information sharing worldwide. But, due to their simplicity, emails are vulnerable to different kinds of attacks, and the most common and dangerous one is spam [1]. No one wants to receive emails not related to their interest because they waste receivers' time and resources. Besides, these emails can have malicious content hidden in the form of attachments or URLs that may lead to the host system's security breaches [2]. Spam is any irrelevant and unwanted message or email sent by the attacker to a significant number of recipients by using emails or any other medium of information sharing [3]. So, it requires an immense demand for the security of the email system. Spam emails may carry viruses, rats, and Trojans. Attackers mostly use this technique for luring users towards online services. They may send spam emails that contain attachments with the multiple-file extension, packed URLs that lead the user to malicious and spamming websites and end up with some sort of data or financial fraud and identify theft [4, 5]. Many email providers allow their users to make keywords base rules that automatically filter emails. Still, this approach is not very useful because it is difficult, and users do not want to customize their emails, due to which spammers attack their email accounts.

In the last few decades, Internet of things (IoT) has become a part of modern life and is growing rapidly. IoT has become an essential component of smart cities. There are a lot of IoT-based social media platforms and applications. Due to the emergence of IoT, spamming problems are

increasing at a high rate. The researchers proposed various spam detection methods to detect and filter spam and spammers. Mainly, the existing spam detection methods are divided into two types: behaviour pattern-based approaches and semantic pattern-based approaches. These approaches have their limitations and drawbacks. There has been significant growth in spam emails, along with the rise of the Internet and communication around the globe [6]. Spams are generated from any location of the world with the Internet's help by hiding the attacker's identity. There are a plenty of antispam tools and techniques, but the spam rate is still very high. The most dangerous spams are malicious emails containing links to malicious websites that can harm the victim's data. Spam emails can also slow down the server response by filling up the memory or capacity of servers. To accurately detect spam emails and avoid the rising email spam issues, every organization carefully evaluates the available tools to tackle spam in their environment. Some famous mechanisms to identify and analyze the incoming emails for spam detection are Whitelist/Blacklist [7], mail header analysis, keyword checking, etc.

Social networking experts estimate that 40% of social network accounts are used for spam [8]. The spammers use popular social networking tools to target specific segments, review pages, or fan pages to send hidden links in the text to pornographic or other product sites designed to sell something from fraudulent accounts. The noxious emails that are sent to the same kind of individuals or associations share regular highlights. By investigating these highlights, one can improve the detection of these types of emails. By utilizing artificial ntelligence (AI) [9], we can classify emails into spam and nonspam emails. This solution is possible by using feature extraction from the messages' headers, subject, and body. After extracting this data based on their nature, we can group them into spam or ham. Today, learning-based classifiers [10] are commonly used for spam detection. In learning-based classification, the detection process assumes that spam emails have a specific set of features that differentiate them from legitimate emails [11]. Many factors increase the complexity of the identification process of spam in learning-based models. These factors include spam subjectivity, idea drift, language problems, overhead processing, and text latency.

One example of learning-based models is extreme learning machine (ELM). This is a modern machine learning model for the feedforward neural networks containing only one hidden layer [12]. It eliminates slow training speed and overfitting problems when compared with traditional neural networks. In ELM, it requires only one cycle of iteration. Because of better generalization potential, robustness, and controllability, this algorithm specifically is now used in many fields. In this paper, we consider different machine learning algorithms for spam detection. Our contributions are delineated as follows:

(i) The study discusses various machine learning-based spam filters, their architecture, along with their pros and cons. We also discussed the basic features of spam email.

(ii) Some exciting research gaps were found in the spam detection and filtering domain by conducting a comprehensive survey of the proposed techniques and spam's nature.

(iii) Open research problems and future research directions are discussed to enhance email security and filtration of spam emails by using machine learning methods.

(iv) Several challenges currently faced by spam filtering models and the effects of those challenges on the models' efficiency are discussed in this study.

(v) A comprehensive comparison of machine learning techniques and concepts that help understand machine learning's role in spam detection is provided.

(vi) The study categorizes different spam detection methods according to machine learning techniques to better understand concepts jointly.

(vii) Various future spam detection and filtration directions are discussed that could be explored to detect spam better and add more security to email platforms.

The rest of the paper is organized into nine sections. Section 2 discusses the comparison of previous surveys that were done on email spam detection. Section 3 discusses the basics of email spam and its effects on the community. Section 4 focuses on basic methods used for spam filtration. Section 5 elaborates on the machine learning background, while Section 6 provides an overview of machine learning algorithms used for spam filtration. This section also reviews various papers and proposed machine learning techniques for spam filtration and detection. Section 7 presents the open issues and research gaps, while Section 8 discusses challenges of spam detection systems. At the end, Section 9 concludes and presents the future directions of email spam detection and filtration. Table 1 presents the list of acronyms used in this article with corresponding definitions.

2. Comparison with Previous Surveys

Email spam is nothing more than fake or unwanted bulk mails sent via any account or an automated system. Spam emails are increasing day by day, and it has become a common problem over the last decade. Email IDs receiving spam emails are typically collected through spambots (a computerized application that crawls email addresses across the Internet). The applications of machine learning have been playing a vital role in the detection of spam emails. It has various models and techniques that researchers are using to develop novel spam detection and filtering models [13]. Kaur and Verma [14] present a survey on email spam detection using a supervised approach with feature selection. They discuss the knowledge discovery process for spam detection systems. They also elaborate various techniques and tools proposed for spam detection. The choice of features based on N-Gram is also addressed in this survey. N-Gram [15, 16] is a predictive-based algorithm used to

Table 1: A-list of acronyms used in this article with corresponding definitions.

Acronym	Description
KNN	K-nearest neighbors
NN	Neural networks
SVM	Support vector machine
MLP	Multilayer perceptron neural network
ECML	European conference of machine learning
AI	Artificial intelligence
CART	Classification and regression tree
TF/IDF	Term frequency/inverse document frequency
PSO	Particle swarm optimization
DTM	Document term frequency
BOG	Bag of words
ML	Machine learning
NB	Na¨ıve Bayes
NB tree	Na¨ıve Bayes tree
LAD tree	Logistic analysis of data tree
REP tree	Reduced error pruning tree
UCI	University of California Irvine repository of machine learning databases
XML	Extensible markable language
ID3	Iterative dichotomizer 3
SOM	Self-organizing maps
DBSCAN	Density-based spatial clustering of applications with noise
ELM	Extreme learning machines
AD tree	Alternating decision tree

predict the probability of the next word occurrence after finding $N-1$ terms in a sentence or text corpus. N-Gram uses probability-based techniques for the next word prediction. They compare various machine learning (multilayer perceptron neural network support vector machine, Na¨ıve Bayes) and nonmachine learning (Signatures, Blacklist and Whitelist, and mail header checking) approaches for email spam detection.

Saleh et al. [17] present a survey on intelligent spam email detection. They discuss various security risks of emails, especially spam emails, the scope of spam analysis, and different machine learning and nonmachine learning techniques for spam detection and filtering. They conclude that there is high adoption of supervised learning [18] algorithms for email spam detection. They state that the high usage of supervised learning is the accuracy and consistency of supervised techniques. They also discussed multialgorithm frameworks and found that multialgorithm frameworks are more efficient than a single algorithm. They found that nearly all research work that uses the content of emails for the identification spam, particularly phishing emails, depends on word-based classification or clustering systems.

Blanzieri and Bryl [2, 19] describe a list of learning-based email spam filtering approaches. In this paper, they addressed the spam problems and provided a review of learning-based spam filtering. They explain various features of spam emails. In this study, effects of spam emails on different domains were discussed. Various economic and ethical issues of spam are also discussed in this study. The antispam approach that is common and learning-based filtering is well developed. The commonly used filters are based on different classification techniques applied to various components of email messages. This study suggests that the Na¨ıve Bayes classifier holds a particular position amongst multiple learning algorithms used for spam filtering. With splendid pace and simplicity, it gives high precision results.

Bhuiyan et al. [20] present a review of current email spam filtering approaches. They summarize multiple spam filtering approaches and sum up the accuracy on various parameters of different proposed systems by analyzing numerous processes. They discuss that all the existing methods are efficient for filtering spam emails. Some have successful results, and others are attempting to incorporate other ways to boost their accuracy performance. Although they are all successful, they still have some issues in spam filtering methods, which is the primary concern for researchers. They are trying to create a next-generation spam filtering mechanism to understand large numbers of multimedia data and filter spam emails. They conclude that most email spam filtering is done by utilizing Na¨ıve Bayes and the SVM algorithm. To test the spam filtration models, these models can be trained on different datasets, such as "ECML" and UCI dataset [21].

Ferrag et al. [13] presented a review of deep learning algorithms of intrusion detection systems and spam detection datasets. They discussed various detection systems based on deep learning models and evaluated the effectiveness of those models. They examined 35 well-known cyber dataset by dividing them into seven categories. These categories include Internet traffic-based, network traffic-based, Interanet traffic-based, electrical network-based, virtual private network-based, andriod apps-based, IoT

traffic-based, and Internet connected device-based datasets. They conclude that deep learning models can perform better than traditional machine learning and lexicon models for intrusion and spam detection.

Vyas et al. [22] present a review on supervised machine learning strategies for filtering spam emails. They concluded that the Na¨ıve Bayes method provides faster results and decent precision over all other methods (except SVM and ID3) from all the techniques discussed. SVM and ID3 offer greater precision than Na¨ıve Bayes but take much longer time to construct a system. There is a trade-off between timing and precision. They conclude that selecting the learning algorithm heavily depends on the situation and the required accuracy and time. They state that all parts of the email should be considered in the future to create a more robust spam filtering framework.

This survey paper discusses three main types of machine learning that can be used for spam filtering. We review various papers, the proposed techniques, and discuss challenges to spam detection and filtration systems. This article also focuses on the advantages and disadvantages of the proposed techniques for spam detection and filtration that is never reviewed in the past.

3. Spam Messages

The email spam definition is ambiguous since everybody has their views on it. At present, email spam is getting the attention of everyone. Email spam ordinarily includes particular spontaneous messages sent in mass by individuals you do not know. The term spam is obtained from the Monty Python sketch [23], in which the Hormel canned meat item has numerous tedious emphases. While the term spam was purportedly first utilized in 1978 to allude to unwanted email, it increased rapidly in the mid-1990s, as we get to turn out to be progressively typical outside scholastic and research circles [24]. A notable model is the development expense trick in which a client receives an email with an offer that should bring about a prize. In the era of technology, the dodger/spammer shows a story where the unfortunate casualty needs forthright financial help so that the fraudster can gain a lot bigger total of cash, which they would then share. The fraudster will either earn a profit or avoid communication when the unfortunate victim completes the installment.

Spam Filtering Methods in Email and IoT Platforms. The number of spam emails is rapidly increasing in mar- keting, chain communications, stock market tips, politics, and education [24]. Currently, various companies develop different techniques and algorithms for efficient spam de- tection and filtering. We address some filtering strategies in this section to understand the filtering process.

The Standard Spam Filtering Method. Standard spam filtering is a filtering system that implements a set of rules and works with that set of protocols as a classifier. Figure 1 illustrates a standard method for filtering spam. In the first

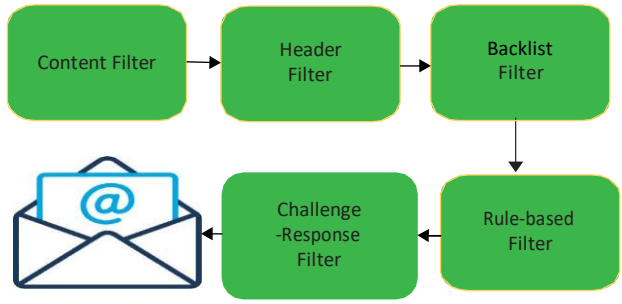

Figure 1: Standard spam filtering.

step, content filters are implemented and use artificial intelligence techniques to figure out the spam [25]. The email header filter, which extracts the header information from the email, is implemented in the second step. After that, backlist filters are applied to the emails to clinch the emails coming from the backlist file to avoid spam emails. After this stage, rule-based filters are implemented, recognizing the sender using the subject line and user-defined parameters. Eventually, allowance and task filters are used by implementing a method that allows the account holder to send the mail [26].

The Client Side Spam Filtering. A client is a person who can use the Internet or email network to send or receivean email [27]. Spam detection at the client point offers different rules and mechanisms to ensure secure commu- nications transmission between people and organizations. For transmission of data, a client should deploy multiple existing frameworks on his/her system. Such systems con- nect with client mail agents and filter the client's mailbox by compositing, accepting, and managing the incoming emails [28, 29].

Enterprise Level Spam Filtering. Email spam detection at the enterprise level is a technique in which various filtering frameworks are installed on the server, dealing with the mail transfer agent and classifying the collected emails into one spam or ham [30]. This system client uses the system consistently and effectively on a network with an enterprise filtering technique to filter the emails. Existing methods of spam detection use the rule of ranking the email. A ranking function is specified in this principle, and a scoreis generated against every post. The junk mail or ham message is given specific scores or ranks [31]. Since spammers use different approaches, all tasks are regularly modified by implementing a list-based technique to block the messages automatically. Figure 2 is reproduced from Bhuiyan et al. [20]. Figure 2 shows the architecture of the client and enterprise level spam filtering process.

Case-Based Spam Filtering. One of the well-known and conventional machine learning methods for spam de- tection is the case-based or sample-based spam filtering system [32]. A typical case base filtering structure is illus- trated in Figure 3. There are many phases to this type of

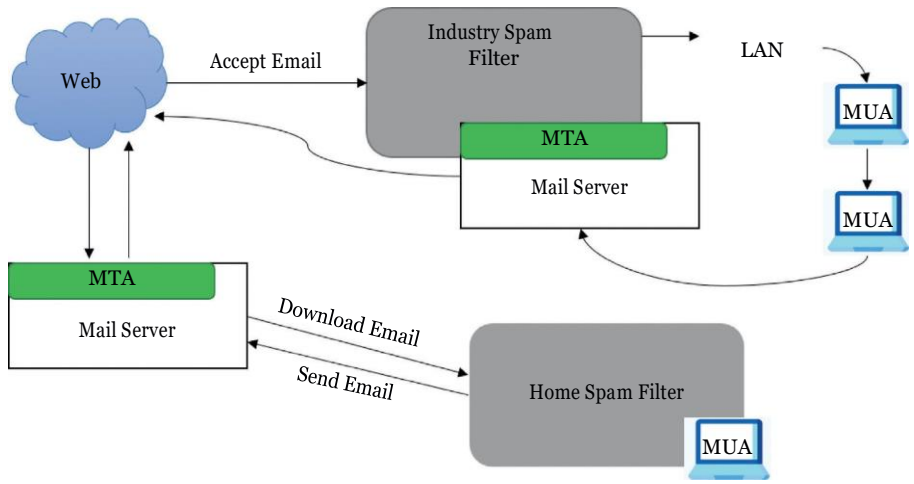

Figure 2: Client based and enterprise level spam filtering [20].

filtering with the aid of the collection method; it collects data (mails) during the first step. After that, the major transition continues with the preprocessing steps through the client graphical user interface, outlining abstraction, and choice of email data classification, testing the entire process using vector expression and classifying the data into two classes: spam and legitimate email.

Finally, the machine learning technique is extended to training sets and test sets to determine whether this is an email. The final decision is made through two steps: self-observation and classifier's result, deciding whether the email is spam or legitimate [32, 33].

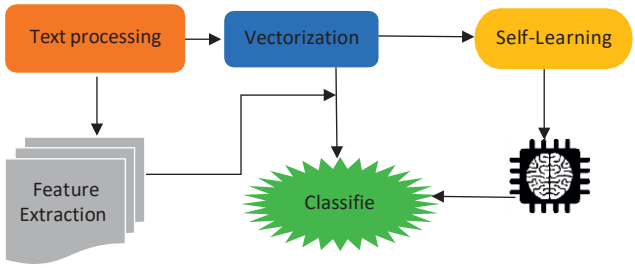

Figure 3: Case-based spam filtering.

4. Internet of Things and Its Attacks (IoT)

The Internet of things (IoT) means a system of interrelated, Internet-connected objects that collect and transfer data over a wireless network without the intervention of humans. IoT enables the integration and implementation of real-world objects regardless of location. In such a scenario, privacy and security techniques are highly critical and challenging in network management and monitoring performance. To solve security problems, such as intrusions, phishing attacks, DoS attacks, spamming, and malware in IoTapplications must protect privacy. Ios systems, including objects and networks, are vulnerable to network and physical attacks and privacy failures. The main types of IoTattacks are illustrated in Figure 4.

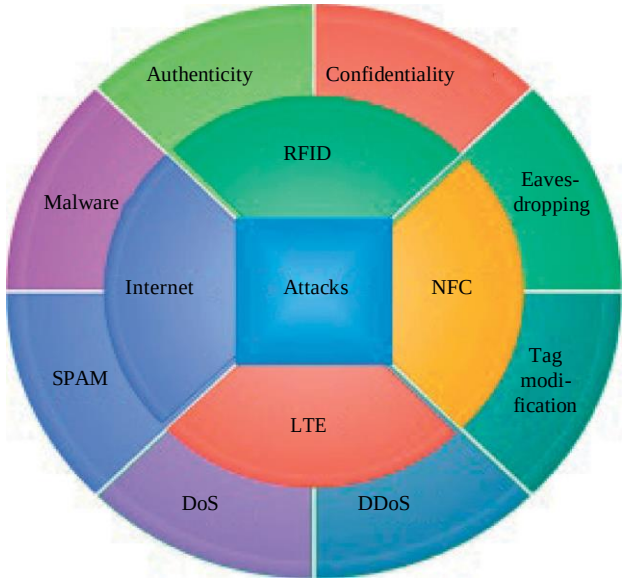

Figure 4: IoT attacks.

The various attacks of IoT systems are listed as follows.

(a) *Self-Promotion Attack*. In this type of attack, the compromised node tries to get importance over the other nodes of the IoT environment for the particular recommendation.

(b) *Bad Mouthing Attack*. In this attack, the compromised node forgave a wrong recommendation; it may execute the trust of the trusted node. It decreased the services of the trusted node.

(c) *Ballot Stuffing Attack*. In this challenge of the IoT environment, the compromised node enhances the other compromised nodes. It is a chance for the compromised node to provide the services. It is also known as the collision recommendation attack.

(d) *Opportunistic Service Attack*. In this type of attack, the compromised node collaborates with the other malicious nodes to build the bad mouthing and ballot stuffing attack.

(e) *On-Off Attack*. In this type of attack, the compromised node provides inadequate services, which means that the compromised node randomly performs a bad service.

(f)) *Node Tempering*. The attacker changes the malicious node and gets specific information such as a security key.

(g) *Malicious Node Attack*. The attacker physically adds the malicious node among nodes.

(h) *Man in the Middle Attack*. The attacker secretly intercepts the communication between two nodes over the Internet in this type of attack. The attacker gets the main information by eavesdropping.

(i) *Sybil Attack*. The compromised node steals the recognition of good nodes and acts as a suitable node.

According to a study from Nozomi Networks, in the first half of 2020, there were increasing attacks and threats on Operational Technology (OT) and the IoT networks. Figure 5 shows the number of attacks in IoT devices in respective years.

Machine learning techniques can be used for the prevention and detection of these attacks with high performance. Various research studies have been carried out to detect and prevent the above issues discussed in Section 5.

5. Machine Learning

Machine learning [34] is one of the most important and valuable applications of artificial intelligence (AI), which gives computer systems the ability of automatically learning and enhancing their functionality without explicit programming [34]. The primary purpose of machine learning algorithms is to build automated tools to access and use the data for training. The learning process starts with learning labeled data, also called training dataset. It can be a real-life experience, review, example, or feedback to recognize trends in the data to make better future decisions based on the user's input. The main objective of machine learning models is to learn automatically without any intervention from humans. Machine learning consists of three major kinds, used for numerous tasks.

For the last decade, researchers have been trying to make email communication better than today. Spam filtering of emails [35] is one of the most critical ways of protecting email networks. Many research articles have been published using various machine learning approaches to identify and process spam emails, but there are still some research gaps. Junk mail is one of the central, attractive research fields for filling the gaps [36]. For this reason, many spam classification studies have already been carried out using several methods to make email communication more trustworthy and valuable for users. That is why, this paper is presented to make a summarized version of different existing machine learning models and approaches that are being used for email spam detection. This paper also evaluates the most common machine learning approaches like KNN, SVM, random forest, and Na¨ıve Bayes.

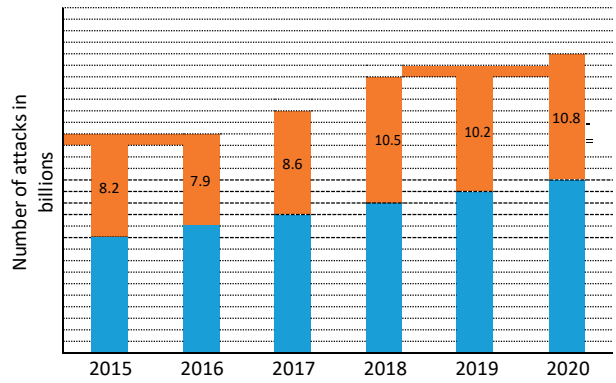

Figure 5: Number of attacks on IoT devices.

Machine Learning-Based Spam Filtering Methods. Machine learning facilitates the processing of vast quantities of data. Though it typically provides faster and more ac- curate results to detect unwanted content, it can also require extra time and resources to train its models for a high level of performance. Integrating machine learning with AI and cognitive computing [37] can make handling massive amounts of data even more powerful. Figure 6 demonstrates various kinds of machine learning.

Supervised Machine Learning. Supervised machine learning algorithms [18] are machine learning models that need labeled data. Initially, labeled training data is provided to these models for training, and after training models predict future events. In other words, these models begin with the analysis of an existing training dataset, and they generate a method to make predictions of success values. Upon proper training, the system can provide [38] the prediction on any new data related to the user's data at the training time. Furthermore, the learning algorithm accu- rately compares the output to the expected output and identifies errors to modify the model.

Supervised learning uses labeled data for training, and then it can predict the new data. This type of learning can be used in solving various problems, i.e., advertisement popularity, spam classification, face recognition, and object classification. The process of supervised learning is illustrated in Figure 7.

Some most commonly used supervised learning techniques are discussed as follows.

Decision Tree Classifier. Decision tree classifier is a machine learning algorithm [39], which has been widely used since the last decade for classification. This algorithm applies a simple method of solving any problem of classi- fication. A decision tree classifier is a collection of well- defined questions about test record attributes. Each time we get an answer, a follow up question is raised until a decision is not made on the record [40]. Tree-based decision algo- rithms define models that are constructed iteratively or recurrently based on the data provided. The decision tree- based algorithms goal is used to predict a target variable's

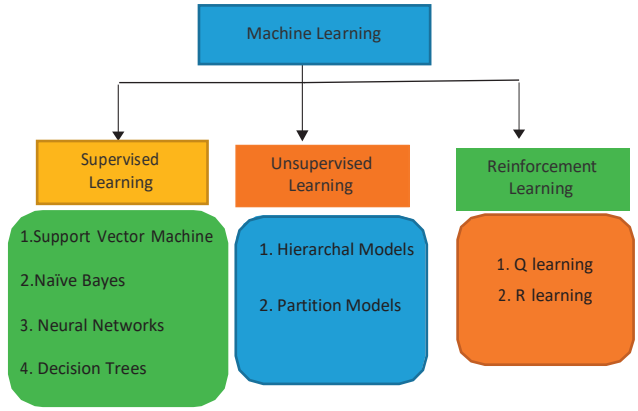

Figure 6: Types of machine learning.

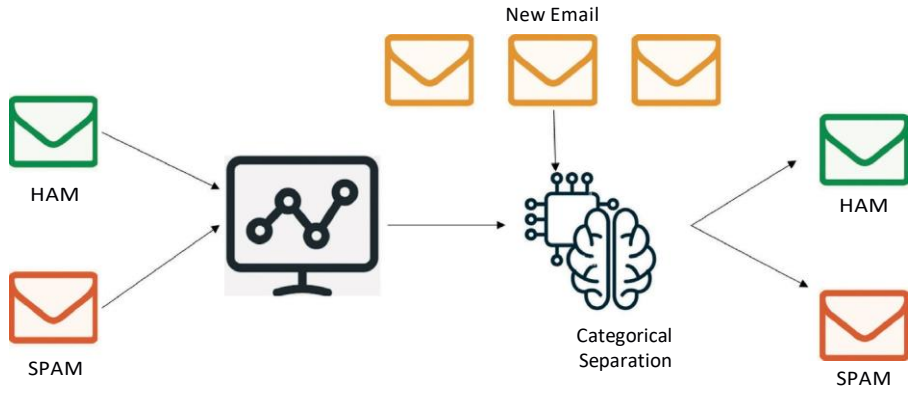

Figure 7: Process of supervised learning.

value on a given set of input values. This algorithm uses a tree structure to solve classification and regression problems [41]. Figure 8 shows the basic structure of the decision tree. Some of the decision tree algorithms are the following:

(i) Random forest
(ii) Classification and regression tree (CART)
(iii) C4.5 and C5.0
(iv) Chi-square.

The following section deliberates some proposed email spam detection and prevention techniques by using decision tree algorithms.

DeBarr and Wechsler [42] discuss a spam filtering technique using random forest algorithms to classify spam emails and active learning to refine the classification [43]. They used the data of email messages from RFC 822 (Internet) [44] and divided each email into two sections. Then, they find term frequency and inverse document frequency of all features of each email (TF/IDF). For the training dataset, they select a set of emails with clustering to label the data. After considering the cluster prototype mails for training, they experiment with supervised machine learning algorithms: random forest, Na¨ıve Bayes, support vector machine, and KNN [45]. The research results show that the algorithm "random forest" classifies data more efficiently with an accuracy of 95.2%.

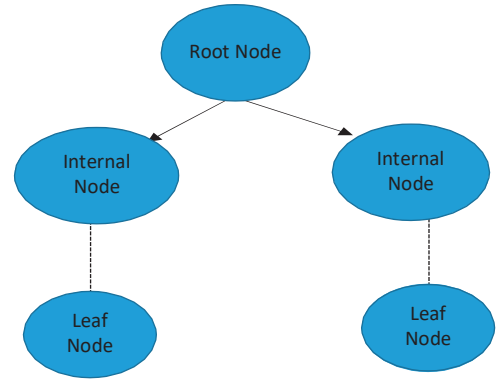

Figure 8: Structure of decision tree.

Takhmiri and Haroonabadi [46] present a different technique to detect spams using a fuzzy decision tree and the Na¨ıve Bayes algorithm. They use the baking voting algorithm to extract patterns of spam behaviour. They do this because obvious characteristics do not exist in the real world. The cross-linking degree for explaining or describing characters is rational and neutral. Decision trees use fuzzy Mamdani rules for the classification of spam and ham email. Then, Na¨ıve Bayes classifier [47] is used by them on the dataset. Finally, the baking method is used by dividing votes into smaller sections. This solution gives them an optimized weight that can be implemented on obtained percentages

that achieve a higher accuracy level. The dataset used in this study contains 1000 emails, from which 350 (35%) were spam and 650 (65%) were ham.

Verma and Sofat [48] used supervised machine learning algorithm ID3 [49] to render the decision trees of the problem and the hidden Markov model [50] to measure the probabilities of events that could occur as a combination to classify the emails as junk mail or ham. The proposed model initially marks all emails as spam or legitimate by measuring each e-mail's total likelihood with the aid of subsequently classified email terms. After that, it makes the decision trees of emails one by one. The Enron dataset [51] is used in this study that contains 5172 emails. From all 5172 emails, 2086 were spam, while 2086 were legitimate emails. Their model can categorize the emails as spam and ham by using the feature set obtained by the Enron dataset. They got an 11% error by using the sklearn library's fitness function in the proposed model. Their model got 89% of accuracy results on the given dataset.

Li et al. [52] proposed an email-classification technique for IoT systems based on supervised machine learning. They use a multiview technique that focuses on the collection of richer information for classification. A double view dataset is created with internal and external feature sets. The proposed approach can be used in both labeled and unlabeled data and was evaluated on two datasets with a real network environment. The results of this study indicate that the multiview model can achieve more accuracy than simple email classification. In the end, the multiview model is compared with various existing models.

A spam filtering approach based on different decision tree algorithms is presented by Subasi et al. [40] to compare the accuracy and find the best one for their dataset. They implement classification and regression tree (CART), C4.5, REP tree, LAD tree, NBT, random forest, and rotation forest algorithm on the dataset to classify emails. Their results show that the proposed modified random forest model got the highest accuracy than other decision tree methods for publicly available datasets.

Support Vector Machine (SVM). The support vector machine (SVM) is an essential and valuable machine learning model [53]. SVM is a formally defined discrimi- native supervised learning classifier that takes labeled ex- amples for training and gives a hyperplane as output, classifying new data [54]. A set of objects belonging to various class memberships are separated by decision planes. Figure 9 shows the classification concept of linear support vector machines. In the figure, some circles and stars are called objects. These objects can belong to any of two classes, i.e., the class of stars or dots. The isolated lines determine the choice of objects between green and brown objects. On the lower side of the plane, the objects are brown stars, and on the upper side of the plane all objects are green dots showingthat two unique objects are classified into two differentclasses. If a new object black circle is given to the model, itwill classify that circle into one of the classes according to the training examples provided in the training phase.

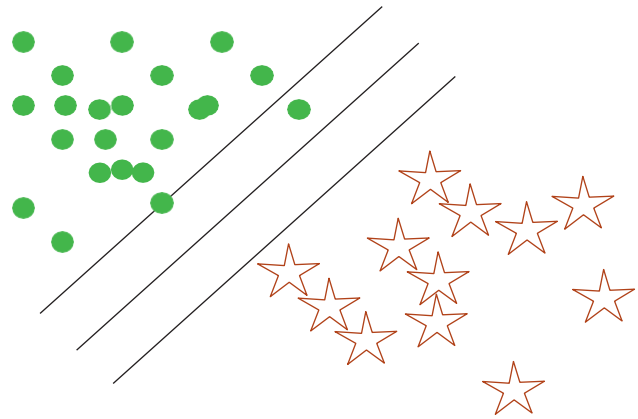

Figure 9: Support vector machine classification.

Banday and Jan [55] present research in which they define the procedure of statistical spam filters. They design those filters using Na¨ıve Bayes, KNN, support vector machines (SVM), and regression trees [56]. They use all these supervised machine learning algorithms and evaluate the results based on precision, recall, and accuracy. Using these machine learning techniques, they found that classification and regression trees (CART) [57] and Na¨ıve Bayes classifiers are the most effective algorithms for the dataset. This approach estimates that, during spam filtering, calculations of false positive are costlier than a false negative.

Zheng et al. [12, 58] present a procedure for detecting spammers and spam messages in any social network. Today, everyone uses social media, and many social media users spend a considerable amount of time communicating with their loved ones. The spammers take advantage of various social media networks and users' posts to send malicious content, advertisements, information, etc., into the social media user's profiles. So, this paper discusses how to detect those posts or malicious content on social media platforms. Their study uses the Sina Weibo social network [59] and machine learning algorithm support vector machine (SVM) for the detection of spammers. The dataset that was used in this study was 16 million messages that were collected from several users. They used 18 features as a feature vector set. The clients of the networks were divided into two categories, legitimate users and spammers. 80% of data was used for the model's training, while 20% was used for testing. For better accuracy, they used 1 : 2 between spammers and non-spammers of the training dataset. With this ratio, the proposed model gives an accuracy level of 99.5% for classifying spammers and nonspammers [60].

A novel fitness framework based on IoT-enabled blockchain technology and machine learning techniques is presented by Jamil et al. [10]. Their proposed model is composed of two modules. The first one is a blockchain-based network used for the security of sensing devices and an intelligent contract-enabled relationship and an inference engine that uncovers hidden insights and usable information from IoT and user device data. The improved smart contract gives users a useful application that allows real-time monitoring, more control, and quick access to several devices distributed across various domains. The inference engine

module attempts to uncover underlying patterns and usable information from IoT environment data, assisting in effective decision-making and providing convenient services. Their proposed model can be used to improve system throughput and resource usage, according to their findings. The proposed system in this article may be used in various fields, including healthcare and smart businesses.

Olatunji [61] developed a spam filtering tool using support vector machine and extreme learning machine algorithms. He used the standard dataset for the development of the spam detection model. SVM got an accuracy of 94.06% in his work, and the extreme learning machine (ELM) model got a 93.04% accuracy level, suggesting just 1.1% performance improvement that SVM achieved over ELM. He indicated that SVM's improvement over ELM accuracy is marginal. It implies that, in situations where detection time is critical, as in real-time systems, the ELM spam detector should be given preference over SVM spam detection. Although SVM got a higher accuracy level in his research, it takes more time for training than the ELM system. Tretyakov [62] also discussed various machine learning techniques for email spam filtering. This paper compared the precision results between false positives and precision results after eliminating false positives. They show the result after eliminating false positives, which were more accurate and reliable than before.

Na¨ıve Bayes Classifier (NB). The Na¨ıve Bayes classifier [47] is based on the Bayes theorem. It assumes that the predictors are independent, which means that knowing the value of one attribute impacts any other attribute's value. Na¨ıve Bayes classifiers are easy to build because they do not require any iterative process and they perform very efficiently on large datasets with a handsome level of accuracy. Despite its simplicity, Na¨ıve Bayes is known to have often outperformed other classification methods in various problems.

Rusland et al. [63] present research on email spam filtering and perform the analysis using a machine learning algorithm Na¨ıve Bayes. They used two datasets evaluated on the value of accuracy, F-measure, precision, and recall. As we know, Na¨ıve Bayes uses probability for classification, and the probability is counting the frequency and combination of values in a dataset. This research uses three steps for the filtration of emails, i.e., preprocessing, feature selection, and, at last, it implements the features by using the Na¨ıve Bayes classifier. The preprocessing step removes all conjunction words, articles, and stop words from the email body. Then, they used the WEKA tool [64] and made two datasets called spam data and spam base dataset. The average accuracy was 89.59% using two datasets, while the spam data got 91.13% accuracy. The spam base dataset got an accuracy of 82.54%. The average precision results for spam data were 83%, while, for spam base, the precision result was 88%. They claimed that the Na¨ıve Bayes classifier performs better on spam base data as compared with spam data.

Arif et al. [11] presented an article on machine learning-based spam detection techniques for IoT devices. They used five ML models and analyzed their results using various performance metrics. A large number of input features were used for the training of proposed models. Each model calculates a spam score based on the input attributes. This score represents the trustworthiness of an IoT device based on a variety of factors. The suggested approach is validated using the REFIT smart home dataset. They claim that their proposed system can detect spam better than currently used spam detection systems. Their work can be utilized in smart homes and other places where intelligent devices are used.

Kumar et al. [14] discussed email spam detection using various ML algorithms. Their article explores ML methods and how to implement them on datasets. The optimal algorithm for email spam detection with the highest precision and accuracy is identified from various ML algorithms. They concluded that the Multinomial Na¨ıve Bayes algorithm produces the best results, but it has limitations due to class-conditional independence, which causes the machine to misclassify some inputs. Ensemble models come after Multinomial Na¨ıve Bayes with the best and reliable results in this study. The proposed system in this study can only detect spam from the body of emails.

Singh and Batra [65] proposed a semisupervised machine learning technique for spam detection in social IoT platforms. They used an ensemble-based framework that is consists of four classifiers. The architecture is based on the use of probabilistic data structures (PDS) such as Quotient Filter (QF) to query the database of URLs, spam users, databases of spam keywords, and Locality Sensitive Hashing (LSH) for similarity search. The proposed model minimizes, so it decides by an adaptive weighted voting approach based on each classifier's output. The hybrid sampling technique minimizes the computational efforts, which sample the data according to each classifier. This study indicates that the proposed model can be used for spam detection on large datasets. The proposed model's efficiency was evaluated by comparing PDS with standard data models and the typical evaluation metrics, including accuracy, recall, and F-score.

Artificial Neural Networks. An artificial neural network (ANN) is a computational model based on the functional aspects of biological neural networks, also known as the neural network (NN) [66]. Many sets of neurons are joined in a neural network, and information is interpreted using a computational approach connection. In most situ- ations, an ANN is an adaptive system, which changes its structure depending on external or internal information flowing through the network during the learning phase. Current neural networks are nonlinear approaches to statistical data processing. These are commonly used when there are complex relationships between inputs and outputs or unusual performance patterns [6]. Figure 10 shows the basic structure of the neural network.

The following section elaborates some proposed email spam detection and prevention techniques by using neural networks.

Xu et al. [67] present a method for the detection of spam in online social networks. Their work focuses on the

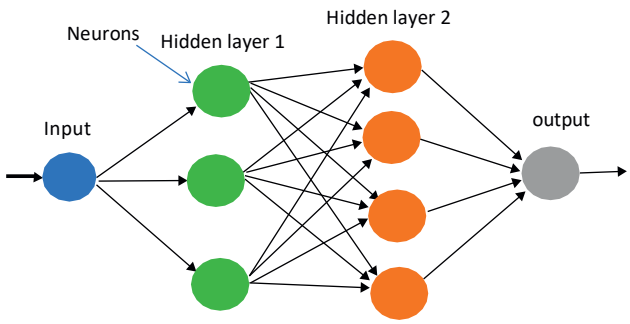

Figure 10: Basic structure of neural network.

combination of spam messages in one social network to another social network. By using Twitter, they gathered 1937 spam and 10943 ham tweets for processing. They also used 1338 spam posts and 9285 ham posts. In TSD, 75.6% of tweets contained URL links for spam tweets, while 24.4% contained different words. Out of 10942 ham tweets, 62.9% contained URL links and words, while 37.1% had only words. For the spam posts of FSD, 32.8% of posts consist of different web links, and the remaining 67.2% of spam posts contain only words [68]. Of 9285 ham posts, 95.1% have web links, and the other 4.9% consist of words. They used the top twenty feature words from Facebook spam data and Twitter spam data. They divide the TSD and FSD into two sets, i.e., training dataset and testing dataset. These datasets were used to train various machine learning classifiers like Na¨ıve Bayes, random forest, logistic regression random tree, and Bayes Net. After analyzing the accuracy of different classifiers, they combine the spam dataset of Facebook into the training dataset of Twitter and the spam dataset of Twitter into the training dataset of Facebook. Then, they used the combined dataset for the training and testing of classifiers. In the end, they compare the results of classifiers on the above-mentioned social networks after measuring the precision, accuracy, recall, and F-1 measure. They found that the accuracy of combined datasets was higher than that of other datasets [68, 69].

Guo et al. [70] proposed a spammer detection technique using a collaborative neural network in IoT applications. They present a novel spam detection mechanism called Cospam for IoT applications. At first, the user and contents of speech at different timestamps are viewed as feature sequences. In the second step, a collaborative neural network model is used. The collaborative model consists of three models: (1) Bi-AE model, (2) GCN model, and (3) LSTM model. These models are used for the identification of the nature of the user. In the end, a series of experiments were conducted for the evaluation of the proposed technique. The proposed model was able to obtain 5% more accuracy than existing spammer detection approaches. Cospam consumes more time than existing techniques because of a large number of parameters.

Makkar and Kumar [71] proposed a deep learning model for web spam detection in an IoTenvironment. Their system enhances the cognitive ability of search engines for the detection of web spam. This model removes spam pages with the help of a web page rank score calculated by a search engine. Their framework uses the extensive features of deep learning. The first time in which the LSTM model was used to detect spam is used for many problems like weather forecasting. In this study, the proposed model is compared with ten different machine learning models. The WEB-SPAM-UK 2007 standard dataset is used in this study. The preprocessing of the dataset is done by a novel technique called “Split by Oversampling and Train by Underfitting.” The accuracy of the proposed model was 95.25%. After the optimization of the system, the proposed model got an accuracy of 96.96%.

Zavvar et al. [72] present a paper on spam detection by considering combined particle swarm optimization and neural networks to select features. They also used SVM for classifying and separating spam. They compared the proposed approach with other approaches such as a self-organizing map and k-means data grouping based on the region under curve parameters. This article uses the UCI base dataset to evaluate spam classification and provide a PSO-ANN and ANFIS algorithm-based approach for spam detection. Seventy percent of data was used for training, and 30 percent was used for testing the models. RMSE, NRMSE, and STD principles were analyzed and got 0.08733, 0.0185, and 0.08742 results in the testing phase. The results show that the proposed method has good accuracy and performance for detecting spam emails. Table 2 summarizes supervised machine learning techniques presented for spam detection.

Discussions and Learned Lessons. Supervised machine learning techniques, i.e., decision trees, random forests, support vector machines, and artificial neural networks, can be used for email spam detection or filtering. Support vector machines classify different objects by using the idea of the hyperplane. Objects are classified into two classes. If a new object is given to the model, it will be classified into one of both classes. Zavvar et al. [12], Garavand et al. [72], and Idris et al. present different techniques for spam detection using the support vector machine (SVM) model. They got a good accuracy level on different spam datasets. Olatunji et al. [73]used the support vector machine and extreme learning machine algorithms on the standard dataset and got 94.06% accuracy using the support vector machine. In their system, extreme learning machines perform better than SVM but take more time, so a time-consuming ELM performs better than SVM. Zheng et al. got the highest accuracy level using Weibo social network dataset. They use two types of features, i.e., content base and user behavior base, to classify spam- mers and nonspammers. Na¨ıve Bayes classification is an- other supervised machine learning technique, which predicts some events based on its na¨ıve theorem. Na¨ıve Bayes classifiers are quite simple, and they do not use an iterative process; they perform very efficiently on large datasets with a handsome level of accuracy. Hijawi et al. [41] use the Na¨ıve Bayes network for the detection of spam. They did not get outstanding results using the spam assassin dataset as their accuracy level was only 89%. Another technique which is widely used in the last decade is

decision

Table 2: Comparison of supervised techniques for spam filtering.

Authors	Algorithm	Dataset	Accuracy	Advantages	Limitations
DeBarr and Wechsler [42]	Random forest	Custom collection	95.2%	They got good accuracy with multiple trees	The dataset that they used was not a standard dataset
Rusland et al. [63]	Modified Na¨ıve Bayes with selective features	Spam base and spam data	88% on spam base 83% on spam data	Selective features are taken that consume less time	They got less accuracy, and their model was not much intelligent
Halu zu et al. [67]	Bayes Net, SVM, and NB	Twitter and Facebook dataset	90% using SVM	They used the combined dataset for the training and testing of classifiers	Multiple algorithms and a combined dataset system take more training time
Hijawi et al. [41]	(MLP), Na¨ıve Bayes, random forest, and decision tree	Spam assassin	99.3% using random forest	They use a list of most common spam features that improve the spam detection rate	They use a significant corpus of 6050 emails, but they use a small number of features extracted from the corpus
Banday and Jan [55]	Na¨ıve Bayes, K-nearest neighbor, SVM, and additive regression tree	Real-life dataset	96.69% using SVM	They make a spam filter based on 8000 real-life spam emails	Their model is not so effective as spammers continuously change the characteristic that they used for making spam filter
Verma and Sofat [48]	ID3 algorithm hidden Markov	Enron dataset	89%	They use a preclassified dataset that uses less time in processing	Their model got an 11% loss that is not too good for spam filters
Subasi et al. [40]	CART, C4.5, REP tree, LAD tree, and NBT	UCI dataset	95.1%	They used 10-fold cross-validation that helps in better evaluation	Less number of features used
Zheng et al. [12]	SVM	Weibo social network data	99.5%	They use both user content and behavior features for detecting spammers	Feature extraction is based on statistical analysis and manual selection
Garavand et al. [72]	SVM, deep learning, and particle swarm optimization	Standard datasets from UCI 70% education data	93% using the support vector machine	They use deep learning models for feature extraction	The neural networks take massive time for training for the extraction of features
Olatunji et al. [5]	ELM and SVM classifier	Enron dataset	94.06 using SVM	They got a high accuracy level as compared to previous studies on the same dataset	For SVM, it takes more time than ELM to gain the accuracy level claimed in the paper
Jamil et al. [10]	SVM, KNN, DT, and LR	Health fitness data	92.1 using SVM	Smart contract-enabled blockchain technique is used that makes the system more secure	Interoperability of proposed model with IoT framework is not evaluated
Arif et al. [11]	XGBoost, bagged model, and generalized linear model with stepwise feature selection	Smart home dataset	91.8 using generalized linear model with stepwise feature selection	PCA was applied that enhances the accuracy of the system	Climatic and surrounding features of IoT devices are not considered

tree. These decision algorithms define models that are constructed iteratively or recurrently based on the data provided. The decision tree-based algorithms goal is to predict a target variable's value on given set of input variables. Subasi et al. [40] used different decision tree-based algorithms for spam detection on the UCI machine learning platform dataset. They used 10-fold cross-validation for the evaluation of decision tree classifiers. They use open-source Weka tools for the development of the model. DeBarr and Wechsler [42] used a tree-based random forest algorithm for email spam detection and active learning for refining the classification. They used the data of email messages from RFC 822 (Internet) and got the highest accuracy level of 95.2% by using the dataset's custom collection of emails. In all supervised machine learning techniques, Zheng et al. [12] got the highest accuracy level among all researchers using the support vector machine (SVM) technique for email spam detection.

Unsupervised Machine Learning. Unsupervised machine learning algorithms are used when we do not have labeled data [74]. Unsupervised learning explores how programs can explain a hidden structure by inferring a feature from un- labeled data [75]. The machine does not evaluate the ap- propriate output but examines the data and can draw inferences from datasets to explain hidden constructs from unlabeled data. Unsupervised learning works on unlabeled

data and makes clusters of the data based on the features of that data. This type of learning can be used for various problems like Recommender Systems, identifying Buying Habits, Grouping User Logs, dimensionality reduction, etc. The process of unsupervised learning is illustrated in Figure 11.

Clustering is the main application of unsupervised learning that has two main types. Different clustering techniques are discussed as follows.

Hierarchical Clustering. Hierarchical clustering identifies clusters with a hierarchy achieved either by iter- atively combining smaller clusters into a more significant cluster or by splitting a more massive cluster into smaller clusters. This cluster hierarchy, generated through a clus- tering algorithm, is called a dendrogram [76]. A dendrogram is one way of representing the hierarchical clusters. The user can understand different clusters based on the level at which the dendrogram is defined. It uses a similarity scale repre- senting the distance between the clusters grouped from the massive cluster. A dendrogram is a visual representation of hierarchical clustering that is illustrated in Figure 12.

Partitional Clustering. A partitional clustering dividesa single set of data objects into nonoverlapping subsets (clusters) so that each data object is in only one subset [77]. Partitional clustering algorithms make different partitions ofdata and then evaluate the required results based on some criteria. Figure 13 illustrates the basic structure of partitional clustering algorithms. In Figure 13, partitions (A, B, and C)are created based on some characteristics. Partitional clus- tering breaks down a dataset into a collection of clusters of disjoints. The partitioning technique forms different parti- tions of data by using the formula K (N/K); each partition represents a cluster based on a set of N points in the data, that is, by fulfilling the following conditions:

(1) Each class contains one point or more

(2) Each point comes as part of exactly one group

Let us discuss some work on filtering email spam using unsupervised machine learning techniques.

Sharma and Rastogi [78] propose a strategy using un- supervised techniques. They performed various experiments on email spam datasets. After data gathering, they use the k-means clustering model for the clustering of emails. They use various distance measures for this purpose. The study's findings show that the proposed model performs well and cluster spam and ham emails are efficient.

Tan et al. [79] developed a reliable model for spam detection. First, they present a Sybil defense-based auto- mated spam detection scheme called SD2, which consid- erably outperforms current techniques by considering the social network relationship. They further developed an unsupervised spam detection system called UNIK to address increased spam attacks effectively. Instead of directly detecting spammers, UNIK operates by intentionally eliminating nonspammers from the network. They used the

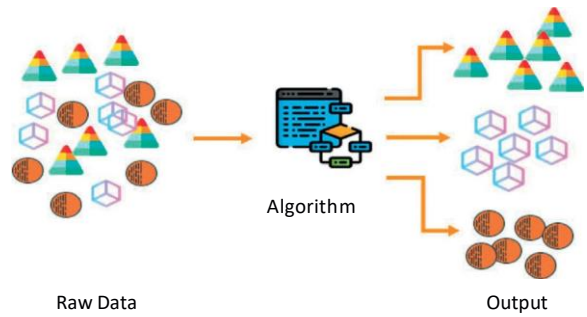

Figure 11: Process of unsupervised learning.

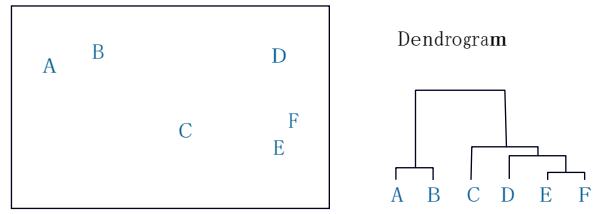

Figure 12: Structure of dendrogram.

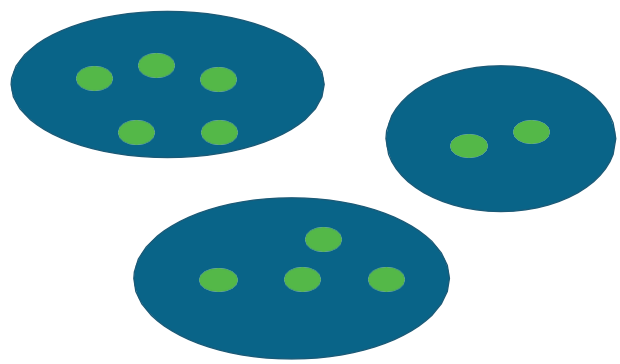

Figure 13: Partitioned clustering structure.

social graph as well as the user-link graph for the detection of the spammer. UNIK's fundamental basis is that spammers actively change their patterns to avoid detection, while nonspammers are not expected to do so. Therefore, we have a reasonably nonvolatile pattern. When tested on a broad network platform, UNIK has a similar performance as SD2 and substantially beats SD2 as spam attack rates go up. They evaluate several known spam activities in the social network platform by the identification of UNIK. Their proposed system, UNIK, can be used for email spam classification. The result shows that various spammer clusters exhibit different characteristics, suggesting the instability of spamming and UNIK's ability of automatically extracting junk mail signatures.

Ahmed [80] used an improved digest algorithm with DBSCAN clustering to classify spam emails. They create a different digest (parts) of emails before clustering. Their proposed model has two key steps. When the system receives emails, it first enters the digest generation phase, where an improved digest algorithm processes it, and the output is the set of digests of each email. These digests are then given to the clustering algorithm, i.e., DBSCAN, in the next phase. In

the clustering phase, similar emails are classified in the clustering process in a cluster of spam mails based on similarities among their digests, where mails that do not look like any other digest are considered noise and not clustered. Such emails that are not clustered are standard (ham) emails.

Using unsupervised artificial neural networks (ANNs), Cabrera-León et al. [81] propose a hybrid antispam filter. Their method contains two main steps. The first step is preprocessing of content, and the second one is actual processing. Each step is based on various models of computation. These models are "programmed and neural (using Kohonen SOM) [55]. This proposed system used the Enron dataset for ham or legitimate emails, while for spam emails they used two distinct sources. The first phase preprocessing was done based on thirteen (13) thematic features found in spam and ham emails. The terms frequency (TF) and inverse term frequency (IDF) were used in their system for the sake of feature extraction. Their results were the same as those of other researchers for the same dataset since they use distinct machine learning techniques and attributes. They evaluated their system with various datasets, defined by interdependent origins, ages, users, and forms like image spam samples. Their system got an accuracy level between 75% and 96%. They show that model performance degradation can vary by variations, in datasets, especially in dates. This phenomenon is known as "topic drift." Generally, it affects all classifiers, but it more affects those classifiers that use offline learning. The same case is with adversarial machine learning problems like spam filtering. Their method is robust to phrase obfuscation, which is commonly used in spam content. It was also independent of the need to use lemmatization or stemming.

Sasaki and Shinnou [82] introduce a new approach for spam detection using the vector-space model of content clustering. Their system automatically calculates disjoint clusters using a spherical k-means technique for all spam and nonspam emails. It collects centroid vectors of clusters for the extraction of vector definition. Each centroid is labeled with spam and nonspam to measure several spam emails in the clusters. The system measures the cosine similarity between the current mail vector and the centroid vector as a new email arrives. Eventually, the new mail is assigned the label of the most appropriate cluster. They obtain several kinds of spam and nonspam email topics by using the proposed approach and effectively identifying the spam emails. They introduce the spam detection framework in this paper and demonstrate the research outcomes utilizing the series of Ling-spam datasets. They got 98.06% accuracy with their model.

Narisawa et al. [83] suggest an unsupervised approach for detecting spam documents from several documents relying on string equivalence. They provide three metrics to quantify a string's alienation, which means how distinct they are inside the documents from other substrings. In their proposed model, a document labeled as spam includes a substring with a significant alien degree in an equivalence class. The proposed approach was unsupervised, independently of language, and scalable. Japanese web forum data were used for computational experiments to show the proposed approach's performance on real data. Table 3 presents comparison of unsupervised learning techniques used for spam filtering.

Discussion and Learned Lessons. Several unsupervised machine learning models are being used for email spam detection and filtering. Hierarchical clustering and parti- tioning clustering are commonly used clustering techniques. Ahmed [80] used DBSCAN clustering and an improved digest algorithm to classify emails. He used the spam assassin dataset for the development of his model. This approach significantly enhances filtering accuracy by 30 percent against the newly proposed algorithms and increases spam detection tolerance against increased spammer's obfuscation effort while maintaining successful email detection at a comparable level of older filtering methods.

Sharma and Rastogi [78] used a machine learning algorithm (k-mean clustering) with local concentration-based content extraction for spam detection and got a handsome accuracy level. Cabrera-León et al. [81] used an artificial neural network that contains two necessary steps. In the first step, they do preprocessing and then in the second step they process cleaned data for computing the results. These steps are based on distinct models of computation. Its accuracy was 95%. Narisawa et al. [83] introduced an unsupervised approach to identify a spam document from a collection of documents based on string equivalence. This solution was a language-independent and scalable method for spam detection. It was tested on the Japanese web forum. Among all the researchers, Sharma Rastogi [78] and Ahmed et al. got the highest accuracy level using DBSCAN and K-mean algorithm, respectively, for the email spam detection. Ahmed [80] used spam assassin dataset for the implementation of his model.

Reinforcement Machine Learning. Reinforcement learning is another type of machine learning which works on reward taken from its environment. It takes suitable actions to make or get the maximum reward in a given situation [84]. Many machines and software employ it to find the optimal path to take in a specific situation.

The main difference between supervised and reinforcement learning is that supervised learning needs training data with correct labels. Simultaneously, there is no correct label in reinforcement learning, but the agent decides what to do to perform the given task. The agent is bound to learn from its experience if there is no training dataset [85].

Figure 14 illustrates the simple reinforcement learning process in which an agent passes an action to the environment. The environment sends back the reward of action and state to the agent. Let us discuss some research work done on email spam detection using reinforcement learning.

Chiu et al. [86] propose an alliance-based approach to classify, identify, and exchange relevant information on spam email contents. Their spam filter consisted of a rough set theory, a machine learning classifier (XCS), and a genetic algorithm. They used several metrics to evaluate the model results. From their paper, two main conclusions can be

Table 3: Comparison of unsupervised learning techniques used for spam filtering.

Authors	Algorithm used	Dataset	Accuracy (%)	Advantages	Disadvantages
Ahmed [80]	Improved digest and DBSCAN	Spam assassin	96.7	The proposed model divides email into fixed-length strings before clustering, which gives better accuracy	The speed of the proposed model depends upon the length of strings
Sharma and Rastogi [78]	K-means clustering	UCI dataset	92.76	It is discretized using supervised attribute filters and also used 10-fold cross-validation	While comparing multiple algorithms, results take a handsome amount of time
Cabrera-León et al. [81]	Unsupervised artificial neural networks	Enron email	95	The system is robust to word obfuscation, used in spam, independently of the use of stemming or lemmatization	Bad false negative and false positive rate are around 11 and 4%, respectively
Sasaki and Shinnou [82]	Spherical k-means algorithm	Ling-spam	96.04	The model uses various contents of spam emails	Updating spam contents and relevance feedback is not in the proposed model
Narisawa et al. [83]	Equivalence relations of strings	Japanese web forums	95	The model was scalable and language-independent	As the model uses N-Gram of documents, so results depend on the value of "*n*"
Tan et al. [79]	UNIK and SD2	Social network sites data	93	It is highly robust to an increased level of spam attacks	The proposed system cannot handle short URLs

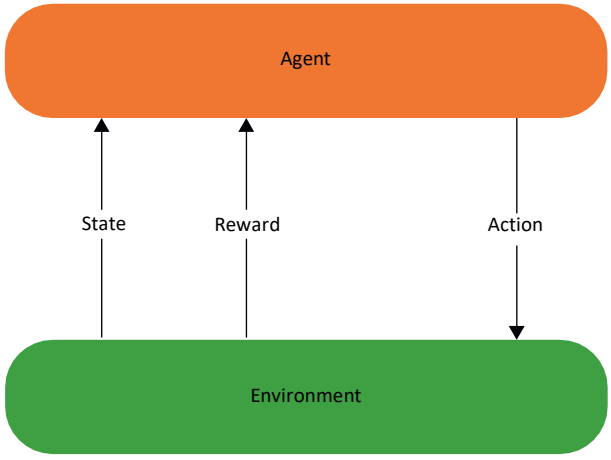

Figure 14: Basic structure of reinforcement learning.

drawn, and they are given as follows: The spam filter is based on a combination of rough set theory, genetic algorithm, and machine classifier XCS. Many metrics are used to assess spam mails filtering results by an alliance-based approach and provide a reasonable output indicator. They may draw two key conclusions which are the following:

(a) The rules that have been shared from many other email servers do help the spam filter to block more spam emails than before

(b) A blend of several techniques increases precision and decreases false positives for the spam detection task

Discussion and Learned Lesson. Reinforcement machine learning is a type of machine learning in which an agent communicates with its environment by producing behaviors and generating results or rewards. This method allows the software agents to find an optimal solution in a specific domain. An agent acts with the environment and gets the error or reward. Chiu et al. [86] used this approach on spam emails. The spam filter was built based on a mixture of rough set theory, genetic algorithm, an XCS classifier system, and good performance measure. Lai et al. [87] propose a practical approach for spam detection using rough set theory and XML format. They use reinforcement learning for the management exchange of spam rules. They suggest that outdated rules should be discarded as spammers are constantly changing their methods for doing spam. They further conclude that the spam filter can block more spam emails than a standalone system by sharing spam rules between the email servers. Samadi et al. [85] and Dou et al. [88] also used reinforcement learning techniques to detect spam and spammers.

6. Overall Insights of the Machine Learning Algorithms for Spam Detection

Figure 15 illustrates the percentage of work on email spam detection discussed in this survey. After discussing the literature, we observed that most of the datasets used to train, test, and implement different models are synthetically created. There is a lack of examples for analysis and the complexity of labeling all the supervised model data. So, the classifiers' results are not 100% trustworthy because of the synthetic datasets used for the models' training. These are not representative of real-world spam reviews as vast numbers of machine learning models are currently used for email spam detection or filtering. The three learning algorithms, logistic regression, Na¨ıve Bayes, and support vector machine (SVM), are widely used, and they outperform the other learning algorithms in most of the discussed studies.

SVM generally gives the best performance; Na¨ıve Bayes and logistic regression commonly beat it. But SVM should not be considered merely as the best algorithm since it is not compared to all others. Multiple learning models on various

datasets should be evaluated in future studies using several different feature engineering methods. This survey paper elaborates the existing machine learning-based spam filtering techniques and models by exploring and observing numerous methods. The conclusions are discussed by the overview of several spam filtering techniques and summarizing the accuracy of different proposed approaches based on various parameters. We conclude that all the spam filtering techniques perform well. Some have outstanding results, while some are trying to use other methods to increase the accuracy level. Though all are effective, the spam filtering system still lacks some, which are the primary concern for researchers. They are trying to generate next-generation spam filtering processes that can work on multimedia data and prominently filter spam emails. Table 4 is reproduced from Awad and Elseuofi [13]. Table 4 summarizes the performance of various machine learning models on 100 selected features.

7. Research Gaps and Open Research Problems

This section discusses the research gaps and open research problems of the spam detection and filtration domain. In the future, experiments and models should be trained on real-life data rather than manually created datasets, because, in the various article, the models trained on artificial datasets perform very poorly on real-life data. Currently, supervised, unsupervised, and reinforcement learning algorithms are used for spam detection, but we can get higher accuracy and efficiency by using hybrid algorithms in the future. Feature extraction can be improved in the future by using deep learning for feature extraction. Using clustering techniques for spam filtering relevance feedback using dynamic updating can better cluster spam and ham. Along with machine learning, blockchain models and concepts can also be used for email spam detection in the future. Experts in linguistics and psycholinguistics can collaborate in the future for manual annotation of datasets, which will result in the development of effective and standard spam datasets with high dimensionality. In future, spam filters can be designed with faster processing and classification accuracy using Graphics Processing Units (GPUs) and Field Programmable Gate Arrays (FPGAs), which offer low energy consumption, flexibility, and real-time processing capabilities. Moreover, future research should concentrate on the availability of standard labeled datasets for researchers to train classifiers and the addition of more attributes to the dataset to improve the accuracy and reliability of spam detection models, such as the spammer's IP address and the location. The following are some other future research directions and open research problems in the domain of spam detection.

(i) Some studies considered header, subject of the email, and message body as a feature for spam classification. While these features are not enough for fully accurate results, manual feature selection and features should also be.

(ii) Almost all researchers presented their results based on accuracy, precision, recall, etc., while the time complexity of machine learning models should be considered an evaluation metric.

(iii) Some researchers show promising results in the process of feature extraction using a bag of words. They claim that the email header is as important for spam detection as the content of the body. So, deep feature extraction of the header line should be considered.

(iv) Fault tolerance, self-learning, and quick response time can be better by using comprehensive feature engineering and an accurate preprocessing phase.

(v) Deep learning models with dynamic updating of feature space are needed to implement for better spam classification. Most of the current filters cannot update their feature space.

(vi) The security of spam detection and filtration system is needed for better accuracy and reliable results.

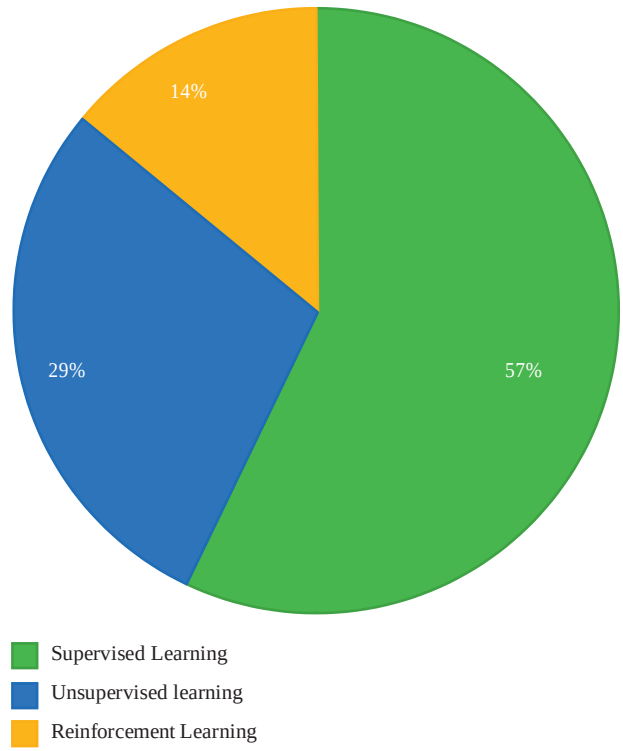

Figure 15: Ratio of machine learning techniques for email spam detection.

Table 4: Performance of various machine learning models on 100 selected features.

Algorithm	Recall	Precision	Accuracy
Naïve Bayes	98.46	99.66	99.46
SVM	95.00	93.12	96.90
KNN	97.14	87.00	96.20
Neural network	96.92	96.02	96.83
AIS	93.68	97.75	96.23
Decision tree	94.36	91.35	93.55

(vii) The false positive rate of many models is still higher than required. It must be reduced to the smallest possible value.

(viii) Few spam filters work on image spam detection and filtration. Expert spammers also use images for spam messages, so it should be considered in detecting spam.

(ix) Real-time spam classification is much needed as most of the proposed models cannot work on real-time data.

(x) Labeled data is one of the major issues in spam detection. There are a few new labeled and up-to-date datasets for this purpose.

(xi) Multilingual spam detection is also a significant research area that can be explored for better spam detection systems. There is less work done on multilingual spam detection using deep learning techniques.

(xii) Semisupervised and federated learning techniques can be used to enhance spam detection in various IoT and email frameworks.

(xiii) A combination of linguistic features for the spam detection approach can also be explored.

(xiv) The research community ignores the identification of spammers and spammer networks.

(xv) Many researchers manually annotate data, using spam features that they think to be accurate. As a result, the evaluation results of the detection systems that they propose are doubted. The ideal solution for this problem has yet to be discovered.

(xvi) There is a lack of a robust method of dealing with challenges regarding the spam filters' security. An attack of this nature can be a casual, exploratory, or targeted attack. The deep learning techniques with blockchain technology can be used for this purpose.

8. Challenges of Spam Detection

Some critical challenges faced by spam filters are discussed as follows:

(i) The growing amount of data on the Internet with various new features is a big challenge for spam detection systems.

(ii) Features' evaluation from several dimensions such as temporal, writing styles, semantic, and statistical ones is also challenging for spam filters.

(iii) Most of the models are trained on balanced datasets, while self-learning models are not possible.

(iv) Many spam detection models face adversarial machine learning attacks that will decrease their effectiveness. Adversaries can throw a variety of attacks during the training and testing of ML models. Adversaries can harm training data to cause a classifier to classify the data incorrectly (poisoning attack), create unfavorable samples during testing to evade detection (evasion attack), and obtain sensitive training data via a learning model (privacy attack)

(v) Deep fake is another big challenge that is being faced by spam detection systems. To generate, modify, and style pictures and videos, neural network models such as GPT-2,3 and image generation models like BigGAN, StyleGAN, and CycleGAN are adopted. Deep fakes can be used to disseminate false information.

9. Conclusion

In the last two decades, 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 Na¨ıve 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.

Code to classify spam emails by python

```
import pandas as pd
from sklearn.model_selection import
train_test_split
from sklearn.feature_extraction.text
import CountVectorizer
from sklearn import svm

spam =
pd.read_csv('C:\\Users\\nethm\\Downlo
ads\\spam.csv')
z = spam['EmailText']
y = spam["Label"]
z_train, z_test,y_train, y_test =
train_test_split(z,y,test_size = 0.2)

cv = CountVectorizer()
features = cv.fit_transform(z_train)

model = svm.SVC()
model.fit(features,y_train)

features_test = cv.transform(z_test)
print(model.score(features_test,y_tes
t))
```

[1]