GOVERNMENT COLLEGE OF ENGINEERING ,

SALEM - 11.

NM ID	NAME
au61772111039	JANANI S

Ramar Bose

AI Master trainer

ABSTRACT

Nowaday, emails are used in almost every field, from business to education. Emails have two subcategories, i.e., ham and spam. Email spam, also called junk emails or unwanted emails, is a type of email that can be used to harm any user by wasting his/her time, computing resources, and stealing valuable information. The ratio of spam emails is increasing rapidly day by day. Spam detection and filtration are significant and enormous problems for email and IoT service providers nowadays. Among all the techniques developed for detecting and preventing spam, filtering email is one of the most essential and prominent approaches. Several machine learning and deep learning techniques have been used for this purpose, i.e., Naïve Bayes, decision trees, neural networks, and random forest. This paper surveys the machine learning techniques used for spam filtering techniques used in email and IoT platforms by classifying them into suitable categories. A comprehensive comparison of these techniques is also made based on accuracy, precision, recall, etc. In the end, comprehensive insights and future research directions are also discussed.

INTRODUCTION

1. PROBLEM STATEMENT

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 . 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. 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 . 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. 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 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. 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, mail header analysis, keyword checking, etc.

---

2. ***SERVICES AND TOOLS REQUIRED***

- Supervised Machine Learning. Supervised machine learning algorithms [18] are machine learning models that need labeled data. ...
- Decision Tree Classifier. ...
- Support Vector Machine (SVM) ...
- Naïve Bayes Classifier (NB) ...
- Artificial Neural Networks. ...
- Discussions and Learned Lessons.

3. PROJECT ARCHITECTURE

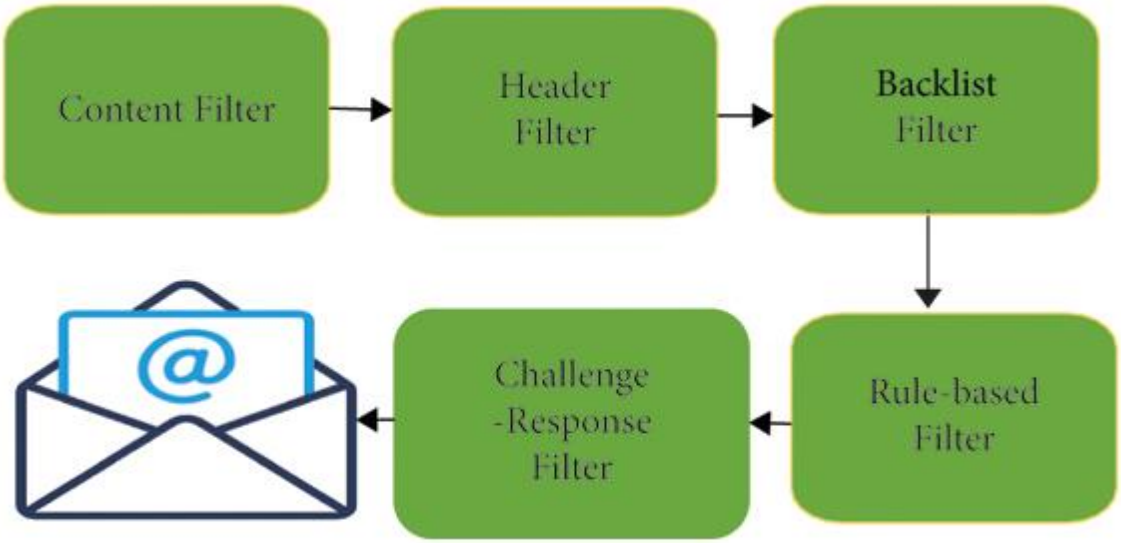
The email spam definition is ambiguous since everybody has their views on it. At present, email spam is getting the attention of everyone. Email spam ordinarily includes particular spontaneous messages sent in mass by individuals you do not know. The term spam is obtained from the Monty Python sketch, 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. 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.

3.1. Spam Filtering Methods in Email and IoT Platforms

The number of spam emails is rapidly increasing in marketing, chain communications, stock market tips, politics, and education. Currently, various companies develop different techniques and algorithms for efficient spam detection and filtering. We address some filtering strategies in this section to understand the filtering process.

3.1.1. 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. The diagram illustrates a standard method for filtering spam. In the first step, content filters are implemented and use artificial intelligence techniques to figure out the spam. 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.

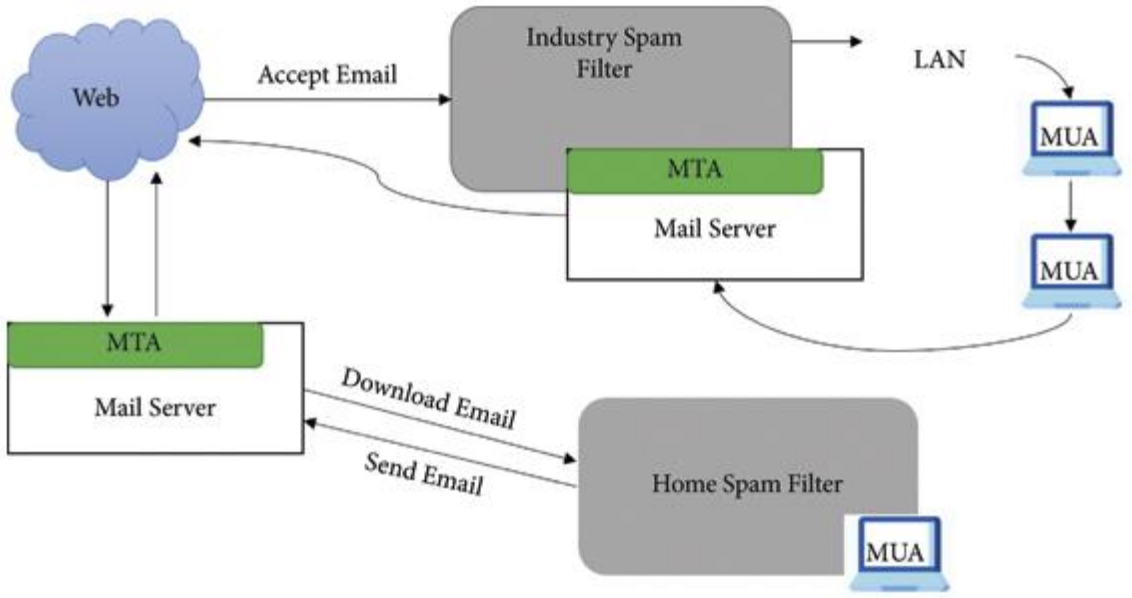


3.1.2. The Client Side Spam Filtering

A client is a person who can use the Internet or email network to send or receive an email. Spam detection at the client point offers different rules and mechanisms to ensure secure communications transmission between people and organizations. For transmission of data, a client should deploy multiple existing frameworks on his/her system. Such systems connect with client mail agents and filter the client's mailbox by compositing, accepting, and managing the incoming emails.

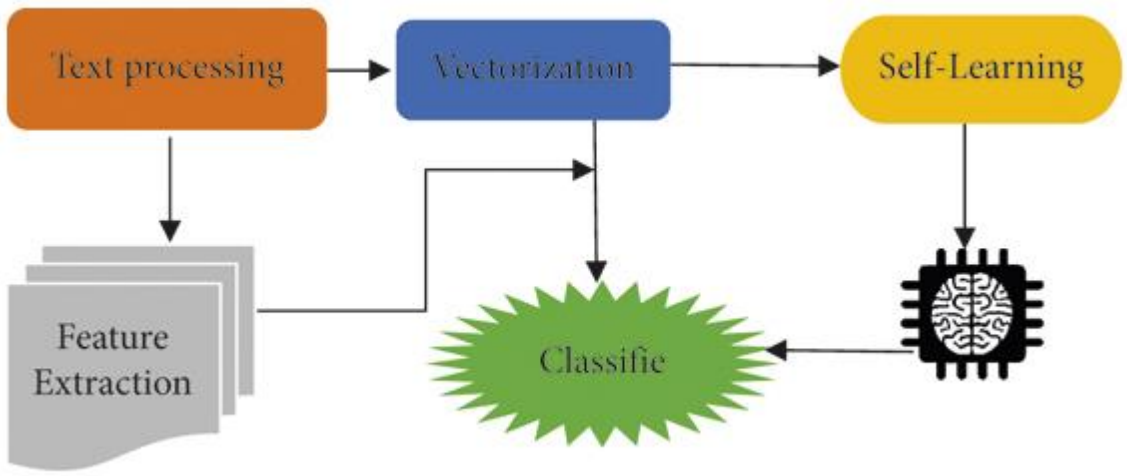
3.1.3. Enterprise Level Spam Filtering

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. 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 score is generated against every post. The junk mail or ham message is given specific scores or ranks.



3.1.4. Case-Based Spam Filtering

One of the well-known and conventional machine learning methods for spam detection is the case-based or sample-based spam filtering system. There are many phases to this type of filtering with the aid of the collection method; it collects data 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.

4. MODELLING AND PROJECT OUTCOME

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 IoT applications must protect privacy. Ios systems, including objects and networks, are vulnerable to network and physical attacks and privacy failures.

The various attacks of IoT systems are listed as follows.

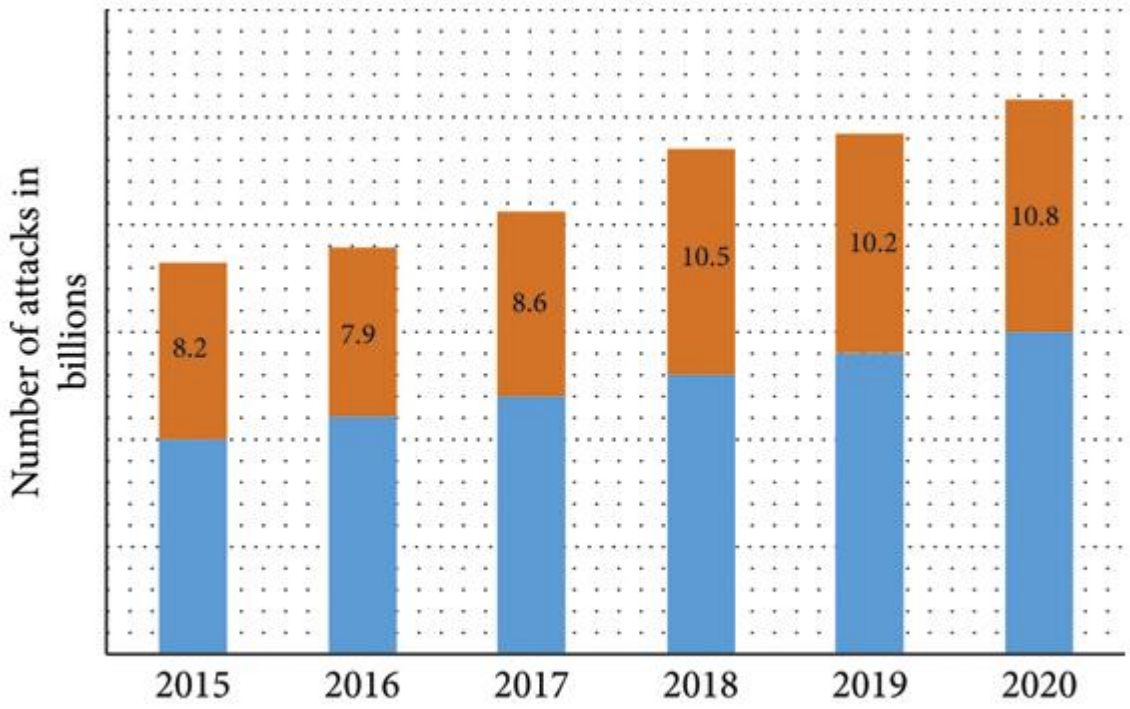
(a)*Self-Promotion Attack*. In this type of attack, the compromised node tries to get importance over the other nodes of the IoT environment for the particular recommendation.

(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.

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.

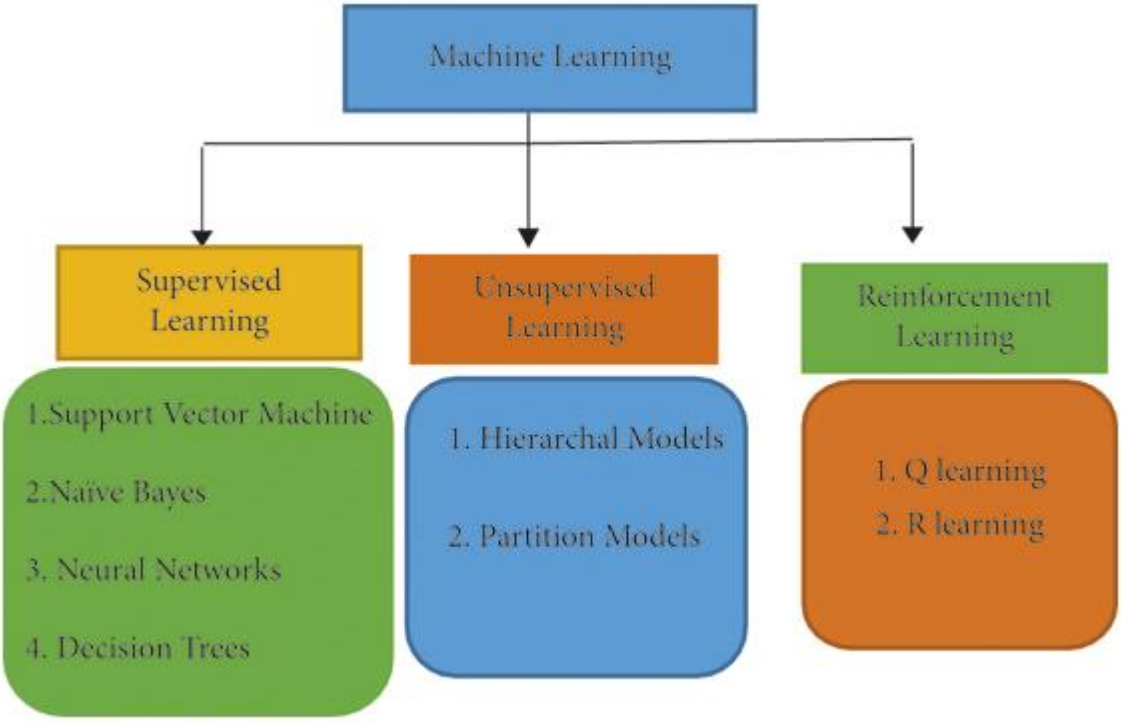
5. MACHINE LEARNING

Machine learning is one of the most important and valuable applications of artificial intelligence, which gives computer systems the ability of automatically learning and enhancing their functionality without explicit programming. 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.

5.1. Machine Learning-Based Spam Filtering Methods

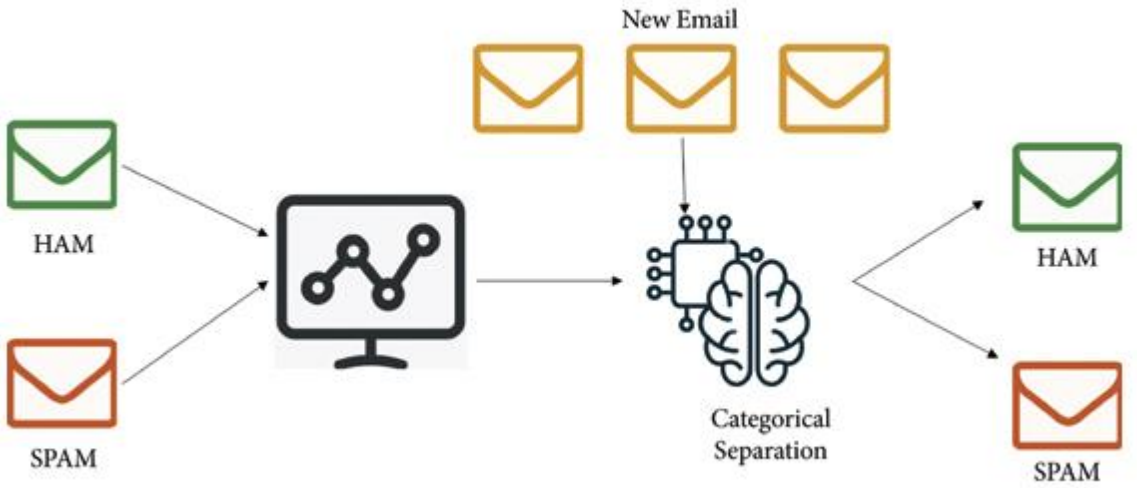
Machine learning facilitates the processing of vast quantities of data. Though it typically provides faster and more accurate results to detect unwanted content, it

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 can make handling massive amounts of data even more powerful. This diagram demonstrates various kinds of machine learning.



5.1.1. Supervised Machine Learning

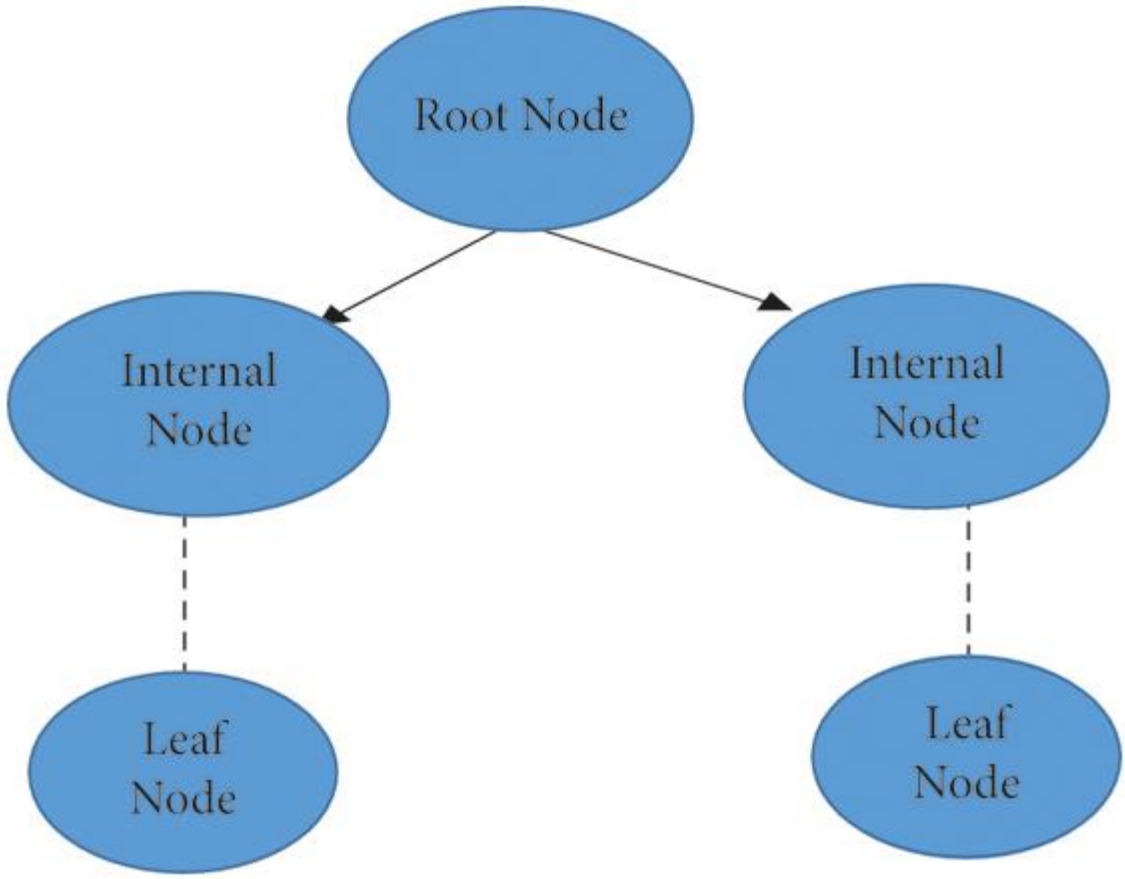
Supervised machine learning algorithms are machine learning models that need labeled data. Initially, labeled training data is provided to these models for training, and after training models predict future events. In other words, these models begin with the analysis of an existing training dataset,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



5.1.2. Support Vector Machine (SVM)

The support vector machine (SVM) is an essential and valuable machine learning model. SVM is a formally defined discriminative supervised learning classifier that takes labeled examples for training and gives a hyperplane as output, classifying new data. A set of objects belonging to various class memberships are separated by decision planes.

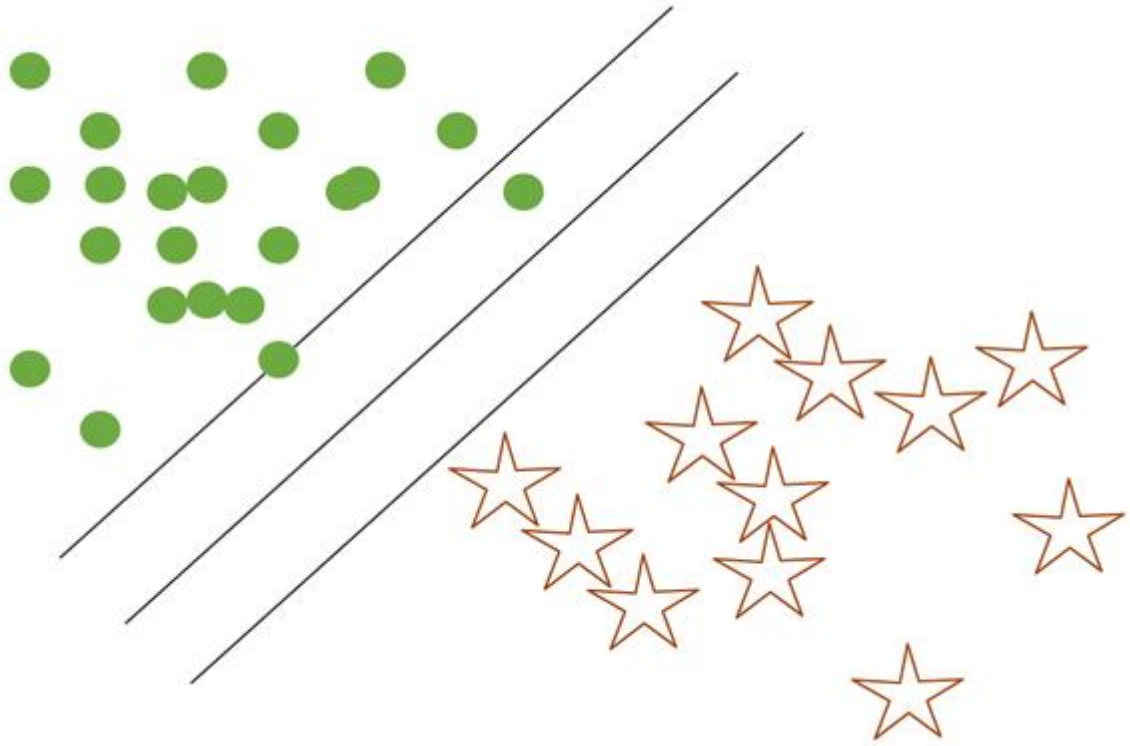
Banday and Jan present research in which they define the procedure of statistical spam filters. They design those filters using Naïve Bayes, KNN, support vector machines (SVM), and regression trees. They use all these supervised machine learning algorithms and evaluate the results based on precision, recall, and accuracy. Using these machine learning techniques, they found that classification and regression trees (CART) and Naïve 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.



Structure of decision tree.

5.1.3. Naïve Bayes Classifier

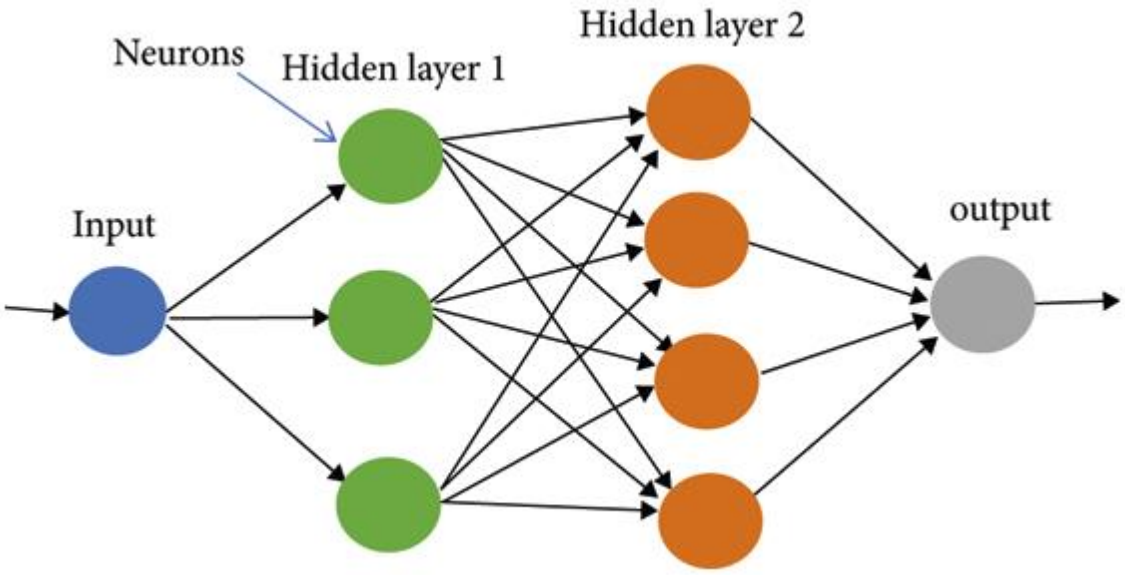
The Naïve Bayes classifier is based on the Bayes theorem. It assumes that the predictors are independent, which means that knowing the value of one attribute impacts any other attribute's value. Naïve Bayes classifiers are easy to build because they do not require any iterative process and they perform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



Support vector machine classification.

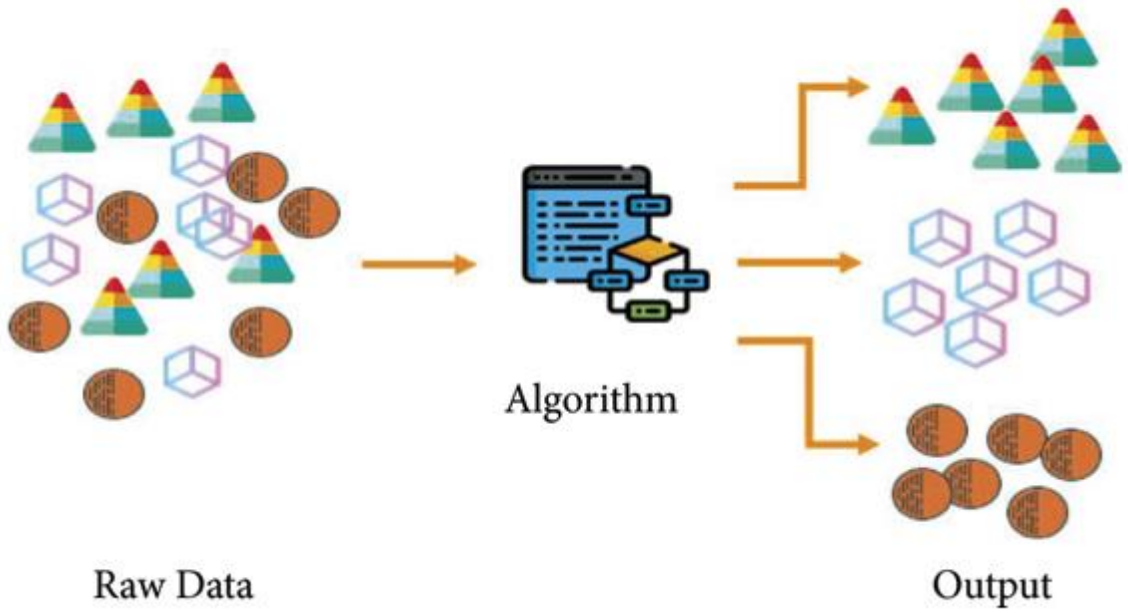
5.1.4. Artificial Neural Networks

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 situations, 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.



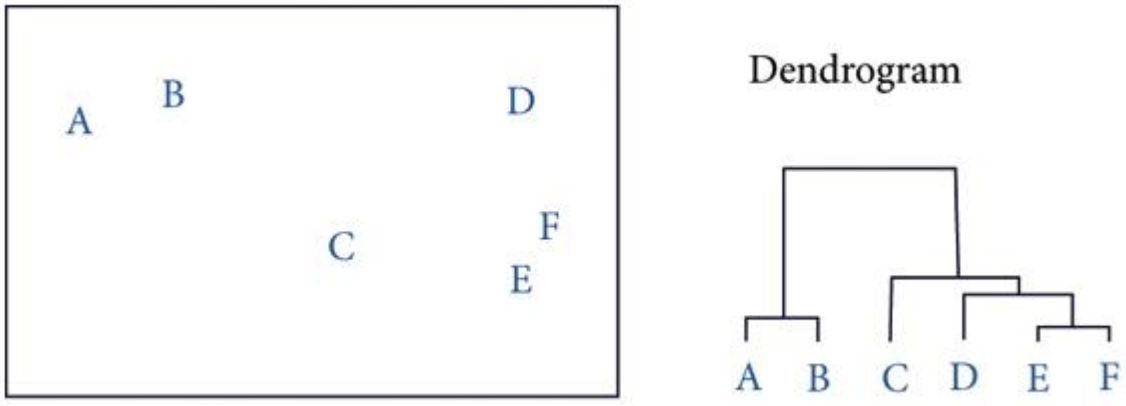
5.2. Unsupervised Machine Learning

Unsupervised machine learning algorithms are used when we do not have labeled data. Unsupervised learning explores how programs can explain a hidden structure by inferring a feature from unlabeled data. The machine does not evaluate the appropriate 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.



5.2.1. Hierarchical Clustering

Hierarchical clustering identifies clusters with a hierarchy achieved either by iteratively combining smaller clusters into a more significant cluster or by splitting a more massive cluster into smaller clusters. This cluster hierarchy, generated through a clustering algorithm, is called a dendrogram.

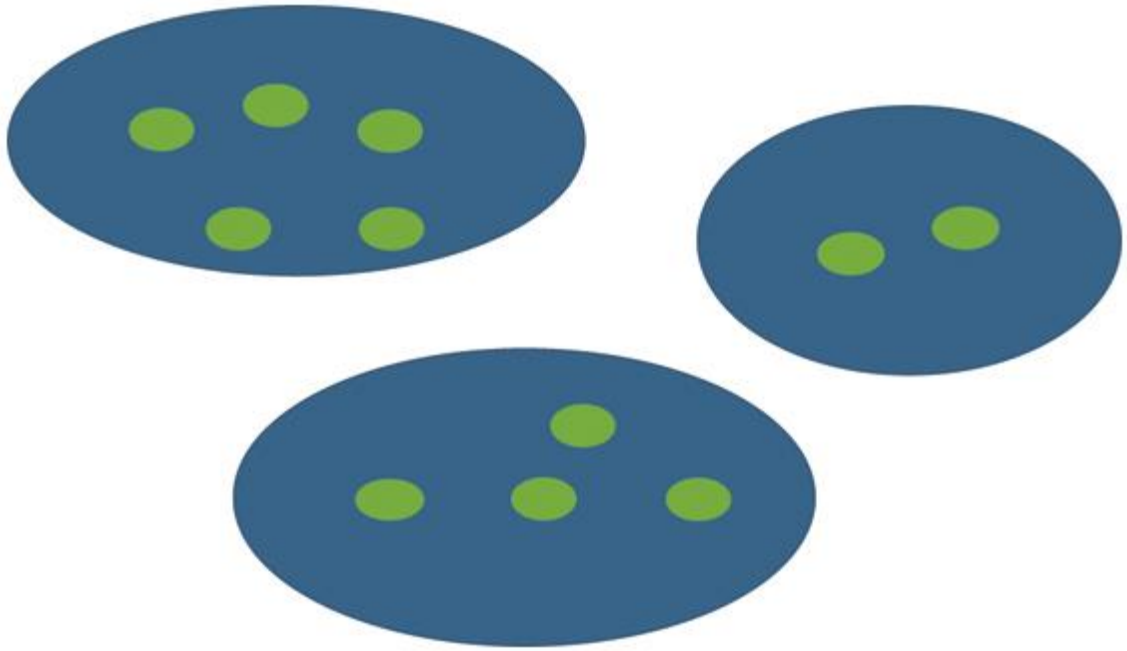


5.2.2. Partitional Clustering

A partitional clustering divides a single set of data objects into nonoverlapping subsets (clusters) so that each data object is in only one subset. Partitional clustering algorithms make different partitions of data and then evaluate the required results based on some criteria.

(1) Each class contains one point or more

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



Partitioned clustering structure.

5.2.3. Discussion and Learned Lessons

Several unsupervised machine learning models are being used for email spam detection and filtering. Hierarchical clustering and partitioning clustering are

commonly used clustering techniques. Ahmed 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.

5.3. 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. 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. This 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.

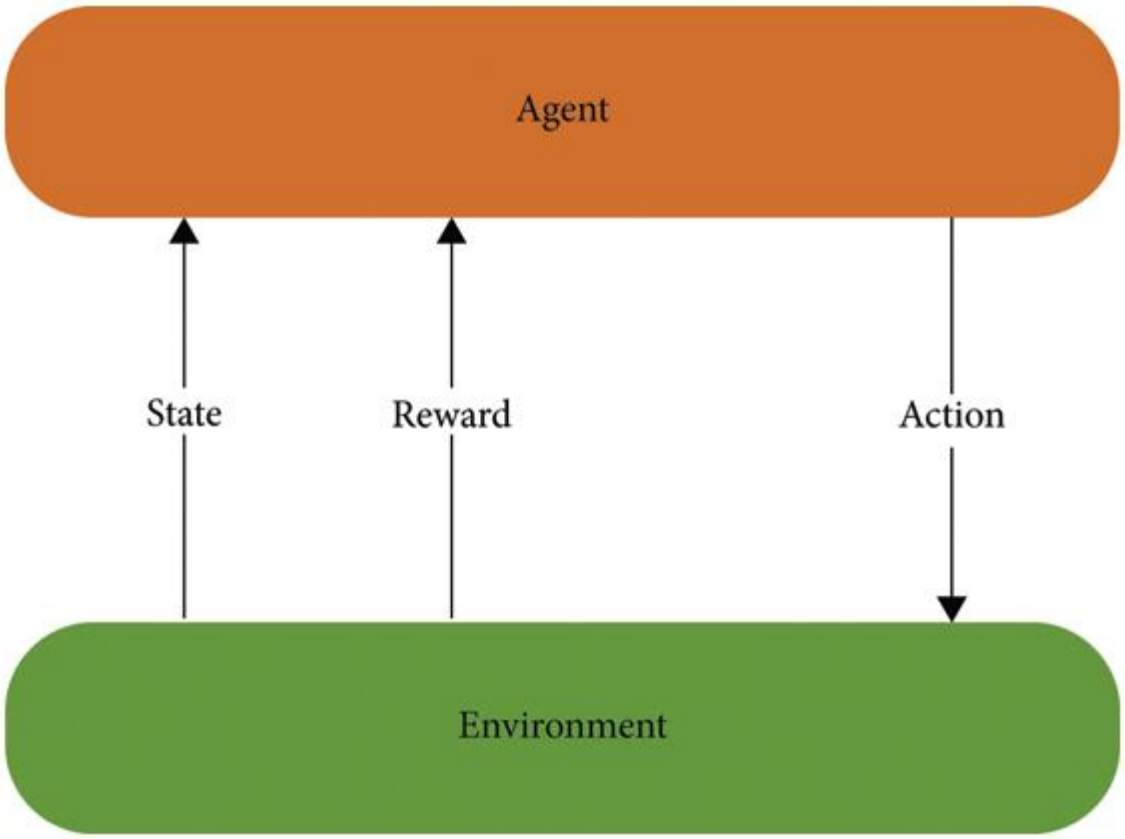
Basic structure of reinforcement learning.

Chiu et al. 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 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



5.3.1. Discussion and Learned Lesson

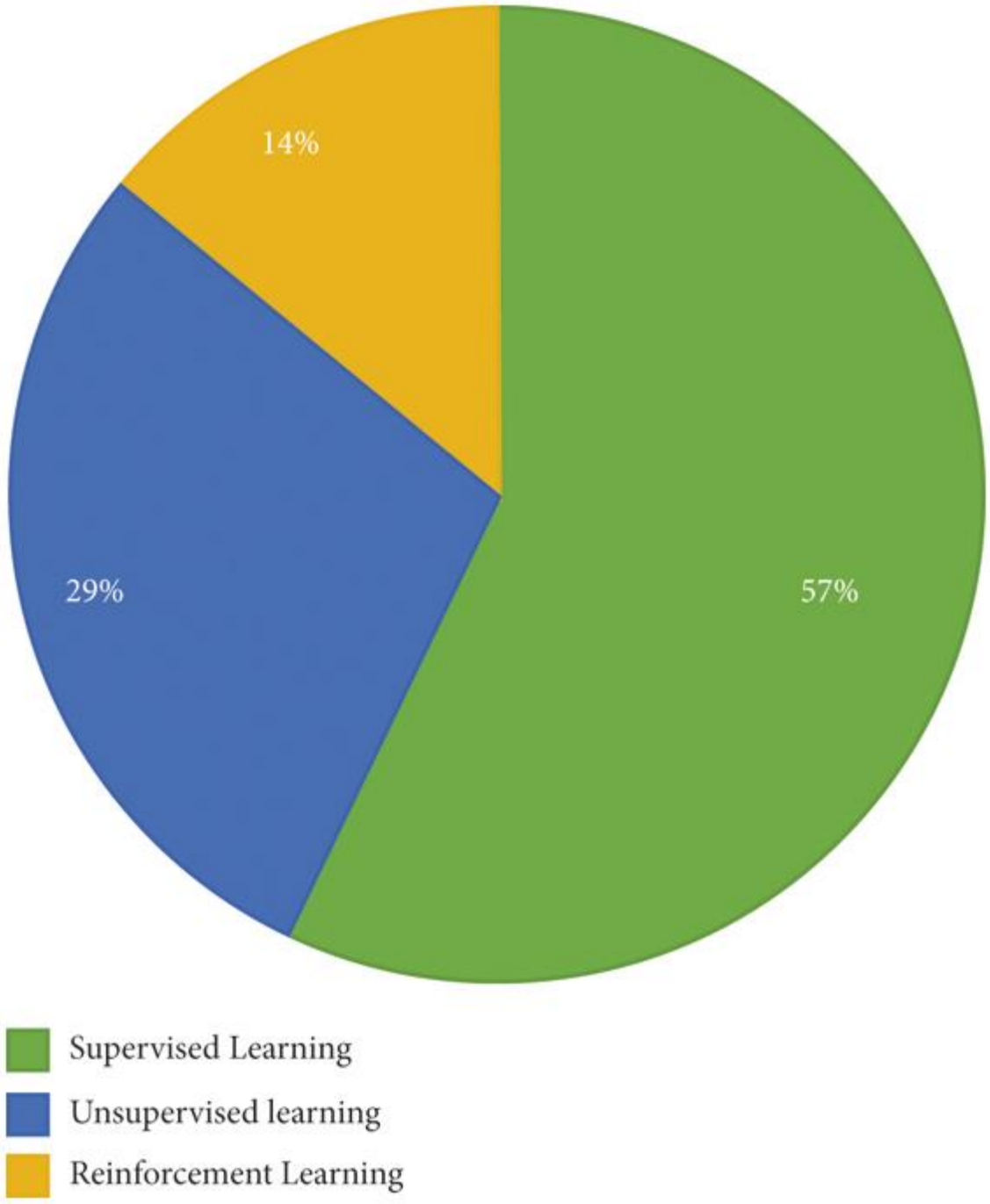
Reinforcement machine learning is a type of machine learning in which an agent communicates with its environment by producing behaviors and generating results or rewards. This method allows the software agents to find an optimal solution in a

specific domain. An agent acts with the environment and gets the error or reward. Chiu et al. 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. 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.

---

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

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, Naïve Bayes, and support vector machine (SVM), are widely used, and they outperform the other learning algorithms in most of the discussed studies.



7. Research Gaps and Future scope

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

---

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.

---

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

REFERENCE

https://www.hindawi.com/journals/scn/2022/1862888/

GIT Hub Link of Project Code
