iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 7 **Issue:** IV **Month of publication:** April 2019

**DOI:** https://doi.org/10.22214/ijraset.2019.4568

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887*
*Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com*

# A Survey on Various Machine Learning and Deep Learning Algorithms used for Classification of Spam and Non-Spam Emails

Shahbaz Ahmad Khanday[1], Suraiya Parveen[2]
[1, 2]*School of Engineering Sciences and Technology (SEST),Jamia Hamdard, New Delhi*

***Abstract:*** ***An email client receives emails from different websites, portals and domains, which can be an advertisement. Receiving a bulk amount of emails can cause serious damages like suspension of a particular email id.Mostly an email client gets exposed to the number of malicious receipts by registering an email account to a web portal, which in turn sends a bulk amount of emails. The email client wants to be decisive about differentiating the useful emails and spam emails. One of the solutions to escape from spam emails is to develop a decision based system which can classify the spam and non-spam emails. This survey gives an overview about different machine learning and deep learning algorithms to classify the spam and non-spam emails by accessing the received emails of an email client. The machine learning approaches and mechanisms like support vector machine, naive Bayesian classifier, artificial neural networks and logistic regression can be of important help to determine spam emails. These approaches use decision trees to run tests on a given sets of data (emails). After classifying a spam email source a user can navigate, block and report the source of the spam email generator. Most of the times the spam emails are generated by the autonomous sources which are called spam-bots***.

***Keywords: Machine learning, decision tree, support vector machine, logistic regression, artificial neural networks, naive Bayesian classifier and spam-bots.***

## I. INTRODUCTION

A common person can receive a huge amount of emails in a day. The email user can receive emails from different sources related to the different day to day activities like social networking, files and sharing, online shopping, e billing, e commerce and applications etc. One should be able to differentiate between important and useful emails over spam or junk emails. Once a user gets exposed to the spam and malicious sources he will receive a large amount of emails from various unknown sources. Therefore it becomes a hectic and time consuming task for an email user to make a selection and difference of all the received emails. The condition becomes very risky when a email client is trapped into a malicious act and then the security and privacy of a system could be breached. It is very hard to recover from such situations and most of the times an email user gets attracted to the spam emails and respond to them. In most of the cases the blocking and reporting of these spam email sources become useless, as the senders change their location continuously. One of the alternatives can be tracking those particular IP addresses from where an email user receives these spam emails, but the task becomes harder when the number of IP addresses are many but not fewer. And the major part is when the senders change their locations and targets.One of the solutions to the email spamming is to access an email id with the number of senders, classify and categorise the received emails into spam and non-spam emails. The classification of received emails can be done by using appropriate and approximate machine learning approaches and some autonomous algorithms like support vector machine, artificial neural networks, logistic regression and naive Bayesian classifier. All of machine learning approaches use decision tree based modelling and testing on a given input data set and produce the result in few classified groups. The decision tree performs testing and examination on data at its nodes. The resultant branch of the tree is the outcome of the test performed at the node of the decision tree. The spam and non-spam emails can be further classified into many more sub groups. A particular email sender which is repeated in the inbox of the email client can be captured and examined. Which in case can be a spam mail source and user can be provided options to report or block such sources.

Literature survey explores many machine learning and deep learning algorithms can be utilized for detection and the classification of spam and non-spam emails. The paper provides the comparison and the basic definitions about the various machine learning and deep learning algorithms used in the suitable environment and platform for classification of datasets. Some of the popular algorithms are:

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887*
*Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com*

### *A. Support Vector Machine*

Support vector machine is a machine learning algorithm which produces two different sets of data from a given input data module, divided by a hyper plane or hyper line. A hyper plane/line is a line dividing two different categories of data, one of which is above the hyper plane and another one is below the hyper plane. Support vector machine can be used for both classification and regression, but it is more efficient in classification purposes. Most of the times the support vector machine takes trained data as an input because it is a supervised learning algorithm. Support vector machine can plot the co-ordinates of any point in a n dimensional space provided if the n number of variables or features are in the input data set. The outcome/output provided by support vector machine are multiple different classes divided by a hyper plane. In the modern complex data science domains support vector plays a vital role in classifying data modules and it is one of the common and fruitful algorithms of machine learning used for segregation of two classes. The support vector machine model was given by Vapnik on the bases of a statistical learning theory. In the early stages support vector machine was used for simple and binary classification cases but it can used for complex and multiple class streams. [14][16].

### *B. Artificial Neural Networks*

Artificial neural network is a simulation model of a human neural network which performs tasks like a human a human brain. Artificial neural network a mathematical model defined by the inventor of first neurocomputer Dr. Hecht Nielsen as a computing system containing multiple interconnected computing phases or layers.

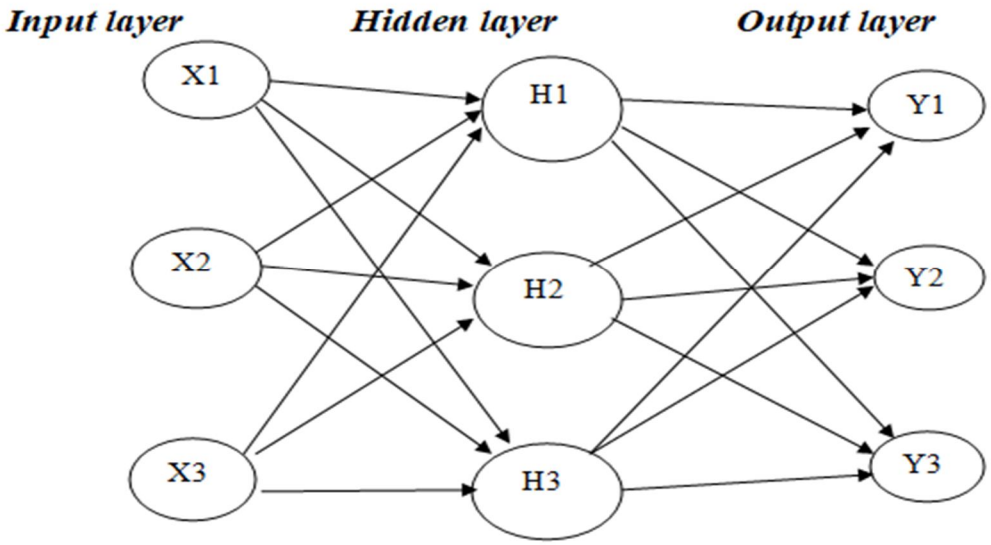



Figure 1: The systematic diagram of artificial neural networks

The information in an artificial neural network is processed by the dynamic state of response to the final output stage.The information in an artificial neural network is processed by the dynamic state of response to the final output stage. Artificial neural networks are based on the context of the human brain, which receives stimuli from various organs of the human body and acts accordingly with the actions.The artificial neural networks models can be formed and achieved by using the quality digital circuits and IC s, which could be used in placed of dendrites and neurons.Artificial neural networks are composed of multiple layers and all these layers are interconnected by the weighted links. Every layer in the artificial neural networks consist of multiple layers and links for the flow of information from one side to another side.The input layer receives input from the external environment at its nodes.The hidden layer receives input from the nodes of input layer, processes it and produces the output at the external or output layer. The procedure is repeated is until the appropriate, exact and approximate result is produced. It is one of the fundamental advantages of using artificial neural networks to produce output with high accuracy.The output layer can further send the feedback to the hidden layerfor the accuracy of output.The hidden layer works on the behaviour of a mathematical activation formula consisting of input and output contents of a data set at different time intervals. For the first transaction the hidden layer receives the input from all the nodes of input layer combined along the weights on the links together. The weights are the parameters or the composition of the numeric values and variables. [16][17]

*C. Naive Bayes classifier*

Naive Bayes classification uses Bayestheorem for the classification task of data sets. Naive Bayes classification is a set or collection of multiple classifying algorithms which have common nature ofworking and share a common working principle. Every classifying algorithm in naive Bayes classification are independent and does not relate to each other. In easier words the algorithms set in naive Bayes classification work separately and are not dependent upon the intermediate values and outcomes of each other. All the algorithms work on a common goal but separately.It is a classification model that uses a probability function for classification task. Its whole crux is based on the Bayestheorem, which is defined as

P(A/B) = P(B/A) P(A) / P(B)

where A and B are the two independent sets.

It states that we can find the probability of set A, if set B has occurred. In that case set B is the evidence and the assumption is that probability of set B is true. Therefore the set A is the hypothesis assuming that set B is the evidence. All the probabilistic outcomes from set A and set B are independent and do not effect each other. [16][17]

*D. Decision Tree*

Decision tree is one of the most powerful machine learning algorithms used for classification, regression and prediction motives. Decision trees generate outcomes and attributes based on the decisive test driven at the nodes of the tree. Decision tree can be viewed as a tree containing nodes and branches or links. Every node represents a test on the any attribute and every branch of the tree are the outcomes of the test driven at the nodes. The leaf nodes or the terminal nodes determine labels.

Decision trees are the learning models which divides the input data or source into a pattern of subsets based on some values on the attributes. The division process of the attributes into subsets is continued until the negligible and desired output is produced. The testing process is repeated on every subset in call back manner called recursive portioning. The advantages of using decision trees are high accuracy of the results as compared to other machine learning algorithms. Decision tree can produce very accurate results in case of multi-dimensional datasets. Consider a common example of testing the gender and the age of a person to vote by means of a decision tree. The structure of the decision tree will be the similar to a tree running tests at its nodes.

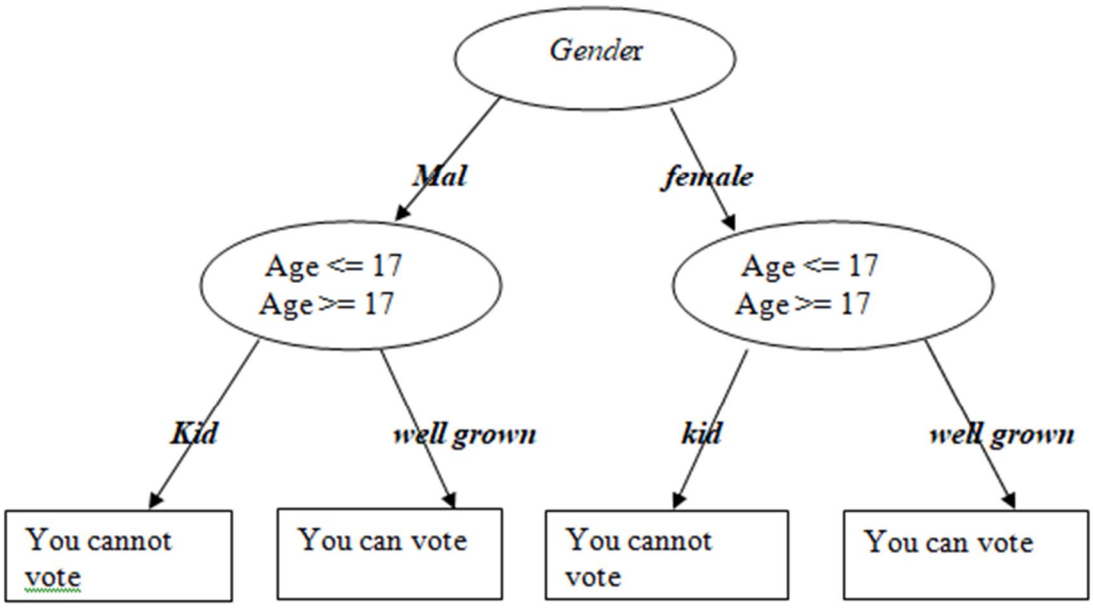



Figure 2; Decision tree

The decision tree model is used to test the age of a person as well as the gender. At the root node the decision trees tests the gender of a person, with outcomes male and female as the labels. The two intermediate nodes tests the age of the both males and females, whether they are eligible to vote or not. The labels at the leaf nodes of the decision tree are the both categories of people which can and cannot vote. [14][16][17]

*E. Random Forests*

Random forests are one of custom and modular machine learning algorithms based on the procedure of the decision tree. Random forest is a collection or hierarchy of decision trees combined to form a forest. As the forest consists of trees a random forest can be constructed by merging multiple decision trees together to produce more accurate and efficient prediction. Random forests are the collection of assembled and ensemble decision trees which could be used as a baseline architecture for predicting models. There is not any restrictions on random forests unlike other tree structures namely Binary search tree, B+ , B- and AVL trees.

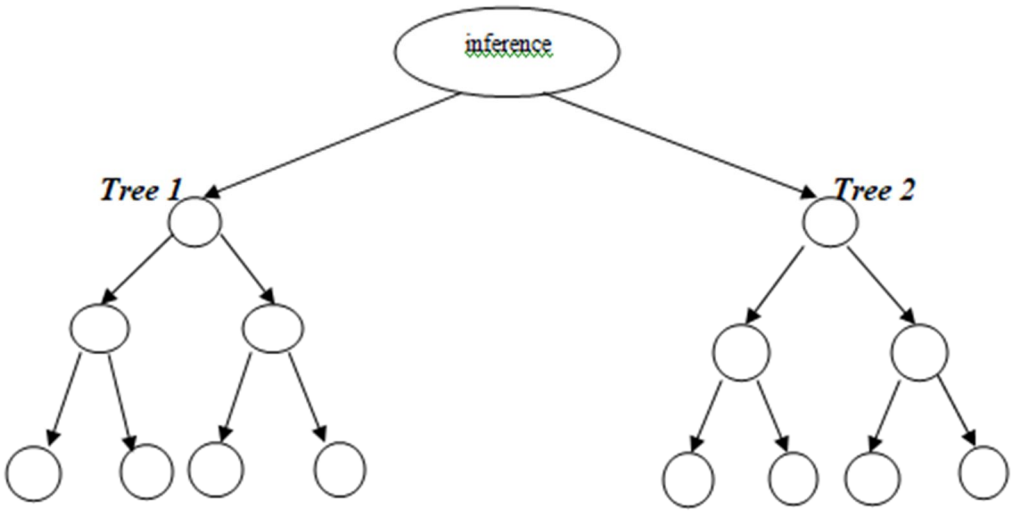



Figure 3: Random forests

Random forests can be viewed as the two or more than two decision trees combined together. Figure 3 is tree based structure of the random forest.[14][16][17]

## II. LITERATURE SURVEY

This section discusses some of the prominent work done in the field of spam detection. Many scholars have realized the need for new methods of detecting spams since social networking is rising and does not offer any mechanism to provide secure identification system.

Nikihilaet. al. paper [1] observes the techniques for reducing the logistic loss function in the spam filtering problem and carries out performance analysis of different techniques. The goal of this paper is to identify if the email is spam or not and recognizes logistic regression as one of the best technique to categorize an email as spam or not spam. Three different type of algorithm for minimization of logistic regression are studied and implemented-Stochastic Gradient Descent Algorithm, Regular Batch Gradient Descent Algorithm, and Regularized Gradient Descent Algorithm. The paper determines that it is unclear to optimally control the weight vector in Stochastic Gradient Descent algorithm which works on simulated annealing technique, whereas performance was upgraded on the test set in normal gradient descent, as it stopped overfitting in the training data.

Qinghaet. al. paper [2] carries out survey on regularly used approaches to thwart e-cheating, and demonstrate how biometrics can be used for this purpose. The author puts forward a new method to observe student activities by using their IP addresses and timestamps to contribute in observing potential cheat behaviour. The outcomes show that the proposed method is effective at recognizing student collision during exam.

Moeinsarviet. al. paper [3] runs a fuzzy expert system that is used for detection of spams. The proposed model uses several email features to prepare a fuzzy model which then results in an expert system followed by defuzzification process. The developed system was tested with sets of 1000, 2000, 3000 4000 messages and the best outcomes were achieved with the set of 3000 messages. The system is measured using Recall and Precision criteria and the best results obtained were 97.4% and 99.3% respectively.

Shadikhawandiet. al. paper [4] share their concern over image spam detection since it has been serious issue over the years and numerous solutions have been provided by different vendors. This paper focuses on the process used for preventing spams while explaining the available solutions for handling spam and image based spam. The paper concludes that the available anti-spam

methods are not sufficient as most of the mail servers count on the blacklists whereas others depend on filters that might convey high false positive rate.

Idris et. al. paper [5] familiarizes us with an email detection system that is considered as an enhancement in the negative selection algorithm (NSA). Particle swarm optimization (PSO) was applied to recover the random detector generation in the negative selection algorithm (NSA). The hybrid which is achieved by combining NSA–PSO practices a local outlier factor in terms of the fitness function designed for the detector generation. The detector generation method is then concluded as the estimated spam coverage is reached. After this step, the enhancement of the uniqueness between the non-spam and spam detectors is carried out by a distance measure and a threshold value. The analysis shows that the accuracy of the proposed hybrid NSA–PSO model is better than the accuracy of the standard NSA model. The proposed model can be used to distinguish between spam and non-spam in a network.

Authors in paper [6] introduce a spam detection technique. This technique uses text clustering method. This method gave an efficient model by realizing contents of various email and detect spam. This technique observes clusters with the help of spherical k means algorithm and clusters into two groups, that is spam and non-spam. Centroid vectors are obtained for extracting the description of clusters. For each vector in centroid, the label , whether the email is spam or not is assigned by checking the number of spam email in the cluster. Finally, the label of the most appropriate cluster is allocated to the new mail. The upshots show accuracy in this model is somewhat near to support vector machine (SVM)

Priyanka Sao et. al. paper [6] compares the performance of Naïve Bayes classifier with support vector machine. The aim of this paper is spam classification and author suggests naïve Bayesian classifier to be one of the simplest and efficient methods for the classification of spam. Results show that naïve Bayesian classifier has more accuracy than support vector machine since the error rate is very low in Naïve Bayes classifier.

The paper [7] targets to investigate available research works in spam detection approaches, the process which is being followed in these methods, and other mitigation systems. Many anti-spam strategies are surveyed for email and social networking in this paper. The author enlightens the importance of working on spam detection for the betterment of the world. This study reveals the new issues and challenges which needs to be addressed and is a big challenge for research.

Paper [8] carries out a study on different algorithms for the purpose of spam detection. These algorithms are studied under two groups, that is, content based filtering and rule based filtering. Several techniques on content based filtering have been calculated and investigated in this paper. It is concluded that rule based filtering is most efficient method to create spam filter since it reduces the filtering time.

In this paper [9], a synopsis of the spam filtering is examined and the methods researchers use for evaluation and comparison purposes of these different methods are analysed. This paper gives the gist of spam detection algorithms which come under the class of content based filtering. The results show that Bayesian classifier correctly classifies at the accuracy of 96.5%, followed by Chi square test, which gives the result at 92%, whereas KNN classifier has the accuracy of 89%.

Paper [10] KrasserSet. al. evaluates classification performance results for C4.5 decision tree and support vector machine for detecting image spam. The analyses conclude that feature extraction is considered to pose a very low computational load and the classification is partial towards a low false positive rate. About 60% of spam images can be eliminated using the techniques with a low false rate of 0.5%. Therefore the model proposed in this paper serves as an efficient first tier framework to detect large amount of spam images without doing expensive calculations.

Paper [11] analyses dataset using TANAGRA data mining tool and explores the efficient classifier for detecting email spams. Feature construction and selection is done to extract relevant features which is followed by classification algorithms and cross validation is done over this dataset. The paper approves Random Forest tree classification as the best classifier since it produces more than 99% accuracy in spam detection. This Random Forest classifier is tested with test dataset and gives accurate results than other classifiers for this spam dataset.

Paper [12] classifies the tweets into spam and non-spam using machine learning techniques. The results show that Naïve Bayes gives better results than Support Vector Machine (SVM). The data set is decomposed into training and testing , 70% for training and 30% for testing. When the data set is applied in SVM , it is first trained and then tested and got 76% accuracy. And then same data set is tested with Naïve Bayes where it is trained and tested and got 92%. The results may vary based on the feature selection.

Paper [13] defines the summary of diagnoses and predicting the factors of multiple diseases in aged and elderly persons. The survey includes the summary of disease called pneumonia using multiple logistic regression algorithms. The authors has also analysed the comparison of pneumonia disease targets differentiating the risk factor of losing lives among younger and elderly people.

Paper [14] differentiates the multiple machine learning and deep leaning functions which could be used for development and improvement in cyber security. Using various machine learning and deep learning the autonomous and smart algorithms could be

developed for the betterment of cyber activities. The author in the survey report points out the similarities and dissimilarities between machine learning and deep learning algorithms which support vector machine, K nearest neighbour, decision trees, deep belief networks, recurrent neural networks and convolutional neural networks. The survey includes a network data set for analysis and decision tree based approach.

Paper [15] sights the survey on the one most crucial disease in the females worldwide. It is the most popular cancer and one of the major concerns regarding the health of the women. The disease is breast cancer which is discussed by the authors in the survey report. Also the authors have studied the effects of the residential areas and society on the breast cancer. Also the authors introduced the difference in the pattern between a urban area and the residence different from the metropolitan cities

## III.TOOLS AND PLATFORMS AVAILABLE FOR IMPLEMENTING MACHINE ALGORITHMS

Machine learning is one of the most dominant and promising field in modern computing age for solving real world problems. The number of tools, platforms and programing languages are multiplied in the field of machine learning and available at ones disposal in solving these problems. The commonly used machine learning tools and platforms in the world are:-

1) *Python:* Python is one of the most popular programing languages used in modern computing age. It provides support to various developing domains like web applications, application software and scripting. Python is used in every scenario when it comes to the development and deployment of server side and client side applications. It also provides support to the dynamic programming contents as well as user interface for application development with the help of various libraries and frameworks. Python provides support to various platforms and libraries for problem solving. [16]
2) *Python Jupyter:* Python Jupyter or Jupyter Notebook is an extension to the python. The difference between JupyterNotebook and python is that the text box is used in earlier python versions, but Jupyter Notebook provides a separate command window and browser enabled window for compiling and executing code. Another major difference is that Jupyter Notebook enables user to run code in blocks. It also saves the projects and programmes in .pynb extension dissimilar than python extension which is .py. it is an open source software which supports healthy number of libraries related to the different fields of computing world namely cryptography, web based application and designing. The code in Jupyter Notebook is written is separate blocks and results can be seen after running each block. It use python for running and executing code. Python Jupyter is one of the most popular platforms used for machine learning by scientists, scholars and students. The code written in Jupyter Notebook is less complex and fewer lines when compared to other platforms. There are many versions of python available in the market with and without the support of Jupyter Notebook and Anaconda.[16]
3) *Google ML Kit:* Google ML kit is a platform developed by the Google for mobile development. It includes a variety of machine learning model and approaches for mobile application in the modern world. The features of the mobile applications can be extended to the higher levels using Google ML kit.[16]
4) *Open NN*: Open Neural Networks is a library for implementing neural networks using C++ programing language. It supports deep learning approaches for implementing neural network structures for the projects and application development. Its architecture is based on neural networks and uses strong syntax as well object oriented concepts for classification, regression, forecasting and association based learning. In general Open NN is strong and powerful library for solving problems using machine learning and data learning algorithms.[16]
5) *Accord NET:* Accord .NET is an application building platform similar to the Microsoft .NET. Microsoft .NET is basically an application development kit for creating graphical user interface (GUI) based applications and software. Accord .NET also supports the layout of .NET which is user interface for start-ups and other business organisations. Accord .NET not only supports graphical user interface architecture but also manages the word and image processing dynamics. One of the key factors of using Accord .NET is that it is written in the one of the powerful and dynamic programing language which is C# in collaboration with .NET. .NET is the one of most leading service by the Microsoft for mobile and computer application development, that is the reason why Accord .NET is a preferable tool used by industries and individuals. [16]
6) *Amazon ML:* Amazon machine learning is a cloud service provided by the Amazon for building and creating predictive machine learning models. One of the main causes of employing Amazon ML service a user can create and build machine learning models without having a prior knowledge about the machine learning model. One of the different advantages of using Amazon ML is that if any user at any point of time want to upgrade or enhance the features of the application Amazon ML provides the privilege to such options like connectivity of the applications using APIs and other customization.[16]
7) *Azure ML work Band[16]:* Azure machine learning work band is a platform provided by the Microsoft in collaboration with python programming language. Azure ML work band is basically a cloud service by the Microsoft, which merges machine

learning as one of its key features for application development, deployment, modelling and publication. Some main services provided by Azure ML work band are software as a service, platform as a service and infrastructure as a service. Some of the main features of Azure ML work band are:-

*a)* Computing resources and services.
*b)* Multimedia services.
*c)* Textual services like messaging.
*d)* IOT (internet of things)
*e)* Machine learning
*f)* Data centre services
*g)* Application development and services

8) *Google Tensor Flow*: Google Tensor flow is a mathematical library which provides support to many machine learning concepts like artificial neural networks. It is an open source library developed by the Google brain in context for the support to the newly developed technologies. Modern day technologies deal with creation and processing of multi-dimensional and large datasets which are real time and unstructured. Some common technologies like big-data analytics are one of the main concerns of the Google Tensor Flow. It provides utilization of various processing schemes like multi CPUs and multi GPUs for processing and analysing real and unstructured datasets. It also provides support to all mobile, computing and clustering schemes. Google Tensor Flow is also compatible with all the operating systems in the market which involves IOS, Microsoft, LINUX, android as well as mobile computing.

9) *Big ML*: Big ML[16] is a company which provides privileges to the researchers and inventors to work with the machine learning concepts and projects. In Big ML platform a user can avail two components of the project. The first one is the production mode which is free of cost and the second part is the development mode which is not free for use. For development mode a user has to pay for enabling different machine learning services. There are many specialized fields provided by Big ML to make things simpler and easier for a user which are :-

*a)* Web based interface
*b)* Command line interface
*c)* API

## IV. ANALYSIS

In this survey report the authors have utilized the logistic regression algorithm which is used for multi-dimensional, complex and multinomial data sets to demonstrate the difference in the patterns. The table given below demonstrates the accuracy of the algorithms in percentage.

Table I. Accuracy of different algorithms

| Algorithm used for classification | Used by the researchers | Definition | Accuracy of classifying objects |
|---|---|---|---|
| Chi square function | Malarvizhi, R | Statistical theory for distribution. | 92% |
| Naïve Bayes classifier | Malarvizhi, R | It uses probabilistic function. | 96.5% |
| Fuzzy systems | Sarvi, M., Mohamadi, M. and Varjani | It uses fuzzy logic and intermediate values. | 94% |
| Decision tree C4.5 | Krasser, S., Tang, Y., Gould, J., Alperovitch, D., & Judge, P | Decision trees run tests at its nodes and the outcome are the labels. | 60% |
| Random Forests | R. Kishore Kumar, G. Poonkuzhali, P. Sudhakar | Collection and gathering of decision trees together. | 99% |
| KNN | Malarvizhi, R | Termed as k-nearest neighbour | 89% |

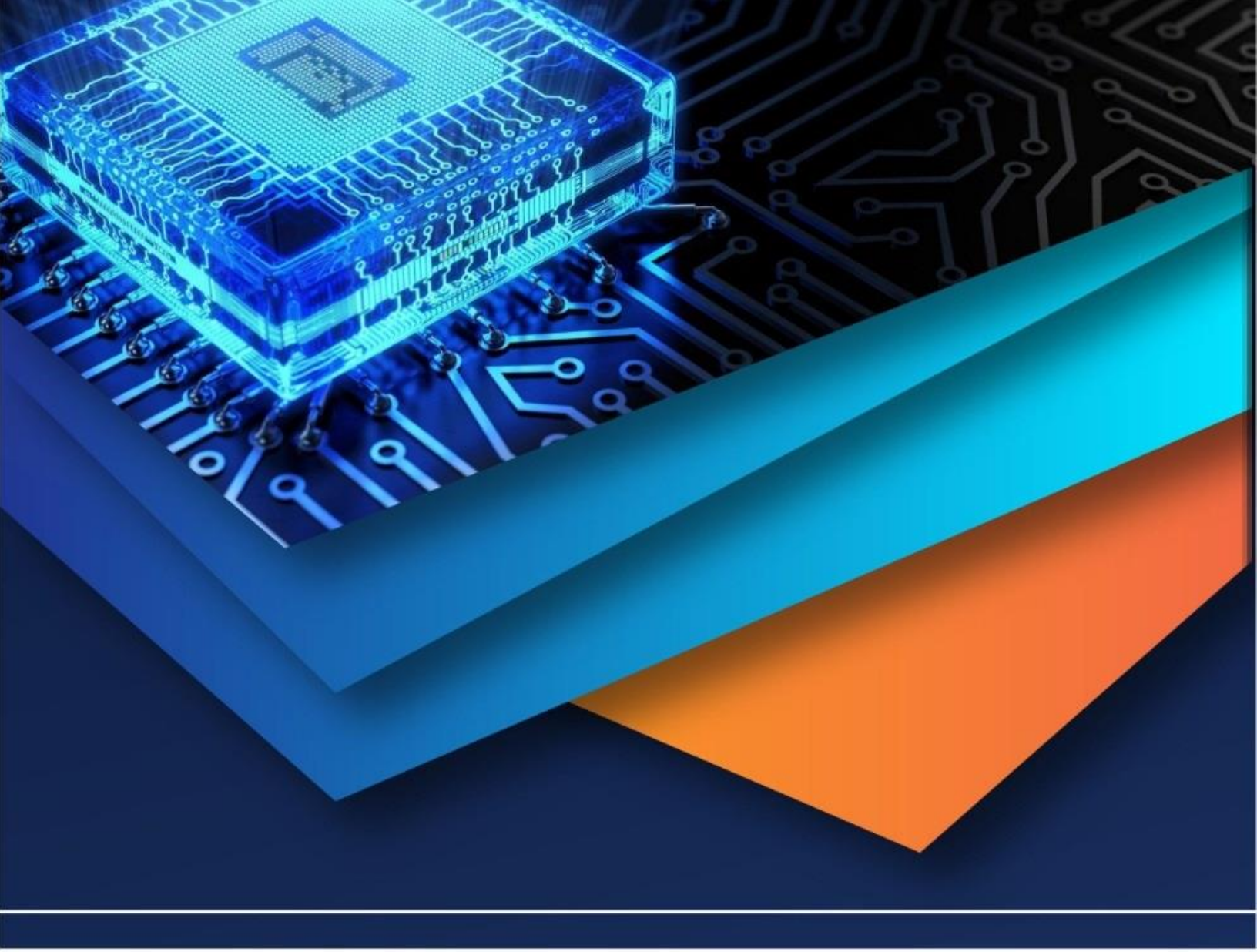
The table provides the overview about the variety of machine learning and deep learning algorithms used for the classification and regression. It also provides the comparison in accuracy for using certain algorithms on some datasets. On the bases of previous survey the outcomes of the decision trees and random forests are more accurate as compared to the other algorithms.

## V. CONCLUSION

The paper focuses on the comprehensive and measured analysis by the various machine learning and the deep learning algorithms for the classification and regression of datasets. In this survey the goal is to find out the most accurate and approximate algorithm in a certain platform for classifying spam and non-spam emails by comparing and contrasts the algorithms with each other. The literature survey in this domain has evolved using many machine learning and deep learning algorithms for the classification of datasets. The goal of the survey report is to compare all there algorithms namely SVM, naïve Bayes classifier, ANN, fuzzy systems, logistic regressionand chi fi function.

## REFERENCES

[1] Kamoru, B.A., Jaafar, A.B., Murad, M.A.A., Ernest, E.O. and Jabar, M.B.A., Spam Detection approaches and strategies: A phenomenon.
[2] Gao, Q., 2012. Using IP addresses as assisting tools to identify collusions. International Journal of Business, Humanities and Technology, 2(1), pp.70-75.
[3] Sarvi, M., Mohamadi, M. and Varjani, A.Y., 2013, August. A fuzzy expert system approach for spam detection. In 2013 13th Iranian Conference on Fuzzy Systems (IFSC) (pp. 1-5). IEEE.
[4] Khawandi, S., Abdallah, F. and Ismail, A., ASurvey ON IMAGE SPAM DETECTION TECHNIQUES. Computer Science & Information Technology, p.13.
[5] Idris, I., Selamat, A., Thanh Nguyen, N., Omatu, S., Krejcar, O., Kuca, K., &Penhaker, M. (2015). A combined negative selection algorithm–particle swarm optimization for an email spam detection system. Engineering Applications of Artificial Intelligence, 39, 33–44
[6] Sasaki, M., &Shinnou, H. (2005). Spam detection using text clustering. 2005 International Conference on Cyberworlds (CW'05).
[7] Kamoru, B.A., Jaafar, A.B., Murad, M.A.A., Ernest, E.O. and Jabar, M.B.A., Spam Detection approaches and strategies: A phenomenon.
[8] Puri, S., Gosain, D., Ahuja, M., Kathuria, I. and Jatana, N., 2013. Comparison and analysis of spam detection algorithms. International Journal of Application or Innovation in Engineering and Management, 2(4).
[9] Malarvizhi, R., 2013. Content-based spam filtering and detection algorithms-an efficient analysis & comparison 1.
[10] Krasser, S., Tang, Y., Gould, J., Alperovitch, D., & Judge, P. (2007). Identifying Image Spam based on Header and File Properties using C4.5 Decision Trees and Support Vector Machine Learning. 2007 IEEE SMC Information Assurance and Security Workshop.
[11] R. Kishore Kumar, G. Poonkuzhali, P. Sudhakar, Member, IAENG, Comparative Study on Email Spam Classifier using Data Mining Techniques, Proceedings of the International MultiConference of Engineers and Computer Scientists 2012 Vol I , IMECS 2012,March 14-16,2012,Hong Kong
[12] VidyaKumari, K. R., &Kavitha, C. R. (2018). Spam Detection Using Machine Learning in R. Lecture Notes on Data Engineering and Communications Technologies, 55–64. doi:10.1007/978-981-10-8681-6_7
[13] Taooka, Y., Takezawa, G., Ohe, M., Sutani, A. and Isobe, T., 2014. Multiple logistic regression analysis of risk factors in elderly pneumonia patients: QTc interval prolongation as a prognostic factor. Multidisciplinary respiratory medicine, 9(1), p.59.
[14] Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H. and Wang, C., 2018. Machine learning and deep learning methods for cybersecurity. IEEE Access, 6, pp.35365-35381.
[15] Breast, S.E.S.O., A Multinomial Logistic Regression Analysis to Study The Influence Of Residence And Socio-Economic Status On Breast Cancer Incidences In Southern Karnataka
[16] www.google.com
[17] www.wikipedia.com



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY