Performance Of Various Computational Intelligence Methods In IDS: An Analytical Perspective

Ch. Raviksihore, P. Sankara Rao, K. Eswara Rao

Abstract: An Intrusion is an unauthorized access to the network and retrieves the secret and confidential information and availability on the computer and network resources. Various solutions are available to overcome this type of access and to identify intrusions; one of the methods is Intrusion detection system. An Intrusion Detection System is efficient whenever its accuracy and detection rate is very high as well as false alarm rate is very less percentage. This paper focuses on apply various methods of classification on data set and calculating performance and identifies the accuracy rate and false alarm rate. We should understand that a single efficient algorithm may not be suitable for all individual attacks. In connection to these results, the authors of this paper focused on the various classifications of algorithms that will be performed well based on kind of attack.

Keywords: KDD Cup, IDS, DoS, R2L, Probe, U2R, Weka, Computational Methods.

1 INTRODUCTION

An huge Internet usage is need for the enormous application running in the network for sharing resources and information, to lead this most big challenge and very important to secure resources and protect from malfunctions like attacks. In the sense an intrusion detection system sometime called like IDS software plays a vital role to protect and detect computer activities and evading computer security policies or standard which can occur in different situations. An intrusion detection system is a technique is to prepare a prophetic system which is capable to identifying those that are “abnormal or “intrusive”. In the present trend network connected altogether various areas and also Internet in all fields for the purpose of sharing the information. It leads to give a chance to an intrusion attack can be perpetrated by an insider or by an outsider [22]. In the “inner attack”, the attack is initiated by an entity inside the security perimeter; the person who has complete authorization involves in the vulnerable activities, i.e., the attacker tries to access some system resources crookedly for which he does not have any authorization. It is very quizzical to find out these types of intruders. An “outer attack” is initiated from the outside that is by an unauthorized or illegitimate user of the system. A computer network contains two components namely hardware and software. These two components may have their own risks and vulnerabilities. Hardware perils are easy to detect and also those cause harm only to the device rather than the data. If the attack is in software, mainly it harms the data.

- *Ch. Raviksihore, Sr. Assistant Professor, Dept. Of CSE, AITAM, TEKKALI, SRIKAKULAM, AP, INDIA-532201. Email: cauchy9@gmail.com*
- *K. Eswara Rao, Sr. Assistant Professor, Dept. Of CSE, AITAM, TEKKALI, SRIKAKULAM, AP, INDIA-532201. Email: eswarkoppala@gmail.com,*
- *P. Sankara Rao, Assistant Professor, Dept.of.CSE, GITAM University, Visakhapatnam, AP, INDIA. 530045. Email: dr.sankararaop@gmail.com*

An Intrusion detection system which have been concentrates on only anomaly based and misuse-based category type of techniques from various types of other categories since past research. For hardware type of intrusions (attacks) misuse-based detection is very supportive and its detection rate is very high and predictability accuracy is also high, and anomaly category based which supportive for investigation upon a scientific inquiry which works in specifics area. Anomaly based is has proven for unreliable accuracy and high detection and less false rate based on applying various machine learning algorithms. But still less concentrate of applying anomaly based techniques on hardware attacks or commercial areas. By studies of above, the authors of the this paper observed several procedures and different machine learning algorithms which need to approach to high detection rate and usage of testing and training datasets, methods for evolution. Their researches express some drawbacks [6] of KDDCup dataset [2] which repeatedly used on various anomaly detection based research. No classification algorithm is best but still each algorithm perform best on some data set so in this study we see how the performance of an algorithm on data set with all attack and we identify which classification algorithm perform best on individual attacks by using weka [1]. This paper alienated into five sections in section 2 Literature review which focuses what works has been done on IDS, section 3 gives a short view about intrusion detection system, section 4 discuses about various classification algorithm under the study, section 5 presents proposed model, section 6 discusses about the experimental results and section 7 presents conclusion and future scope.

2 LITERATURE REVIEW

Before start our research we flow journey on various researches which is done by the several fact finding works on intrusion detection area, few analysis on this. Initially, In 2017 Bambang and Djanali [4] has been studied more number of research papers on Intrusion detection. They

study clearly specifies research differences between hybrid machine learning and centroid based learning by considering several things like datasets are used, features, preprocessing data and evolution methods considered. And also propose research moves through data preprocessing can gets more accuracy comparatively existing research because of imbalance of KDD dataset. In 2013 K. Wankhade[11] implements an IDS system based on misused and deviates from intrusions, it's practically proves very high accuracy reported from known patterns and includes misused and deviate intrusions. While consider misused and deviate attacks possible detection rate is good but at the same time false alarm rate is very high. These research report good for false alarm rate. In 2011 P. Amudha et.al .[20], proposed a model of Intrusion Detection system with takes support of various types of classification algorithms named J48, NaiveBayes, Nbtree and Random forest also takes input from KDD dataset implements calculates performance of each classifier based on correlation feature selection(CFS) measure. Their research observes detection rate of Probe and DoS attacks is good for Random forest comparatively other classifications, and also naive bayes classifier is good for U2R, R2L category attacks. From another aspect Naive Bayes execution time is good comparatively from other type of classifiers. Agarwal and Joshi [3] is one of the fact finders to get well accuracy detection rate and also shows how individual attacks get how much accuracy rate acquired with his practice approached by their rule-based model (PNrule) build by two stage structures. His flow of research PNrule concentrates classifier models and evaluates different class distribution in the KDD dataset was not mention in actual KDD dataset like Local to Remote attacks. His ground work reports effective accuracy and also identifies 96.9%, 73.2%,6.6%,10.7% accuracy rate for DoS,Probe,U2R, R2L attacks, and detects only 10% false alarm rate for all types of attacks excludes only U2RAnd Xin Xu [10] in 2006 build a temporary machine learning adaptive detection system implemented by a classifier like multi class support vector machine. Support of randomly selected records from original KDD dataset and gets the Results on these type of design is very satisfactory and analysis of results on various category type of attacks detection rate of 76.7%, 81.2%, 21.4% and 11.2% for DoS, Probe, U2R, and R2L. Less difference between these categories'. Results are good but it's a complicated to understand.

3 BASIC PRELIMINARIES AND BACKGROUND STUDY ABOUT IDS

An Intrusion Detection System is concentrates on detection of nasty actions. These detection system mainly classified into anomaly based detection and signature detection approaches. Both are which is used to protect system from malicious attacks. Systems monitoring incoming connections and examine traffic of the network and identify suspicious connections or packets. Because an intruder may attack through like this. Whenever a detection system is says best is depends on how much rate of connection detected, and also best detection system also classified whenever it's found denied connections. Mainly these types of attacks fall in one of the following category [6].

3.1 Denial of Service (DoS) Attack

Denial of Service is an attack where an intruder tries to access the services of system illegally. These type of attacks identified whenever an intruder sends an inundate message or mail to the network or servers to properly with invalid return addresses. In this situation the network fails to identify return address of the intruder before authorization of inundate message or mail from attacker. It leads to network wait for some moment to close the intruder request, at this stage they can send more number of valid messages with invalid return addresses. In the sense the server or network waiting keeping server is busy. Detection system may identify these type of malfunctions by "Apache2, arppoison, Back, Crashiis, dosnuke, Land, Mailbomb, SYN Flood, Ping of Death, Process Table, selfping, Smurf, sshprocesstable, Syslogd, tcpreset, Teardrop, Udpstorm" attacks.

3.2 User to Root (U2R) Attack

U2R is an attack which an intruder starts attack to a system from normal user account. Through valid account like passwords, device programs, system related programs, etc.. it take a chance to enter susceptibility into system and access the system privileges to gain main functioning in the system. The U2R attacks categorized into "anypw, casesen, Eject, Ffbconfig, Fdformat, Loadmodule, ntfsdos, Perl, Ps, sechole, Xterm, yaga".

3.3 Remote to Local Attack (R2L)

These types of attacks occurs whenever an intruder sends a request from remote system to local system and gains unauthorized access from remote machine and gets rights. These types of attacks come under "Dictionary, Ftpwrite, Guest, Httptunnel, Imap, Named, ncftp, netbus, netcat, Phf, ppmacro, Sendmail, sshtrojan, Xlock,Xsnoop".

3.4 Probing Attack

This type of attacks raises initially scans network system and target short IP addresses. An unused port number mark and an attack by using short IP to enable gather information about a network of computers for the obvious purpose of getting around its security controls and this attack is followed by "insidesniffer, Ipsweep, ls_domain, Mscan, NTinfoscan, Nmap, queso, resetscan, Saint, Satan".

4 METHODS OF COMPUTER INTELLIGENCE

There are many mining intelligences for intrusion detection such as pattern mining, support vector machine (SVM), classification, clustering, etc. Let us see some of them here.

4.1 Classification

Classification is a mining intelligence which predicts categorical labels, is a method takes number of instances in a collection simply called dataset considered and assign it to a known model. A classifier classifies data and predicts particular class label normal or abnormal. Classifiers are suitable for best in misuse and little in anomaly detection methods. Datasets are converted into predetermined sets by applying classification. Different classification techniques

such as LBk, Naivebayes classifier, J48, Random forest, Hoeffding Tree etc. are used in IDS.

4.1.1 LBk

K-nearest neighbor classifier is LBK [17] an intelligence technique in weka tool. This intelligence method uses similar distance metric. In object editor, the amount of nearest neighbor can be defined clearly. It can be automatically determined by using leave-one-out cross-validation emphasize to an upper bound given by the particular value. Dissimilar searching algorithms are employed to find the nearest neighbors at high speed. Linear search is used as a default search, but there are further options containing "KD-trees," "ball trees" and "cover trees" [19]. As a parameter, distance function may be employed. The behind thing is similar as "IBL." That is a Euclidean distance; further choices are "Chebyshev," "Manhattan," and "Minkowski distances." Predictions/Guesses can be weighted, from one or more neighbor, following their distance from the test examples. Distance is converted into weights by applying two dissimilar formulas. Training examples that are held by the classifier can be limited by "window size" choice. When novel examples are included, previous examples are deleted to retain the training examples at this size [18].

4.1.2 Hoeffding Tree

Most streaming algorithm suffers with two major problems first one the order in which input appears second slower in batch processing .to overcome these problems in the construction of decision tree using Hoeffding [23] follows incremental fashion and assumes the distribution of data not change over time . and splitting of node is only possible when the sufficient statistical evidence exist with attribute associated with that node. Decision tree learned by Hoeffding is asymptotically identical to non-incremental learner if training data is large enough.

4.1.3 J48 Model

The j48 [15] algorithm is an easy Java program of a C4.5 decision tree for classification. The developer of the C4.5 algorithm is Ross Quinlan. It is employed to produce decision tree. In classification problems, decision tree algorithm is very useful. To build a simple model process like tree is constructed, by using the J48 algorithm. When a tree is build, it is in use to each row of data- set. J48 algorithm avoids missing values (item value can be forecasted by what is known about the attributes values for other records) during tree construction. Core intention is that data is partitioning into range. It relies on value for that entry that is in training example data. J48 algorithm permits classification through "DT" or rules produced from them [13], [12].

4.1.4 NaiveBayes

The Naive Bayes [14] classifier is a probabilistic classifier based theorem. Naïve Bayes algorithm computes probability of each class and out puts a class with max probability .For the computation complexity Naïve Bayes assumes that all attributes are independent of each other for a given class C1, C2, C3… CN are different class and X is unseen input then the input assigned class label Cj is $P(x/Cj)>P(X/Ci)$ for $i\neq j$ for i=1,2…….N

4.1.5 Random Forest

Random forest [5] algorithm is a supervised classification algorithm. It follows the idea of construction of decision tree instead it creates a collection of decision trees .It assigns the class label for unseen input based on the class labels obtained from decision trees by taking majority voting.

Algorithm for Random forest

1. Randomly select "t" attributes from total "n" attributes where t<< m
2. Among the "t" attributes, calculate the node "d" using the best split point
3. Split the node into d nodes using the best split
4. Repeat the 1 to 3 steps until "l" number of nodes has been reached
5. Build forest by repeating steps 1 to 5 for "k" number times to create "k" number of trees

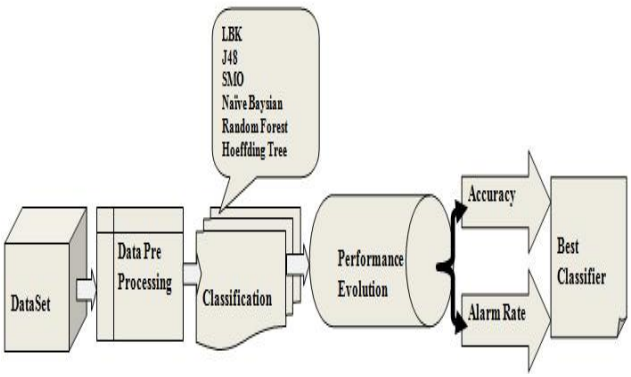
4.1.6 SMO

Sequential Minimal Optimization (SMO) [16] is used to train support vector machine. The SMO algorithm contains many optimizations designed to speed up on large datasets.

5 Proposed Framework

Different classification algorithms are focused in the area of intrusion detection system to verify the efficiency between the methods. An effective model was implemented step by step shown in Figure1. In this study we consider KDD cup data set which contains samples of different attacks. In pre processing phase we sample some data and also prepared data set of individual attacks from the original dataset. This experimental study focuses on six frequently used best methods named J48, LBK, SMO, Naïve Bayes, Hoeffding tree and Random forest. For performance evaluation this study considers two measures accuracy and failure rate. Based on performance evolution measures compare the accuracy for all six classifier methods and proposes which is best classifier and also propose which classification algorithm is best one for different types of attacks, in this study each classification technique is applied on each category of attack and identify the accuracy and error rate of each one.

Figure 1: *Experimental evolution*



5.1 Discussion and Evaluation

Since 1999 The KDD Cup Dataset has been using for various research works mainly in data mining research. Actually Original KDDCup dataset has contained huge set of well instances which have nearly 5 lacks of records collected from practically ruled on the network traffic. In these dataset contain millions of records each record called a connection it maintained by serial TCP message transfer from source node to destination node. While message transfer from source to destination each connection has 41 features identified by Data extraction is based on some basis like initially randomly extracted 10% of instances (Table 1: (a)) of each label next extraction (Table1: (b)) done for only 49596 instances which also 10 % from first extraction.

Table1 (b) shows normal instances 9841, DoS labeled instances 39092, Probe labeled instances 437, U2R labeled instances 13 and R2L labeled instances 213. By using these two datasets present in Table 1 uses effectively evaluate the accuracy of different intelligence methods.

***Table 1**: Randomly selected sample dataset twice from KDDCup*

Class	a)10% of training KDD dataset		b) Actual Training Data for Classification	
	Number of Records	% of Occurrence	Number of Records	% of Occurrence
Normal	97277	19.691	9841	19.842
DoS	391458	79.240	39092	78.821
U2R	52	0.011	13	0.026
R2L	1120	0.227	213	0.429
Probe	4107	0.831	437	0.881
Total	494014	100	49596	100

5.2 System Setup

Every part of experiment be execute in high build up system with the Intel(R) Core i3 Processor, 3 Gigabytes RAM and an independent system contains windows 7 operation system. For experimental execution purpose takes support of an open source machine learning package tool Weka version 3.6. Weka is a software tool contains collection of machine learning models for mining tasks like data preprocessing, classifiers for classification, clustering, association rules and visualization. All the intelligence methods that were used in this paper are implemented in Weka so that everything be easy and every result is fair compared to each other

6 DESCRIPTIONS ABOUT THE RESULT

From Table 2 and Figure 2 we conclude that all most all classification performing well on overall data set under the study except naive base algorithm which behaved poorly.

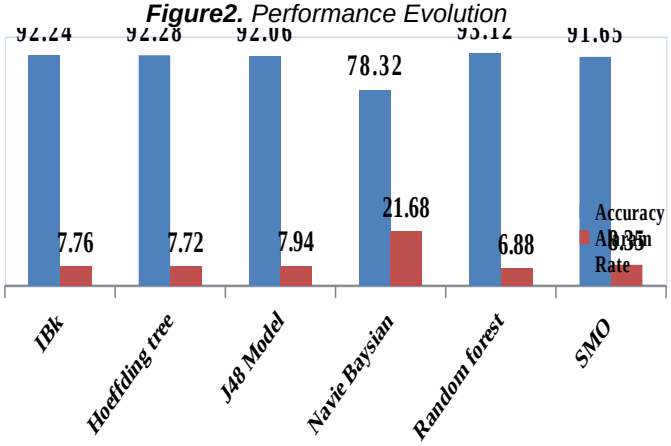
***Table 2** Performance Evolution*

Classification Model	Accuracy	Alarm Rate
LBk	92.24	7.76
Hoeffding tree	92.28	7.72
J48 Model	92.06	7.94
NaiveBaysian	78.32	21.68
Random forest	93.12	6.88
SMO	91.65	8.35

6.1 Analytical Description about the Result

From Table-3 and Figure-3 it is evident that for Dos attack Hoeffding tree gives classification accuracy of 99.201 which very high compare to reaming algorithms under the study. From Table-3 and Figure-4 it is evident that for U2R and LBk gives classification accuracy of 22.361 which very high compare to reaming algorithms under the study

***Figure2.** Performance Evolution*



From Table-3 and Figure-5 it is evident that for R2L and LBK gives classification accuracy of 7.81 which very high compare to reaming algorithms under the study.

Figure 3 *DoS Category Attacks*

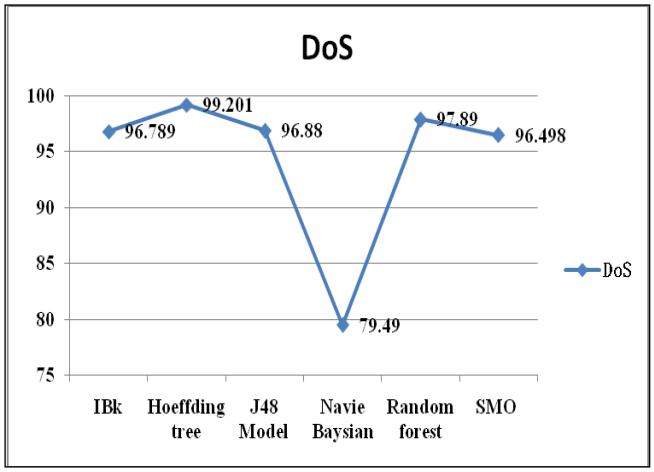


Figure 4 *U2R Category Attacks*

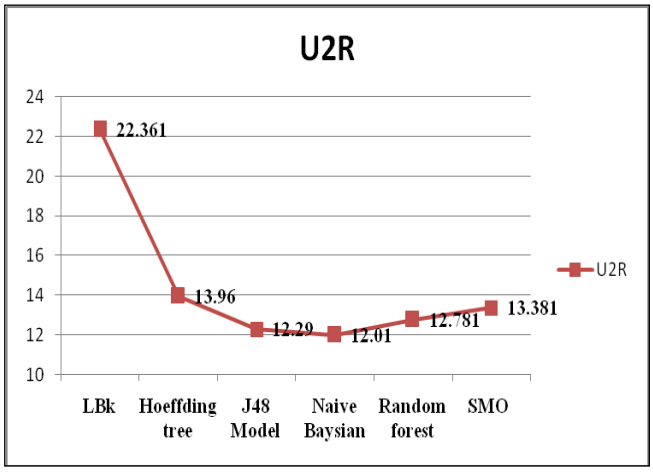


Table 3 *Intelligence Evolution for Intrusion Attacks*

Classification Model	Attack Type				Accuracy
	DoS	U2R	R2L	Probe	
LBk	96.789	22.361	7.81	72.49	92.24
Hoeffding tree	99.201	13.96	0.1	74.8	92.28
J48 Model	96.88	12.29	0.1	75.28	92.06
Naive Bayesian	79.49	12.01	0.1	94.843	78.32
Random forest	97.89	12.781	0.14	75.43	93.12
SMO	96.498	13.381	0.12	74.3	91.65
Average	94.458	14.464	1.4	77.857	89.945

Figure 5 *R2L Category Attacks*

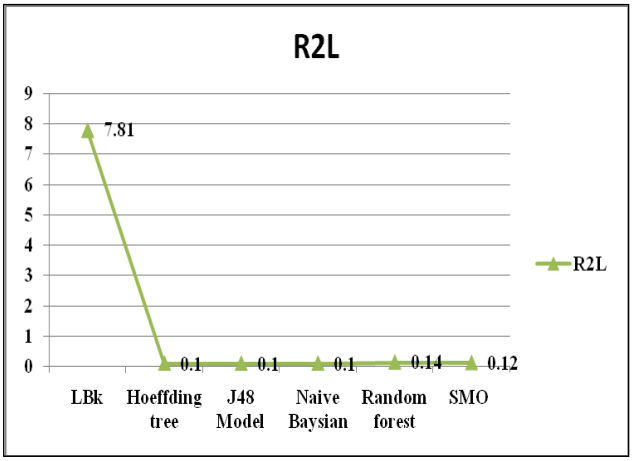
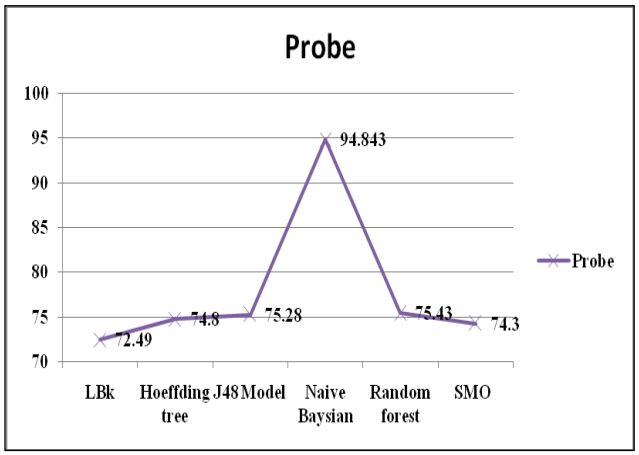


Figure 6 *Probe Category Attacks*



From table 3 and figure 6 it is evident that for Probe Naivebayes gives classification accuracy of 94.843 which very high compare to reaming algorithms under the study. Even though it gives poor performance on overall dataset it gives best performance on probe attacks.

6 CONCLUSION WITH THE FUTURE WORK

Best motivation of write this paper is due to huge research on Intrusion detection system using data mining classification, most of the researchers used common dataset like KDDCup. Weka also supposes to reduce the complexities of execution for various types of data mining methods. Took reports from weka and compare the performance among classifier algorithms, data set with all attacks all most all algorithms perform equally except Naïve Bayes so we further continued our study of classification algorithms how these perform on individual attacks . This research feel as baseline and in future research going on IDS with merged classification algorithms, and we would like to move my research on another way is get same performance by using less number of attributes in KDD dataset.

REFERENCES

[1] Weka – Data Mining Machine Learning Software. http://www.cs.waikato.ac.nz/ml/weka/

[2] KDD Cup 1999 Data. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[3] Agarwal, R., Joshi, M.V.: PNrule: A New Framework for Learning Classifier Models in data Mining. Tech. Report, Dept. of Computer Science, University of Minnesota (2000)

[4] Setiawan, Bambang, Supeno Djanali, and Tohari Ahmad. "A Study on Intrusion Detection Using Centroid-Based Classification." Procedia Computer Science 124 (2017):672-681.

[5] J. Zhang, M. Zulkernine and A. Haque, "Random-Forests-Based Network Intrusion detection Systems," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 5, pp. 649-659, Sept. 2008.doi:10.1109/TSMCC.2008.923876.

[6] Tavallaee, Mahbod, et al. "A detailed analysis of the KDD CUP 99 dataset." Computational Intelligence for Security and Defense Applications, 2009.CISDA 2009. IEEE Symposium on IEEE, 2009.

[7] Sabhnani, M., Serpen, G.: Why Machine Learning Algorithms Fail in Misuse Detection on KDD Intrusion Detection Dataset. In: Intelligent Data Analysis, vol.6 (June 2004) Extra Data

[8] Bolon-Canedo, Veronica, Noelia Sanchez-Marono, and Amparo Alonso-Betanzos. "Feature selection and classification in multiple class datasets: An application to KDD Cup 99 Dataset ".Expert Systems with Applications 38.5 (2011): 5947-5957.

[9] Nguyen H.A., Choi D. (2008) Application of Data Mining to Network Intrusion Detection Classifier Selection Model. In: Ma Y., Choi D., Ata S. (eds) Challenges for Next Generation Network Operations and Service Management. APNOMS 2008. Science, vol 5297. Springer, Berlin, Heidelberg

[10] Xu, X.: Adaptive Intrusion Detection Based on Machine Learning: Feature Extraction Classifier Construction and Sequential Pattern Prediction. International Journal of Web Services Practices 2(1-2), 49–58 (2006).

[11] K. Wankhade, S. Patka and R. Thool, "An efficient approach for Intrusion Detection using data mining methods," 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Mysore,2013, pp. 1615-1618.doi: 10.1109/ICACCI.2013.6637422

[12] Patil, Tina R.,and S. S. Sherekar. "Performance analysis of Naïve Bayes and J48 classification algorithm for data classification." International Journal of Computer Science and Applications 6.2 (2013): 256-261.

[13] Sharma, Aman Kumar, and SuruchiSahni. "A comparative study of classification algorithms for spam email data analysis." International Journal on Computer Science and Engineering 3.5 (2011): 1890-1895.

[14] John, G.H., Langley, P.: Estimating Continuous Distributions in Bayesian Classifiers. In: Proc. of the 11th Conf. on Uncertainty in Artificial Intelligence (1995)

[15] Quinlan, J.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)

[16] Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)

[17] Aksoy, S.: k-Nearest Neighbor Classifier and Distance Functions. Technical Report, Department of Computer Engineering, Bilkent University (February 2008).

[18] Vijayarani, S., and M. Muthulakshmi. "Comparative analysis of bayes and lazy Classification algorithms." International Journal of Advanced Research in Computer and Communication Engineering 2.8 (2013): 3118-3124

[19] Jacob Goldberger, SamRoweis, and RuslanSalakhutdinovGeoffHinton."Neighbourhood components analysis." NIPS'04 (2004).

[20] P. Amudha and H. Abdul Rauf, "Performance Analysis of Data Mining Approaches in Intrusion Detection," 2011 International Conference on Process Automation, Control and Computing, Coimbatore, 2011, pp. 1-6.

[21] Kevric, J., Jukic, S. & Subasi, A. Neural Comput & Applic (2017) 28(Suppl 1): 1051.https://doi.org/10.1007/s00521-016-2418-1

[22] S. Latha and S. J. Prakash, "A survey on network attacks and Intrusion detection systems," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, 2017, pp. 1-7

[23] Yin, Chunyong & Feng, Lu & Ma, Luyu. (2015). An improved Hoeffding-ID data-stream classification algorithm. The Journal of Supercomputing. 72. .10.1007/s11227-015-1573-y.