International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 12– No.23, August 2019– www.ijais.org

Performance Analysis of Machine Learning Techniques for Intrusion Detection

Aftab Ahmad Malik, PhD
Department of Computer Science,
Lahore Garrison University

Muhammad Bilal Butt
Department of Computer Science,
Lahore Garrison University

Rabia Aslam Khan
Department of Computer Science,
Lahore Garrison University

ABSTRACT
During the recent years, there has been tremendous development in the area of Computer Networks. This paper deals with the important area that is performance analysis of techniques used in machine learning. One of the major problems in Network Security is "intrusion detection system", which is software, remains active during processing. The intrusion detection system helps in monitoring computers and computer networks, vulnerabilities or malicious activities. The attacks or malicious activities censor information and then corrupt the system networking protocols. In this paper, different machine learning techniques and their performance are compared and discussed. How machine learning techniques can ideally help in developing efficient "Intrusion detection system".

General Term
KNN Algorithm

The K-nearest algorithm is very useful in pattern recognition. While handling the regression and classification, the K closest training examples from' feature space' are used as input. The Algorithm is applicable in case of S.O.M self-organizing map, where every node serves as the center of cluster of similar points.in every field for sharing data, accessing, manipulating data, business purpose and many other. As internet is used in everyfield and important data is exchanged and shared over the internet. Hence, data over the internet should be secure. The unauthorized user should not be allowed to access. The major concern is internet security.

According to [1], there is risk of vulnerabilities over internet a system should be designed in order to secure data. To fulfill the requirement of data security "Intrusion Detection System" is developed. This system is adapted by the network administrator so that they can prevent malicious activities and attacks. Which in return makes intrusion detection system the key part of the security management. "Intrusion Detection System" identifies intrusion on the network and generates report.

According to [2], an experienced analysis of the security is made by the system. IDS; which is capable to detect malicious attacks and block them. IDS is most secure, active and efficient technology, which ensures availability, confidentiality and integrity IDS do not allow stalkers and intruders to bypass mechanism of security in a network.

According to [3] the approaches of "Intrusion Detection System" can be classified in two categories; anomaly detection and detection for misuse. Anomaly detection tries typical norm of intrusion whereas, misuse detection is used against attacks that are well known.

Keywords
Machine Learning Algorithm, Security, weka, Classification, Intrusion Detection, Decision tree.

1. INTRODUCTION
In this era of technology, the Internet is the most widely used as essential source of information. Nowadays, internet is used In machine learning, tasls of supervised and unsupervised learning are important having bearing on prediction, knowledge extraction, the detection outlayer, pattern recognition and reinforcement learning. The systemis shown in Figure 1.

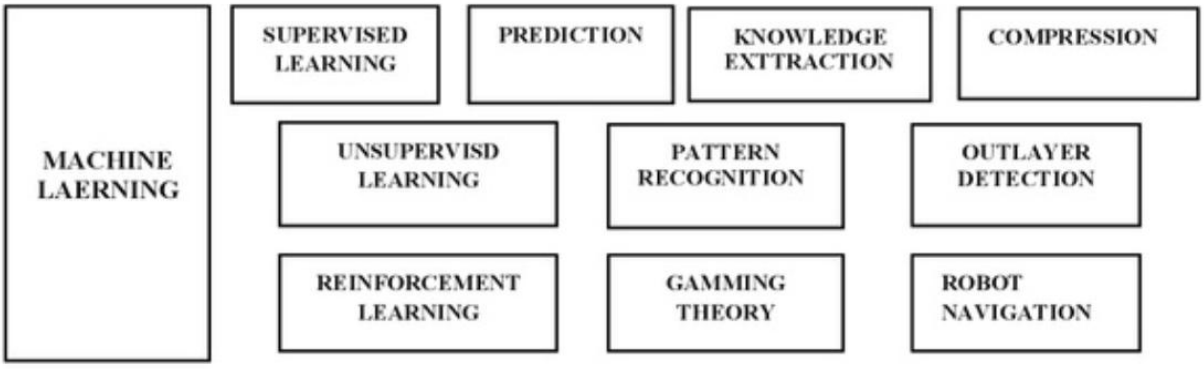

Figure 1. Machine Learning Categories

2. MACHINE LEARNING TECHNIQUES
Machine learning is a field that studies about algorithms, which improve their performance by experiencing and exercise automatically or computerized. Consider, machine learning as a form of artificial intelligence, which provides ability in computers to be able to learn without programming or predefined structure.

According to [4], there are many machine learning techniques that may be supervised, reinforcement or unsupervised depending on absence or presence of data while labeled.

The Weka contains machine learning algorithms. This Software tool is useful for machine learning, consists of several algorithms. The algorithms help in the task of data preparation, clustering, regression, visualization, follow association rules and classification. It is available from open source. The Machine Learning enables the set of technologies. Also it supports the advances in (A.I) artificial intelligence.

Analyzing the work done on "intrusion detection system", it is clear that the pivotal components are main classifiers; ensemble classifiers, single classifiers and hybrid classifiers.

2.1 Single classifiers:

There are number of single classifiers which are as follows:

- Fuzzy logic
- Decision Trees
- Support Vector Machine
- Genetic algorithms
- Self-Organizing maps
- Artificial Neural Networks(ANN)[5]

In order to analyze the data, the decision trees are helpful and useful. The task of identification of malicious and 'misused detection' for example, random forest activities, becomes convenient.

The idea of support vector machine has its own benefits. It is a learning Algorithm. In this algorithm the Data is sorted into 2 groups at the time of initial training. It is also termed as support vector Network. The task of regression analysis and classification can be easily performed.

According to [5], as an example of (ANN) Artificial Neural Network, a "self organizing map" (SOM) is used. It works with unsupervised learning to prepare a two-dimensional map of the problem space. The self organizing map uses competitive learning technique. One of the aspect of the (SOM) "self organizing map" is that it produces 2-dim discretized representation of the input.

3. AI BASED MACHINE LEARNING APPROACHES

There are two types of approaches used in machine learning; Techniques that are based on AI and Computation Intelligence. AI based technique uses comparative analysis of clustering (unsupervised learning) and classification (supervised learning).

4. COMPUTATION INTELLIGENCE:

Computational Intelligence is based on different algorithms or techniques; artificial neural network, genetic algorithm, artificial immune-system and fuzzy logics[9].

4.1 Ensemble Classifies

According to [6],, ensemble classifiers are used so that the single classifiers can be improved. Weak-Single Classifiers are combined with ensemble classifiers in order to generate better results collectively.

4.2 Types of Ensemble Classifies

- Multiple classifier system
- MLP
- SVM
- Statistical-Rule Based Methods(SRBM)
- Clustering techniques
- Standard Machine learnings
- Neuro Tree
- Density estimation
- Neutrosophic-Logic Classifiers

4.3 Hybrid Classifiers:

According to [3], the work that is done is mostly to improve the existing system to build improved and better version of the system, which results in Hybrid Classifiers. In hybrid classifiers different techniques of machine-learning are combined used together in order to improve performance of the system. For example, DT is combined with GA or SVM is combined with KNN. In these hybrid techniques, first one works on raw data then it generate the immediate results while the other one takes input of immediate results and generates the final results.

4.4 Machine learning based "intrusion detection System"

According to [7], in any intrusion detection system, there are two types of data sets; training set and testing set. Features selection method is used in selection of feature for classifiers. Significant features is used for selecting training classifiers.

5. DATA SETS AND WIRESHARK

Wireshark is useful tool by means of which Data on a Network can be viewed as well as captured for analysis. The Data can be viewed on front and backend. It is therefore, extensively popular and user friendly. The Data is captured in the form of Data Sets. The integrated decryption tools can be used. In the IDE (Integrated Development Environment) of wireshark, it supports files, edit, view, capture, analyze, Statistics and other tools. It provides the list of received packets, time of capturing and the address of the packet along with protocol, for example TCP.

The Data packets details are provided with 'Packet Bites' in Hexadecimal. However, the color rules are applied when Wireshark captures the filters; colors being categorized according to their Hue. The statistical information provided by Wireshark is useful for application to the CSV, XML and TXT files.

The illustration of mode of working with Wireshark for capturing the Data Sets has been shown in Figure 1. The examples of working with data sets are reported in Tables 1, Table 2 and Table 3.

After capturing data set using Wireshark data sets are exported to CSV file and then convert that CSV file into arff file using tool named as weka.

According to [8], there are diverse purposes for techniques of machine learning i.e. association findings, classification and clustering. It helps to implement more than one classifier. Weka is the tool that gathers all the machine-learning techniques. These techniques in weka can be implemented on Data using java code. Given below is the screen shot of weka explorer.

The Figure 3 shows the experiment format for producing Table1, Table2 and Table 3

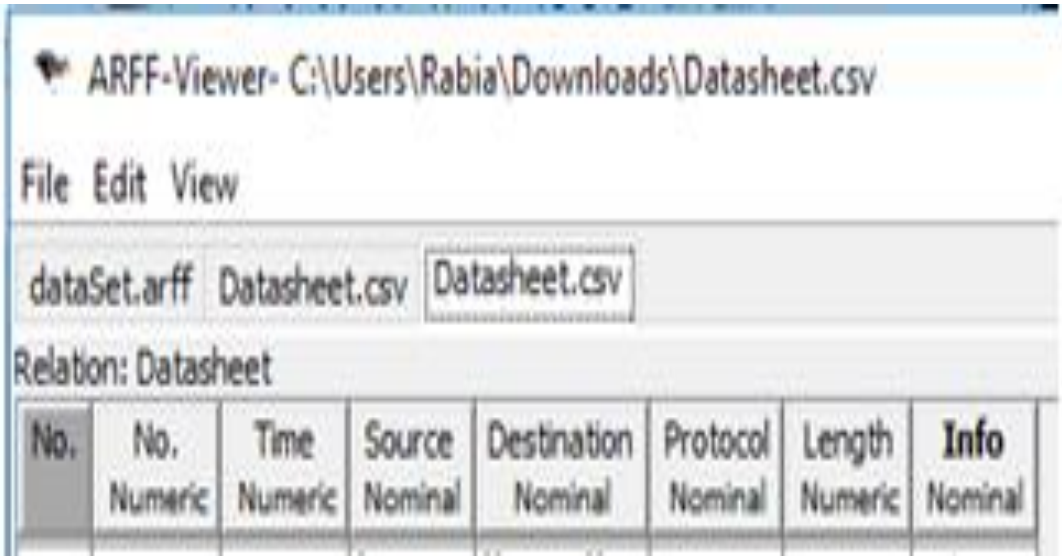

Figure 2 ARF viewer

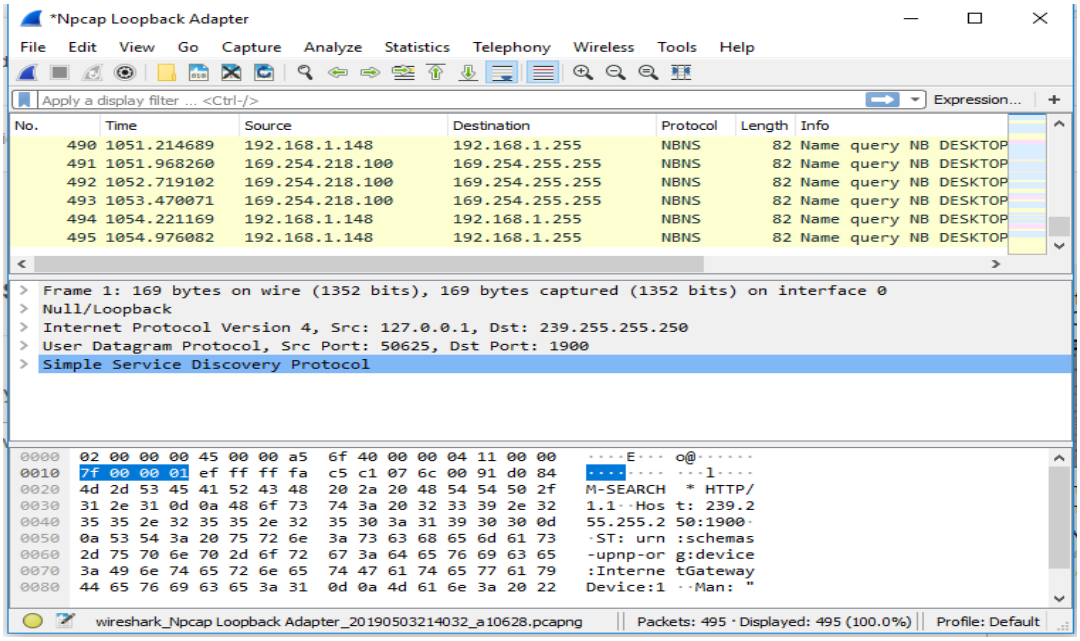

Figure 3: Working with Wire shark for Capturing Data Sets

The data sets used are shown as follows:

Table 1. Intrusion Detection Data Sets

No.	Time	Source	Destination	Protocol	Info
1	0	127.0.0.1	239.255.255.250	SSDP	M-SEARCH * HTTP/1.1
2	3.015019	127.0.0.1	239.255.255.250	SSDP	M-SEARCH * HTTP/1.2
3	6.016576	127.0.0.1	239.255.255.250	SSDP	M-SEARCH * HTTP/1.3
4	7.833611	0.0.0.0	255.255.255.255	DHCP	M-SEARCH * HTTP/1.4
5	7.833739	0.0.0.0	255.255.255.255	DHCP	M-SEARCH * HTTP/1.5
6	9.017679	127.0.0.1	239.255.255.250	SSDP	M-SEARCH * HTTP/1.6

7	15.05556	fe80::7cf1:5c7e:1004:ff23	ff02::1:3	ICMPv6	M-SEARCH * HTTP/1.7
8	20.05716	fe80::7cf1:5c7e:1004:ff23	ff02::1:ff04:ff23	ICMPv6	M-SEARCH * HTTP/1.8
9	21.55899	fe80::7cf1:5c7e:1004:ff23	ff02::c	ICMPv6	M-SEARCH * HTTP/1.9
10	22.90921	0.0.0.0	255.255.255.255	DHCP	M-SEARCH * HTTP/1.10

Table 2. Intrusion Detection Data Sets

No.	Time	Source	Destination	Protocol	Info
11	22.90933	0.0.0.0	255.255.255.255	DHCP	M-SEARCH *HTTP/1.11
12	54.94189	127.0.0.1	239.255.255.250	SSDP	M-SEARCH HTTP/1.12
13	57.94017	127.0.0.1	239.255.255.250	SSDP	M-SEARCH *HTTP/1.13
14	60.94513	127.0.0.1	239.255.255.250	SSDP	M-SEARCH *HTTP/1.14
15	63.95789	127.0.0.1	239.255.255.250	SSDP	M-SEARCH *HTTP/1.15
16	66.95957	127.0.0.1	239.255.255.250	SSDP	M-SEARCH *HTTP/1.16
17	69.95986	127.0.0.1	239.255.255.250	SSDP	M-SEARCH *HTTP/1.17
18	74.55987	192.168.1.148	224.0.0.251	IGMPv2	M-SEARCH *HTTP/1.18
19	77.55554	192.168.1.148	239.255.255.250	IGMPv2	M-SEARCH HTTP/1.19
20	80.05767	192.168.1.148	224.0.0.252	IGMPv2	M-SEARCH *HTTP/1.20

Table 3. Intrusion Detection Data Sets

No.	Time	Source	Destination	Protocol	Info(MetaData)
21	110.6813	169.254.218.100	239.255.255.250	SSDP	M-SEARCH *HTTP/1.21
22	110.6816	192.168.1.148	239.255.255.250	SSDP	M-SEARCH * HTTP/1.22
23	111.6882	169.254.218.100	239.255.255.250	SSDP	M-SEARCH * HTTP/1.23
24	111.6884	192.168.1.148	239.255.255.250	SSDP	M-SEARCH * HTTP/1.24
25	112.6902	169.254.218.100	239.255.255.250	SSDP	M-SEARCH * HTTP/1.25
26	112.6903	192.168.1.148	239.255.255.250	SSDP	M-SEARCH * HTTP/1.26
27	113.6938	169.254.218.100	239.255.255.250	SSDP	M-SEARCH * HTTP/1.27
28	113.694	192.168.1.148	239.255.255.250	SSDP	M-SEARCH * HTTP/1.28
29	140.0572	fe80::7cf1:5c7e:1004:ff23	ff02::fb	ICMPv6	M-SEARCH * HTTP/1.29
30	141.0555	fe80::7cf1:5c7e:1004:ff23	ff02::1:ff04:ff23	ICMPv6	M-SEARCH * HTTP/1.30

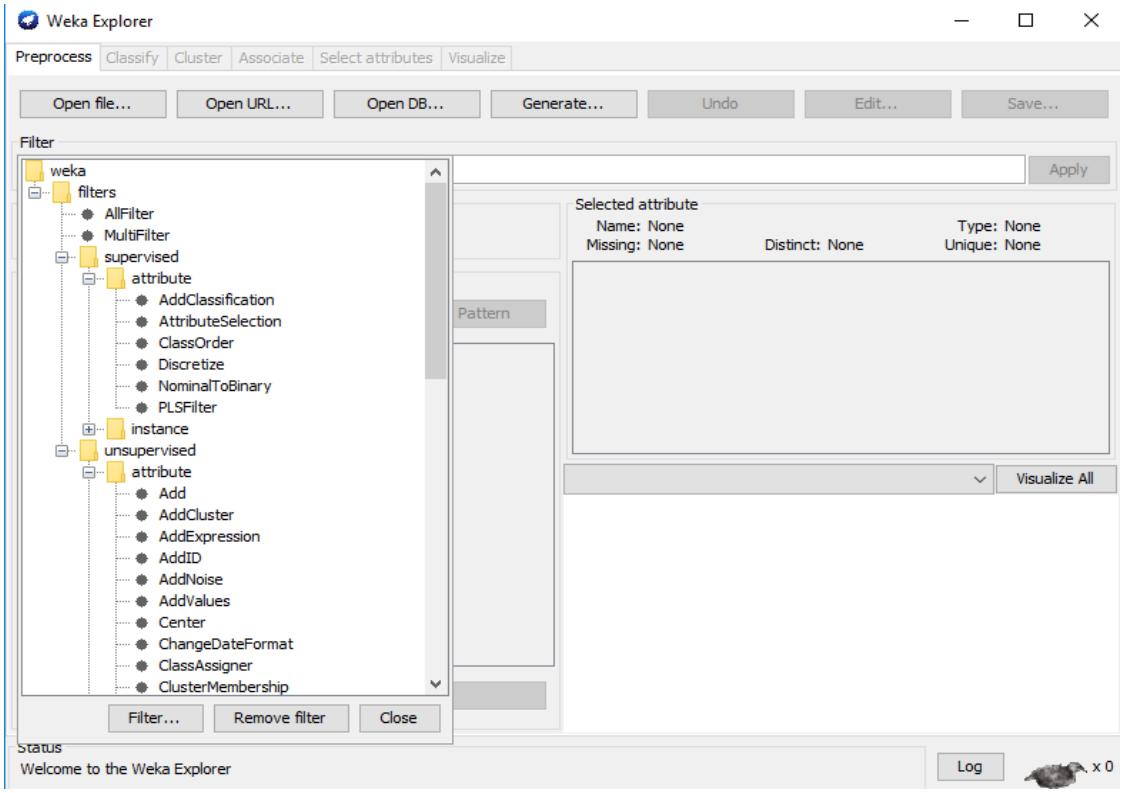

Figure 5. Weka explorer

Following screen shows the implementation of Naive Bayes

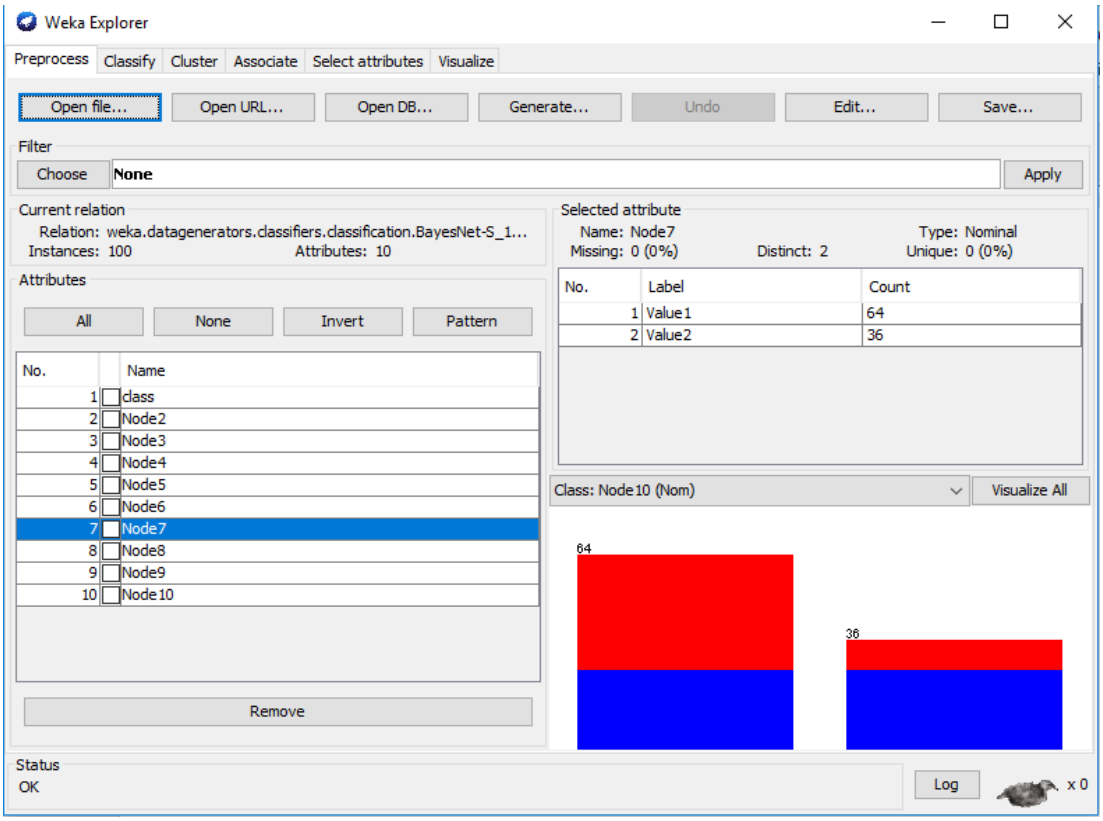

Figure 6. Weka preprocess

Following screen shows classification using cross validation technique and decision tree algorithm.

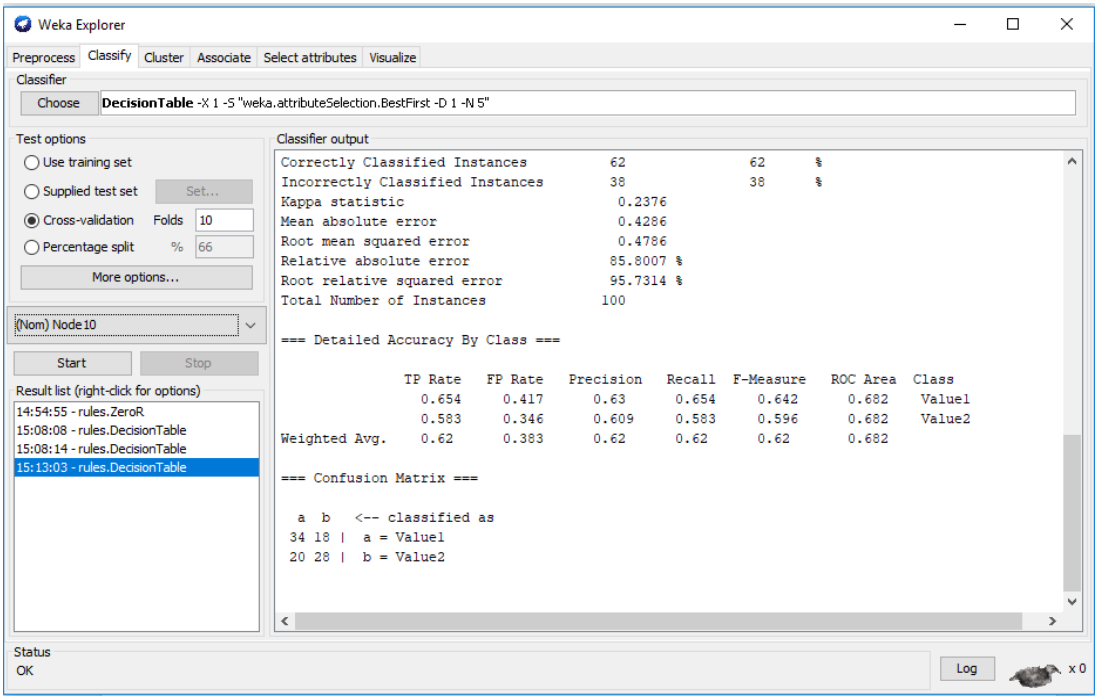

Figure 7. Classification using the technique Cross Validation and Decision tree

Following screen shows the classification result using training set and decision tree.

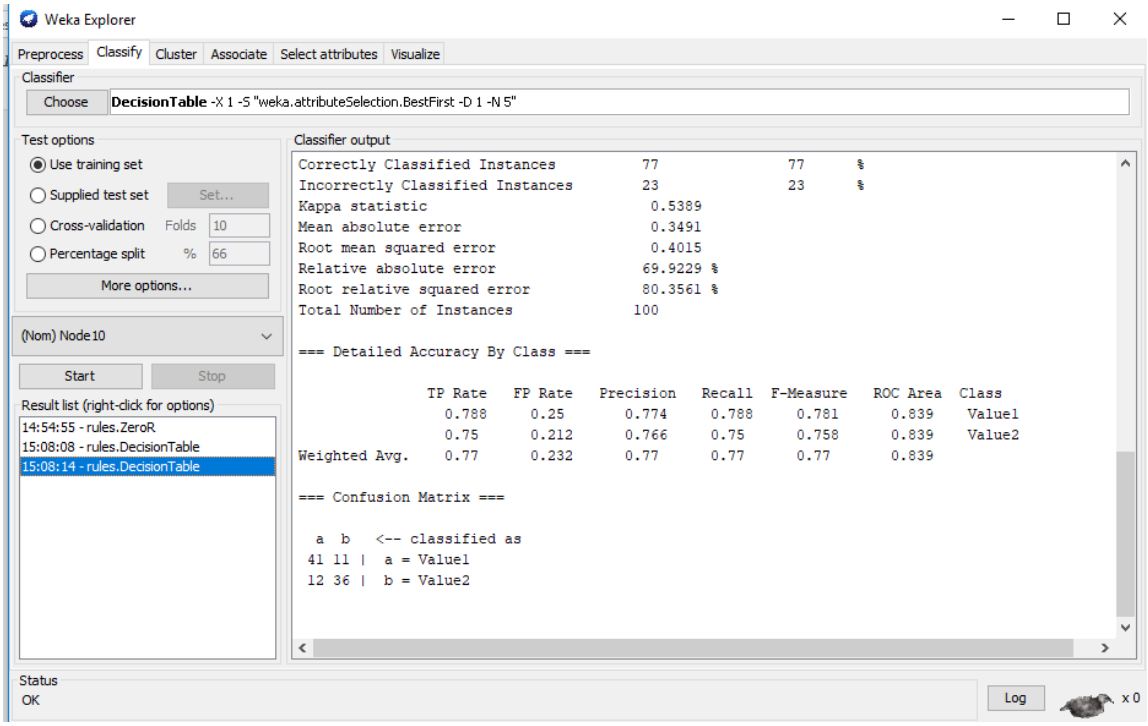

Figure 8. Classification using Decision Tree

Following screen shows classification using cross validation technique and decision tree algorithm.

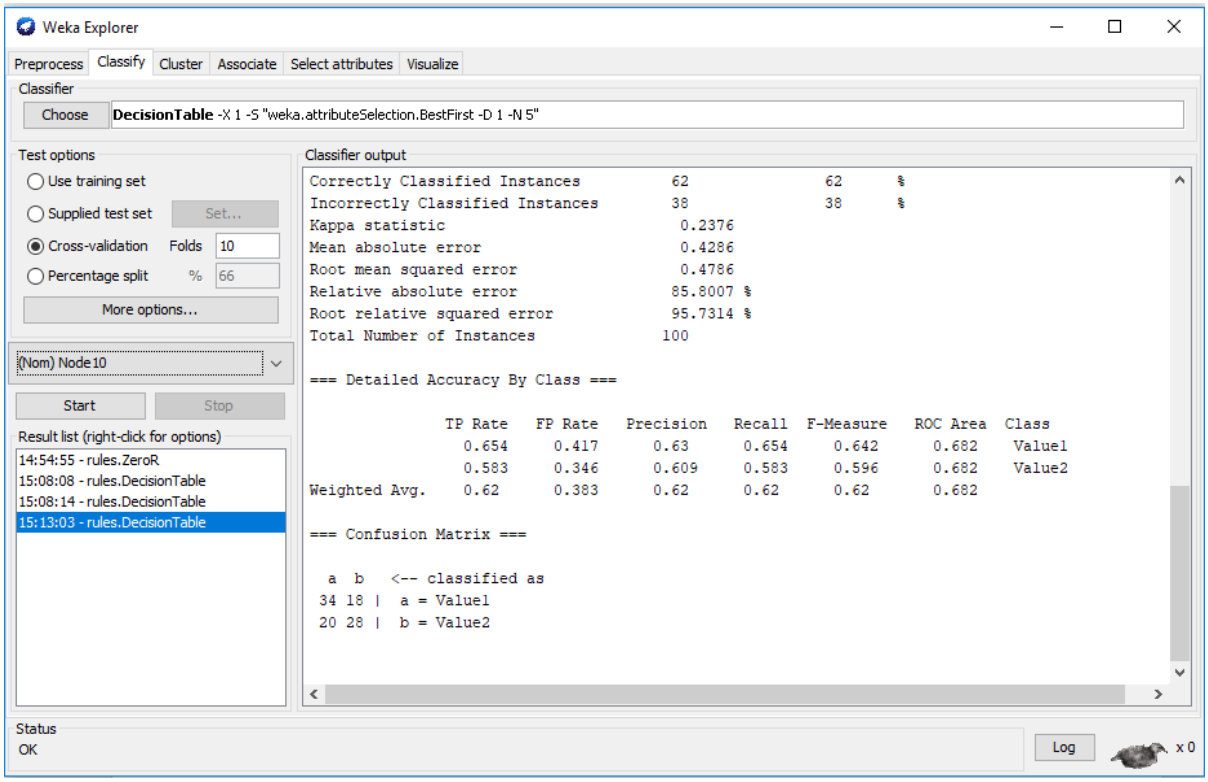

Figure 9. Classification using the technique Cross Validation and Decision tree

Following screen shows the result of machine learning technique named as clustering using training set.

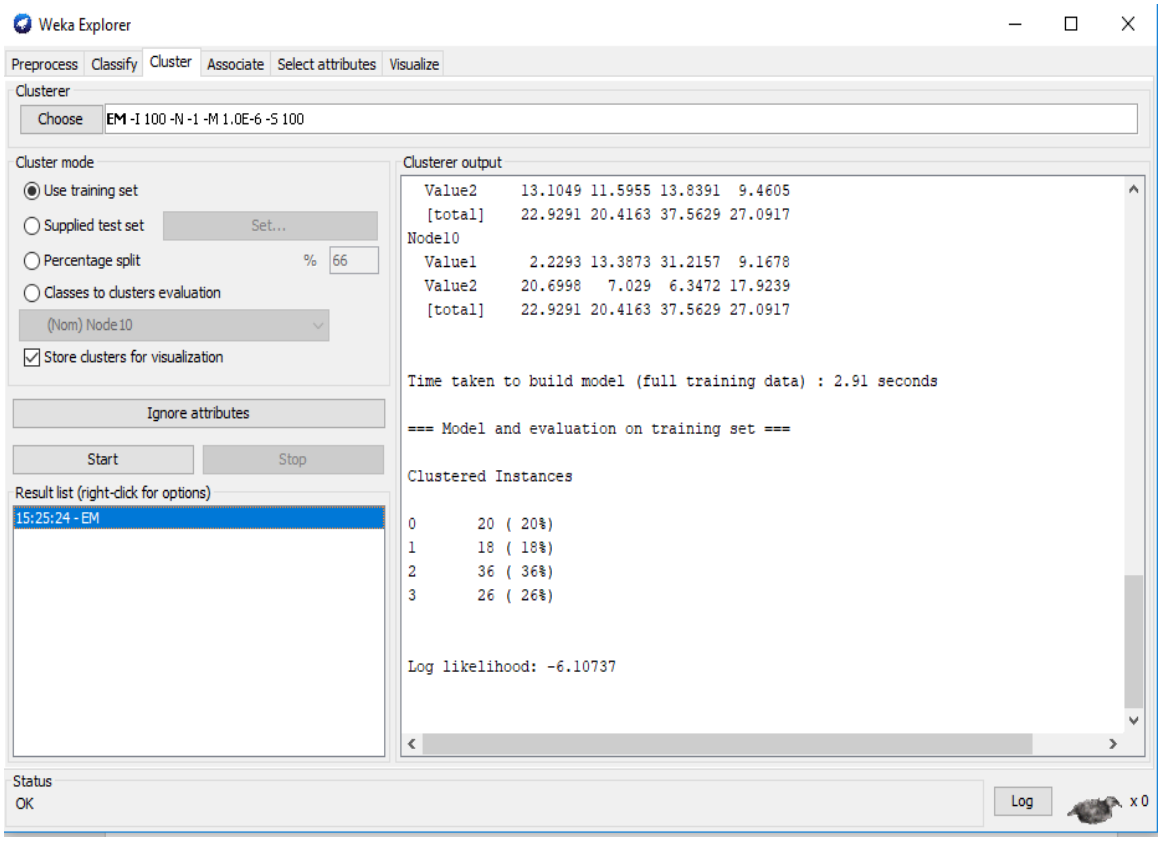

Figure 10. Clustering using training

Table 4. Actual and predicted class

		Predicted Class	
		Yes	**No**
Actual Class	**Yes**	True Positive(TP)	False Negative(FN)
	No	False Positive(FN)	True Negative(TN)

6. IMPORTANCE OF PERFORMANCE MATRICES IN MACHINE LEARNING ALGORITHMS

Every technique that is used is checked on the basis of some performance metrics derived from using confusion matrix. There are different learning algorithm which use Python Ecosystem, logistic regression, visualization or area under curve matrices can also be used for evaluation.

Following are the techniques and their description

Table 5 Accuracy precision and TP Rate

S. No	Technique	Description
1	Accuracy	Proportion of all classification that are accurate.
2.	Precision	Proportion of accurate classification that are positive.
3.	TP Rate	It is used to proportion of accurate classification, which are recognized as positive.

7. REFERENCES

[1] N. F. Haq, "Application of Machine LearningApproaches in Intrusion Detection System?: A Survey," vol. 4, no. 3, pp. 9–18, 2015.

[2] S. Juma, Z. Muda, M. A. Mohamed, and W. Yassin, "MACHINE LEARNING TECHNIQUES FOR INTRUSION DETECTION SYSTEM?: A REVIEW," vol. 72, no. 3, 2015.

[3] A. A. Shah, "Analysis of Machine Learning Techniques for Intrusion Detection System?: A Analysis of Machine Learning Techniques for Intrusion Detection System?: A Review," no. June, 2015.

[4] M. Zamani, "Machine Learning Techniques for Intrusion Detection," no. December 2013, 2014.

[5] B. Zhang, "Network Intrusion Detection Method Based on PCA and Bayes Algorithm," vol. 2018, 2018.

[6] R. R. Chaudhari and S. P. Patil, "A Study on Data Mining & Machine Learning for Intrusion Detection System," vol. 6, no. 2, pp. 114–118, 2017.

[7] U. R. Salunkhe, "Security Enrichment in Intrusion Detection System Using Classifier Ensemble," vol. 2017, 2017.

[8] M. Alkasassbeh and M. Almseidin, "Machine Learning Methods for Network Intrusion Detection," no. October, 2018.

[9] Shaoqiang Wang, DongSheng Xu , ShiLiang Yan "Analysis and application of Wireshark in TCP/IP protocol teaching", Publisher: IEEE

[10] Mark Hall, Eibe Frank etal (2009), "WEKA data Mining Software", ACM SIGKDD Explorations, Volume 11 Issue 1, ACM New York, NY, USA doi>10.1145/1656274.1656278