e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:12/December -2020 Impact Factor- 5.354 www.irjmets.com

NETWORK INTRUSION DETECTION SYSTEMS COMPARITIVE ANALYSIS USING MACHINE LEARNING

Nikhil Kumar H S *1, Mehtab Mehdi*2, Karthik Srivathsa D S*3

*1*2*3 MCA, Jain University, Bengaluru, Karnataka, India.

(Mehtab Mehdi, Jain University, Bengaluru, Karnataka, India)

ABSTRACT

In this fast-growing world, Network security is the main concern for every organization. There is a need for constant observation for identifying skeptical traffic or violation of rules. Thus there is a necessity to build an effective model that can detect lethal activities and prompts the organization. the way toward recognizing obscure assault designs stays an uncertain issue. Nonetheless, a few explorations results as of late have indicated that there are possible arrangements to this issue.

In this research using Machine learning technology, we've tried to automate the detection of intrusion by comparing different models with different accuracies and listing the best model according to their results.

Keywords Network Intrusion,, KNN , Logistic regression, decision tree classifier, Machine learning.

I. INTRODUCTION

In the present world as and how organizations are growing even the data exchange between the organizations are also happening at a larger rate so there is a necessity for networks to handle large data traffic that may include malicious content. Simple firewall and antivirus bundles are not, at this point adequate in shielding an association's organization from digital assaults. We need to have a generally excellent thought of what's going on our organizations and frameworks and supportive of effectively control it, i.e. we need to screen the traffic moving over our organizations, break down it with the capacity to identify any oddities or vindictive exercises, and react to such occasions.

There are two kinds of intrusion detection system, Network based and Host based intrusion detection system. In network based intrusion detection system ,they're placed in the nodes within the network so that they can observe the traffic and analyze the subnet and match it with the predefined attacks when there is suspicious traffic analysis in the network system prompts the authority which is handled. Whereas on the Host based intrusion detection system they run on networking devices or individual hosts look for suspicious data packets and prompts the administration. Network intrusion based detection systems are of two types Anomaly based and Signature based. A Signature based framework is predefined for a specific weakness, so it has a diminished number of false positives, consequently offering less adaptability. In an Anamoly based framework, it is more dynamic in nature and will look for potential assaults that are out of the predefined ones, henceforth bringing about a more prominent number of false positives.

An Intrusion Detection System framework arranges the organization traffic, for normal and anomalous. Organization traffic is viewed as anamolous when the conduct of network traffic strays from the standard organization traffic design. The adequacy of the NIDS relies upon the arrangement calculation being utilized. The exactness and time utilization of the algorithm are significant parameters in the choice cycle of an algorithm.

In this paper ,we've compared two dataset for our comparative research which are UNSW and NSL-KDD dataset. Classification algorithms are used to differentiate between the normal and anamolous network traffic. We've tried to draw the comparison among four models which can detect Intrusion in the network and given our opinion with the best model to use.

Bayesian Network : It is a type of probabilistic graphical model comprised of nodes and directed edges. This model capture both conditionally independent and dependent connections between arbitrary factors.

Logistic: Logistic this strategy uses regression to foresee the probability of an outcome which can have only two values. One or a few indicators are used to make the forecast. Logistic regression conveys a logistic bend that is kept to values somewhere in the range of 0 and 1. The bend is developed utilizing the natural logarithm of the chances of the target variable and not the likelihood.

KNN: K nearest neighbors is a straightforward algorithm that stores every single accessible case and classifies new cases dependent on a similitude measure. KNN has been used in measurable assessment and pattern recognition in the beginning of 1970's as a non-parametric system.

Random forest: The overall thought of the bagging technique is that a combination of learning models expands the general outcome. Simply saying, random forest forms multiple decision trees and integrates them to get a more exact and stable prediction.

Random tree: It creates a tree by randomly choosing branches from a potential arrangement of trees. The trees are circulated in a uniform manner so odds of getting examined are equiprobable.

II. PERFORMANCE MEASURES

	Predicted		
Actual		Normal	Anomaly
	Normal	TP	FN
	Anomaly	FP	TN

True positive (TP)— It indicates the instances which are anticipated as expected correctly.

False negative (FN)— It denotes wrong prediction i.e. it detects instances which are attacks in reality, as normal.

False positive (FP)— It gives a trace of the quantity of identified attacks which are ordinary as a general rule.

True negative (TN)— It denotes instances which are correctly detected as an attack.

ROC (Receiver operating characteristics)— To plan the curve between true positive rate (TPR) and false positive rate (FPR), this term is required. The area below the curve is termed as AUC, which gives the estimation of ROC. The more noteworthy the area the curve possesses, more prominent will be the estimation of ROC.

Sensitivity=TP / (TP+FN)

Precision=TP / (TP+FP)

Accuracy= (TP+TN) / (TP+TN+FP+FN)

Mean absolute error (MAE)— This mistake should be least for an algorithm to be the best in execution. It is the mean of the general blunder which a classification algorithm makes.

F1 score— It is characterized as the harmonic mean of sensitivity and precision. The test performance can be assessed by means of this performance metric.

F=2*TP / (2TP+FP+FN)

False positive rate (FPR)— It demonstrates the chance of an algorithm to predict instances as assaults which are really typical.

FPR=FP / (TN+FP) =1-SPC

False discovery rate (FDR)— It determines the chance of a positive prediction made being inaccurate FDR=FP/ (FP+TP) =1-PPV

Negative predictive rate (NPV)— It demonstrates the chance of a calculation to accurately recognize instances as attack.

NPV=TN / (TN+FN)

Training time--It is the time the classifier devours to build the model on the dataset. It is generally estimated in a flash. The lesser the estimation of this parameter, the better will be the classifier

III. DESCRIPTION OF THE DATASET

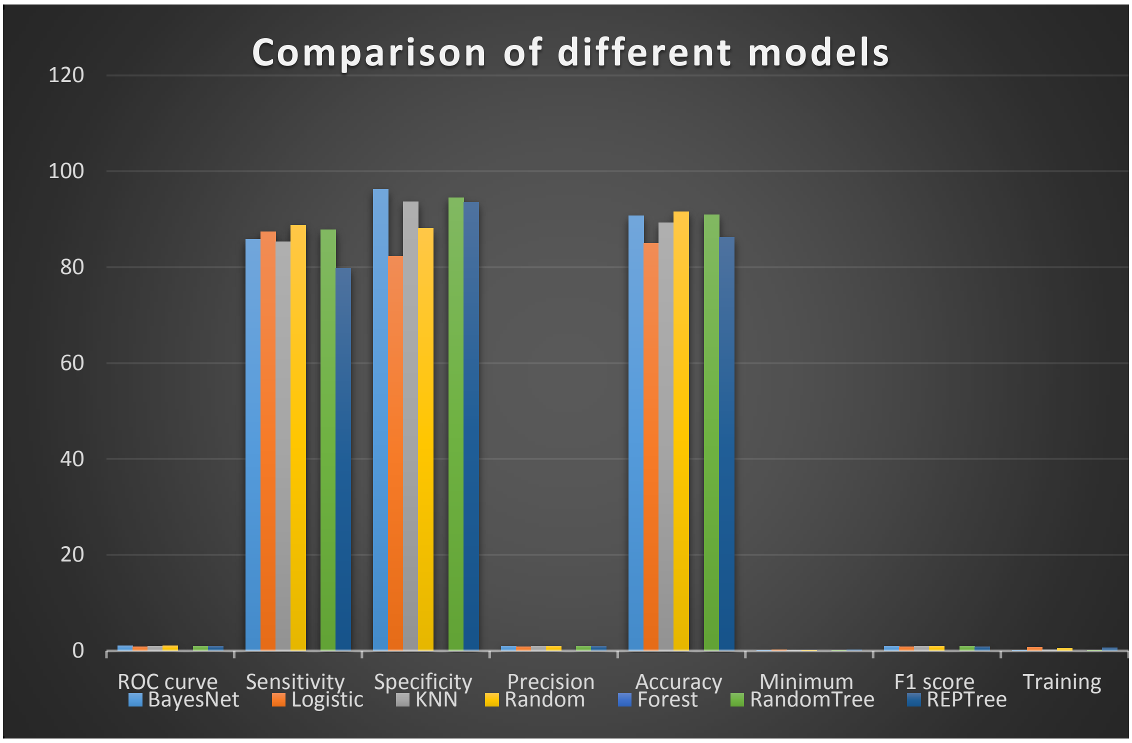
The dataset that is been used in this paper is UNSW dataset and NSL-KDD dataset. The training set involves 42 attributes and 1166 instances while the testing set involves 38 properties and 6486 instances. This dataset has certain favorable circumstances over the unique KDD informational index, so we have picked the above mentioned referenced dataset

IV. RESULTS AND DISCUSSION

Among Bayes net , Logistic , KNN, Random forest, Random tree Logistic is the most noticeably awful classifier as its worth is most extreme for time and least for specificity, accuracy, F1 score, and precision. The great classifiers noted from the above outcomes are BayesNet and RandomForests with BayesNet i.e. 96.18 %, RandomForest i.e. 88.68%,KNN ie 93.63%, Random tree ie 87.74% and logistic 82.2%.

Classifiers	ROC curve	Sensitivity (%)	Specificity (%)	Precision	Accuracy (%)	Minimum Absolute Error	F1 score	Training Time (seconds)
BayesNet	0.976	85.8	96.18	0.962	90.66	0.1002	0.907	0.02
Logistic	0.806	87.34	82.2	0.848	84.9651	0.1542	0.860	0.71
KNN	0.933	85.32	93.63	0.938	89.2167	0.1095	0.893	0.23
Random Forest	0.984	88.68	88.68	0.951	91.5236	0.1135	0.917	0.49
RandomTree	0.912	87.74	87.74	0.947	90.8798	0.0932	0.911	0.01

Comparison of different Machine Algorithms with their Results



Comparison of models

V. CONCLUSION

In this paper we've tried to compare the results of different papers results with selected Algorithms and required parameters to ensure the best choice to be made when selecting the model, compared the results for choosing the right model for network intrusion detection. From these results, it can be considered for efficient network intrusion detection mechanism which maybe used for secure Network for an organization which deals with large amount of data transfer.

ACKNOWLEDGEMENTS

I would like to express my sincere feeling and obligation to Dr. MN Nachappa and Asst. Prof Mehtab Mehdi and project coordinators for their effective steerage and constant inspirations throughout my analysis work. Their timely direction, complete co-operation and minute observation have created my work fruitful..

VI. REFERENCES

[1] Tanya Garg, "Analysis of Various Features Selection Techniques for Network Intrusion Detection Dataset in WEKA",CT International Journal of Information & Communication Technology,2014,Vol 2,Issue 1.

[2] Ms S.Vijayarani, Ms M.Muthulakshmi,"Comparative Analysis of Bayes and Lazy Classification Algorithms", International Journal of Advanced Research in Computer and Communication Engineering, 2013, Vol.2, Issue 8, August 2013

[3] UNSW NB15 Dataset

[4] NSL-KDD dataset

[5] Nour Moustafa, IEEE student Member, Jill Slay School of Engineering and Information Technology

[6] Ms S.Vijayarani, Ms M.Muthulakshmi,"Comparative Analysis of Bayes and Lazy Classification Algorithms", International Journal of Advanced Research in Computer and Communication Engineering, 2013, Vol.2, Issue 8, August 2013

[7] Choudhury, S., & Bhowal, A. (2015, May). Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection. In 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM) (pp. 89-95). IEEE.

[8] Sinclair, C., Pierce, L. and Matzner, S., 1999, December. An application of machine learning to network intrusion detection. In Proceedings 15th Annual Computer Security Applications Conference (ACSAC'99) (pp. 371-377). IEEE.

[9] Sommer, R. and Paxson, V., 2010, May. Outside the closed world: On using machine learning for network intrusion detection. In 2010 IEEE symposium on security and privacy (pp. 305-316). IEEE.