IJCSE International Journal of Computer Sciences and Engineering Open Access

Review Paper Vol.-7, Issue-7, July 2019 E-ISSN: 2347-2693

Performance Analysis of Different Machine Learning Algorithm on Intrusion Detection System

Ritu Ganeshe [1*], Manish Kumar Ahirwar[2], Rajeev Pandey [3]

[1,2,3]Dept. of Computer Science & Engineering, University Institute of Technology, RGPV, Bhopal

Corresponding Author: ganesheritu@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i7.8386 | Available online at: www.ijcseonline.org

Accepted: 17/Jul/2019, Published: 31/Jul/2019

Abstract- There are rapidly increasing attacks on computers creates a problem for network administration for averting the computer from these attacks. There are many conventional intrusion detection systems (IDS) is present but they are unable to prevent computer system completely. These IDS finds the spiteful actions on net traffics and they find the anomalies in network system. But in numerous instances they are unable for detecting spiteful actions in the networks. There is human interaction is also required to process the network traffic or detect malicious activity. In this paper we present various data mining algorithms helps in machine learning to detect intrusion accurately.

Keywords— *Intrusion Detection system, Anomaly detection, deep belief network, state preserving extreme learning machine.*

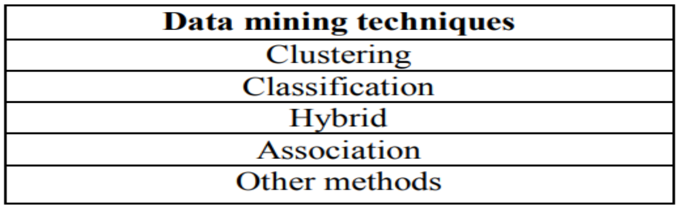
I. INTRODUCTION

An intrusion detection system is used to check spiteful actions or guidelines violations and produce reports to a administration Station. Intrusion detection is primarily based on distinguishing doable incidents, work data regarding them, and coverage makes an attempt. Moreover IDS is used for alternative functions, like distinguishing issues with security policies, presenting already current threats, and deterring people from violating safety policies. IDS (Intrusion Detection system) became an essential addition to the protection infrastructure of almost each organization.

II. DATA MINING ALGORITHMS FOR INTRUSION DETECTION

There are many categories varieties of algorithms drawn from areas as pattern recognition, machine learning and database analysis. Algorithms can be applied to mine audit data. The analysis result are later used by the algorithm to define optimal parameters to create the selected mining model. The parameters are practical across the dataset, jointly with chosen patterns and detailed statistics [14].

Table 1. DATA MINING TECHNIQUES



Data mining techniques
Clustering
Classification
Hybrid
Association
Other methods

I. Numerous studies indicate that classification techniques and clustering are by far the mainly used data mining techniques. The hybrid technique is considered shortly after together with the Association technique.

Clustering is the method of splitting data into clusters based upon the features of the data. This clustering divides data into groups of alike stuff. Each member within the cluster is alike to one another.Also quite a few clustering, classification algorithms are identified, though the mainly broadly used seemed to be the k-means classification.

Classification and prediction are seen as the accepted mining techniques,it allows extraction of models ,telling significant data module and helps in predicting future trends. It classifies data into normal or abnormal in intrusion detection systems. More importantly it also tells that classification maps data items to one of many predefined categories. The classifier's output be capable to guess a model that may forecast future trends, when sufficient normal and abnormal behavior is gathered in audit data. The classification algorithm may be capable for predicting new unobserved data classifying it by using pre-existing information. The widely used approaches is data classification; Bayesian classification, decision tree induction, neural network and statistical learning. Classification algorithms need information in both typical and identified assault facts to separate modules throughout recognition.

The Association technique discovers anomalies by using association rule algorithms, signifying with the intention of the use of the method is checking defects to be detected fault inside data evaluating outcomes throughout inspections and analysing reasons for anomalies persistent inside data. This method is suitable for forensic investigation and not actual instance attacks.

A. Machine Learning Approach

Machine learning might be a division of AI that acquires information from training information supported known facts. Machine learning is outlined as a study that enables computers to be told information while not being programmed mentioned by Arthur prophet in 1959.Machine learning has the main focuses on prediction. Machine learning techniques square measure classified into 3 categories –

1) Supervised Learning

Supervised learning as well recognized as classification. In supervised learning data, instances are labeled in the training phase.

2) Unsupervised Learning

Data instances in this learning are unlabeled. Clustering is best in this learning.

3) Reinforcement Learning

Reinforcement learning means computer interacting with an environment to achieve a certain goal. A reinforcement approach can ask a user (e.g., a domain expert) to label an instance, which may be from a set of unlabeled instances.

B. Hybrid Classifiers

A hybrid classifier offers mixture of over one machine learning algorithms or techniques for up the intrusion detection system's presentation immensely. Victimization a few clustering-based techniques for preprocessing samples in coaching knowledge to eliminate non-representative coaching samples after that the consequences of the bunch are used as coaching samples for pattern recognition so as to style a classifier. Therefore, either supervised or unsupervised learning approaches is the primary stage of a hybrid classifier.

C. Ensemble Classifiers

The classifiers playacting slightly higher than a random classifier are called weak learners. Once multiple weak learners are combined for the bigger purpose of rising the performance of a classifier considerably is understood as Ensemble classifier. Majority vote, cloth and boosting square measure some common methods for combining weak learners.Though it's better-known that the disadvantages of the element classifiers get accumulated within the ensemble classifier, however it's been manufacturing a awfully economical performance in some combination. thus researchers has turn out to be a set of inquisitive about ensemble classifiers day by day.

III. LITERATURE REVIEW

According to (Yaping Chang et al., 2017) [2], The network intrusion detection techniques are significant to avert our system and net from spiteful behaviours. To improve exactness of network intrusion detection, machine learning, feature selection and optimization methods have been used, and the result tell us that the combination of machine learning and feature selection can improve accuracy. In this study,they developed a new machine learning approach for predicting network intrusion based on random forest and support vector machine. Since there were many potential features for network intrusion classification, random forest were used for feature selection based on variable importance score. The presentation of the support vector machine which used the 14 selected features on KDD 99 dataset has been evaluated by comparing it with the total(41) features and popular classifiers. The result showed to facilitate the selected featuresbe able to accomplish high attack detection rate and it may be lone viable classifier for net intrusion detection.

According to (David Ahmad Effendy, Kusrini Kusrini, Sudarmawan Sudarmawan et al, 2017) they created a manual solution for detecting intrusion attacks and also prevent the PC from various kind of network attacks. In this they present a pattern based variance revealing, so first they collect some related pattern and based on these pattern they learn the model and after learning they spot the spiteful actions. They only detect the normal attacks because the model is not strongly enough to detect latest intrusions or malicious activity so they need to improved the model so it will work on latest anomalies and detect intrusion. So they uses machine learning based algorithm to enhanced thier model for detecting the anomalies, so for this first the model gets trained on various predefined malicious activities and before giving these datasets to the machine learning algorithm first they need to pre\process the dataset and remove noise if present, attributes selection for better learning and after the pre-processing is complete we can get the data to the machine learning algorithm so the model get trained efficiently. In these they uses NSL-KDDcup99 datasets and uses naive bayes algorithm to categorize the net packets into various category of attacks like either dos, probe, r2l or u2r kind of attacks.

According to (Amreen Sultana, M.A.Jabbar et al, 2016) when the web is growing rapidly and millions of internet users are present which uses the internet intended for socially connected to each other, online transactions. So the privacy and security is main issue in that so for that they working on it to detect various kind of attacks and also detect malicious activities.

So they proposed an intrusion detection system based on AODE algorithm to detect the attacks and spiteful actions within network. In these they uses NSL-KDDcup99 datasets and uses AODE algorithm to classify the net packets into various category of attacks like it is a either dos, probe, r2l or u2r kind of attacks. In these they pre-process the data by using AODE algorithm for better learning the algorithm, because they can effort with numerical data. After that they can evaluate various concert actions headed for checking how accurate his algorithm is working.

Table 1 Related work of different authors and comparison

Author	**Objective**	**Tool Used**	**Algorithm Used**	**Accuracy**	**Result**
Guoquan Li	Data Fusion for Network Intrusion Detection:A Review	-	-	-	The techniques of NIDSs i.e feature fusion and decision fusion is presented.
Dogukan AKSU	Detecting Port Scan Attempts with Comparative Analysis of Deep Learning and Support Vector Machine Algorithms	-	Deep Learning and SVM	97.80% and 69.79% Precision 99% and 80%	Results show that the deep learning algorithm performed significantly better results than SVM
Amreen Sultana	Intelligent Network Intrusion Detection System using Data Mining Techniques	Weka tool	AODE algorithm	97.19% Detection rate 98%	Result prove that accuracy, DR and MCC for four types of attacks are increased by our proposed method.
RAJAN GOSWAMI	Intrusion Detection In Computer Networks By using Decision Tree Algorithm	Python	Decision tree	98.04% Precision 68% Recall 61%	After testing the results of different classification results we can say that Decision Tree model takes less time for training because it creates a tree to handle attributes for prediction outcomes and affects the final classification results
Chidananda Murthy P	Predicting Unlabeled Traffic For Intrusion Detection Using Semi-Supervised Machine Learning	Weka tool	Random tree, J48 and Naïve Bayes	99.7666, 99.7785 and 90.4384	Semi supervised learning was used to predict and classify the unlabeled data and hence decrease the work of network administrator in identifying the threats.
L.Haripriya	Role of Machine Learning in Intrusion Detection System: Review	-	Various ML Algorithms	99% Detection rate 99%	Discussed various Machine Learning (ML) techniques for detection of IDS

IV. PROBLEM DEFINITION

The lone main problems for IDS is to build effective behaviour models or patterns to differentiate standard behaviours from abnormal behaviours by analyzing composed network dataset. To resolve the trouble, former IDSs typically depends on safekeeping experts for analyzing the dataset and build intrusion detection system based on supervised algorithm. As the quantity of records increases very rapidly, it turn in a less accurate and tedious task for supervised algorithm to analyse and extract attack signatures or detection rules from dynamic, huge volumes of network data.

V. CONCLUSION

Traditional IDS are suffering from various different problems such as accuracy and efficiency. There are many conventional intrusion detection systems (IDS) is present but they are unable to prevent computer system completely. These IDS finds the spiteful behaviour lying on net traffics and they find the anomalies in network system. But sometimes they are unable to check spiteful behaviour in the networks . There is human interaction is also required to process the network traffic or detect malicious activities. IDS can help cover the gap in traditional network systems, which considers a good move to integrate computer network with advanced data mining techniques.

REFERENCES

[1] Rahul Vigneswaran K, Vinayakumar R and Prabaharan Poornachandran, "*Evaluating Shallow and Deep Neural Networks for Network Intrusion Detection Systems in Cyber Security*" in IEEE 2018.

[2] Yaping Chang ; Wei Li ; Zhongming Yang, " *Network Intrusion Detection Based on Random Forest and Support Vector Machine*" in IEEE 2017.

[3] David Ahmad Effendy, Sudarmawan Sudarmawan, "*Classification of Intrusion Detection System (IDS) Based on Computer Network*" in 2017 IEEE.

[4] Amreen Sultana, M.A.Jabbar, "*Intelligent Network Intrusion Detection System using Data Mining Techniques*" in IEEE 2016.

[5] James P. Anderson, "*Computer security threat monitoring and surveillance*," in USA, April 1980.

[6] Nawfal Turki Obeis, Wesam Bhaya, "*Review of Data Mining Techniques for Malicious Detetion*", in RJAS, 2016.

[7] Jau-Hwang WANG and Peter S. DENG, "*Virus Detection Using Data Mining Techniques*", in Taiwan.

[8] Chi Zhang, Jinyuan Sun, "*Privacy and Security for Online Social Networks: Challenges and Opportunity*", in University of Florida and Xidian University.

[9] Uma Salunkhe, Suresh N. Mali, " *Enrichment in Intrusion Detection System Using Ensemble*", in JECE.

[10] Q.S. Qassim, A. M. Zin and M. J. Ab Aziz, "*Anomalies classification approach for network- based intrusion detection system*", in IJNS, 2016.

[11] O.Y.Al-Jarrah, P.D.Yoo, K.Taha and K. Kim, *" Data Randomization and Cluster-based Partitioning for botnet intrusion detection*", in IEEE, 2016.

[12] Solane Duque, Dr. Mohd. Nizam Bin Omar, "*Using Data Mining Algorithm for Developing a Model for Intrusion Detection System(IDS)*", in procedia Computer Science, 2015.

[13] Abhaya, K. Kumar, S. Afroz, "*Data Mining Techniques for Intrusion Detection: A Review,*" in IJARCCE, 2014.

[14] R.J. Manish, H.T. Hadi, "*A review of network traffic analysis and prediction techniques".*

[15] S. Choudhury, A.Bhowal,"*Comparative Analysis of Machine Learning Algorithms along with Classifiers for Network Intrusion Detection*." in IEEE, 2015.

[16]] S.B. Kotsiantis, P.E. Pintelas, "*Machine Learning: a Review of Classification and combining Techniques*," in Artificial Intelligence Review, 2006.

[17] I. Witten, E. Frank, M. Hall, "*Data mining: Practical Machine Learning Tools and Techniques."* in 2011.

[18] M. Masud, L. Khan, B. Thuraisingham, "*Data mining tools for malware detection*," in 2012.

[19] R.S. Wahono, "*A Systematic Literature Review of Software Defect Prediction: Research Trends, Datasets, Methods and Frameworks*," In 2015.

[20] M.H. Haratian, "*An Architectural Design for a Hybrid Intrusion Detection System for Database*,".

[21] S. Zargari, D. Voorhris, "*Feature Selection in the Corrected KDDdataset,*" in 2012.

Authors Profile

Ms. Ritu Ganeshe is currently pursuing Master of Engineering Post Graduation Programme in Computer Science and Engineering from University Institute of Technology RGPV, Bhopal (M.P.),India. Her research area is cyber security. She received her bachelor's degree in computer science & Engineering from RGPV University Bhopal.

Prof. Manish Ahirwar is an Assistant Professor in Department of Computer Science and Engineering, University Institute of Technology RGPV, Bhopal, (M.P.) since july 2007. He has 12 years of academic experience. He received his Bachelor's degree in Computer Science and Engineering in the stream of Information Technology. He has done Ph.D from University Institute of Technology RGPV, Bhopal, (M.P.) in stream of computer science. He is famous for academic, administrative and motivational skills. His motive is to spread practical knowledge to develop students and institute as a whole.

Dr. Rajeev Pandey is an Assistant Professor in Department of Computer Science and Science and Engineering, University Institute of Technology RGPV, Bhopal (M.P.) since july 2007. He has 12 years of academic experience. He received his Bachelor‟s degree in Computer Science and Engineering from IET, DR. B.R.A. University, Agra (U.P.). He has done M.E. in Computer Science and Engineering in 2004 & Ph.D in 2010 from DR. B.R.A. University, Agra (U.P.), India.