CLASSIFICATION OF CYBER ATTACKS USING MACHINE LEARNING TECHNIQUE

Kaveri Mirji[1], Suma S[2]

[1]Post Graduate Student, Dept. of Computer Applications, DSCE, Bangalore.

[2]Asst. Professor, Dept. of Computer Applications, DSCE, Bangalore.

Abstract—Security of significant data is constantly a basic issue for present day advanced world. Intrusion Detection System (IDS) and numerous security procedures is broadly utilized against digital assaults. Data mining and AI strategies have likewise been utilized by specialists to acquire high discovery rate and false alarm rate. Proposed work aims to structure and advancement of a methodology for improving cyber-attack identification framework utilizing cloud. Employments of distributed computing is increments continuously. Utilizing customary ML Techniques don't bolster well preparing of enormous datasets, so new methodologies and stages are required. This paper recommends that cloud-based AI system can be utilized so as to arrange assault into a cloud-based AI stage. The work proposes an attack classification framework utilizing NSL KDD Cup99 dataset. The classifier is assembled which depends on 'Multiclass Decision Forest' Machine Learning Algorithm and is sent on Microsoft's Azure Machine Learning (Azure ML) platform, is an open source platform. The outcomes got by proposed model are assessed as far as exactness and the correlation is obtained with benchmarks. The outcomes got are promising and the paper additionally coordinates the future research work in the field.

Keywords— Cloud Computing, Cyber Attack, IDS, Classification, Machine Learning, Microsoft Azure Cloud, Cyber threats, Multicast Decision Forest, Regression.

1. INTRODUCTION

1.1 Machine Learning and IDS:

Machine learning is a type of artificial intelligence (AI) that provides computers with the capacity to learn without being explicitly customized. Machine Learning centers around the advancement of computer programs that can instruct themselves to develop and change when exposed to new information. Machine Learning strategies have capacity to execute a framework that can gain from information. For instance, a machine learning framework could be prepared on approaching bundles to figure out how to recognize nosy and ordinary parcel. In the wake of learning, it can at that point be utilized to order new approaching packets into nosy and typical packets.[1]

Intrusion Detection System (IDS) is a functioning procedure or device that breaks down framework and network activity for unauthorized action [2]. An ID is a hardware or programming or a blend of both which is utilized to screen a framework or system of frameworks against any malicious or unauthorized activities [2]. Intrusion Detection Systems (IDSs) are utilized to improve organize security. An ID improves the security of the network by recognizing, surveying, and revealing unauthorized network activities. IDS are classified into two classes: network based and host based.

System based Intrusion Detection Systems investigations network packets recovered from the system. Host-based Intrusion Detection System examinations system calls created by individual hosts [2]. The information moves through a system is extremely huge and it is hard to analyze and recognize the attacks using custom strategies. Today we have number of Machine learning strategies accessible which are exceptionally helpful for analyzing the information and identifying the attacks. In this paper we have used various machine learning techniques for network intrusion detection[2].

1.2 IDS in Cloud:

Intrusion detection system assumes a significant job in the security and constancy of dynamic protection system against intruder hostile attacks for any business and IT organization. IDS execution in distributed computing requires an effective, versatile and virtualization-based methodology. In distributed computing, client information and application is hosted on cloud specialist organization's remote servers and cloud client has a restricted authority over its information and assets. In such case, the organization of IDS in cloud turns into the responsibility of cloud provider. In spite of the fact that the executive of cloud IDS ought to be the client and not the provider of cloud services [8].

1.3 Microsoft Azure Cloud Computing Environment for Machine learning:

Microsoft's Azure Machine Learning (Azure ML) [3] is a cloud administration that empowers execution of machine learning process. Microsoft Azure is an open cloud platform. The advantages of utilizing open distributed computing platform (Azure ML) incorporates: taking care of large information and access from anyplace on the planet. The procedure of Azure ML is appeared in Figure – 1, which is same as that of fundamental procedure of ML. Sky blue ML gives a graphical device to dealing with the ML procedure, a lot of information pre-handling modules, a lot of ML algorithms, and an API to dispatch a model to applications. ML Studio is a graphical tool that is utilized to control the procedure from start to finish for example from information pre-handling to run tests utilizing an ML algorithm, and test the subsequent model. ML Studio additionally enables its clients to send that model on genuine cloud.

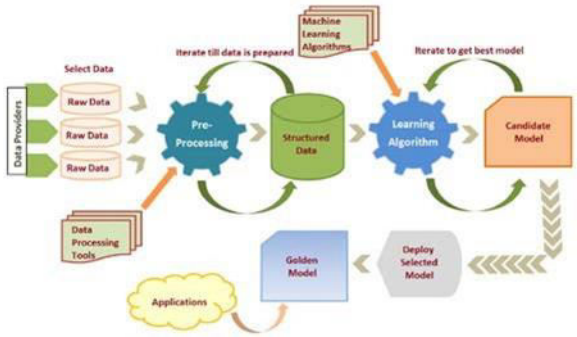

Figure 1: Machine Learning Process

2. LITERATURE REVIEW (NEED OF CLOUD PLATFORMS FOR IDS)

Traditional Intrusion detection system using data mining and machine learning methods are take a shot at data framework they are not taking a shot at cloud condition. Here gives some literature about Intrusion detection system and utilizing cloud for classification with ML algorithms[13]. Numerous decisions of distributed computing models are accessible for various outstanding burden the executives, execution and computational prerequisites. The mainstream statistical tools and environments like Octave, R and Python are currently implanted in the cloud also [5]. The significant discoveries of work [6] show the region of client maintenance got most research attention.

2.1 Machine Learning on Cloud environment for Fast Prediction in Big Data:

As the data is growing at quicker rate and turning out to be "Large Data", the calculation speed for prediction and different tasks is inevitable. This paper [7] concentrated on the particular issue of classification of network interruption traffic which is a Big Data.

Authors [3] dealt with IDS for web proxy, taking motivation from Intrusion Detection Systems that utilize AI abilities to improve anomaly detection accuracy, this paper suggests that cloud based AI can be utilized so as to recognize and characterize web intermediary use by catching bundle information and taking care of it into a cloud based machine learning web service. Authors [1] said about the cloud-based attack system. Authors add new esteemed component to the cloud-based sites and simultaneously presents new dangers for such administrations. DDoS attack is one such genuine risk. Covariance network approach is utilized to recognize such attacks. The outcomes were empowering, as indicated by confusion matrix and ROC descriptors.

Authors [8] proposed cloud IDS handle enormous progression of information bundles, classify them and produce reports effectively by coordinating information and conduct investigation to distinguish interruptions.

Authors [9] proposed Anomaly Intrusion Detection System utilizing machine learning approach for virtual machines on distributed computing. Right now feature selection from Virtual Machine Monitor to recognize anomaly in corresponding to training the system so it will learn new threats and update the model. The Proposed explore has been done on NSL-KDD'99 datasets utilizing Naïve Bayes Tree (NB Tree) Classifier and hybrid approach of NB Tree and Random Forest.

3. PROPOSED FRAMEWORK FORCLASSIFICATION:

The Proposed system which uses basic ML model with little change. The info KDDcup99 dataset is appropriately prepared and changed over into a suitable format. The AI calculations are iteratively applied in the following stage, and competitor model is resolved. These ML calculations commonly apply some statistical analysis like regression or progressively complex methodologies like decision forest. Here in the proposed structure, the ensemble methods [12] are additionally applied to the model for better precision. Finally, the model is deployed and tested on test information the depiction of actual model form using specified steps, at Microsoft Azure ML platform (Figure -2).

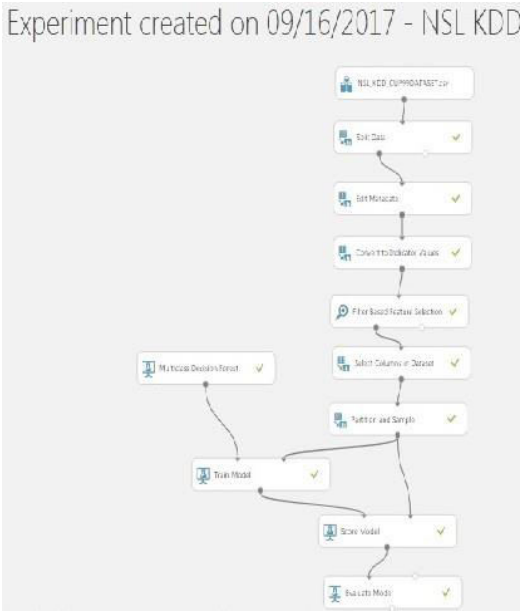

Figure 2: Model built using ML

4. SIMULATION ENVIRONMENT SETUP AND RESULT ANALYSIS:

Azure ML gives ML studio, a graphical device that can be utilized to control the procedure from start to end. It incorporates: a

lot of information pre-handling modules; a lot of ML algorithms; An Azure ML API to get to display sent on Azure. ML Studio permits a client to import datasets and information pre-handling strategies.

4.1 NSL KDD CUP 99 Dataset:

To understand these issues, another informational collection as, NSL-KDD [10] is proposed, which comprises of chosen records of the total KDD informational index. The upside of NSL KDD dataset is 1. No excess records in the train set, so the classifier won't produce any one-sided result 2. No copy record in the test set which have better decrease rates. 3. The quantity of chose records from each troublesome level gathering is conversely relative to the level of records in the first KDD informational index. The preparation dataset is comprised of 21 distinct attacks out of the 37 presents in the test dataset. The realized attack types are those present in the preparation dataset while the novel attacks are the extra attacks in the test dataset for example not accessible in the preparation datasets. The subtleties of attack classifications and explicit sorts are appeared in Table1. As per Table1, there are four attack classes in NSL KDD99 dataset:

(1) Probing/Site scanning: Scan networks to gather deeperinformation
(2) DoS: Denial ofservice
(3) U2R (User to Root): Illegal access to gain super userprivileges
(4) R2L (Remote to Local): Illegal access from a remotemachine.

4.2 Execution of Implemented Work (Experiment Steps):

The test steps that are and spoken to in Figure–2, are clarified beneath:

1)Create New Resource: Machine Learning Analytics arrangement.

2)Import/Upload the dataset.

3)Pre-process the dataset.

4)Randomly split and parcel the information into 70% preparing and 30% testing, utilizing the 'Split Data' module.

5)Identify all out properties and cast them into straight out highlights utilizing the 'Alter Metadata' module.

6)Convert to Indicator Values module to change over segments that contain all out qualities which can all the more effectively be utilized as highlights.

7)Select Columns in Dataset those are pertinent.

8)Apply Ensemble Method.

9)Apply Machine Learning Algorithm to Train the model.

10)Now Score and Evaluate the Model. The 'Assess model' likewise envisions the outcomes through confusion matrix.

4.3 Experimental Results Analysis and Discussion:

The analysis is assessed on a straightforward multicast decision forest grouping exactness parameter. Exactness is characterized as the quantity of accurately ordered cases separated by the complete number of cases:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Number of instances}}$$

The outcomes got utilizing the benchmark code by setting the multicast decision forest model got the precision of 0.9633 in explore, while the benchmark results given by rivalry overseers with is 0.50241. Here we have performed explore at cloud stage with Multiclass choice timberland strategies with

an outfit strategy. The assessment results are construed from disarray framework appeared in Figure – 3. A confusion matrix otherwise called error matrix and is utilized to portray the exhibition of a classifier (grouping model). The general accuracy acquired with our results is 0.9633, which is higher than the benchmark given.

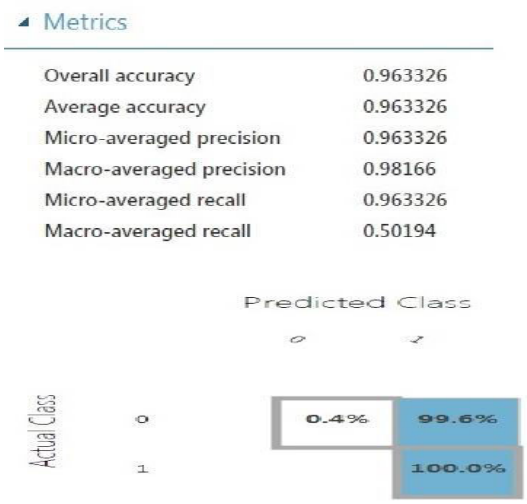

Figure 3: Confusion Matrix with Multicast Decision Forest

Figure 4: Comparison for Accuracy.

5. CONCLUSION AND FUTURE WORK:

Here, Machine Learning system has been proposed the extent that precision and acknowledgment rate for four categories of attack under different degree of run of the mill data. The purpose behind this proposed technique viably bunch unpredictable and standard data by using huge enlightening record and identify interruptions even in gigantic datasets with short training and testing times. With proposed system we get high accuracy for certain classes of attacks and distinguishing proof rate with low false alarm. At this moment, proposed an Azure ML based model for attack order. The model used Multicast Decision Forest calculation to prepare the classifier. The evaluation results show that the proposed classifier performs better in regards to exactness. We have performed try different things with multicast decision forest system. Our examinations showed the favoured exactness over benchmark. The proposed research can give potential approach to manage preparing and testing of huge data for tending to multi-class characterization issues. Right now, research will survey the structure with different ML algorithms. In future the model can be streamlined to manage imbalanced datasets from various sources and regions. In like manner, the model can be adjusted for applying on Hadoop MapReduce [11] system.

REFERENCES

[1] Anku Jaiswal, Chidananda Murthy P, Madhu BR “Prevent DDOS Attack in Cloud Using Machine Learning” Volume 6, Issue 6, June 2016 ISSN: 2277 128X

[2] Ch. Ambedkar, V. Kishore Babu, “©ARC Page 25 Detection of Probe Attacks Using Machine Learning Techniques” International Journal of Research Studies in Computer Science and Engineering (IJRSCSE) Volume 2, Issue 3, March 2015, PP 25-29 ISSN 2349-4840 (Print) & ISSN 2349-4859(Online)

[3] Shane Miller, Kevin Curran, Tom Lunne” Cloud- based machine learning for the detection of anonymous web proxies” ISSC2016.

[4] David Chappell, “introducing azure machine learning: a guide for

technical professionals", Sponsored by Microsoft Corporation, 2015 Chappell &Associates.

[5] https://portal.azure.com

[6] Daniel Pop, "Machine Learning and Cloud Computing Survey of Distributed and SaaS Solutions", https://www.researchgate.net/publication/257068169.

[7] E.W.T. Ngai, Li Xiu, D.C.K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification", Expert Systems withApplications 36 (2009)2592–2602,Elsevier

[8] Ms. Parag K. Shelke, Ms. Sneha Sontakke, Dr. A. D. Gawande "Intrusion Detection System for Cloud Computing" International Journal of Scientific & Technology Research Volume 1, Issue 4, May 2012 ISSN 2277-8616

[9]Amjad Hussain Bhat1, Sabyasachi Patra2, Dr. Debasish Jena3 "Machine Learning Approach for Intrusion Detection on Cloud Virtual Machines" Web Site: www.ijaiem.org, Volume 2, Issue 6, June 2013.

[10] MariaMuntean,HonoriuVălean,LiviuMiclea,Arpad Incze "A Novel Intrusion Detection Method Based on Support Vector Machines" IEEE2010.

[11] https://www.MulticlassDecisionForest.html

[12] Apache Hadoop Websitehttp://hadoop.apache.org/

[13] https://www.semanticscholar.org/paper/CLASSIFICATION-OF-CYBER-ATTACK-USING-MACHINE-AT-Chourasiya-Patel/8bb9a44d268200d6e83faa086ca0f6a36a85145c