iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XI **Month of publication:** November 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI November 2017- Available at www.ijraset.com

Anomaly and Missuses Detection Using IDS Technique

Anu Devi[1], Sandeep Garg[2]
[1]*Research Schollar CSE Department RPIIT Karnal Haryana*
[2]*Assistant Professor CSE Department RPIIT Karnal Haryana*

Abstract: ***In the recent years, as the second line of defense after firewall, the intrusion detection technique has got fast development. a mixture of data mining techniques such as clustering, categorization and association rule detection are being used for intrusion detection This research proposed IDS using cloud computing by integrated signature based (Snort) with abnormality based (Naive Bayes) to enhance system security to detect attacks. This research used Knowledge Discovery Data Mining (KDD) CUP 20 dataset and Waikato Environment for Knowledge Analysis (WEKA) program for testing the proposed hybrid IDS. Accuracy, detection rate, time to construct model and false alarm rate were used as parameters to evaluate performance with Naïve Bayes, Snort with J48graft and Snort***
Keywords: Data Mining, IDS, KDD, Native Bays

I. INTRODUCTION

Intrusion detection technologies began in the early of 1990's. Haystack Labs were the first one to work upon these technologies. They invented tools not only for intrusion detection technologies, also various host based products too. There were various organizations that were developing the IDS system according to their applications. But the end of this era, ASIM became successful in developing solutions related to hardware and software for protection network.

With the leap of time intrusion detection technologies become so much commercial that now day's customers have started developing Intrusion Detection technologies for their personal usage. The aim of Intrusion detection System is to defend the security to the Computer system by a layer over the defense system. IDS systems sense the misuse, breach in the security system and also the malicious or unauthorized access to the system. Although Firewalls works for the same reason but the major difference between firewalls and the IDS is IDS suspect the source of the attack and signals the alarm to the system but a firewall directly stops the communication without informing the system.

These attacks requires true concerns as they harm the data stored in system and also effect the network traffic, data packet etc. G.V. Nadiammai showed the setup phase by following diagram.

II. DATA MINING TECHNIQUES, KDD

A. knowledge discovery

Volume data used . The Knowledge discovery in databases (KDD) process is used . step in this process. analyzing data from the data mining. Summarizing it into useful information of data miming. It is the process of finding correlations or patterns amongst dozens of fields in large relational databases. The figure bellows show the different steps for extracting useful data from volume data [8].

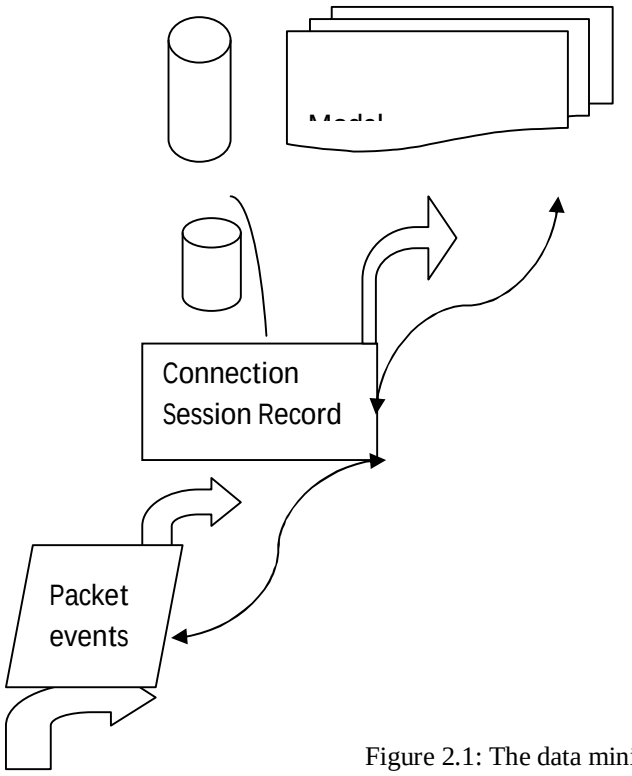


Figure 2.1: The data mining process of building ID models [3]

B. Data-mining technique

basically Some techniques such as association rules are only one of its kinds to data-mining, are pattern discovery algorithms but most In this part we used in IDS. are related information to pattern recognition..

C. KDD Cup'99 DATA SET

The proposed method is evaluated data record of extracted features from a network connection gathered during the replicated intrusions over the KDD Cup1999data. It contains CP packets to and from various IP addresses. . A connection is a sequence of TCP connection documentation consists of 41 fields. The KDD'99 used for The Third worldwide Knowledge Discovery and Data Mining Tools opposition, was simulated in a military network environment and which was held in conjunction with KDD-99.
or a classifier accomplished of distinguishing between legitimate and dishonest connections in a computer network. The contest task was to learn a projecting model This data set contains one type of normal data and 24 different types of attacks that are characterized into four types such:

1) *Denial of Service Attack (DoS):* that flood it with a waste of time traffics by the utilization of resources .
2) *Users to Root Attack (U2R):* user relation on the system with target to get to root access the attacker login a normal to the classification.
3) *Remote to Local Attack (R2L):* a piece of equipment on a network which he doesn't have any accurate to access to system is when the attacker challenge to get a local access as a user of a piece of equipment on a network which he doesn't have any accurate to access to system.
4) *Probe Attack:* This attack is in relation to collecting in turn from a network of computers for a later use.

III.RELATED WORK

A. *Subaira. A.S et al [2014]:* On the top of JXTA, a layer of middleware is recognized and it consists of four components: security agent module, task scheduling module, data distribution module and task coordinator module. Cloud SEC architectural topographies for the peers of task coordinator and security agent are provided by the first two modules. Data sharing facility

and task scheduling are provided by the latter two modules. The architecture is organized in a virtualized, extremely configurable trial cot. Applied and fault-tolerant are the introductory results of the Cloud SEC prototype [7].

B. *Charles A. Fowler et al [2014]:* Different modes of data transfer and communication [48means (e.g. satellite communication) might need to take into account. Huge amount of data transfer is a common anticipation in a cloud environment; the communication technology used along with the security concerns of the adapted communication technology also becomes a security concern for the cloud computing approach. The broadcast nature of some communication technology is a core concern in this regard.[5]

C. *S.V. Shirbhate et al [2014]* : The study, analysis clustering is one of the needs for machine learning algorithms to be useful to large scale data and exploration of recent Development of data mining application such as classification and will lead to obtain the direction of future research.. The aim methods for a set of large data are to investigate the performance of different clustering. The algorithms are tested on intrusion detection data set. The KDD data set subsequently, clustering technique that has the potential to significantly get better the conventional method is used for this purpose. will be suggested for the use in intrusion detection in mobile network data. The aim of this Different clustering methods using WEKA tool for intrusion detection. Intrusion is defined work is to investigate the performance of as “the act of wrongfully incoming upon, grasping, or taking possession of the property of another”. Intrusion the computers from unauthorized or malicious actions detection system protects the computers from unauthorized or malicious actions. [11].

D. *Nadya El MOUSSAID et al [2015]*: At that time to discriminate between the genuine packets and the dose packets a filter methodology was used. In this methodology all provision requirements are initially impelled to SBTA (SOA based traceback approach) for marking them. Then a haze trace back mark tag within the header of the message is placed by SBTA and this will be referred to the Web Server. If the communication found to be usual, for processing, it will be directed to the application manager. In this method, to filter out attack messages, a cloud filter was employed. Then false alarm ratio and detection proportion are considered. The outcome demonstrates that Cloud filter have a great detection ratios and less alarm ratio. Thus, grouping of SBTA and Cloud filter over cloud system can be effective technique for distinguishing and recognizing dose messages. The number of attack packet increases protective performance decreases progressively, is the deficiency of this method.[2]

IV. INTRUSION DETECTION SYSTEM BASED ON DATA MINING

A. *Security and Privacy*

It deals with securing the stored data and to monitor the use of the cloud by the service providers. This challenge can be addressed by storing the data into the organization itself and allowing it to be used in the cloud. So the security mechanisms between the organizations and the cloud need to be robust.

B. *Service Delivery and Billing*

The service level agreements (SLAs) of the provider are not adequate to guarantee the availability and scalability as it is difficult to assess the cost involved due to dynamic nature of services.

C. *And Portability*

As the cloud environment is highly dynamic to user requests and due to the concept of virtualization, the leverage of migrating in and out of the resources and applications should be allowed. Also, switching providers should switch between clouds as per their need, and no lock-in period should exist.

D. *Reliability and Availability*

Cloud providers still lack in round-the-clock service which results in frequent outages. Therefore, it becomes important to monitor the service being provided using internal or third party tools.

E. *Automated service provisioning*

A key feature of cloud computing is elasticity; resources can be allocated or released automatically. So a strategy is required to use or release the resources of the cloud, by keeping the same performance as traditional systems and using optimal resources.

F. *Performance and Bandwidth Cost*

Businesses can save money on hardware but they have to spend more for the bandwidth. This can be low cost for smaller applications but can be significantly high for the data-intensive applications.

G. Virtual Machines Migration

With virtualization technology, an entire machine can be taken as a file or set of files. To unload a heavily loaded physical machine, it is required to move a virtual machine between physical machines. The main objective is to distribute the load in a datacenter or set of datacenters. Then a strategy is required to dynamically distribute load when moving virtual machine to avoid bottlenecks in Cloud computing system.

H. Energy Cost

Cloud infrastructure consumes enormous amounts of electrical energy resulting in high operating costs and carbon dioxide emissions [11].

V. PROPOSED WORK

Due to the increase of internet technology in the past few years network traffic has also been increased to a great extent. Data travelling over the network has become a hot topic for the researchers because security is concerned for this data.

Data travelling over the network is broadly classified in to two categories, normal data and anomaly. Anomaly detection is the goal of Intrusion Detection System (IDS). Different patterns can be drawn on the basis of the user's usage over the network. These patterns can be grouped together according to the similarities in them. Data clustering. So putting similar data into groups is called there are many techniques; we are using k-mean clustering which is an unsupervised learning algorithm

A. J48 decision tree classifier

C4.5 algorithm is the decision tree based. With this technique a tree is decision tree the internal nodes of the tree denotes a the topmost node is the root node Algorithm test on leaf node holds a class label and [1] J48:

PARTICIPATION
R // Training data
PRODUCTION
P // Decision tree
R
{
P = Null;
P = Create root node
P = Add arc to root node
R = splitting predicate to R;
P'= Create label with appropriate class leaf node
Else
P' = KTBUILD (R);
P = Enhance P' to arch;

B. Naïve Bayesian classifier:

Classification of Bayesian represents classification a supervised learning method statistical method for as well as. Strong independence assumption simple probabilistic classifier based on Bayesian theorem with. It is particularly suited when the dimensionality of input is high.

Bayesian formula can be written as [4]: $Q(P \mid A) = [Q(A \mid P) * Q(P)] / Q(A)$ The Bayes's rule is work a hypothesis or an event (H) can be predicted based on some evidences (E) that can be observed from the Bayes's rule.

Percent data set available for research on network interruption used KDD Trains 20 detection. 1999 KDD intrusion detection competition and is called the KDD Cup 99 data. And for analyzing data we have used WEKA, open source software. WEKA" short for the Waikato Environment for Knowledge Analysis, WEKA a practical machine learning tools package ". Set of machine learning algorithms for solving real-world data WEKA is extensible and has become a mining problem. It runs on almost every platform. intrusion detection system (IDS the core problem of an the exactness of the detection results, in this Dissertation K-Mean Algorithm for Network Intrusion exposure a Data Clustering Using that possess highly accurate intrusion detection capability.

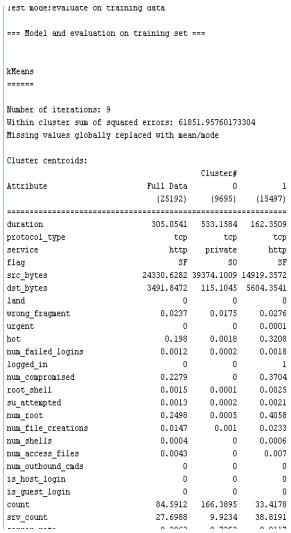
1) It first describes a detailed description of the IDS components the information set used and IDS architecture, followed by. Then, it illustrates the K-mean algorithm
2) the dataset used for the large training data set for intrusion detection is presented evaluation is described .
3) Describes how the TC-K-Mean based IDS the intrusion-tolerant framework, can be complete intrusion-tolerant by introducing an intrusion-tolerant mechanism.
4) Activities in intrusion detection are evaluated by experiments. algorithm based on the *k*-mean clustering for analyzing program
5) KDD Trains 20 Percent audit data have shown preliminary experiments.
6) Effectively detect intrusive program behavior

VI. SIMULATION RESULT

A. *Analysis Data Set*

The main aim of the thesis was to evaluate to find out all the attacks in a database of transaction. This started with studying various existing techniques. After literature review it was clear that every algorithm has applied, but combinations are still on study. Many combinations have applied, but somehow efficiency always lags behind. So Aim decided as designing and development of a prototype application for intrusion detection system. Capturing logs was the initial and most difficult task, because to implement design we should have enough databases so that we can easily measure its efficiency. Another challenge came in between thet type of input to be used. In thesis [they have mentioned a database centric architecture using KDD dataset. KDD data set is considered as dummy dataset for the transactions It provides both training data as well as test data. TP is the number of positive cases classified correctly, FN is the number of positive cases classified as negative, FP is the number of negative cases classified as positive, and TN is the number of negative cases classified correctly. In Table 2, the TP is 10298, FN is 1445 FP is 1177 and TN is 12272

Accuracy = (TN+TP/ (TN+TP+FN+FP))
recall = TP/(TP+FN)
F-measure = (2*recall *precision) I (recall + precision),
sensitivity = TP/(TP+FN) = recall,
specificity = TN/(FP+ TN).
Step1 Now we find out Centroid,We select weka 3.6.2 in cluster.



```
Test mode:evaluate on training data

=== Model and evaluation on training set ===


kMeans
======

Number of iterations: 9
Within cluster sum of squared errors: 61851.95760173304
Missing values globally replaced with mean/mode

Cluster centroids:
                                        Cluster#
Attribute                  Full Data          0          1
                             (25192)     (9695)    (15497)
=============================================================
duration                   305.0541   533.1584   162.3509
protocol_type                   tcp        tcp        tcp
service                        http    private       http
flag                             SF         S0         SF
src_bytes                24330.6282 39374.1009 14919.3572
dst_bytes                 3491.8472   115.1045  5604.3541
land                              0          0          0
wrong_fragment               0.0237     0.0175     0.0276
urgent                            0          0     0.0001
hot                          0.198      0.0018     0.3208
num_failed_logins            0.0012     0.0002     0.0018
logged_in                         0          0          1
num_compromised              0.2279          0     0.3704
root_shell                   0.0015     0.0001     0.0025
su_attempted                 0.0013     0.0002     0.0021
num_root                     0.2498     0.0005     0.4058
num_file_creations           0.0147      0.001     0.0233
num_shells                   0.0004          0     0.0006
num_access_files             0.0043          0      0.007
num_outbound_cmds                 0          0          0
is_host_login                     0          0          0
is_guest_login                    0          0          0
count                        84.5912   166.3895    33.4178
srv_count                    27.6988     9.9234    38.8191
```

Fig6.1: Select Cluster

Perform robust tests based on the custom configurations; and detection rate, accuracy, false alarm The performance of intrusion detection systems of and time build to model.
Result for classification using j48 j48 working for module. When applying j48 on kdd dataset result are as given below

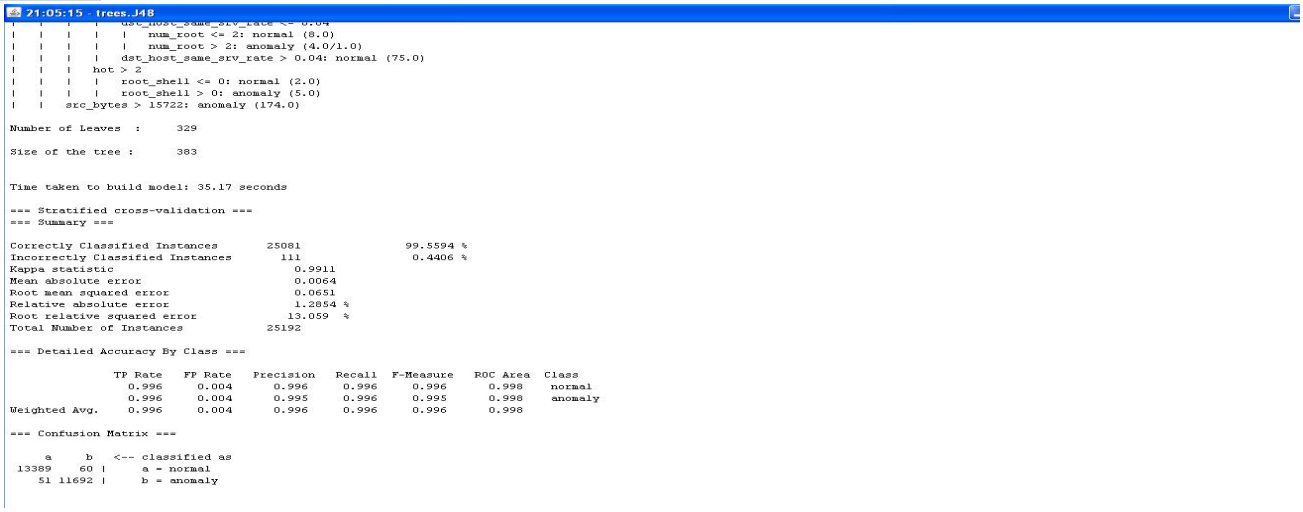


Figure 6.2: Confusion matrix for J48

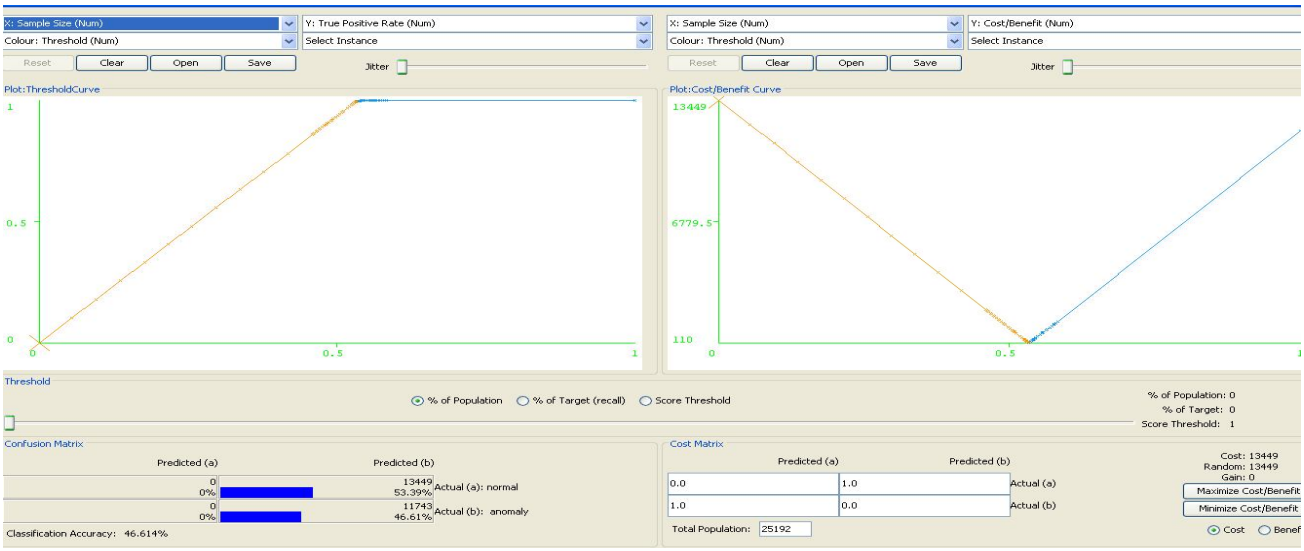


Figure6.3: Cost analysis of J48 for class Normal

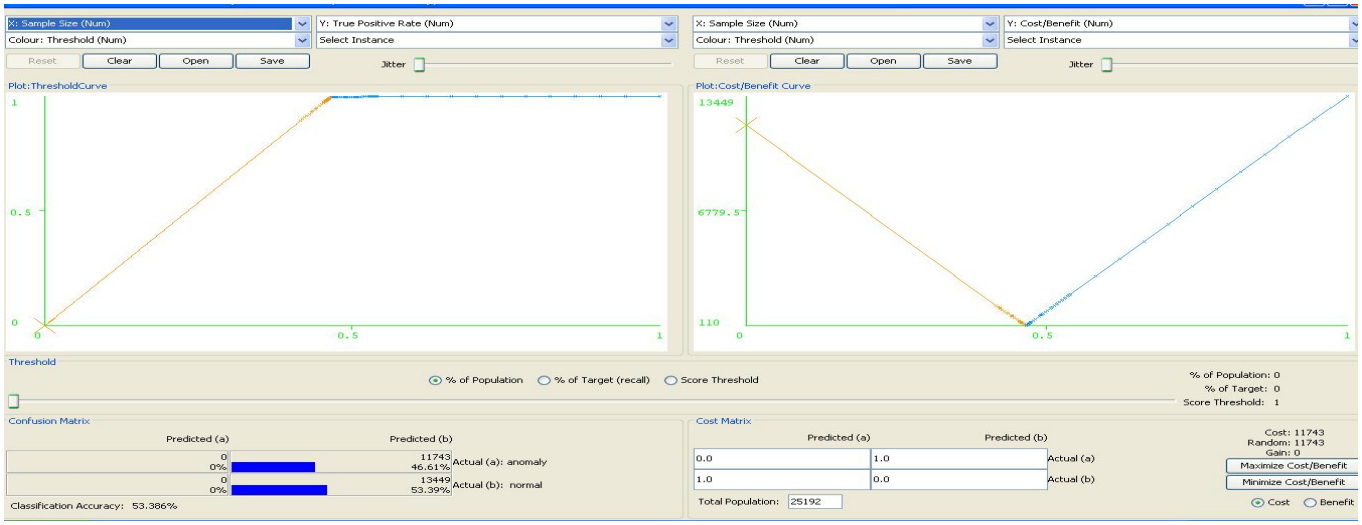


Figure6.4: Cost analysis of J48 for class Anomaly

B. Result For Classification Using Naïve Bayes

Classifier Naïve Bayes algorithm is applied on KDD dataset . result shown as below on figure .

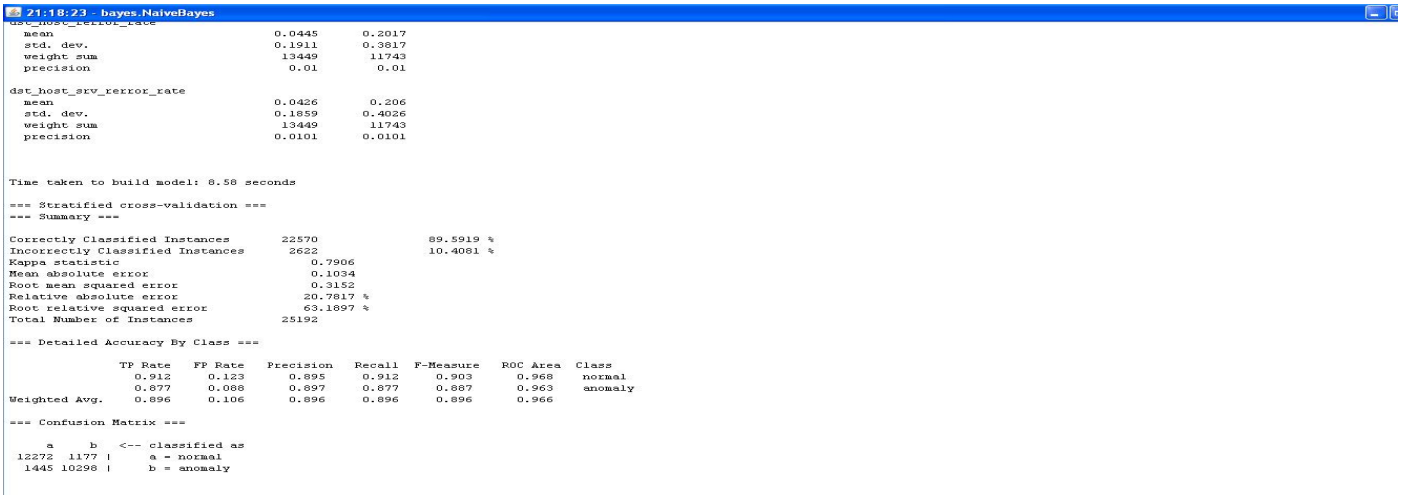


Figure6.5: Confusion matrix for native bays

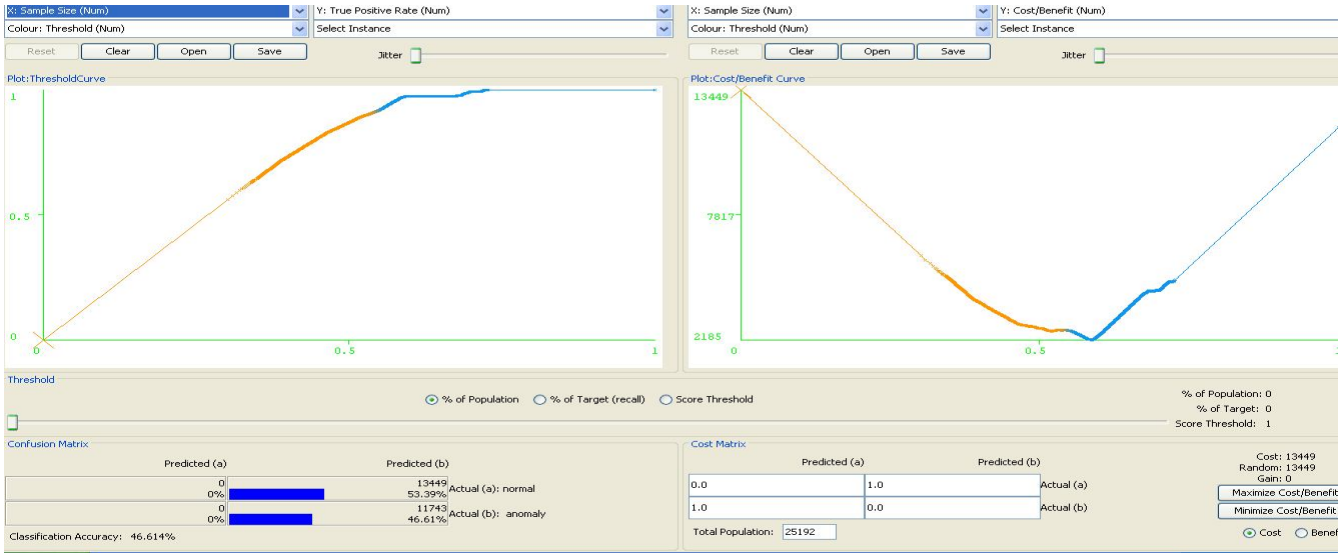


Figure6.6 : Cost analysis of Native Bays for class Normal

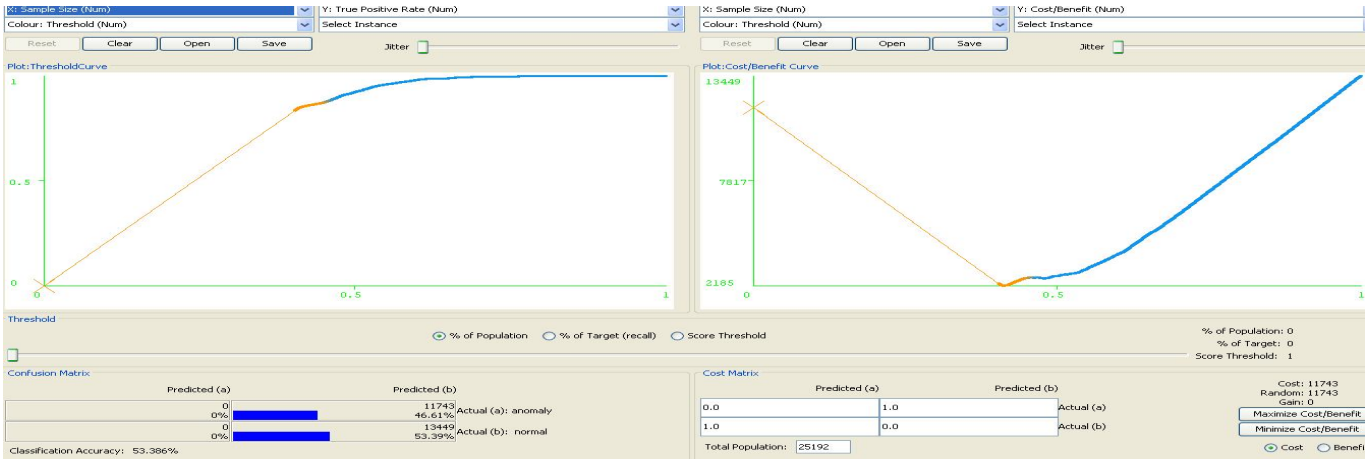


Figure6.7 : Cost analysis of Native Bays for class anomaly

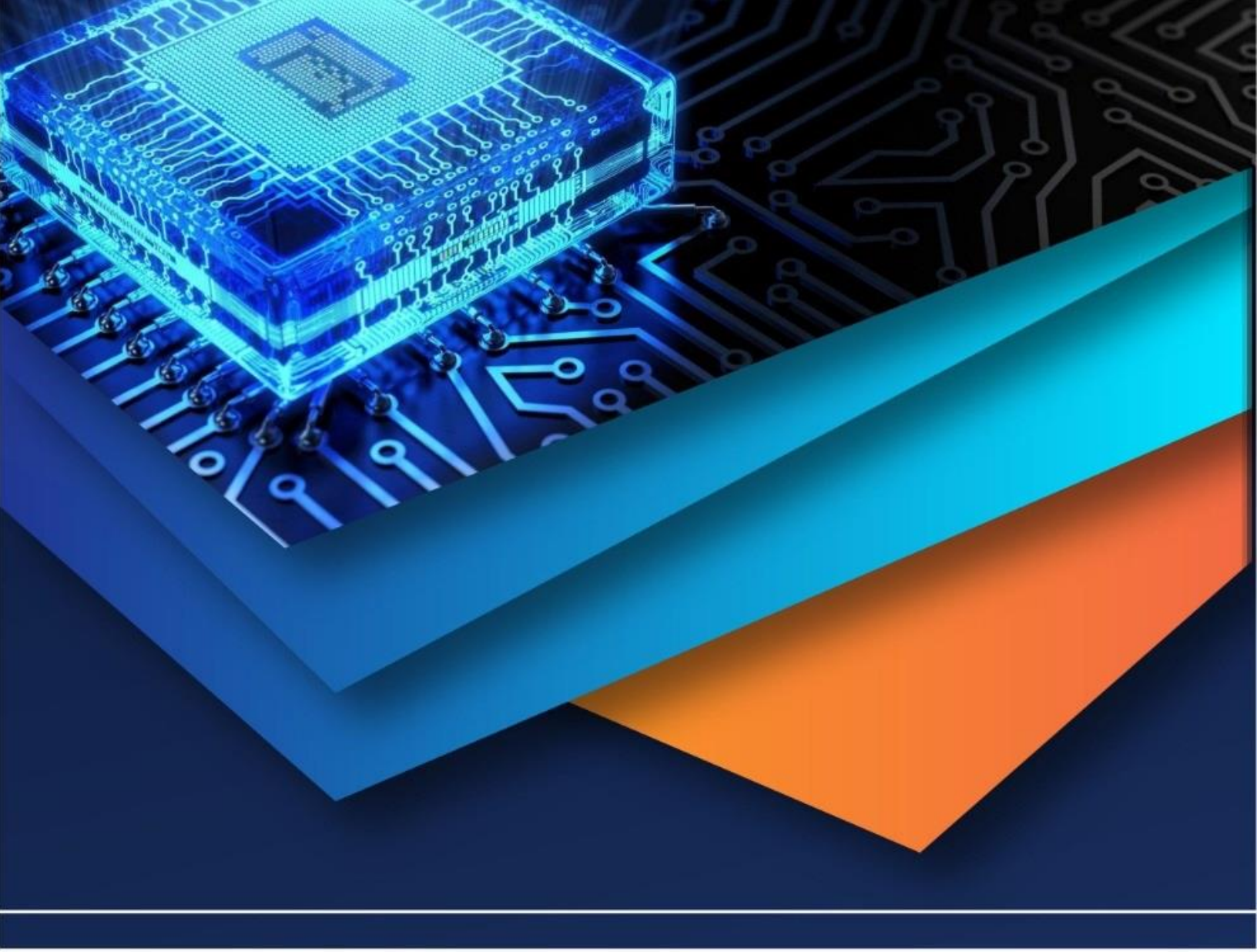
generate classifier using Bayesian technique and J48 classifiers with default settings .data mining tools [19] are used and five-fold cross-validation. Bayesian classifier, shows its confusion matrix and shows the results provide more informative evaluation of classifier performance is calculate in in classifier that the following when dealing with confusion matrix formula : recall, precision (prec.), F-measures, sensitivity, and specificity [5], which are defined as]

VI. CONCLUSION

In this paper provided a detailed study of IDS that occurred in Data Mining. . Thus there will likely be obstacle in developing an effective solution Intrusion detection systems using have been an area of active research for above fifteen years. Current industrial intrusion detection systems make use of misuse detection. As such, they completely are short of the ability to detect new attacks. It is impossible to prevent security violation completely by using the exciting security technology. Accordingly, Intrusion Detection is an important component of network security

REFERENCES

[1] Su-Yun Wua, Ester Yen " Data mining-based intrusion detectors" Crown Copyright _ 2008 Published by Elsevier Ltd. All rights reserved Corresponding author "IEEE 2008
[2] G.V. Nadiammai, " Effective approach toward Intrusion Detection System using data mining techniques "IEEE 2013.
[3] Ahmed Patel, MonaTaghavi, KavehBakhtiyari "An intrusion detection and prevention system in cloud computing: A systematic review "IEEE 2012.
[4] Kalpana Jaswal, Seema Rawat,Praveen Kumar "Design and Development of a prototype Application for Intrusion Detection using Data mining" ©2015 IEEE.
[5] Charles A. Fowler and Robert J. Hammell "Converting PCAPs into Weka Mineable Data "IEEE 2014.
[6] Sunil Kumar Khatri "Intrusion Detection Using Data Mining "IEEE 2014.
[7] Subaira.A.S, Mrs. Anitha.P " Efficient Classification Mechanism for Network Intrusion Detection System Basedon Data Mining Techniques"2014.
[8] Nadya EL Moussaid, Ahmed Toumanari Essi, "Overview of Intrusion Detection Using Data-Mining and the features selection"IEEE 2015.
[9] Bace, Rebecca G."NIST special publication on intrusion detection systems"2002
[10] Ertoz, L., Eilertson, E., Lazarevic, A., Tan, P., Srivastava, J., Kumar, V."The MINDS – minnesota intrusion detection system. Next generation data mining. 2004.
[11] S.V. Shirbhate, Dr.S.S. Sherkar ,Dr. V.M.Thakare " Performance Evaluation of PCA Filter In Clustered Based Intrusion DetectionSystem"2014. 2014 International Conference on Electronic Systems, Signal Processing and Computing Technologiessan
[12] Chirag N. Modi, Dhiren R. Patela, Avi Patelb, Muttukrishnan Rajaraja " Integrating Signature Apriori based Network Intrusion DetectionSystem (NIDS) in Cloud Computing"2012
[13] Shengyi pan, Thaomas marris " Developing a Hybrid Intrusion Detection System Using Data Mining for Power Syetem"IEEE 2015. 2015 IEEE.
[14] Mr. Mohit Sharma, Mr. Nimish Unde, Mr. Ketan Borude, A Data Mining Based Approach towards Detection of Low Rate DoS Attack"2014. T.R. Gopalakrishnan Nair, K.Lakshmi Madhuri" Data Mining USsing Hierachical Virtual K-Means Apporaoch Intergrating Data Fragments In Cloud Computing Envioroment "IEEE 2011.
[15] Kapil Wankhade, Sadia Patka " An Efficient Approach for Intrusion Detection Using Data Mining Methods"IEEE 2013.
[16] Ketan Sanjay Desale, Chandrakant Namdev Kumathekar, Arjun Pramod Chavan "Efficient Intrusion Detection System using Stream Data Mining Classification Technique"IEEE 201
[17] R. Robu and V. Stoicu-Tivadar," Arff Convertor Tool for WEKA Data Mining Software" IEEE 2010
[18] Amrit Priyadarshi M.M. Waghmare " Use of rule base data mining algorithm for Intrusion Detection" 2015 IEEE.
[19] Byung-joo Kim " Kernel Based Intrusion Detection System" IEEE 2005.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY