International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Performance Behavior of Intrusion Detection System (Ids) Based on Ensemble Base Classifier (EBC)

Rajesh Phursule

Abstract*: *There is a tremendous growth in the area of information technology due to which, network defence is also facing major challenges. The conventional Intrusion Detection System (IDS) is not able to prevent the recent attacks and malwares. Hence, IDS which is an essential component of the network needs to be protected. Data mining introduce to the process of separate hidden, previously unknown and useful information from huge databases. Data Mining based Intrusion Detection System is combined with Multi-Agent System to improve the presentation of the IDS. We combine the classifiers which is the widespread approach, to increase the accuracy of a single classifier. For experimentation purpose, we use a benchmark intrusion detection dataset, which is KDDCup'99 and the accuracy of the classifiers were estimated using 10-fold cross validation method. In this work, we use the feature selection methods, namely Flexible mutual information based feature selection (FMIFS) and hybrid feature selection algorithm (HFS) to evaluate the importance of features. This work provides Support Vector Machine (SVM), Nave Bayes (NB) and Feed Forward Neural Network (FFNN) to classify attack and normal threads as well as to improve the accuracy we ensemble all classifier into single hybrid classifier using Bagging algorithm. The proposed hybrid approach achieves an accuracy rate of 95.11

Keywords* : *IDS, KDDCup'99, FMIFS, HFS, SVM, FFNN

I. INTRODUCTION

An intrusion detection system observes network traffic for suspicious activity and alerts the system or network administrator in order to take evasive action. In recent years, intrusion detection method and key technology has become one research focus in network security field.

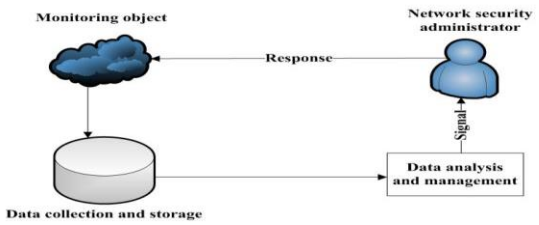

Fig. 1. IDS structure

Manuscript published on 30 September 2019.
*****Correspondence Author(s)
Dr. Rajesh Phursule, Associate Professor, Department of Computer Science, ICOER, Wagholi, Pune, India.

A. Privacy Preserving Data Mining P P DM

In recent years, many PPDM methods have been but there is no standardization in these appear optimized results while

preserving the subjects efficiency
1) The distribution of the basic data
2) The modi cation of the basic data
3) Mining method being used
4) If basic data or rules are to be hidden an
5) Additional methods for privacy preserve

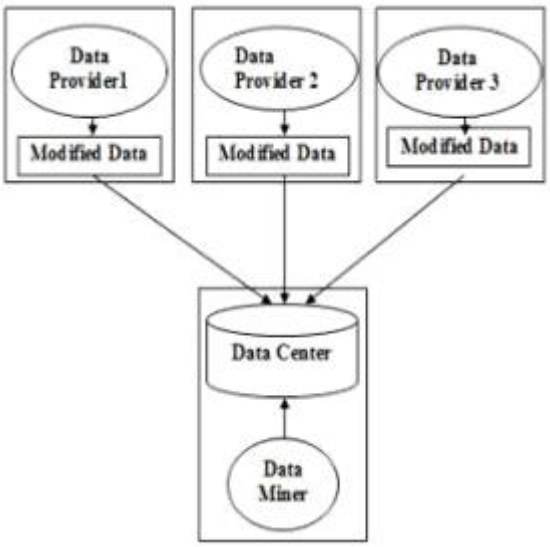

Fig. 2. PPDM based on data publishing scenario

This shows that from a technical viewpoint methods and procedures in the perspective of be used.

B. The ensemble framework

A typical ensemble technique for classification undertakings contains the accompanying building squares:
1) Preparing set - A marked dataset utilized for group preparing. The preparation set can be depicted in an assortment of dialects. Most much of the time, the occurrences are portrayed as quality esteem vectors. We utilize the documentation A to mean the arrangement of information traits containing n properties: A = f a1,...,ai, g and...y,anto speak to the class of variable the measures or and what reason every one in every of them the objective quality.
2)Base Inducer - The inducer is an acceptance life dataset calculation received from the human resource subject and is that acquires a preparation set and structures verified on set a of classibenchmarker datasets.

that speaks to the summed up connection J.Vaidyabetween.al[9]. This paper presents an green protocol for information characteristics and the objective securely determining trait the scale.
Given a set of intersection, and shows a chance to speak to an inducer. We how utilize this could be the used to generate association regulations documentation M = I(S) for speaking to where in a classifier multiple events M have one of a kind (and personal) which was incited by inducer I on a preparation records approximately set S the identical set of individuals.

II. LITERATURE SURVEY

Data mining can damage singular security [2]. This is it's utilized for research pr another objectives because of potential abuse of private or delicate data construed C.
Clifton et. al [11], this paper present some additives of from data mining comes about. Protection conservation in datathis sort of toolkit, and said that how they can utilize to solve numerous security-protecting data mining problem. W. Fan et.al [12], in this innovation work, the author investi-gates a simple model depends on summary of the training datasets by using a set of the random decisions trees. It required the lower data of the user, however it's applicable for the each.
We have practical's on a large challenges involves that proba-bility method working very well, most selection of benchmarks datasets for non-parametric regression, and larger non-linear stochastic issue
Mining has developed as a noteworthy exploration field in light of the omnipresence of demographic and touchy information [5-10].

S. Mukherjee et.al [5], in this paper they suggest technique function vitality primarily based discount approach, to identify vital decrease input feature. The author experiments on the green classifier na¨ıve bayes on decreased datasets for IDS. The experiments output display that, determines on reduced attribute provide better allover performance to scheme IDS this is green and efficient for network IDS
M. Aghamohammadi et.al [6], this paper is set Intrusion Detection. The principle intention of IDS (Intrusion Detection gadget) is to shield the device by analyzing user's behaviors and conduct whilst they're working with device, come across behaviors that don't match with formerly discovered regular behaviours styles and raise a warning. Support Vector system (SVM) is a category method that used for IDS in many researches. They evaluate overall performance of SVM and Multi-degree support Vector system (MLSVM) as a new version of SVM on difficult intrusion detection information set primarily based onKDD'99 with call NSL-KDD. Our experiments indicate that MLSVM is extra appropriate for this records set in place of SVM. Q.Liu et.al [7], on this paper, creator recommends a progressed classifier the usage of a single-hidden layer feed forward neural community (SLFN) skilled with severe learning machine. The unconventional classifier first makes use of essential thing analysis to reduce the feature measurement after which selects the most excellent structure of the SLFN based on a new localized generalization blunders model in the predominant aspect space. Experimental and statistical consequences at the NSL-KDD.

III. OBJECTIVE

To study and investigate an ensemble approach of base classifiers - An ensemble consists of a bunch of individually trained classifiers whose predictions are integrated when classifying novel instances.

- To study bagging and boosting methods for producing ensembles
- To study Intrusion Detection Systems, Direct Marketing, and Signature Verification using existing classification algorithms.
- To study the threats imposed by data mining techniques to privacy/security and possible remedies.
- To study the statistical effect of distributed data sources to privacy and security.
- To study Data quality, privacy, and security measures.
- To study the relationship between data mining and data warehousing.
- To solve the current research problem of an analytical study on classifier based text approaches for data mining methods, this research work focused on three important aspects such as

1) Data Acquisition and Normalization
2) Data Mining Text Feature Extraction and Classifier SVM
3) Classifier Nave Bayes and FFNN

IV. METHODOLOGY

The purpose of the present research is to study the classifier based text approaches for data mining methods. The researcher will identify techniques that were developed. Hence the purpose of this methodology is illuminating the concept of classifier based text approaches for data mining methods. The overall system architecture as shown in figure 3

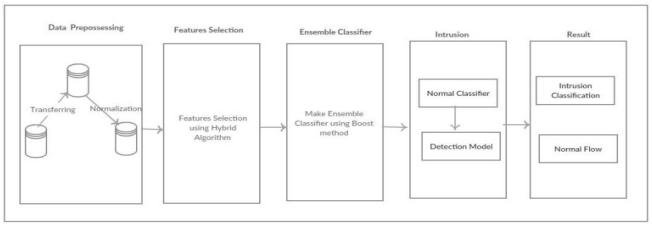

Fig. 3. System Architecture

A. Ensemble Methods

A sequence of classifiers that complement one another is provided by this system. The algorithm considers that the education set consists multiple of m, classified as -1 or +1. The classification of a brand new instance is made with the aid of the use of vote record on all classifiers Ci where i =1..t, each having a weight of fit. The fundamental AdaBoost, offers with binary classification. Variations of the AdaBoost set of rules (AdaBoost.M1, AdaBoost.M2) Freund and Schapire (1996) describe, which may be vary of their coping with of multiclass classification difficulties and equivalent for binary classification.

The classification of a modern day example is finished in keeping with the following equation.

$\sum_t(X) = sign(t)Ct(x)$

B. The AdaBoost Algorithm

Input: N (the total number of iteration), I (Inducer), S (training dataset) Output: Cn; n; n = 1; ::; N
Build Classifier Ct using I and distribution Dt. Normalize Dn + 1 to be a proper distribution n++
until n
N
end AdaBoost

C. Bagging Algorithm

Bagging is a vote casting scheme wherein n fashions, usually of same kind, are constructed. For an unknown example, every version's predictions are recorded [7, 9].
Input: N (the total number of iteration), I (Inducer), S (training set), d (weighting distribution) Output: M_n; $n = 1$; ::::; N
S_n n with random weight drawn from d Build Classifier M_n using I and distribution S_n n + + until n
N
end Bagging

D. Data processing and mutual information based features selection - (Hybrid FSA)

Input:Feature set F = f i; i = 1; ::::; n, Dataset d
Output : S=selected feature subset

F. Support Vector Machines

The history and improvement of assist vector machines can be observed in [394]. SVMs have attracted interest from the clinical imaging community because of a number of theoretical and computational merits derived from the statistical getting to know theory. The mathematical background of SVM is as follows, do not forget $f(x_i; y_i)g^N{}_{i=1}$ and $y_i = f\,1; +1g$
A linearly separable binary class hassle in which i x is an n-size vector and that i y is the label of the class that the vector belongs to. SVM separates the two instructions of factors through a hyper plane wT x + b = 0

G. Artificial Neural Network

They're Feed beforehand neural community, Elman neural community, Radial foundation neural community and Generalized Regression neural network. Feed in advance neural community additionally may be defined due to the fact the multilayer perceptions. Arithmetically the functionality of a hidden neuron is categorized with the useful resource of Σn ($_{j=1}$ $w_j x_j$ $+b_j$) Wherein weights (w_j ; b_j) are denoted with the arrows feeding into the neuron. It is most common and broadly used characteristic. Network is east to keep. A feed forward neural community is a synthetic neural community wherein connections a number of the gadgets do no longer form a cycle. [1]

H. Na¨ıve Bayes

The work said in investigate the situation beneath which the Na¨ıve bayes classifier plays nicely and why. It states that the mistake is a result of three elements: bias, variance and training statistics noise. Education information noise can most effective be minimized by using the usage of way of choosing

V. RESULTS AND DISCUSSION

The performance of the classifier has been firstly evaluated using three different dataset method and algorithm runs on each dataset. The experimental results of the classification algorithms with using feature selection method and ensemble three classifiers for a KDD datasets are presented below graphs.

A. Analysis of SVM Classifier

After classification is done we see the performance of SVM using graphs. As a part of this work, the SVM classifiers are used by applying SVM classification algorithm and the effectiveness of features obtained by FMIFS. All result rep-resents and discuss below. Attack classification graph show classification of attacks in terms of normal, probe, dos, r2l, u2r.

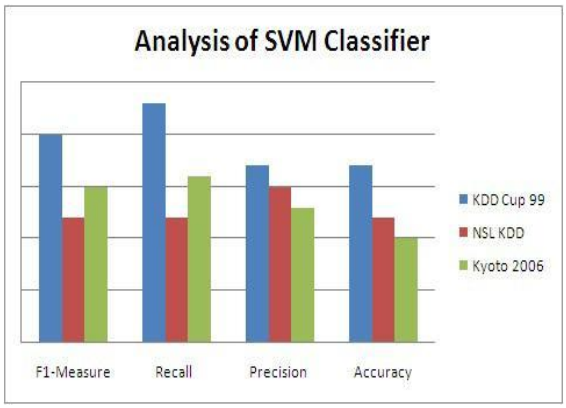

Fig. 4. Analysis of SVM Classifier

From figure 4, F-measure performance of SVM with KDD Cup 99 Dataset, NSL KDD and Kyoto as 90%, 82 % and 85% respectively, recall performance of SVM with KDD Cup 99 Dataset, NSL KDD and Kyoto as 93%, 82% and 86% respectively, the precision performance of SVM with KDD Cup 99 Dataset, NSL KDD and Kyoto as 87%, 85 % and 83% respectively and accuracy performance of SVM with KDD Cup 99 Dataset, NSL KDD and Kyoto as 87%, 82% and 80% respectively.

B. Analysis of Na¨ıve Bayes Classifier

From figure 5 f-measure performance of Na¨ıve Bayes with KDD Cup 99 Dataset, NSL KDD and Kyoto as 90%, 88% and 80% respectively, the precision performance of Na¨ıve Bayes with KDD Cup 99 Dataset, NSL KDD and Kyoto as 84%, 87% and 84% respectively, recall performance of Na¨ıve Bayes with KDD Cup 99 Dataset, NSL KDD and Kyoto as 98%, 84% and 87% respectively and accuracy performance of Na¨ıve Bayes with KDD Cup 99 Dataset, NSL KDD and Kyoto as 85%, 80% and 84% respectively.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

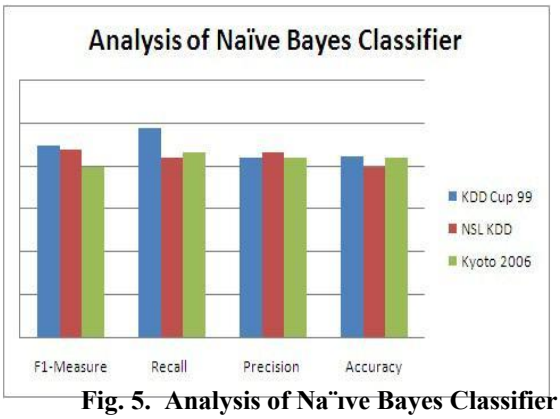

Fig. 5. Analysis of Na¨ıve Bayes Classifier

C. Analysis of Ensemble Classifier

Now, this section discuss performance of ensemble classifier with the comparison of SVM, NB, and FFNN. Refer

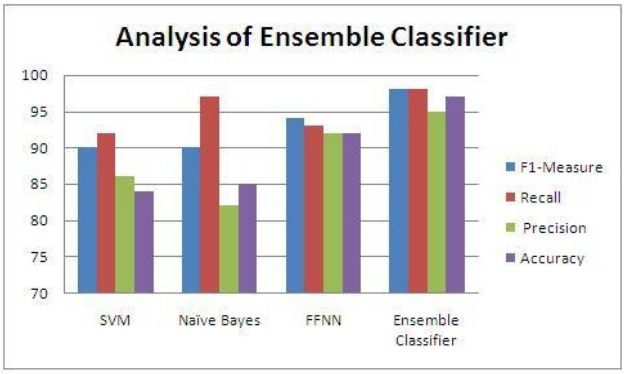

Fig. 6. Analysis of EC Classifier

D.. Contribution of features in enhancing the IDS performance

The post computations after creating effective and efficient IDS is to defend the proposed reduced subset of 12 features against the rest of the 29 features having a selection count of six. Selection count is the frequency of selection of features from six correlation based feature selection algorithms. Refer figures 7 and 8

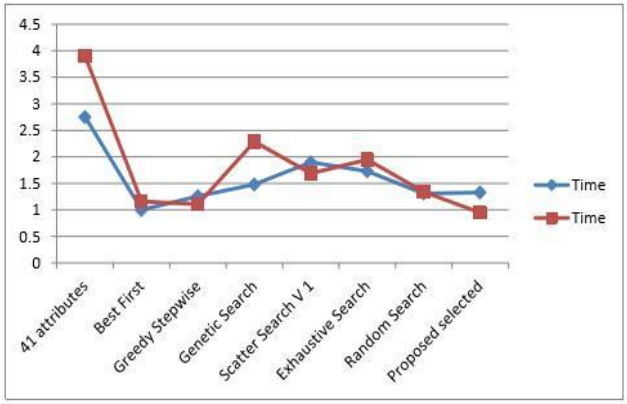

Fig. 7. Graphical Representation of Accuracy

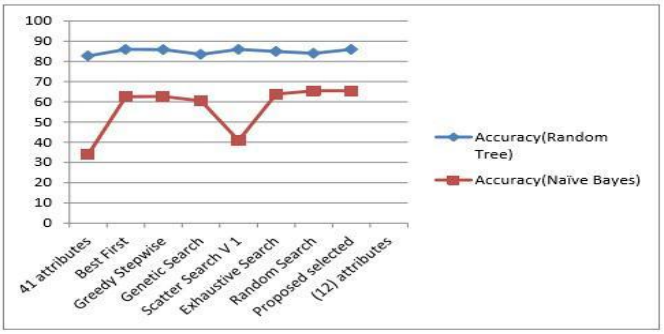

Fig. 8. Graphical Representations of Computational Time

The planned hybrid classifier produces best results using the features of hybrid methods. Also, the performance of proposed method is compared with traditional classifiers: Nave Bayes N B, Support Vector Machine SV M, Feed Forward Neural Network . The proposed hybrid approach achieves an accuracy rate of 95.11 %, detection rate of 98.67 %. Future work can be extended using various bio-inspired algorithms for feature selection and classification with real-time network datasets. The effectiveness of IDS can be still improved to handle newly rising attacks for achieving 100 % detection rate.

VI. CONCLUSION AND FUTURE WORK

This work spotlight on classification algorithms and feature selection methods to get better the detection presentation and to decrease the time required to carry out the computations for intrusion detection systems.

This work dealt with the setback of feature selection, which is of great significance in intrusion detection due to towering dimensional data. To get better the accurateness rate of the detection system, the classifiers are hybridized and weigh up on the standard intrusion detection dataset, KDDCup'99 from UCI machine learning warehouse. Data mining methods and classification advances have been applied for intrusion detection system to differentiate normal and irregular performance.

REFERENCES

1. S. Pontarelli, G. Bianchi, S. Teofili, "Traffic-aware design of a high speed fpga network intrusion detection system" Computers, IEEE Transactions on 62 11 2013 2322-2334.
2. D. S. Kim, J. S. Park, "Network-based intrusion detection with support vector machines, in Information Networking", Vol. 2662, Springer, 2003, pp. 747-756.
3. S. Mukkamala, A. H. Sung, A. Abraham, "Intrusion detection using an ensemble of intelligent paradigms" Journal of network and computer applications 2282005 167-182.
4. F. Amiri, M. RezaeiYouse , C. Lucas, A. Shakery, N. Yazdani, "`Mutual information-based feature selection for intrusion detection systems", Journal of Network and Computer Applications4201134 1184-1199.
5. S. Mukherjee, N. Sharma, "Intrusion detection using naive bayes classi?er with feature reduction",Procedia Technology2012119-4128.
6. M. Aghamohammadi, M. Analoui, "A comparison of support vector machine and multi-level support vector machine on intrusion detection", World of Computer Science and Information Technology Journal7 2 2012 215-219.
7. Q. Liu,J. Yin,V. C. Leung,J. H. Zhai,Z. Cai,J. Lin, "Applying a new localized generalization error model to design neural networks trained with extreme learning machine", Neural Computing and Applications 2014 1-8.

8. Arbel R, Rokach L, "Secure Set Intersection Cardinality with Appli-cation to Association Rule Mining", Pattern Recognit14:1619Lett-27 1631, 2006.
9. J. Vaidya and C. Clifton, "Classi er evaluation under limited resources", published in J. Computer Security, 2005.
10. Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining", J. Cryp-tology, vol. 15, no. 3, 2002, pp. 177-206; http://www.cs.biu.ac.il/ lin dell/abstracts/id3.html.

AUTHORS PROFILE

Dr. Phursule Rajesh, Associate Professor, Department of Computer Science, ICOER, Wagholi, Pune, India.His research interest includes: Data Mining, System Programming.
He has total 15 years of teaching and research experience.