ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN659

Intrusion Detection System with Ensemble Machine Learning Approaches using Voting Classifier

Karuna G. Bagde[1] (Research Scholar)
Department of Computer Science & Engg.
Sant Gadge Baba Amravati University
Amravati, India

Atul D. Raut[2]
Department of Computer Science & Engg
P. R. Pote Patil College of Engg & Mgmt
Amravati, India

Abstract:- Internets have become a part of our everyday life due to the advancement in the electronics and signal processing technologies during past decades. The tremendous growth of internet leads towards the network threats. Many times firewalls and anti-viruses fails to manage the network because of this Intrusion Detection System (IDS) comes to assists us. In this paper we use IDS with Ensemble methodologies utilized in machine learning involve the fusion of multiple classifiers to improve predictive performance, while voting classifiers combine predictions from individual models to reach conclusive decisions. The paper employs a voting ensemble method combing decision tree, logistic regression and support vector machine classifier models. We test our proposed model to classify the NSL-KDD dataset. Our ensemble methodologies of proposed algorithmproduce a good result.

Keywords:- *Intrusion Detection System, Ensemble Algorithm, Machine Learning.*

I. INTRODUCTION

Web changes our life; due to its keenness the computer structures are uncovered to an expanded number of dangers. The inquire about and mechanical developments in are advancing quickly, a supreme Cyber security remains a challenge.

The Intrusion Detection Systems (IDSs)detect attacks against a given set of computer assets from a single desktop PC to a major corporate enterprise network. The attacks are detected by looking for a predetermined set of criteria that is not present during normal daily use.

Intrusion detection systems observe and analyze network traffic to identify anomalies in network behavior and potential unauthorized access. IDSs are designed to constantly monitor the network, resulting in resource usage even when there are no attacks.

➢ *Previous Work:*

The paper [1] presents a cloud-based intrusion detection model using random forest and feature engineering, achieving high accuracy in detecting abnormal activities in network traffic.

The paper [2] proposes a prediction-level fusion model for intrusion detection and classification using machine learning techniques.

The paper [3] proposes a combination of ant colony optimization and the firefly approach for feature selection in intrusion detection using machine learning algorithms such as AdaBoost, gradient boost, and Bayesian network.

The paper [4] proposes a combination of ant colony optimization and the firefly approach for feature selection in intrusion detection using machine learning algorithms such as AdaBoost, gradient boost, and Bayesian network. Gradient boost performs better in recognizing and classifying intrusions.

The paper [5] explores the use of machine learning algorithms for intrusion detection systems, specifically focusing on dataset selection, machine algorithms, and performance metrics.

The paper [6] discusses the development of an Intrusion Detection System (IDS) that uses machine learning techniques such as Support Vector Machines, Random Forest, and K-Nearest Neighbor to automatically identify attacks on complex networks and systems.

II. DATASET

➢ *NSL-KDD Dataset:*

The NSL-KDD data set is an improved version of the KDD'99 intrusion data set. Data were captured from an evaluation test bed and included large numbers of virtual hosts and user automata. NSL- KDD is a randomly selected subset of KDD'99 after redundant data were removed and is a widely used benchmark for evaluating anomaly detection techniques. NSL-KDD dataset captures TCP, UDP, and Internet Control Message Protocol (ICMP) traffic collected using the tcpdump utility. It contains four types of intrusion attacks: DoS, U2R, R2L, and Probe described in Table [1]

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN659

Table 1 NSL-KDD Data Set

Type	Intrusion attacks
DoS	back, land, neptune, pod, smurf, teardrop, mailbomb, processtable, udpstorm, apache2, worm
U2R	buffer-overflow, loadmodule, perl, rootkit,sqlattack, xterm, ps
R2L	fpt-write, guess-passwd, imap, multihop, phf, spy, warezmaster,xlock, xsnoop, snmpguess, snmpgetattack, httptunnel, sendmail,named
Probe	ipsweep, nmap, portsweep, satan, mscan, saint

III. EXISTING SYSTEM

➢ *Let's Delve into the Available Methods:*

- *Support Vector Machines (SVM):*

Support Vector Machine (SVM) presents itself as a classification algorithm designed to identify the hyper plane that maximizes the margin between distinct classes within the dataset. This technique proves effective in handling linear and non-linear data through the utilization of kernel functions, such as linear, radial basis function, polynomial, or sigmoid, which aid in the transformation of data into a higher-dimensional space. The computational complexity of SVMs is notable, necessitating the consideration of feature reduction methods, such as Principal Component Analysis, to enhance operational efficiency. It is essential to adjust certain hyper parameters when employing SVM, including the kernel type (for instance, linear, rbf, poly, sigmoid) and the regularization parameter (C value).

- *Decision Trees (DT):*

Decision Trees (DTs) are algorithmic models rooted in tree structures, which iteratively partition the dataset according to distinct features in order to formulate decision criteria. These models exhibit Proficiency in addressing tasks related to classification as well as regression analysis. Decision Trees have a tendency to excessively fit the training data, thus prompting the necessity for ensemble methodologies to alleviate this particular drawback.

- *Logistic Regression (LR):*

Logistic Regression (LR) represents a straightforward yet efficient classification methodology. The goal is to predict the likelihood of a binary outcome by analyzing different input features. LR is renowned for its interpretability and performance, particularly in scenarios where the association between predictors and the response variable is close to being linear.

➢ *Proposed System:*

In this work, to improve the efficiency of intrusion detection system an ensemble algorithm based on the decision tree, Support vector machine and linear regression is used. The result shows that Ensemble methods work best when the predictors are as independent from one another as possible. To get diverse classifiers is to train those using very different algorithms. This increases the chance that they improve the ensemble's accuracy. The NSL-KDD data set is used to verify the superiority of the algorithm.

The proposed intrusion detection system which used ensemble method. The method we uses the combination of best available algorithm. Ensemble learning is a powerful technique that combines multiple machine learning models to create a stronger, more robust predictor.

- Proposed Algorithm 1: Intrusion Detection model using Ensemble Method.
- Input: Dataset
- Output: Model for Intrusion Detection

- Take the Dataset.
- Data preprocessing.
- Feature Selection.

➢ *Cc = Find Correlation on Data components to select high correlation values.*

- Classify Cc using train data

➢ *Logistic Regression, Decision Tree and SVM*

- Use Ensemble Voting algorithm
- Propose the Ensemble model
- Test the proposed Ensemble model by using test data
- Compute the accuracy, precision, Recall
- Return the model

➢ *Performance Analysis:*

- True Positive (TP). A true positive outcome is one where the model predicts a positive outcome correctly.
- False Positive (FP). A false positive outcome is one where the model predicts a positive outcome incorrectly.
- True Negative (TN). A true negative outcome is one where the model predicts a negative outcome correctly.
- False Negative (FN). A false negative outcome is one where the model predicts a negative outcome incorrectly.
- Accuracy. Accuracy is simply the measure of how correctly the model predicts a data given.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

➢ *Precision.*

Precision is the proportion of positives out of the total number of positives.

$$\text{Precision} = \frac{\text{Total Number of correct predictions}}{\text{Total Number of positive predictions}}$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN659

$$= \frac{TP}{TP + FP}$$

➢ *Recall.*

Recall is the proportion of positives that was identified correctly

$$\text{Recall} = \frac{TP}{TP + FN}$$

➢ *F1-Score:*

F1 Score is similar to accuracy but is a better metric because it seeks to create a balance between precision and recall especially when there is an uneven class. F1 Score is given by :

$$\text{F1 Score} = \frac{Precision \times Recall}{2 \times Precision + Recall}$$

IV. RESULTS

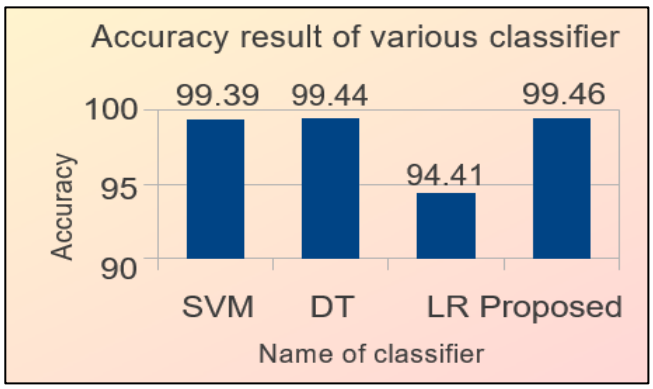

Fig 1 Accuracy Result of Various Classifier

The machine used to run the above algorithm was Intel® Core™ i5-5200U CPU @ 2.20GHz × 4, 7.7 GiB,Ubuntu 20.04.6 LTS machine.

The existing classifier algorithm Logistic Regression, Support Vector Machine and Decision Tree are train and tested with NSL_KDD dataset, The result are shown in fig 1. The Proposed ensemble algorithm achieved the accuracy score 99.46 followed by Decision Tree was 99.44 then SVM 99.39 and Logistic Regression was 94.41. Thus, showing that our ensemble model was able to achieve the best result .

The proposed ensemble model shows the promising results with comparison to SVM+RF, IntrudTree and PCA-FELM techniques. The Precision of the proposed model is 99.64% and Recall rate is 99.1% which is quite good as compared to existing methods.

Table 2 Comparison with Existing Work

	Accuracy	Precision	Recall
Proposed	0.9946	0.9964	0.991
SVM+RF	0.675	0.636	0.426
IntrudTree	0.98	0.98	0.98
PCA-FELM	0.998	0.92	

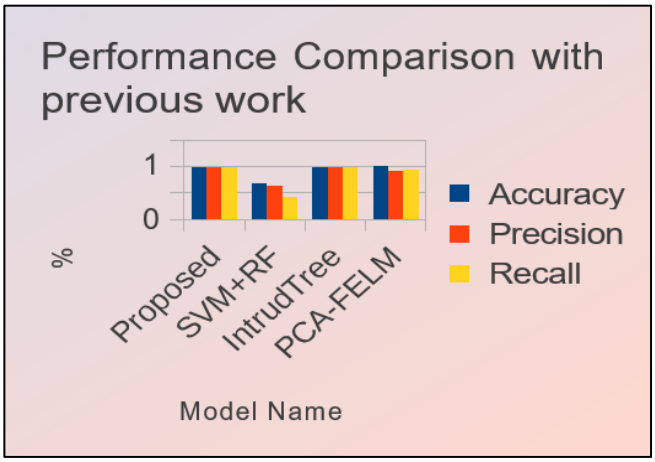

Fig 2 Performance Comparison with Previous Works

V. CONCLUSIONS

In this study, we proposed a ensemble intrusion detection model for detecting widely known attacks in networks. Our model uses correlation methods to select the best feature. Then applied ensemble classification algorithm, i.e., Decision Tree, Logistic Regression and SVM for better accuracy rate. Results show our model's better performance on NSL-KDD datasets in comparison to existing methods.

REFERANCES

[1]. Hanaa, Attou., Azidine, Guezzaz., Said, Benkirane., Mourade, Azrour., Yousef, Farhaoui (2023), "Cloud-Based Intrusion Detection Approach Using Machine Learning Techniques. Big data mining and analytics", doi: 10.26599/bdma.2022.9020038

[2]. Ramesh, Boraiah. (2023), "Network intrusion detection and classification using machine learning predictions fusion", Indonesian Journal of Electrical Engineering and Computer Science, doi: 10.11591/ijeecs.v31.i2.pp1147-1153

[3]. Mutyalaiah, Paricherla., Mahyudin, Ritonga., Sandip, R., Shinde., Smita, M., Chaudhari., Rahmat, Linur., Abhishek, Raghuvanshi. (2023), "Machine learning techniques for accurate classification and detection of intrusions in computer network", Bulletin of Electrical Engineering and Informatics, doi: 10.11591/beei.v12i4.4708

[4]. "Machine learning techniques for accurate classification and detection of intrusions in computer network", Bulletin of Electrical Engineering and Informatics, doi: 10.11591/eei.v12i4.4708

[5]. Pierpaolo, Dini., Abdussalam, Elhanashi., Andrea, Begni., Sergio, Saponara., Qinghe, Zheng., Kaouther, Gasmi. (2023), "Overview on Intrusion Detection Systems Design Exploiting Machine Learning for Networking Cybersecurity", Applied Sciences, doi: 10.3390/app13137507

[6]. Ch. Sai Sampath, Dr. P. Anuradha (2023), "Intrusion Detection using Machine Learning: A Random Forest-based Approach", International Journal For Multidisciplinary Research, doi: 10.36948/ijfmr.2023.v05i03.3408

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN659

[7]. D. Xuan, H. Hu, B. Wang and B. Liu , "Intrusion Detection System Based on RF-SVM Model Optimized with Feature Selection", 2021 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), Beijing, China, 2021, pp. 1-5, doi: 10.1109/CCCI52664.2021.9583206.

[8]. Sarker, I.H.; Abushark, Y.B.; Alsolami, F.; Khan, A.I., "IntruDTree: A Machine Learning Based Cyber Security Intrusion Detection Model", Symmetry2020,12,754 https://doi.org/10.3390/sym12050754

[9]. E. Vishnu Balan, M.K. Priyan, C. Gokulnath, G. Usha Devi, "Fuzzy Based Intrusion Detection Systems in MANET" Procedia Computer Science, Volume 50,2015,Pages 109-114,ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.04.071.