IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

A Comparative Study of Algorithms for IDS

Nupoor Rajput[1], Riddhi Jain[2], Yashika Khatri[3], Praveen Gupta[4]
Students, Department of Computer Science and Information Technology[1,2,3]
Professor, Department of Computer Science and Information Technology[4]
Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India

Abstract: *Nowadays, it is very important to maintain a high level of data security to ensure safe and reliable transfer of data between different organizations. Cyber Attacks, or attacks on computer networks, are already widespread and impact almost everyone and every internet-connected device. To avoid these attacks there are various approaches available but they are not quite efficient, therefore machine learning and deep learning are now being used by organizations to prevent these kinds of attacks because they are successful without requiring human intervention. The primary advantage of machine learning is its inherent ability to recognize, stop, prevent, recover and even cope up with various types of threats without the need of explicit programming. This work is discussing various algorithms available to prevent such cyber attacks. Here we include the following algorithms: linear support vector machine, quadratic support vector machine, K- nearest- neighbor, linear discriminant analysis classifier, quadratic discriminant analysis classifier, multilayer perceptron classifier, auto encoder. The work focuses on providing more accurate algorithms among these to improve the performance. The dataset used for this work was KDD. The datasets will be processed using the modified methodology based on the number of features.*

Keywords: Intrusion Detection System, KDD, Machine Learning, Deep Learning, Algorithm, Computer Network, Support Vector, Cyber attacks

I. INTRODUCTION

Digital content has exploded in popularity across many industries, which has raised concerns about cyberattacks. This perspective has resulted in the creation of systems and methods to which companies, institutions, and people can all be tracked and protected. The common security features that a network must have are privacy, authentication, integrity, Non-repudiation and availablity[1]. For this one of the most crucial approaches for preventing and monitoring infiltration in computer networks to make them more secure is the intrusion detection system (IDS) [2].Monitoring the network, confirming network activity, and reporting events that don't comply with the network administrator's security policies compose the IDS work mechanism. The intrusive and non-intrusive network packet detection and identification techniques are included in the IDS system. Currently, human analysts analyze system logs for intrusion detection systems to differentiate

between intrusive and non-intrusive network traffic . As a result, we observe that the majority of these systems depend heavily on humans for the majority of data analysis tasks. Two categories can be used to classify intrusion detection[3] . Both the signature-based detection system and the anomaly-based detection system are keen on investigating network data for specific byte or packet sequences.[4].The fact that signatures are relatively straightforward to create and comprehend is one of the drawbacks and observations of this variety. A key component of network security is the anomaly network intrusion detection, where the behavior of an abnormality can be compared to the regular use of data. How to categorize normal and abnormal activities so that we can successfully distinguish between the processes is one of the obstacles that must be conquered in an intrusion detection system based on anomaly detection. Recently, the majority of efficient intrusion detection systems depend on machine learning, whose mechanisms are highly functional and provide an excellent probability of detecting intrusions in the network [5].

Intrusion detection systems can aid in identifying network users' malicious intentions. There are a variety of machine learning algorithms available that can generalize when exposed to new, untrained data. For study on intrusion detection systems, a common data collection is the KDD data set [6].In this work KDD Dataset is being used and here itrustion is proposed into two categories that is normal and abnormal Network events. And then we had followed various steps that

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

included data preprocessing, feature extraction and then last is classification. To gain the accuracy in predicting the intrusion the algorithms that we used are linear support vector. machine, quadratic support vector machine, linear discriminant analysis classifier, quadratic discriminant analysis classifier, multilayer perceptron classifier, auto encoder [7].

II. RELATED WORK

Numerous surveys on intrusion detection have been conducted over the last ten years. Bishop gave one of the first presentations on trends in vulnerability analysis and intrusion detection. Trends in intrusion detection are infrastructure-based protocols and strategies needed to create intrusion detection systems[8]. Another well-known study by Kabiri and Ghorbani presented IDS trends as well as examined some issues with intrusion detection[8]. Traditional IDS encounters issues with accuracy, time usage, log-file updating, statistical analysis, and rule-based analysis.A review paper based on some significant machine learning-based algorithms used in intrusion detection was published by Zamani and Movahedi[8].According to Zamani's research, using a machine learning method for intrusion detection allows for a high detection rate, a low rate of false positives, and the capacity for fast adaptation to shifting intrusive behavior. In this review paper's analysis of algorithms, grounds for cognitive computing and artificial intelligence (AI) have been distinguished[8].For intrusion detection, Agrawal and Agrawal [8] looked at a variety of data mining methods.

Numerous machine learning methods, either separately or in combination, have been extensively used for IDS selection of features and dimensionality reduction in addition to clustering and categorizationFor the purpose of training and testing their models, intrusion detection experts can use benchmark datasets, which Hamid et al. reviewed[8]. The analysis of several datasets, including ADFA-WD (Australian Defense Force Academy Window Dataset), Caida DDoS (Caida Distributed denial of Service) Dataset, KDD'999[9],NSL-KDD, UNM-Dataset, and UNSW-NW15, gave information on classes, attributes, and instances..In the most recent publication, Mishra also suggested a thorough study and analysis using machine learning approaches for intrusion detection[8]. This survey is dependent on the classifiers being divided into four categories: single classifiers with all features in the dataset, single classifiers with chosen features in the dataset, multiple classifiers with all features in the dataset, and multiple classifiers with selected features in the dataset. This analysis also shows that an intrusion detection technique that performs well for one kind of attack may not work well for other kinds of attacks.

III. PROPOSED WORK

The KDD dataset utilized in this study is initially introduced in this section. The proposed methodology is then detailed, along with the pre-processing, feature extraction, and classification phases.

3.1 Source of Data

The database archive at UC Irvine houses the information. The Fifth International Conference on Knowledge Discovery and Data Mining, which took place together with KDD-99[9], The Third International Knowledge Discovery and Data Mining Tools Competition, used this dataset.

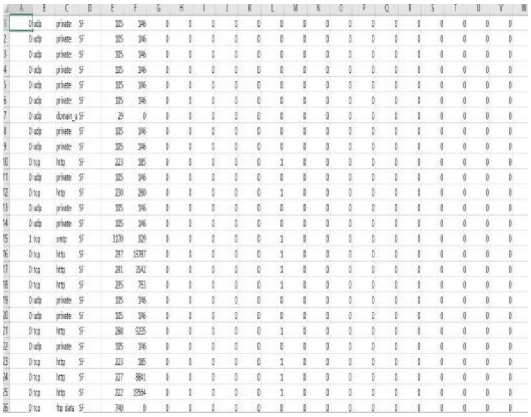

Figure 1: KDD dataset

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

The goal of the challenge was to create a network intrusion detector—a predictive model that can predict with better accuracy whether the system has undergone an intrusion or not using various machine learning algorithms. A standard set of auditable data, including a broad range of simulated intrusions into a military network's surroundings, can be found in this database. The KDD99 dataset[9] has been around for more than 15 years, but based on the amount of studies that have been published, it is still the most popular dataset for IDS and machine learning

3.2 Dataset Pre-processing

Pre-processing is done to transform unprocessed data into a format that can be used by machine learning. A data scientist can use an applied machine learning algorithm to achieve more accurate outcomes by using structured and clean data[2]. The method involves cleaning, and sampling of the data. For pre-processing of dataset one can apply data normalization, one hot encoding, binary classification, multi class classification and feature extraction.

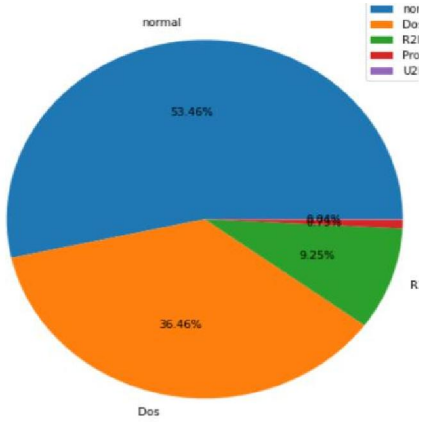

Figure 2: Pie chart distribution of multi-class labels

3.3 Data Splitting

For splitting of data we have divided the data into two subset training and testing dataset. For this splitting we have divided the data in 1:4 ratio. Out of 97 attributes, 93 were chosen to be removed from binary classification in order to include the target attribute. As the target attribute, intrusion attribute was chosen. Out of 100 attributes, 93 were chosen to remove the target attribute (encoded, one-hot-encoded, original) from the multi-class classification process.

3.4 Model Training

A data scientist can start building a model afterpre-processing the gathered data and separating it into train and test sets. This procedure involves "feeding" training data to the algorithm. An algorithm will process data and produce a model that can locate a goal value (attribute) in new data, an answer you are looking for with predictive analysis. To create a model that can predict the target value is the goal of model training.

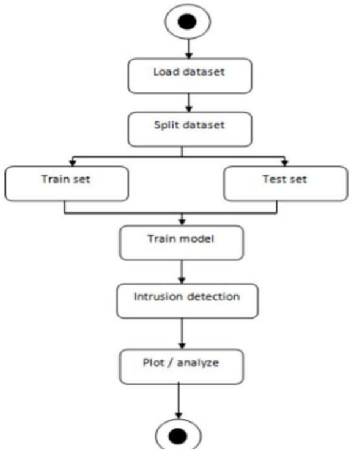

Figure 3: Flowchart

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

3.5 Model Evaluation and Testing

The objective of this stage is to create the most basic model that can adequately and quickly construct a target value. Through model tuning, parameters are optimized in this manner. Data analysts can accomplish this objective. To attain an algorithm's highest efficiency, the model is crucial to remember that the test data comes from a different probability distribution than the training data and contains particular attack kinds that weren't present in the training data. The job becomes more doable as a result.

3.6 Method Used

For achieving the accuracy of various algorithms which can predict accurately whether the system has undergone an intrusion or not we can use various algorithms such as linear support vector.machine, quadratic support vector machine, K- nearest- neighbor, linear discriminant analysis classifier, quadratic discriminant analysis classifier, multilayer perceptron classifier, auto encoder[7]..

IV. RESULT ANALYSIS

The experimental study is carried out by assessing our intrusion detection system using the prestigious KDDCUP99 data set. The coding is done using Python on the Windows platform. This study project calls for an enormous quantity of reliable test data. For our test in this study, we took advantage of the KDDCUP99 dataset. The collection consists of network data from a US Air Force LAN that was recreated over the course of nine weeks.. Fulldata. Corrected. Each dataset entered into KDDCUP99 has a collection of forty-one fixed characteristics as well as a class label. There are more specific distinctions between the four kinds of attacks: U2R (unauthorized access to local super user), R2L (unauthorized access from a remote computer), DOS (denial-of-service), and probing (surveillance and other probing). All four kinds of attacks have now been fully detected through experimentation.

Classification of attacks	Subclasses
DOS (Denial of service attack)	land, back, pod, neptune, teardrop, smurf,mailbomb, apache2
R2L (Illegal access from remote machines)	imap, multihop, phf, spy, warezmaster, Xlock, warezclient, snmpgetattack, ftp_ write
U2R (Unauthorized access of ordinary users) To privileges of administrator)	buffer_overflow, loadmodule, perl, rootkit,guess_passwd
Probing (Monitoring and other detection activities)	Ipsweep,nmap, portsweep,satan,saint

Source: https://www.researchgate.net/publication/349029421_Decision_Tree_A_Machine_Learning_for_Intrusion_Detection

After classifying the attacks in classes, we have applied various algorithms and achieved the maximum accuracy of 98.55% using K- nearest-neighbor binary classification classifier. Other than this we have used various other algorithms such as linear support vector. machine, quadratic support vector machine,linear discriminant analysis classifier, quadratic discriminant analysis classifier, multilayer perceptron classifier, auto encoder.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

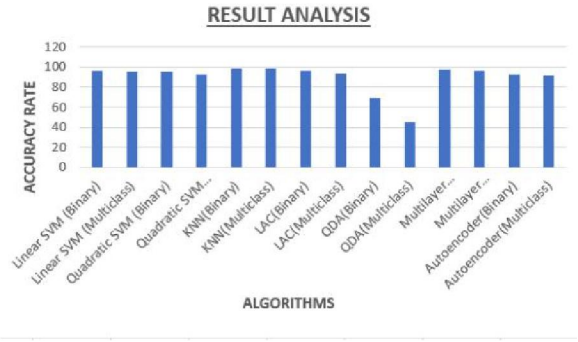

Figure 4: Result Analysis

V. CONCLUSION

In this work the KDD dataset produced an accuracy of 98.55% by fine tuning the epochs batch size and by using a suitable activation function, which is a significant improvement over earlier studies.

REFERENCES

[1]. Different Attacks and their Defense Line in Mobile Ad hoc Networks: A Survey , P. Gupta, P. Bansal in IJCSE,2018,Vol.-6, Issue-8, Aug 2018

[2]. Decision Tree: A Machine Learning for Intrusion Detection,Shilpashree. S, S. C. Lingareddy, Nayana G Bhat, Sunil Kumar G in IJITEE ,2019 ,Volume-8, Issue- 6S4, April 2019

[3]. Survey of intrusion detection systems: techniques, datasets and challenges,Ansam Khraisat, Iqbal Gondal, Peter Vamplew Joarder Kamruzzaman in SpringeOpen ,2019

[4]. A Signature-based Intrusion Detection System for the Internet of Things, Philokypros P. Ioulianou, Vassilios G. Vassilakis, Ioannis D. Moscholios†, Michael D. Logothetis in ICTF in 2018

[5]. A review of intrusion detection system using machine learning approach,SH Kok, Azween Abdullah, NZ Jhanjhi, Mahadevan Supramaniam in International Journal of Engineering Research and Technology, 2019, Volume 12

[6]. Analysis of KDD Dataset Attributes - Class wise For Intrusion Detection,Preeti Aggarwala,, Sudhir Kumar Sharma in ICRTC 2015

[7]. Comparative Analysis of Intrusion Detection System Using Machine Learning and Deep Learning Algorithms,Johan Note Maaruf Ali in AETiC Vol. 6, No. 3, 2022

[8]. Research Trends in Network-Based Intrusion Detection Systems: A Review,Satish Kumar,Sunanda Gupta,Sakshi arora in IEEE,2022

[9]. KDD Cup 1999 Data. [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. [Open Source]