iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: IV Month of publication: April 2023

DOI: https://doi.org/10.22214/ijraset.2023.51086

A Machine Learning Approach for Intrusion Detection

Prof. Dipali Mane[1], Chaitanya Chaudhari[2], Saurabh Shitole[3], Mubin Shaikh[4], Shivam Sashte[5]
[1]Assistant Professor, Department Of Computer Engineering, ZCOER, Pune
[2, 3, 4, 5]BE Students, Zeal College of Engineering and Research, Pune, Maharashtra, India

Abstract: Computer networks and virtual machine security are very essential in today's era. IDS monitors a network or system for malicious action and protects a computer network from unofficial access from users, including perhaps insiders. Various existing systems have already been developed to detect malicious activity on target machines; sometimes any external user creates some malicious behavior and gets unauthorized access to victim machines to such a behavior system considered as malicious activities or Intruder. Machine Learning (ML) algorithms are applied in IDS in order to identify and classify security threats. Numerous machine learning and soft computing techniques are designed to detect the activities in real-time network log audit data. KKDDCUP99 and NLSKDD most utilized data sets to detect the Intruder on the benchmark data set. In this paper, we proposed the identification of impostors using machine learning algorithms. Two different techniques have been proposed a signature with detection and anomaly-based detection. The experimental analysis demonstrates SVM, Naïve Bayes, and ANN algorithms with various data sets and demonstrates system performance in the real-time network environment.

Keywords: Intrusion Detection System, SVM, Naïve Bayes, Host Intrusion Detection System, Network Intrusion Detection System. Perimeter intrusion detection

I. INTRODUCTION

By definition, A security event, or a mixture of multiple security events, constitutes a security incident within which associate personality non grata gains, or tries to achieve, access to a system or system resource while not having the authorization to try and do thus. IDS may be a security event, or a mixture of multiple security events, that constitutes a security incident within which associate persona non grata gains, or tries to achieve, access to a system or system resource while not having the authorization to try and do thus. Signature-based IDS detects the intrusion on the idea of the already famed malicious instruction classification that's utilized by the malware.

They noticed patterns within the IDS area unit referred to as signatures. Signature-based IDS will simply observe the attacks whose pattern (signature) already exists within the system however it's quite troublesome to observe new malware attacks as their pattern (signature) isn't famed. Anomaly detection: - during this technique, network traffic or host OS behavior is analyzed supported by varied parameters, and compared with the traditional behavior. If the system detects any deviation from traditional behavior, it raises the associate alarm. Intrusion Detection Systems area unit is loosely classified into 2 categories:

1) *Network Intrusion Detection System (NIDS):* It captures the packets from network traffic. The header of the captured packets is analyzed based on various constraints to detect malicious activities. It can be set up in the network support, server, switches, and gateways.
2) *Host Intrusion Detection System (HIDS):* It is installed on a separate system to detect intrusion or misuse. HIDS analyses the key system files, process behaviors, rare resource utilization, unauthorized access, etc

II. LITERATURE SURVEY

According to [1], In this paper, Monika D. Rokade and Yogesh Kumar Sharma used machine learning algorithms to detect unauthorized users in a network. They proposed two different techniques one is signature-based and the other is anomaly-based detection. They used various algorithms like naïve Bayes, ANN, and SVM on various Datasets to detect the intrusion. The results they got after the implementation showed that the SVM algorithm has better accuracy than the other two algorithms like NB and ANN.

According to [2], A system developed for anomaly-based or pattern-less intrusion detection of an unauthorized user in a real-time network dataset. This system is good at finding intrusion even at large datasets easily. This system also checks the working condition of the Network.

According to [3], Distributed Intrusion Detection System (DIDS) uses technologies like Blockchain and Cloud Computing to give better results than present IDS methods like Host Intrusion Detection System (HIDS) and Network Intrusion Detection System(NIDS). Amazon Cloud services are used for performing the experiment. DIDS is good at detecting anomaly-based abnormalities in a network.

According to [4], Kaggle Dataset and NSL KDD datasets are used. Hybrid Anomaly Based Intrusion Detection is used instead of NIDS. They found that using Classifiers and Boosting increases the intrusion detection rate in a dataset. According to their results, Random Tree has great accuracy and minimum false alarm rate in a dataset. They concluded that ADA boost gives very optimal results with Decision Tree.

According to [6], Pattern Based Intrusion Detection(PBID) is proposed. Rules are defined for pattern-recognized engines. The experiment has concluded that implementing PBID and SBID helps in detecting intrusion at a huge level. Anomaly Based Intrusion Detection is used with already existing methods.

According to [7], WLAN is used for intrusion detection. Challenges of blockchain network intrusion are discussed. The Characteristics of blockchain are discussed. For experiments tools in the Kali, system are used. It increases the detection of abnormalities in the block and the dataset.

According to [8], the decision tree model has increased the accuracy of detecting outliers in a network dataset. They have also used the KNN model and Random Forest Model for detecting the outliers with more than 90% accuracy.

III. INTRUSION DETECTION MODEL

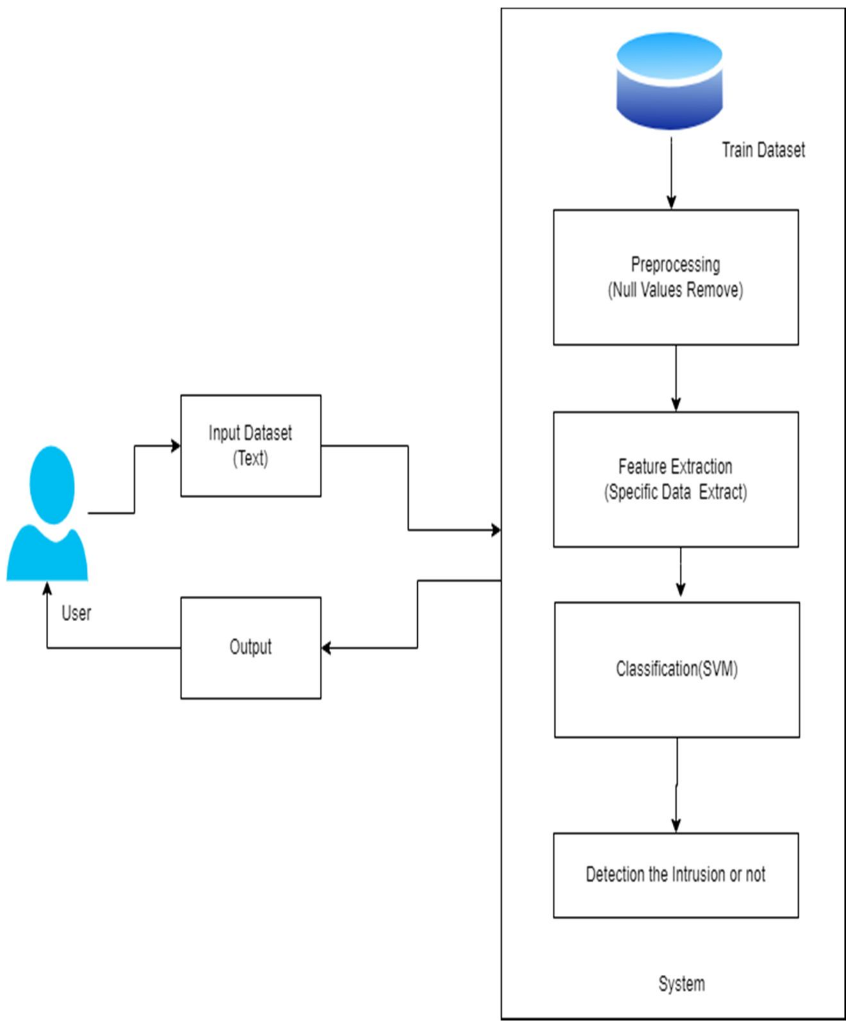


Figure 1: System Architecture

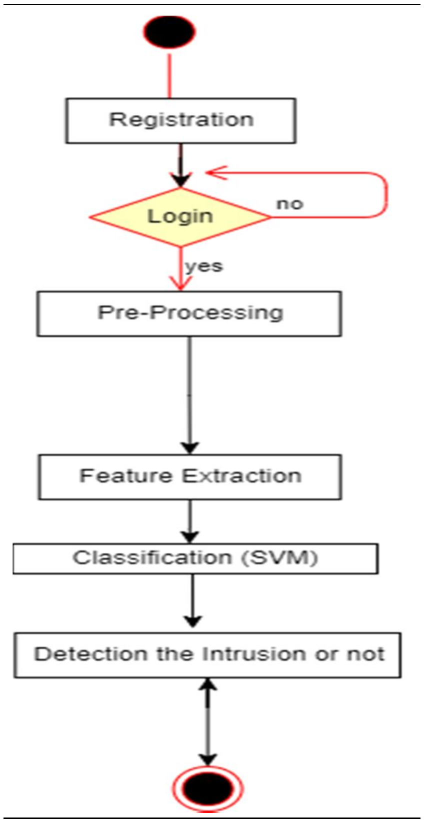


Figure 2: Activity Diagram

A. About intrusion KAGGLE dataset

This is another type of Dataset for Intrusion Detection System (IDS) assessment. We have taken this Dataset for finding irregularities over the network. The Irregular instances contain attacks like Dos, normal, probe, r21, and u2r. The number of Data instances is 10000. The attributes present in this dataset are 43 in number.

B. About NSL KDD dataset

It is one of the versions in the KDD dataset. NSL-KDD is often used for the assessment of new Intrusion Detection Systems (IDS). We have taken this Dataset for network Intrusion Detection analysis and comparison. It consists of 42 attributes and instances which are envisioned in single vector planes. The vectors are classified as Normal and Anomaly. The Dataset is discretized and boosted with classification. The number of Data points in the Dataset is 22544.

C. Naïve bayes Classification

It is one of the classification techniques based on Bayes Theorem. Naïve Bayes assumes the presence of a specific feature in a class is unconnected to the presence of any other feature. It is actually based on probabilistic models which have strong independence assumptions. Each class is calculated from the training data and is independent of the others. This probabilistic model is fast and easy to calculate so it has been shown to be an effective classification algorithm. It makes predictions for classes with higher probability. Only in large training datasets, it gives great accuracy results

1) *Anaconda:* Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment.
2) *Spyder:* Spyder is a powerful scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. It offers a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package.

IV. SVM ALGORITHM

One of the most popular supervised learning algorithms, Support Vector Machine, or SVM, is used to solve Classification and Regression problems. However, it's largely employed for Machine Learning Classification problems.

SVM selects the sharp vectors and points that make it easier to create the hyperplane. The algorithmic software is called a "Support Vector Machine" because these extreme examples are known as support vectors. Consider the diagram below, in which there are two distinct classes that are separated by a call boundary or hyperplane:

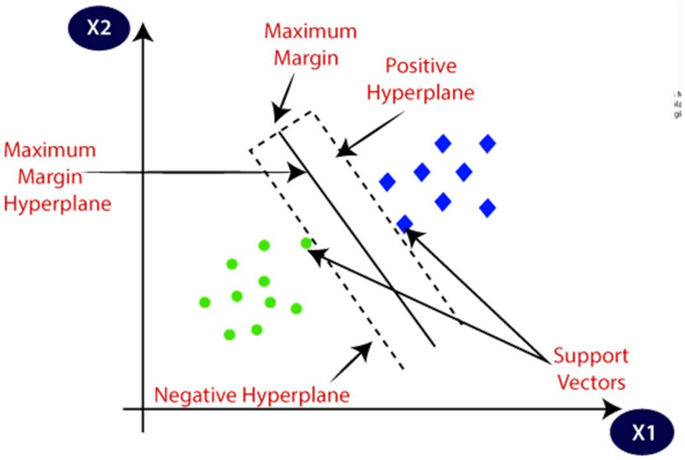


Figure 3: SVM

A. Types of SVM

1) *Linear SVM:* Linear SVM is used for linearly separable data, which is defined as data that can be divided into two classes using a single straight line. The classifier used for such data is called the Linear SVM classifier.
2) *Non-Linear SVM:* Non-Linear SVM is used for non-linearly separated data, which implies that if a dataset cannot be categorized using a straight line, it is non-linear data, and the classifier employed is referred to as a Non-linear SVM classifier.

V. SCOPE

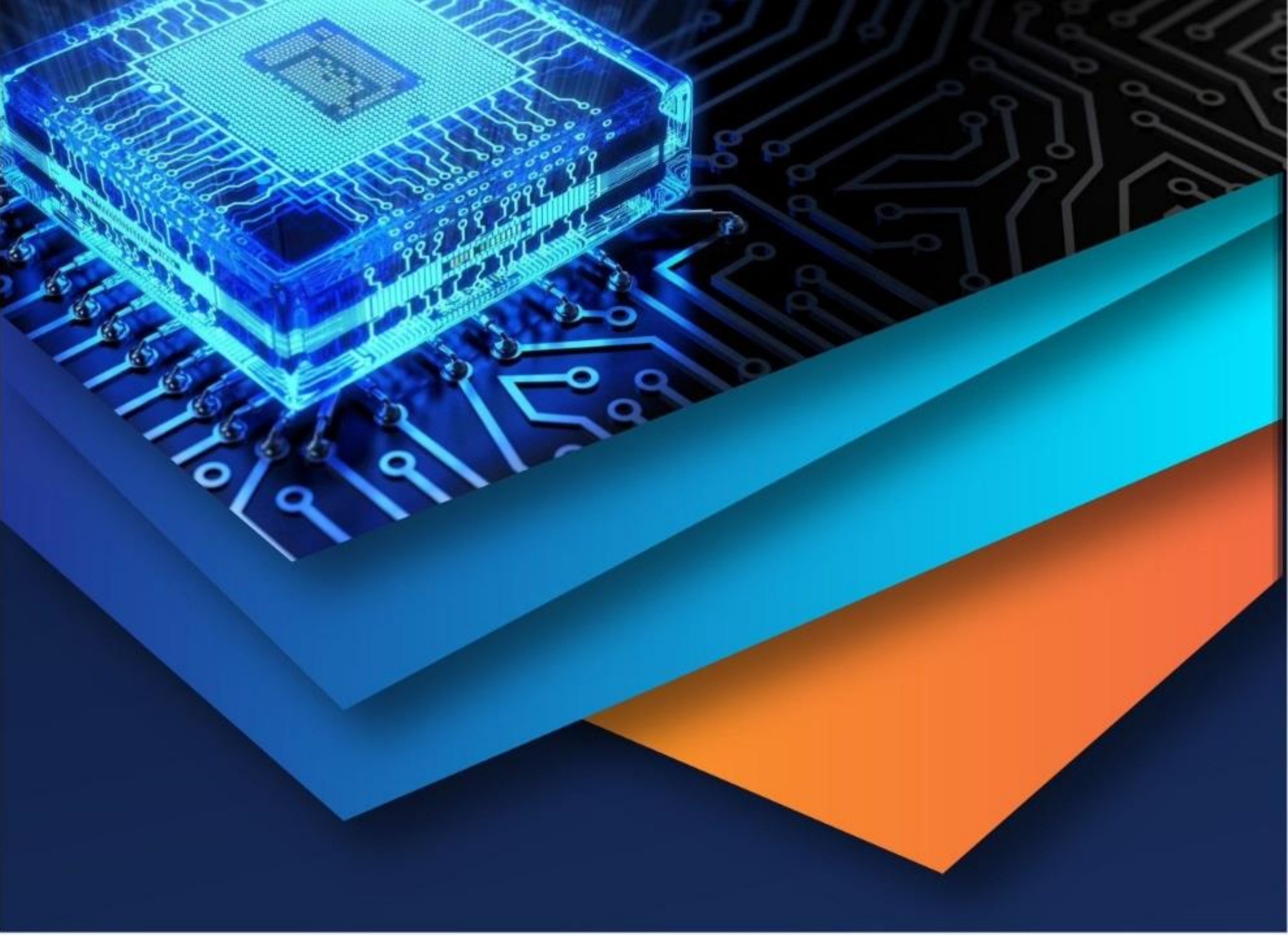
With the help of machine learning algorithms, we can increase the rate of intrusion detection in the real-time Network dataset.
We can use it in the Banking sector, Healthcare sector, IT sector, etc.

VI. CONCLUSION

This study proposed an SVM-IDS approach based on deep learning to suggest an efficient system of IDs. To test anomaly detection accuracy, we used the synthetic-based intrusion dataset - NSL-KDD. We plan to implement IDS into the cloud environment in the future using the deep learning technique. We also analyze and compare various methods of deep learning, namely. To detect intrusions in the network, NB ANN, RF, and SVM on the NSL-KDD dataset.

REFERENCES

[1] Monika D.Rokade, Yogesh Kumar Sharma," MLIDS: A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset",2021
[2] Stephen D. Donald, Rabert V McMillen, David K Fard, and John C. McEachen," Therminator 2: a thermodynamics-based method for real-time patternless intrusion detection",
[3] Dr. Manish Kumar, Ashish Kumar Singh," Distributed Intrusion Detection System using Blockchain and Cloud Computing Infrastructure",202
[4] S. Sivanantham, R. Abirami, R. Gowsalya," Comparing the Performance of Adaptive Boosted Classifiers in Anomaly-based Intrusion Detection System for Networks",2019
[5] Meng W, Tischhauser E W, Wang Q, et al. When Intrusion Detection Meets Blockchain Technology: A Review[J]. IEEE Access, 2018
[6] Zakiyabanu S. Malek, Bhushan Trivedi, Axita Shah," User behavior Pattern -Signature based Intrusion Detection",2020
[7] Zhan Xin, Wang Xiaodong, Yuan Huabing," Research on Block Chain Network Intrusion Detection System",2019
[8] Ajay Shah, Sophine Clachar, Manfred Minimair, Davis Cook," Building Multiclass Classification Baselines for Anomaly-based Network Intrusion Detection Systems",2020
[9] Dong YuanTong," Research of Intrusion Detection Method Based on IL-FSVM",2019



doi crossref

10.22214/IJRASET

INDEX COPERNICUS

45.98

ISRA JIF

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY