e-ISSN:2582-7219

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 6, Issue 6, June 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.54

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 6, Issue 6, June 2023 |

# An Intrusion Detection System Using Machine Learning Techniques

[1]M Venkataiah, [2]Shaik Subhani

[1]PG Scholar, Dept. of CSE (AIML), St Marys group of Institutions Guntur, AP, India

[2]Assistant Professor, Dept. of CSE, St Marys group of Institutions Guntur, AP, India

**ABSTRACT:** As of late, the information stream over the web has dramatically expanded because of the monstrous development of PC networks associated with it. A portion of these information can be named a pernicious movement which can't be caught by firewalls and hostile to malwares. Because of this, the interruption identification frameworks are earnest need to perceive noxious action to keep information trustworthiness and accessibility. In this review, an interruption discovery framework in view of group highlight ideas and KNN classifier has been recommended to deal with the different difficulties issues in information like deficient information, blended type, and clamor information. To fortify the proposed framework an extraordinary sort of examples similitude measures are upheld to manage these kinds of difficulties. The exploratory outcomes show that the order exactness of the recommended framework is superior to K-closest neighbor (KNN) and support vector machine classifiers while handling deficient informational collection, inspite of droping down the general location precision.

**KEYWORDS**: Intrusion Detection, Support Vector Machine Naive Bayes, Machine Learning.

## I. INTRODUCTION

In this day and age, there has been extraordinary advancement and improvement in correspondence innovations and the Web, and quite possibly of the main region wherein it has seemed is network security. It utilizes instruments like firewalls, antivirus programming and interruption discovery frameworks to guarantee an organization security and all its connected assets in the Web (IDS). These methodologies safeguard networks from both homegrown and outside dangers. An IDS is a distinguish gadget that tracks the condition of an organization's product and equipment and safeguards network safety. Nonetheless, a few Interruption location frameworks actually have a high deceptions, making various admonitions for low-danger cases, adding to security examiners' responsibility and possibly making serious assaults slip by everyone's notice. As an outcome, obscure assaults should be recognized by IDSs.

Scientists have begun to zero in on the development of AI (ML) strategies since it is a wise innovation to recover valuable information consequently from monstrous datasets. While adequate preparation information is accessible, IDSs can accomplish great degrees of detecting and AI models are adequately summed up to distinguish assaults. Also, ML doesn't depend to a great extent on the space information, making it simple to plan and fabricate. Profound learning (DL) can create phenomenal outcomes. An unmistakable component of DL is the profound construction that involves a few secret layers. Then again, run of the mill models are either without stowed away layers or have only one. This article makes three significant commitments. We have directed a precise survey of IDS and how they are utilized with the ML-DL calculations that have been finished during the most recent two years and examined each article as far as strength, shortcoming and assessment measures utilized, then, at that point, we applied the calculations lastly tracked down a distinction in exactness between them.

To perceive strange way of behaving that happens in a PC or organization, Interruption discovery framework (IDS) is utilized. IDSs can be portrayed in more than one way, among them abuse based and peculiarity based IDSs are the most well-known. To recognize realized assault like grunt, Abuse based IDS can perform capably. This sort of IDSs has less misleading problem rate. It unable to perceives new goes after which customizes no guidance in data set. In Irregularity based IDS, it fosters a model of customary conduct after that; it isolates any fundamental deviations

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 6, Issue 6, June 2023 |

from this model and think about that deviation as interruption. This sort of IDS can identify both known and obscure assaults, however experiences a high deception rate. Different AI procedures are integrated to diminish deception rate.

A. Intrusion Identification Framework

A particular presence of interruption can take or kill data from PC or organization frameworks in restricted length. Consequently interruption is one of the significant issues in network security. Framework equipment likewise gets hurt because of interruption[2-5]. Different strategies of interruption location are performed; but exactness is one of the serious issues. Location rate and misleading problem rate assumes a fundamental part for the investigation of exactness. Interruption recognition should be improved to lessen phony problems and to build the identification rate. Subsequently, Backing Vector Machine (SVM) and Guileless Bayes are applied. Order can be tended to by these calculations. Aside from that, Standardization and Element Decrease are likewise applied to make a relative investigation.

B. Machine Learning

AI is utilized to mechanize logical model structure. It is a method of information examination. It is one of the parts of Man-made brainpower which chips away at the idea that a framework gets prepared, decide and figure out how to distinguish designs with less mediations of people. Managed and Solo learning are the two most broadly utilized AI methods[1]. Marked models like a contribution with favored yield are taken for preparing calculations. Occasions without verifiable names get prepared utilizing unaided learning. To find some design inside the information and to investigate the information are the two fundamental goal of unaided learning. Aside from these strategies, approaches like Semi supervised learning and Support learning are utilized.

For preparing reason, semi supervised learning utilizes less measures of marked information and gigantic measures of unlabeled information. Experimentation technique is utilized in Support Learning in which the activities yield the best rewards. Grouping, relapse and forecast are utilized. Specialist, climate and activities are the three essential part utilized in this kind of learning. That's what the objective is, the specialist needs to choose those activities, which exploit the anticipated award. By applying great strategy, the specialist ready to arrive at the objective a lot quicker.

## II. LITERATURE SURVEY

Protecting computer and network information of an organizations and individuals become an important task, because compromised information can cause huge loss. Hence, intrusion detection system is used to prevent this damage. To enrich the function of IDS, different machine learning approaches get developed. The main objective [11] is to address the problem of adaptability of Intrusion Detection System (IDS).The proposed IDS has the proficiency to recognize the well-known attacks as well as unknown attacks. The proposed IDS consist of three major mechanisms: Clustering Manager (CM), Decision Maker (DM), Update Manager (UM). NSL-KDD dataset is applied to estimate the working of the proposed IDS. Both supervised and unsupervised techniques were accompanied. The information received to the system is grounded on the education of an agent who disregards the correction proposals presented by IDS. This technique is applied on supervised mode. Both known and unknown traffics can be detected by the system, when they work under unsupervised mode. After updating recently arrived data from both supervised and unsupervised modes, the function of the system has been improved. Performance of the system gets improved, when it runs in unsupervised mode. By incorporating machine learning techniques like, SVM and Extreme Learning Machine (ELM), a hybrid model get developed. Modified K-means is used to construct high quality dataset. It builds small dataset that denote overall original training datasets. By this step, the training time of the classifier gets reduced. KDDCUP 1999 is used for implementation. It shows accuracy of about 95.75 percentages. Various machine learning techniques like SVM, Random Forest (RF) and ELM are examined to report this problem. ELM shows better result when compared to other techniques in accuracy. Datasets get divided into onefourth of the data samples, half of the dataset and full datasets. However, SVM produces better results in half of the data samples and one-fourth samples of data. ELM is the best method to handle the huge amount of data of about two lakh instances and more. A new hybrid classification algorithm on Artificial Bee Colony (ABC) and Artificial Fish Swam (AFS) is proposed [12]. Nowadays computer system is prone to different information thefts due to the widespread usage of internet, which leads to the emergence of IDS. Fuzzy CMeans Clustering (FCM) and

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 6, Issue 6, June 2023 |

Correlation-based Feature Selection (CFS) is applied [6] for separating training datasets and to eliminate irrelevant features. If-then rues are generated by using CART technique, which is applied to differentiate normal and anomaly records according to the selected features. Correlation-based feature selection method which is a simple filter-based model is used in the proposed system. Datasets containing the features, highly correlated with the class, yet uncorrelated with the others are applied. By using NSL-KDD and UNSW-NB15 datasets this approach get achieved 99 percentages of detection rate of anomalies and 0.01 percentages of false positive rate. A hybrid method for A-NIDS using AdaBoost algorithms and Artificial Bee Colony to obtain low false positive rate (FPR) and high detection rate (DR).

## III. PROPOSED SYSTEM

Research approach continued in this venture is trial research. In like manner, a technique is proposed to have proficient method for recognition of DDoS assaults. It depends on administered learning approach and furthermore gathering learning. The system additionally incorporates highlight choice as the component determination strategy could further develop preparing quality during the time spent distinguishing DDoS assaults. In light of regulated learning contains stages like preparation and testing. The primary design is as delineated in Figure 1.

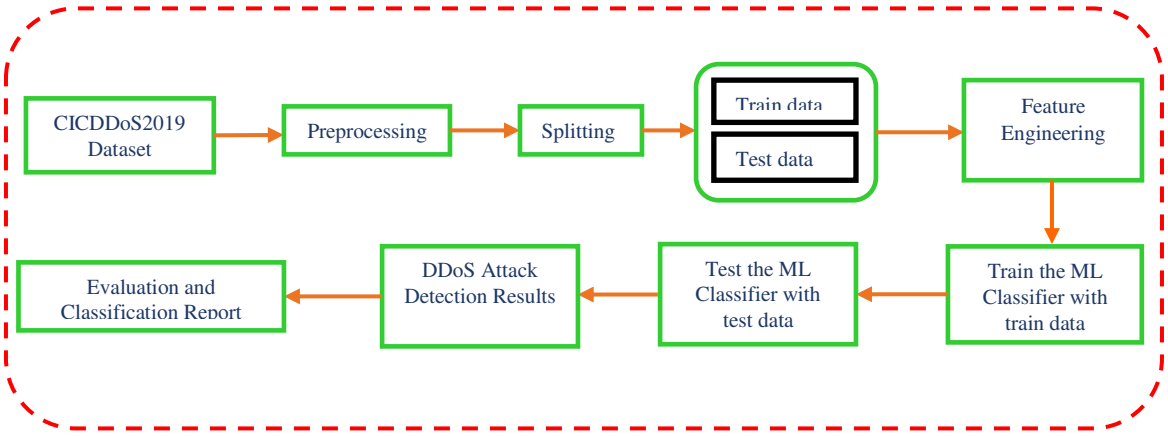



**Figure 1:** Proposed framework for DDoS attack detection

As introduced in Figure 1, CICDDoS2019 dataset is utilized for tests. The dataset is exposed to pre-handling and parting into 75% preparation and 25% testing information. Highlight designing is the most common way of tracking down contributing elements that assume a part in class name determination. After include designing, the preparation dataset is utilized to work on the nature of preparing. Different ML classifiers are prepared and models are put something aside for reuse later. Then the test information is exposed to the prepared models to identify DDoS assaults.

Different ML models are utilized in the proposed system. Every last one of them has its inside working and each model showed different degree of execution. The plan to utilize different models exclusively to distinguish DDoS assaults is to recognize the best models that display 90% of more elevated level of precision and make them into gathering growing experience for additional improvement in discovery execution.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 6, Issue 6, June 2023 |

Algorithm

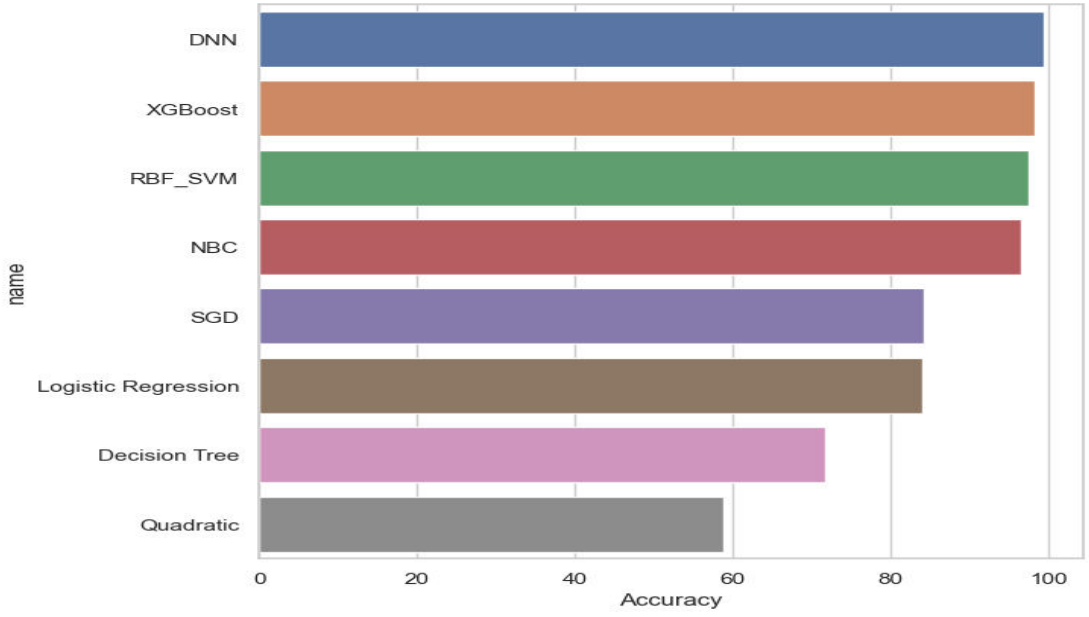
**Algorithm 2:**Automatic DDoS Attack Detection (ADAD)

**Inputs:**CICDDoS2019 dataset D, ML pipeline P

**Output:**DDoS attack detection results R

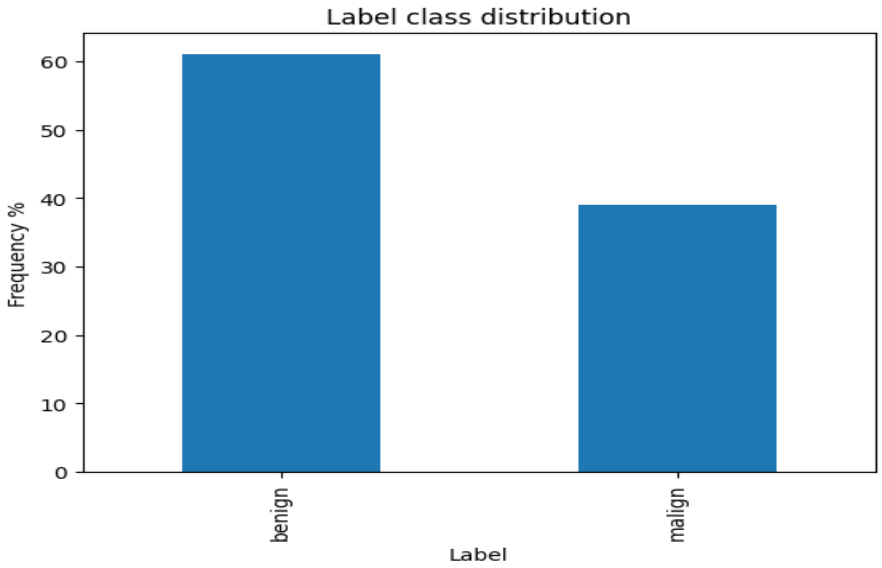
1. Begin
2. Initialize results map M
3. (*T1*, *T2*)←PreProcess(*D*)
4. *F*←FeatureSelection(T1)
5. For each detection model m in pipeline P
6. Use F for training m
7. R←testModel(m, T2)
8. Add m and R to M
9. End For
10. While M is not Empty
11. Display m and R
12. End While
13. End

Results:



| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 6, Issue 6, June 2023 |




## IV. CONCLUSION

The analysis of the intrusion detection data set based on machine learning techniques is a challenging task due to its massive size, mixed-type attributes, and the redundancy of data. Besides that, the data may be incomplete and noisey. In this study, an intrution detection system has been proposed to tackle these issues, it connsists of two phases: the learing phase and testing phase. The learning phase supports the cluster feature concept to summarize the data set and special kind of similarity measures to deal with mixedtype attributes and incomplete data. While the testing phase uses the KNN calssifer due to its low computational cost. The experimental results shows that the proposed classifier has a higher classification accuracy and lower running time in actual data and incomplete data when randomly remove 5, 10, and 15% persantege of data inspite of droping down the overall detection accuracy as compared with SVM and KNN classifier.

## REFERENCES

[1] Gao, Xianwei; Shan, Chun; Hu, Changzhen; Niu, Zequn and Liu, Zhen (2019). An Adaptive Ensemble Machine Learning Model for Intrusion Detection. IEEE Access, 1–1. http://doi:10.1109/ACCESS.2019.2923640

[2] Zhou, Yuyang; Cheng, Guang; Jiang, Shanqing and Dai, Mian (2020). Building an efficient intrusion detection system based on feature selection and ensemble classifier. Computer Networks, 174, 107247–. http://doi:10.1016/j.comnet.2020.107247

[3] Hajisalem, Vajiheh and Babaie, Shahram (2018). A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection. Computer Networks, 136, 37–50. http://doi:10.1016/j.comnet.2018.02.028

[4] Gauthama Raman, M.R.; Somu, Nivethitha; Kirthivasan, Kannan; Liscano, Ramiro and Shankar Sriram, V.S. (2017). An Efficient Intrusion Detection System based on Hypergraph - Genetic Algorithm for Parameter Optimization and Feature Selection in Support Vector Machine. Knowledge-Based Systems, S0950705117303209–. http://doi:10.1016/j.knosys.2017.07.005

[5] Ali, Mohammed Hasan; AL Mohammed, Bahaa Abbas Dawood; Ismail, Madya Alyani Binti and Zolkipli, Mohamad Fadli (2018). A new intrusion detection system based on Fast Learning Network and Particle swarm optimization. IEEE Access, 1–1. http://doi:10.1109/ACCESS.2018.2820092

[6] Sultana, Nasrin; Chilamkurti, Naveen; Peng, Wei and Alhadad, Rabei (2018). Survey on SDN based network intrusion detection system using machine learning approaches. Peer-to-Peer Networking and Applications. http://doi:10.1007/s12083-017-0630-0

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 6, Issue 6, June 2023 |

[7] Alazzam, Hadeel; Sharieh, Ahmad and Sabri, Khair Eddin (2020). A Feature Selection Algorithm for Intrusion Detection System Based on Pigeon Inspired Optimizer. Expert Systems with Applications, 113249–. http://doi:10.1016/j.eswa.2020.113249

[8] Shah, Syed Ali Raza and Issac, Biju (2017). Performance comparison of intrusion detection systems and application of machine learning to Snort system. Future Generation Computer Systems, S0167739X17323178–. http://doi:10.1016/j.future.2017.10.016

[9] Khraisat, Ansam; Gondal, Iqbal; Vamplew, Peter and Kamruzzaman, Joarder (2019). Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity, 2(1), 20–. http://doi:10.1186/s42400-019-0038-7

[10] da Costa, Kelton A.P.; Papa, João P.; Lisboa, Celso O.; Munoz, Roberto and de Albuquerque, Victor Hugo C. (2019). Internet of Things: A Survey on Machine Learning-based Intrusion Detection Approaches. Computer Networks, S1389128618308739–. http://doi:10.1016/j.comnet.2019.01.023

[11]Setareh Roshan, Yoan Miche, Anton Akusok, Amaury Lendasse; "Adaptive and Online Network Intrusion Detection System using Clustering and Extreme Learning Machines", ELSEVIER, Journal of the Franklin Institute, Volume.355, Issue 4,March 2018,pp.1752-1779.

[12]Pinjia He, Jieming Zhu, Shilin He, Jian Li, and Michael R. Lyu; "A Feature Reduced Intrusion Detection System Using ANN Classifier", ELSEVIER, Expert Systems with Applications,Vol.88,December 2017 pp.249-247

Impact Factor
7.54

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com