iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XI **Month of publication:** November 2023

DOI: https://doi.org/10.22214/ijraset.2023.56689

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

Safeguarding Connections: Machine Learning Powered Intrusion Detection

Somasekhar T[1], Moniesh S[2], Monika N[3], Pavithra R[4], Sindhura H[5]
[1]*Associate Professor, Dept. of Computer Science, K S Institute of Technology, Bangalore, Karnataka*
[2, 3, 4, 5]*Dept. of Computer Science, K S Institute of Technology, Bangalore, Karnataka*

Abstract: ***It's crucial to have reliable intrusion detection systems. a cutting-edge method of machine learning-based intrusion detection. Our solution uses cutting-edge algorithms to detect and eliminate any threats instantly, acting as a preventative measure against a wide range of cyberattacks. Since the model has been trained on a large number of datasets, it can eventually strengthen network security by evolving and adapting to new threats. Naive Bayes (NB) classifiers and correlation-based feature selection (CFS) methods are used to reduce the amount of data. For attack classification, the Intrusion Detection System recommends using an Instance-Based Learning algorithm (IBK) in combination with a Multilayer Perceptron (MLP).***
Keywords: Support Vector Machine (SVM), Multilayer Perceptron (MLP), Correlation Based Feature (CFS), Classifier subset evaluation, Intrusion Detection System (IDS), and Instance-Based Learning algorithm (IBK).

I. INTRODUCTION

The dynamic and expansive nature, has revolutionized the way we connect, communicate, and conduct business. With this paradigm shift, the security of networks and the data they handle has become a critical concern. The influx of smart devices, coupled with the constant exchange of information, underscores the imperative need for a sophisticated security infrastructure.

A. Machine Learning

Machine Learning is a type of data analysis that falls under Artificial Intelligence. It's like teaching a system to learn, make decisions, and recognize patterns without needing a lot of human instructions. Mainly two types: Supervised and Unsupervised. There are also other methods like Semi-supervised and Reinforcement Learning. Machine Learning is like training a computer to be smart on its own. Where the system learns from data, makes decisions, and figures out patterns without being explicitly told everything. Think of it as two main types: one where it learns from examples with labels (Supervised), and the other where it explores data to find patterns without labels (Unsupervised). There are also other ways, like using a bit of labeled 2 data with lots of unlabeled data (Semi-supervised), or learning through trial and error for the best outcomes (Reinforcement Learning). The end goal is to make the system smart enough to make good decisions on its own.

B. Intrusion Detection System

Intrusion, where unauthorized access can steal or damage computer and network data quickly, is a big problem in network security. It can even harm the hardware. While many techniques try to detect intrusion, getting it right all the time is tough. Accuracy, which depends on how well it detects real intrusions without giving false alarms, is a big challenge. Naïve Bayes and Support Vector Machine (SVM) are examples of clever algorithms that are used. This also offers feature reduction and normalization approaches so you can compare and determine which performs the best.

C. Naive Bayes

Bayesian classifiers are like statistical wizards in the computer world. They can predict the chance that a certain model belongs to a specific group.

They rely on something called Bayes' theorem, which is a bit like a math magic trick. The idea behind Bayesian classifiers is that, for a certain group, the characteristics of the thing we are looking at don't depend on each other. It's like assuming that different features don't really influence each other in a group. This assumption is called class conditional independence. So, these classifiers use the probability of features to figure out the likelihood of something belonging to a particular category. It is a smart way of making predictions in the world of data.

D. Support Vector Machine

Support Vector Machine (SVM) is like a smart learner in the world of computers, especially in cases where we have different types of data from different subjects. It's a bit like drawing lines or planes in a high-dimensional space to best separate different groups of data. The goal is to find the best line or plane (called a hyperplane) that creates the largest gap between the different groups. SVM uses something called kernel functions, which are like special tools to help draw these separating lines or planes. These tools can be linear, polynomial, radial basis, or sigmoid in nature. The main job of these tools is to maximize the space between the lines or planes, making the separation as clear as possible. Developers and researchers love SVM because it's great for image processing and recognizing patterns in data. When you are dealing with SVM, you often have two sets of data: one for training the computer and another for testing how well it learned. The training data has labels like "this is what we're looking for," and the computer tries to learn and then recognize these things in the testing data. It is like teaching the computer to recognize patterns on its own.

II. LITERATURE SURVEY

Cyber threats are a big problem worldwide due to the growing use of the internet. To tackle this 4 issue, many people have explored various approaches and solutions.

1) Title: System for detecting intrusions using machine learning

Author: Anish Halimaa A, Dr. K.Sundarakantham

Published: 2019

The Network Intrusion Detection (NID) system is an essential tool for maintaining the security of computer networks, which is a major concern. With the aid of machine learning algorithms, these systems—which are intended to identify and stop network attacks—are becoming more intelligent. The great thing about machine learning in this situation is that it doesn't require specialized expertise to function, unlike black- or white-list models, which require expert knowledge. In this paper, the focus is on a specific machine learning method called equality constrained optimization-based extreme learning machine. The idea is to make the learning process adaptively incremental, meaning it figures out the best number of hidden neurons on its own. The paper introduces a smart strategy to optimize the learning process, and they tested this approach on network intrusion detection. The results from their experiments show that this approach is effective. It not only builds models that detect attacks well but also learns really quickly.

2) Title: The use of machine learning techniques in an intrusion detection system.

Author: Mandeep Kaur, Usman Shuaibu Musa, Aniso Ali, Megha Chhabra.

Published: 2020

With the widespread use of computer networks, keeping them secure is a big challenge. To maintain the availability, integrity, and secrecy of the network, network managers employ intrusion detection systems, or IDS, to scan network traffic for any unauthorized or malicious activity. In this sense, intrusion refers to a malicious security breach. An IDS keeps an eye on network traffic, looking for signs of malicious activities or known threats. It alerts administrators when it spots something suspicious. Two types are Misuse or Signature-based Detection method collects information, analyses it, and compares it to a database of known attack signatures and Anomaly Detection method sees any action deviating from normal behaviour as potentially malicious. The proposed paper gives an overview of different efforts to build efficient IDS using machine learning. It explores single, hybrid, and ensemble ML classifiers, evaluating them across seven datasets.

3) Title: Network Intrusion Detection using Machine Learning Technique along with Feature Selection.

Author: Md. Mahbubur Rahman, Billal Mohammed Yasin Jisan, Kazi Abu Taher

Published: 2020

A novel machine learning system called Novel Machine Learning System for Network Traffic has been developed to determine if network traffic is safe or hazardous.

Combining a feature selection technique with a supervised learning algorithm to produce the best model with an emphasis on potential problem detection was the aim. In this study, they found that the support vector machine (SVM) strategy is not as effective as artificial neural network (ANN) based machine learning with a wrapper feature selection method for classifying network traffic. Their suggested model performs better than other current models in terms of successfully detecting intrusions in network traffic, according to the findings of their performance test conducted using the NSL-KDD dataset.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

III. OBJECTIVES

1) Using machine learning algorithms to create an effective Intrusion Detection System (IDS) for network security.
2) Correlation-based Feature Selection (CFS) and Naïve Bayes (NB) should be used for dimensional reduction.
3) Use the Instance-Based Learning algorithm (IBK) and Multilayer Perceptron (MLP) to assess the IDS performance.
4) Detect intrusions with excellent accuracy, especially when using fewer features.
5) Compare the proposed model with existing methods for intrusion detection efficiency.

IV. METHODOLOGY

We are investigating a particular class of cyberattack on the CIC IDS-2017 dataset known as a port scan. In this assault, a computer is targeted in order to determine whether ports are accessible or open. Our method, which is depicted in Fig. 1, consists of several steps. First, we use CFS and Classifier subset assessment for dimensionality reduction, which is akin to selecting the salient features from our dataset, in order to deconstruct the data and concentrate on what matters most. Subsequently, we utilize MLP and IBK ML techniques for model construction and training. Equipment and surroundings: Weka 3.8.4 is a potent data mining and machine learning technology that we are utilizing. Weka is a data preparation and classification tool that works similarly to a Swiss army knife.

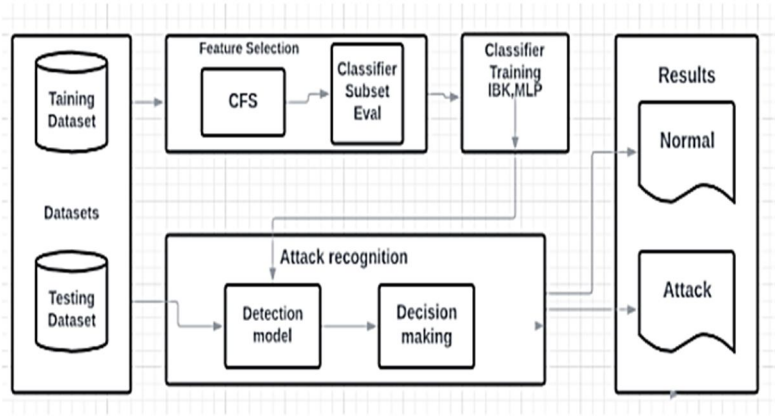


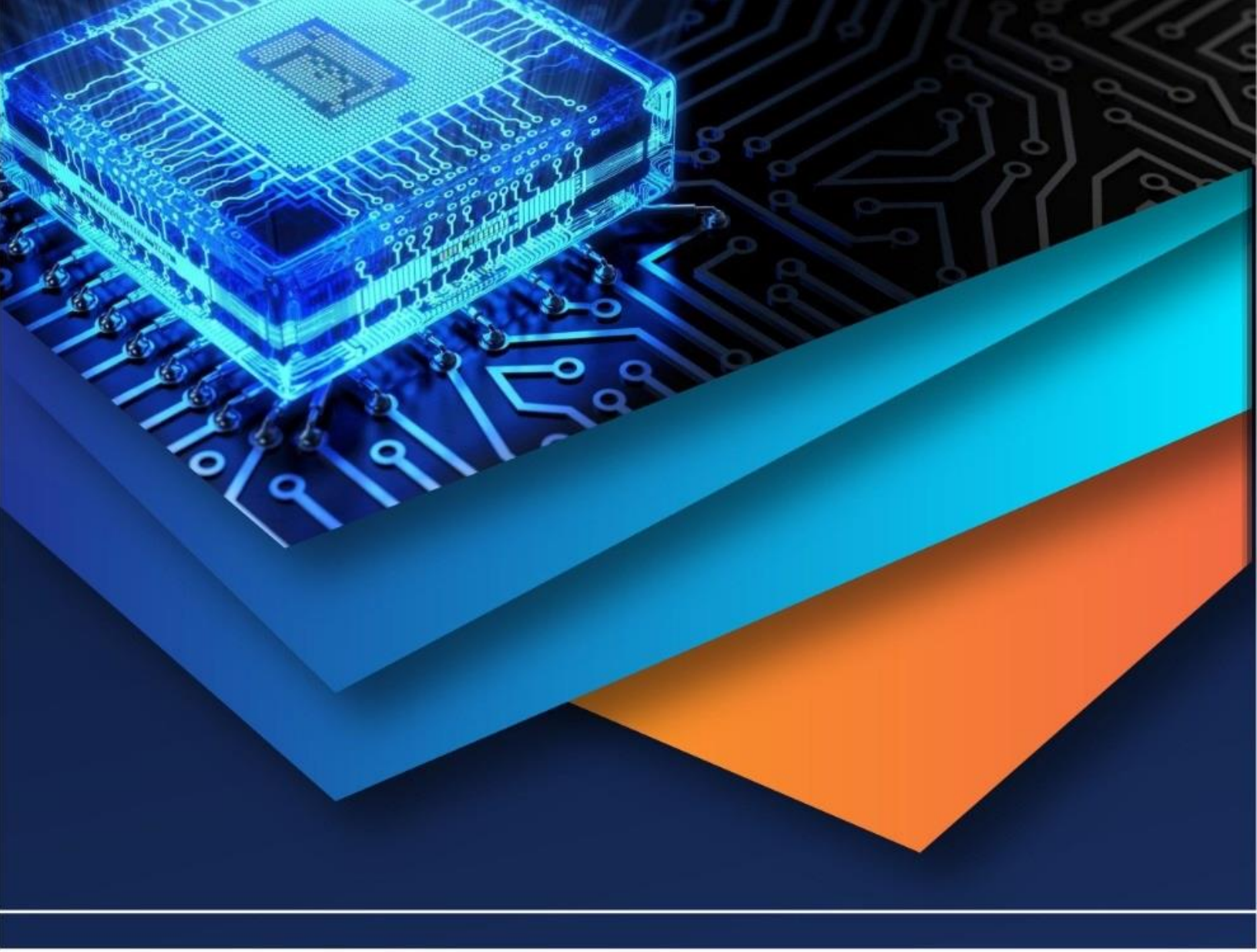
Figure 1: Simple Framework

V. SELECTION OF FEATURES

1) *Correlation-based Feature Selection:* The CFS technique looks for the most pertinent features in the data set by applying a Best-First search strategy. When evaluating attribute subsets, CFS takes into account each feature's unique predictive power as well as the correlation across features. By using a heuristic assessment function based on correlation, CFS prioritizes subsets where there is less than 1% intercorrelation between characteristics.
2) *Classifier Subset Evaluation (Naive Bayes):* Best-First is another search strategy for feature selection in classifier subset evaluation. Using either a different testing set or training data, it tests attribute subsets. Classifier subset assessment uses a classifier to determine how accurate a subset of attributes is. In this work, we evaluated the selected attribute set's accuracy using the Naive Bayes (NB) classifier. Since NB is a statistical classifier, it is predicated on the Bayes theorem and asserts that every feature in the data set is independent of every other feature.
3) *Multilayer Perceptron:* The most popular machine learning classifier is the feed-forward 1 and output layers. Imagine a network with multiple layers. The input layer would be the first layer, the output layer would be the last, and the hidden levels would be the middle layers.
4) *Instance Based Learning Algorithm:* IBK has two applications: regression and classification. A component of the lazy learning strategy is the IBK machine learning classifier, often known as the K Nearest Neighbours classifier (K-NN). The training model of such an ML classifier doesn't need to learn anything; instead, it uses the raw training cases to generate predictions. However, the basic principle of K-NN is that it uses a majority poll to ask questions of both the new instances and the majority of the k most comparable cases; the similarity between the data artificial neural network, or MLP, which has one or more layers separating the network's input vectors is determined by their distance from each other.

REFERENCES

[1] Larijani, H., Ahmad, J. and Mtetwa, N., 2019, July. A heuristic intrusion detection system for Internet-of-Things (IoT). In Intelligent computing proceedings of the computing conference (pp. 86-98). Springer, Cham.
[2] Garcia-Teodoro, P., Diaz-Verdejo, J., Macia-Fern ´ andez, G. and V ´ azquez, ´ E., 2009. Anomaly-based network intrusion detection: Techniques, systems and challenges. computers security, 28(1-2), pp.18-28.
[3] Tang, Y. and Chen, S., 2007. An automated signature-based approach against polymorphic internet worms. IEEE Transactions on Parallel and Distributed Systems, 18(7), pp.879-892.
[4] Liao, H.J., Lin, C.H.R., Lin, Y.C. and Tung, K.Y., 2013. Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications, 36(1), pp.16-24.
[5] Sarnovsky, M. and Paralic, J., 2020. Hierarchical intrusion detection using machine learning and knowledge model. Symmetry, 12(2), p.203.
[6] Salo, F., Nassif, A.B. and Essex, A., 2019. Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection. Computer Networks, 148, pp.164-175.
[7] Qureshi, A.U.H., Larijani, H., Mtetwa, N., Javed, A. and Ahmad, J., 2019. RNN-ABC: A new swarm optimization based technique for anomaly detection. Computers, 8(3), p.59.
[8] Mukkamala, S., Sung, A.H. and Abraham, A., 2005. Intrusion detection using an ensemble of intelligent paradigms. Journal of network and computer applications, 28(2), pp.167-182.
[9] A. Yulianto, P. Sukarno, and N. A. Suwastika, "Improving ad boost based intrusion detection system (ids) performance on cic ids 2017 dataset," in Journal of Physics: Conference Series, vol. 1192, p. 012018,IOP Publishing, 2019.
[10] A. Gharib, I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Anevaluation framework for intrusion detection dataset," in 2016 International Conference on Information Science and Security (ICISS), pp.1–6,IEEE, 2016.
[11] G. Holmes, A. Donkin, I.H. Witten, Weka: a machine learning workbench, in: Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on, IEEE, 1994, pp. 357–361.
[12] Noviyanto, A., Isa, S.M., Wasito, I. and Arymurthy, A.M., 2011. Selecting features of single lead ECG signal for automatic sleep stages classification using correlation-based feature subset selection. IJCSI International Journal of Computer Science Issues, 8(1-5).
[13] Hall, M.A. and Smith, L.A., 1998. Practical feature subset selection for machine learning.
[14] Hao, H., Liu, C.L. and Sako, H., 2003, August. Comparison of genetic algorithm and sequential search methods for classifier subset selection. In Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. (pp. 765-769). IEEE.
[15] Koc, L., Mazzuchi, T.A. and Sarkani, S., 2012. A network intrusion detection system based on a Hidden Na¨ıve Bayes multiclass classifier. Expert Systems with Applications, 39(18), pp.13492-13500.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY