International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

DETECTION OF WIRELESS NETWORK ATTACKS USING SUPERVISED MACHINE LEARNING TECHNIQUE

Rajesh Kumar K[#1], Edison Singaraj M [#2], Naveen V[#3], Umesh K[#4] Arun Pillai K.V[#5]

[#1,2,3,4] Department of Computer Science And Engineering, Jaya Sakthi Engineering College, Chennai,602024,Tamil Nadu ,India

Corresponding Author:Rajeshkumar.jsec@gmail.com, vnaveenchn@gmail.com

Abstract: ***The security and integrity of computer networks are seriously threatened by network assaults. Keeping a safe network environment requires the capacity to anticipate and stop these threats. Supervised machine learning methods have become powerful instruments for attacking network traffic and spotting patterns that point to malicious behaviour. We provide an in-depth examination of supervised machine learning methods for network attack prediction. We gather the data, preprocess it, extract pertinent features, and structure it so that machine learning algorithms may use it.We assess these algorithms' performance. To understand the fundamental patterns and traits of network assaults, we look at how interpretability the trained models are. This enables network managers to comprehend the types of threats and create suitable.***

Keywords - Network Attacks, Supervised Machine Learning, Network security,Network traffic analysis, Malicious behavior detection, Feature extraction,Machine learning algorithm,Model interpretability,Network attack patterns,Network defense strategies,Naive Bayes Algorithm, Ridge Classfier ,Random Forest Classfier

I. INTRODUCTION

Attacks by malware This entails setting up ransomware, spyware, viruses, or worms as well as other dangerous software on a computer. These programmes have the ability to eavesdrop on user activities, corrupt files, encrypt files for ransom, and steal data.

Attacks known as denial-of-service (DoS) are designed to flood a system with traffic so that it cannot be used by authorised users. This indefinite quantity be achieved by sending a sizable amount of data grouping via the network or by taking advantage of system flaws.Attacks using a man-in-the-middle In this kind of assault, a third party eavesdrops on or intercepts communication between two parties with the intent to steal information or change messages.

Phishing attacks: These are attempts to victimise citizenry into discover reclusive information, view credit card numbers, usernames, or information Emails, texts, or phoney websites can be used for this. Attacks using SQL injection an individual identity. The picture shows a Sybil assault where an attacker node (AD) has numerous identities. .SQL injection attacks use database software vulnerabilities to inject malicious code that can steal data or change the database. Zero-day attacks are those that exploit software flaws that the vendor is unaware of. These attacks are particularly risky because no patch is available to address the vulnerability. Firewalls: Firewalls serve as a barrier between a trustworthy and an entrusted network, such as the internet. They can filter data and prevent unauthorized access. Intrusion Detection Systems (IDS): These systems scan network traffic and system activity for indicators of malicious activity. When an IDS identifies a possible attack, it can inform system administrators. Intrusion Prevention Systems (IPS) detect and prevent Signature-based intrusion detection systems Compare network traffic against a database of previously known attack signatures. When a match is identified, the IDS generates an yell Anomaly-based IDS: These systems examine network traffic for patterns that differ from usual behaviour. This is more successful at

identifying zero-day threats, although it can produce false positives.
Early detection of assaults: An IDS may identify attacks in real time, allowing administrators to respond before they cause major harm.
Improved security posture: Using an IDS can dissuade attackers and make it more difficult for them to breach a network.
Overview Patterns or abnormalities in wireless network traffic are often identified using algorithms trained on labelled data.The attacker gathers information from wireless network traffic. ThisDistributed denial-of-service (DDoS) assaults are rapidly increasing, affecting Internet Service Providers (ISPs) and persons in a clandestine way. Thus, intrusion detection systems (IDSs) must evolve to deal with these more complex and demanding security threats. Traditional IDSs are vulnerable to zero-day attacks since they are frequently signature-based detection systems. However, in the absence of up-to-date tagged training datasets, these ML/DL-based IDSs become useless. The privacy issues of massive datasets, along with the widespread emergence of adversarial attacks, make it hard for large businesses to share sensitive data. Federated Learning (FL) is gaining traction in both academia and business as a new subfield of machine learning that aims to train a global statistical model among numerous remote users.

II. MATERIALS AND METHODS

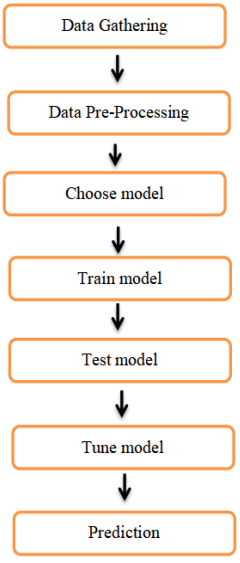


Fig. 1. Block diagram showing the outline of the proposed model

2. **List Of Modules**

- Preparing the data
- Analyzing it for visualization
- Ridge Classifier Algorithm(RC)
- Random forest Algorithm(RF)
- Bernoulli naive bayes Algorithm(BNB)
- Deployment Using Result

1.Preparing the data

The error rate of the Machine Learning (ML) model, which is believed to be relatively comparable to the actual error rate of the dataset obtained using automated validation approaches. It's possible that the validation approaches won't be necessary If the quantity of data is sufficient to represent the population. That being said, working with data samples that might not accurately reflect the It is typical to populate a given dataset in real-world settings. Identify duplicate values, missing values, and the data type**.** —integer or float—please refer to the data description. The dataset that is utilised to objectively assess a model's fit while fine-tuning its hyper parameters on the training dataset.

2.Analyzing it for visualization

Visualisation is a Key skills in applied statistics and machine learning. Statistics is primarily concerned with quantitative data explanation and figure. Data visualisation delivers an efficient set of approaches to gaining qualitative knowledge. This is approaches for exploring and learning about a dataset, as well as spotting trends,Outliers and faulty data and other issues. With a little domain knowledge, data visualisations may be used to explain and demonstrate important relationships in plots and charts that are more visceral and meaningful to stakeholders than measures of association or significance. Data visualisation and exploratory data analysis are two independent disciplines, and we recommend going deeper into some of the books listed at the end.

3.RIDGE CLASSIFIER:

Like many other classification algorithms, the Ridge Classifier is based on a linear model. IT anticipate that location is a linear state between the sign property and the target classes. In other words, it

assumes that the decision boundary separating different classes is a linear hyperplane.

Regularization: The term "Ridge" in Ridge Classifier comes from Ridge Regression. It means that the algorithm uses L2 regularization to prevent over fitting. Regularization helps in controlling the complexity of the framework and cut down the danger of it adjustment interference in the education data.

The Ridge Classifier seeks to minimise an objective function that is divided into two parts:

The loss function calculates the model's inaccuracy in predicting class labels. Typically, it employs a loss function suited for classification issues, such as the cross-entropy loss or the log-loss.

The L2 regularisation term is applied to the loss function. This term prevents the model from overemphasising any specific characteristic, resulting in a stable and well-behaved model**.** Training: To train the Ridge Classifier, you must feed it with a labelled dataset including input features and their labels. During training, the model's parameters (coefficients) are adjusted using optimisation techniques such as gradient descent to minimise the objective function.

Prediction: After training the Ridge Classifier, you can use it to generate predictions about fresh, previously unknown data. It assigns class labels using a linear combination of input characteristics and previously learnt coefficients. The projected class is determined by the highest score.

Multi-Class Classification: The Ridge Classifier is naturally suited to multi-class classification jobs. It commonly adopts a one-vs-rest (OvR) or one-vs-one (OvO) technique to extend binary classification to several classes**.**

4.NAIVE BAYES ALGORITHM

The collection of rules is referred regarded as "naive" because it makes a strong and frequently unrealistic assumption: it implies that the capabilities utilised to create predictions are conditionally independent, resulting in the elegance label. This implies that it considers each function as if it had no link to another function, significantly simplifying the computations.

Binary Features: BernoulliNB is well-suited for datasets with binary features, which can take on values of 0 or 1. These binary features are frequently utilised to express the existence (1) or absence (0) of specific properties or characteristics in the data.

Independence Assumption: Like other Naive Bayes algorithms, BernoulliNB uses the "naive" assumption that features are independent. The presence or absence of one characteristic is presumed to have no bearing on the presence or absence of any other attribute. This simplifying assumption helps the software to compute probabilities faster.

.

The class with the highest posterior probability determines the projected class for the data point. In binary classification, you choose the class with the highest probability as the expected class.

5.RANDOMFOREST CLASSIFIER:

A Random Forest Classifier is an intricate machine learning technique that may be employed in both classification and regression applications. It is a system of ensemble learning that combines the predictions of many decision trees to produce a more accurate and robust model. The following is an explanation of how the Random Forest Classifier acts.

Decision Trees: Decision trees constitute the base for the Random Forest Classifier. A decision tree is a fundamental tree-like structure that relies on a chain of binary
judgements to categorise data. Each internal node in the tree reflects a feature-based decision, whereas each leaf node a combination class label or a numerical value (for regression)**.**

Ensemble Learning: A random forest is a collection of decision trees. Instead Random Forests can give information on feature relevance, allowing you to understand which features are most significant in creating predictions. This might help with feature selection or recognising the significance of different variables in your dataset.

Input data, expected outcome is accuracy.

6. Deployment

Deploying the mathematical framework in the Django Framework and forecasting output

In this module, the taught deep learning model is translated into a hierarchical format for data file (.h5), which is then deployed in our Django framework to give a better user experience and forecast if the picture we provide is CKD or not.

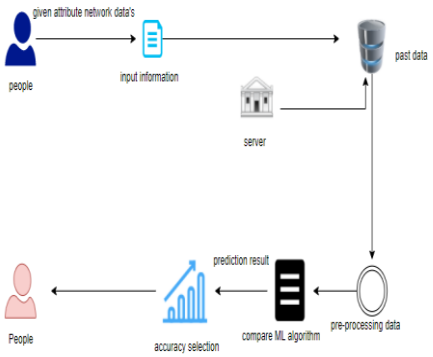
6.1 Django

Django is a Python web framework for creating secure, maintainable websites. Django, built by expert developers, handles the most of the pain of developing for the web, allowing you to focus on designing your project instead of reinventing the wheel. It is free and open source with an active community.excellent documentation, and offers both free and paid support options.

III MODELING AND ANALYSIS

3.1.SYSTEM ARCHITECTURE

Architectural design is the process of identifying the subsystems that make up a system, as well as the framework for subsystem control and communication.This design technique creates a software infrastructure description..It establishes an abstraction level from which designers may identify the system's functional and performance properties.It checks all high-level designs. creates and shows general designs for internal and external It publishes all early versions of user documentation. It determines and publishes preliminary testing guidelines. as well as a software integration timetable.A system construction is an abstract picture of a system's structure, conduct that and other traits. An explanation of architecture is a formal account and image of a system scheduled in a way that allows for reasoning about it.



3.2 Preparing Dataset:

This Dataset carry 374662 platters. It is confidential into 5 classes.

1. Blackhole
2. Flooding
3. Grayhole
4. Normal
5. TDMA

3.3 import dataset

	id	Time	Is_CH	who_CH	Dist_To_CH	ADV_S	ADV_R	JOIN_S	JOIN_R	SCH_S	SCH_R	Rank	DATA_S	DATA_R	Data_Sent_To_BS	dist_CH_To_BS	send_code	E:
0	101000	50	1	101000	0.00000	1	0	0	25	1	0	0	0	1200	48	130.08535	0	
1	101001	50	0	101044	75.32345	0	4	1	0	0	1	2	38	0	0	0.00000	4	
2	101002	50	0	101010	46.95453	0	4	1	0	0	1	19	41	0	0	0.00000	3	
3	101003	50	0	101044	64.85231	0	4	1	0	0	1	16	38	0	0	0.00000	4	
4	101004	50	0	101010	4.83341	0	4	1	0	0	1	25	41	0	0	0.00000	3	

The dataset can change into numbers using the python, sklearn is a machine learning package

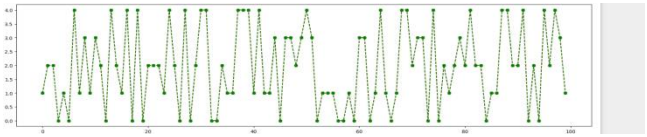
It is a scientific part of Python which gives lightning-fast mathematical operations for computations.

It is used to acquire information extracted from the numpy library structures and perform manipulations.

3.4.Training algorithm steps

Data preparation: Collect network traffic statistics, including both regular and attack occurrences. This information may originate from network traffic capturing techniques or publicly available databases. Feature Engineering: Extract useful features from network traffic data that might help in attack detection. Data Preprocessing: Ensure that all features are in a format appropriate for the algorithm model. This might involve: Scaling: Using approaches such as z-score normalisation or min-max scaling, numerical characteristics are standardised or normalised to a common scale. Encoding: Use techniques such as one-hot encoding to convert categorical characteristics (for example, protocol type) into numerical values. Model training: Labelling: Label your data as "normal" or a specific form of assault (for example, DoS attack, port scan).Model training involves putting the algorithm into action.

IV result and discussion



What sort of algorithm has been used for preprocessing with the model? Training and testing this model works and predicting accurately using

Random classification algorithm to get the accuracy 99% we discuss Accuracy: The proportion of total correct predictions; in other words, On average does the model correctly forecast defaulters

V CONCLUSION

The process of analysis commenced with data purification and analysing, followed by lacking value analysis, model creation, and evaluation. The top accuracy score on the public examination collection will be calculated by juxtaposing each algorithm to the sorts of all WSN hits for future prognosis possibilities.

This provides the following information into diagnosing the network assault of each new connection To provide a prediction model using artificial intelligence to increase human accuracy and expand the area of early detection. It can be deduced from this model that, area analysis and the usage of machine learning approach is effective in constructing prediction models that may assist network sectors minimize the long process of of diagnosis and eradicate any human error.

REFERENCES

1. Time-Aware Gradient Attack on Dynamic Network Link PredictionJinyin Chen , Jian Zhang, Zhi Chen, Min Du, and Qi Xuan , SeniorMember, IEEE FEBRUARY 2023.

2. FooBaR: Fault Fooling Backdoor Attack on Neural Network Training Jakub Breier , Xiaolu Hou , Martı´n Ochoa , and Jesus Solano MAY/JUNE 2023.

3. MiTFed: A Privacy Preserving Collaborative Network Attack Mitigation Framework Based on Federated Learning Using SDN and BlockchainZakaria Abou El Houda , Abdelhakim Senhaji Hafid , and Lyes Khoukhi , Senior Member, IEEE JULY/AUGUST 2023.

4. Evaluation of Machine Learning Algorithms for Intrusion Detection System in WSN Mohammed S.Alsahli, Marwah M.Almasri, Mousa Al-Akhras,Abdulaziz I.Al-Issa, Mohammed Alawairdhi, 2021.

5. Machine Learning in Wireless Sensor Networks: Algorithms, Strategies,and Applications Mohammad Abu Alsheikh, Shaowei Lin, Dusit Niyato and Hwee-Pink Tan 2015.

6. Performance Evaluation of Machine Learning Techniques For Dos Detection In Wireless Sensor Network Lama Alsulaiman and Saad Al-Ahmadi, 2021.

7. Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications
Mohammad Abu Alsheikh, Shaowei Lin, Dusit Niyato and Hwee-Pink Tan, 2015.

8. Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey Dr. E. Baraneetharan, 2020.