JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

IMPLEMENTATION PAPER ON CYBER THREATS DETECTION APPLYING RANDOM - FOREST ALGORITHM

Divya D P[1], Divyashree S[2], Ananya Vinayak Shanbhag[3] , Bhuvana H R[4] , Deepthi M[5]

[1]Student ,[2]Assistant Professor, [3]Student , [4]Student,

[5]Student

ABSTRACT - Intrusion-Detection Systems (IDS) play a vital part in preserving computer systems and networks from unauthorized access, attacks, and malicious activities. With the ever-evolving landscape of cyber threats, the requirement for resilient and flexible intrusion detection approaches becomes paramount. One such approach involves employing artificial intelligence algorithms, and in this instance, the Random-Forest algorithm appears as a powerful tool. Utilizing machine learning (ML) to detect intrusions involves leveraging computational algorithms and statistical models to automatically identify abnormal activities or potential security threats within a computer system or network. Conventional intrusion detection methods frequently rely on rule-based approaches, but machine learning introduces a more adaptive and data-driven paradigm. Random Forest is a popular and efficient choice for intrusion-detection due to several key advantages that address matters pertaining to the nature of cybersecurity data and the requirements of intrusion-detection systems. The utilization of a Random Forest-based Intrusion-Detection System (IDS) is a noteworthy development in enhancing the cybersecurity posture of a network or system. Utilize ensemble learning, feature importance analysis, and adaptability to diverse patterns positions Random Forest as a robust and efficient solution for identifying both known and novel threats.

1. INTRODUCTION

Computer networks and systems must be protected from malicious activity, unauthorised access, and attacks using Intrusion-Detection Systems (IDS). Because cyber-threats are constantly changing, it is critical to have intrusion-detection methods that are both sturdy and flexible. Using machine learning algorithms is one such method; in this setting, the Random Forest algorithm seems to be an effective instrument. The ensemble learning technique Random Forest is renowned for its adaptability and effectiveness in managing intricate data patterns. Utilising an assortment of variables and patterns found in the data, it provides the capacity to distinguish between typical and unusual network activity When it occurs to intrusion detection. The intention of this computer learning additionally cybersecurity integration is to cut back false positives, improve proactive threat detection, and offer a flexible and intelligent defence system against ever- evolving cyberthreats. This investigation into the field of "Intrusion- System of Detection Employing Random Forest Algorithm" aims to strengthen network security by utilising machine learning's advantages and providing a proactive approach against the ever-evolving and complex terrain of cyber threats.

2. OBJECTIVES

This project aims to produce an ensemble of decision trees for reliable and accurate intrusion detection by developing and implementing the Random Forest algorithm.

3. EXISTING SYSTEM

Conventional intrusion detection techniques include anomaly-based detection, which finds departures from predetermined baselines of typical behaviour, and signature-based detection, which uses prepared attack signatures to swiftly identify known threats. While anomaly-based detection excels at recognising fresh assaults but has a higher tendency to produce false positives, signature-based detection is good for known threats but susceptible to evasion strategies. The goal of hybrid approaches is to improve accuracy by combining the best features of both systems. Network behaviour analysis examines linkages in network traffic, heuristic-based detection makes use of flexible criteria, and stateful inspection keeps a watchful active connections. The basis for intrusion detection is founded on traditional methods like rule-based detection, which depend on predefined

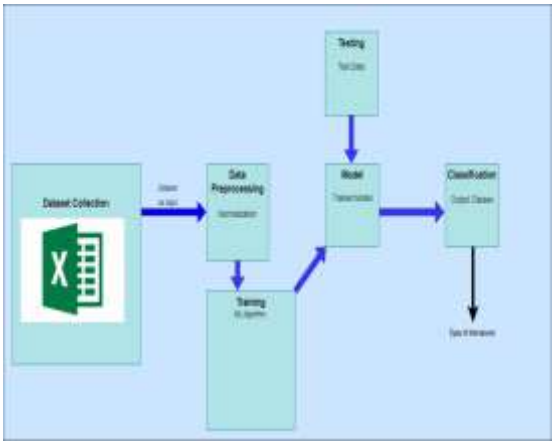
conditions. However, because of the intricacy of modern cyber threats, these methods may not be as effective as they once were, which is why there is a growing interest in advanced techniques like machine learning-based approaches. The basis for intrusion detection is founded on traditional methods like rule-based detection, which depend on predefined conditions. However, because of the intricacy of modern cyber threats, these methods may not be as effective as they once were, which is why there's a growing interest in advanced techniques like machine learning-based approaches.Limitations of Existing System:

There are various problems with traditional approaches.

1. Because signature-based detection depends on a database of known attack signatures, it can only identify dangers that have already been discovered.
2. Attackers can use evasion strategies to avoid being detected by signature-based systems, such as changing or obscuring known signatures.
3. Since alerts may be triggered by legitimate activities that depart from predetermined baselines, anomaly-based detection typically results in greater false-positive rates.

4. PROPOSED SYSTEM

Using statistical models and computational techniques, machine learning (ML) for intrusion-detection identifies anomalous activity or possible security risks in a computer system or network automatically. While rule-based techniques are often employed in traditional intrusion-detection systems, machine- learning offers a more flexible and data-driven paradigm. Because of a several significant benefits that address issues connected to the characteristics of cybersecurity data and intrusion-detection system needs, Random-Forest is a well- liked and successful option for intrusion-detection.



5. ADVANTAGES OF PROPOSED SYSTEM

1. Inconsistent Forest's ability to adjust to a diversity of intricate patterns within network information makes it useful for spotting both known and unknown dangers. It’s able to identify minute departures from typical behaviour because of its broad feature analysis capabilities.

6. Unbalanced datasets, which frequently occur in intrusion-detection and involve a large proportion of legitimate traffic relative to criminal activity, can be handled by Random-Forest. Because the technique is ensemble-based, biases are reduced and occurrences of minority classes can be accurately detected.

RESULTS

Test cases:

Test Case Id	Test Case Name	Test Case Description	Test Steps				Test Status P/F
			Steps	I/P Given	Expected O/P	Actual O/P	
TC01	Registration	To verify that the user has registered by entering valid detail	Enter all valid user detail	Valid detail	Registered successfully	Registered successfully	Pass
	Registration	To verify that the user has registered by entering valid detail	If entered details are not valid or some of the details are missing	Invalid detail	Registered unsuccessfully	Registered unsuccessfully	Pass
TC02	Login	To verify that the user has entered valid username/ email and password	Login with valid user name/email and password	Valid user name/email and password	Login successful	Login successful	Pass
	Login	To verify that the user has entered valid username/ email and password	Login with invalid user name/email and password	Invalid user name/email and password	Login unsuccessful	Login unsuccessful	Pass
TC 03	Train the data	To train data	Implement algorithm	Train button is clicked	Training successful	Training successful	Pass
TC 04	Intrusion Detection	Test the trained algorithm	Upload test dataset	Select dataset file from the system	Intrusion detected successfully	Intrusion detected successfully	Pass

Admin login

Admin has the responsibility of overseeing the training phase of the application. Once logged in successfully admin can upload the training dataset, initiate the training process using the RF algorithm, create a model, and view the training results.

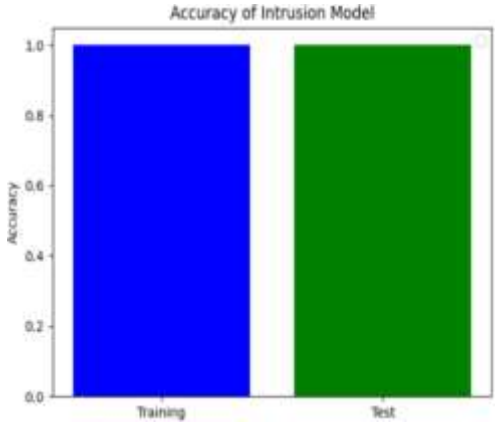
User login

User has the responsibility of overseeing the testing phase of the application. He has to first register with his details. After successful login, he can input test data and check the type of intrusion.

Accuracy of Intrusion Model:

The proportion of correctly categorised instances (both normal and intrusive) to the total-number of instances in testing dataset is used to calculate accuracy in a Random Forest intrusion detection model. The following is the accuracy formula:

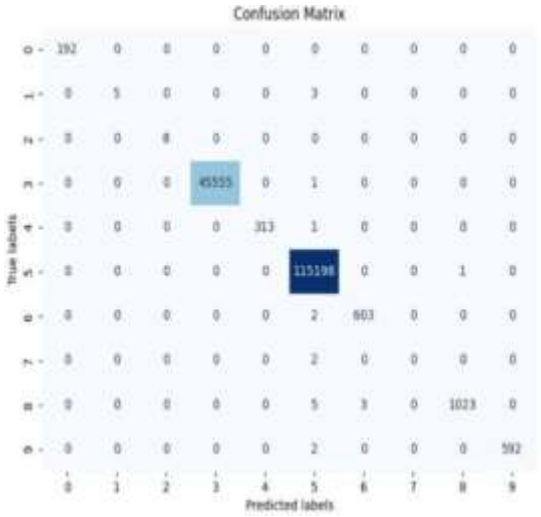
$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$



To further describe accuracy in binary classification, where examples classified as either a normal or invasive, the terms True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) can be used:

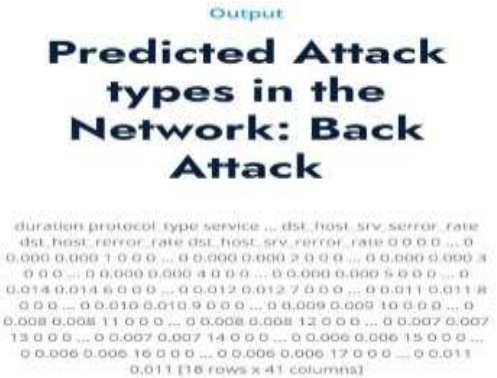
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Confusion Matrix:



A table that aids in evaluating a classification model's performance is called a confusion matrix. To illustrate how successfully the model categorises examples, it shows the counts of true-positive, true-negative, false-positive, and false- negative predictions. Accurate positive forecasts are known as true positives, accurate negative forecasts are known as true negatives, and inaccurate positive and negative forecasts are known as false positives and false negatives, respectively. These counts provide insight into the model's accuracy, precision, recall, as well as additional important performance-metrics that are essential to assess the model's efficacy in data classification.

Output Screen:



7.CONCLUSION

In summary, a major step towards improving a network or system's cybersecurity posture is the deployment of a intrusion-detection system (IDS) based on a random forest. Because Random Forest can adapt to different patterns, use feature importance analysis, and ensemble learning, It's an reliable and effective method for recognising dangers that are both known and unknown. Important issues in intrusion-detection are handled by the ability to manage unbalanced datasets, withstand overfitting, and attain excellent precision and minimal false positives.

8.REFERENCES

1. Bace, R.; Mell, P. *Intrusion Detection Systems*; NIST Special Publication on Intrusion Detection Systems; NIST: Gaithersburg, MD, USA, 2001.

2. Mbona, I.; Eloff, J.H.P. Detecting Zero-Day Intrusion Attacks Using Semi-Supervised Machine Learning Approaches. *IEEE Access* **2022**, *10*, 69822–69838.

3. Buczak, A.L.; Guven, E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 1153–1176.

4. Mishra, P.; Varadharajan, V.; Tupakula, U.; Pilli, E.S. A Detailed Investigation and Analysis of Using Machine Learning Techniques for Intrusion Detection. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 686–728.

5. Aldweesh, A.; Derhab, A.; Emam, A.Z. Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. *Knowl.-Based Syst.* **2020**, *189*, 105124.

6. Dhanabal, L.; Shantharajah, S.P. A Study on NSL- KDD Dataset for Intrusion Detection system Based on Classification Algorithms. *Int. J. Adv. Res. Comput. Commun. Eng.* **2015**, *4*, 446–452.

7. Olouhal, O.U.; Yange, T.S.; Okerekel, G.E.; Bakpol, F.S. Cutting Edge Trends in Deception Based Intrusion Detection Systems-A Survey. *J. Inf. Secur.* **2021**, *12*, 250–269

8. Shitharth, S.; Kshirsagar, P.R.; Balachandran, P.K.; Alyoubi, K.H.; Khadidos, A.O. An Innovative Perceptual Pigeon Galvanized Optimization (PPGO) Based Likelihood Naïve Bayes (LNB) Classification Approach for Network Intrusion Detection System. *IEEE Access* **2022**, *10*, 46424–46441.

9. Prashanth, S.K.; Shitharth, S.; Praveen Kumar, B.; Subedha, V.; Sangeetha, K. Optimal Feature Selection Based on Evolutionary Algorithm for Intrusion Detection. *SN Comput. Sci.* **2022**, *3*, 439.