IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134127

INTRUSION DETECTION WITH MACHINE LEARNING COMPARISON ANALYSIS

PROF.BHARATH M B[1], AMAR DADGE[2], B RAJASEKHAR[3], SANJAY D B[4]

Asst prof. Department of CSE, Dayananda Sagar University Bengaluru, India[1]

Computer Science Of Engineering Dayananda Sagar University Bengaluru, India[2-4]

Abstract: Machine learning techniques have brought about a revolution in various fields, with a significant impact on cyber security. In the face of growing cyber threats, the need for effective intrusion detection systems (IDS) has become more crucial than ever. These systems play a vital role in the timely and automatic detection and classification of cyber attacks, at both the network-level and the host-level. However, traditional IDS, which rely on conventional machine learning methods, often fall short in terms of reliability and accuracy.As the number of network-related applications, programs, and services continues to grow, so do the associated network security issues. Safeguarding the network against malicious activities is a challenging and critical task. In order to maintain a secure network environment, an effective system for detecting and identifying any suspicious activity is essential. This system is commonly known as an Intrusion Detection System (IDS).

I. INTRODUCTION

In the fast-paced world of cyber security, the ever-evolving landscape of threats requires a constant evolution of Intrusion Detection Systems (IDS) to safeguard digital environments. While traditional rule-based IDS are somewhat effective, they often struggle to keep up with the changing nature of cyber threats. To tackle this challenge, incorporating Machine Learning (ML) algorithms into intrusion detection methodologies has become increasingly popular. This article delves into the application of ML algorithms in intrusion detection and offers a comparative analysis of their effectiveness.

1.1 EASE OF USE

Exploring the Intersection of Intrusion Detection and Machine Learning:

The fusion of Intrusion Detection and Machine Learning aims to provide a comprehensive understanding of how ML algorithms can enhance the security posture of networks. The focus is on examining various ML algorithms, including Supervised Learning models like Support Vector Machines, Random Forests, Decision Trees, and Neural Networks, as well as Unsupervised Learning techniques such as k-means and hierarchical clustering.

Types of intrusion detection System

An Intrusion Detection System primarily monitors network processes and analyzes them to identify deviations from normal operations or any abnormalities. It scans network activities to detect suspicious behavior and policy violations. IDS are categorized into five main types: Network Intrusion Detection System, Host Intrusion Detection System, Protocol-based Intrusion Detection System, Application Protocol-based Intrusion Detection System, and Hybrid Intrusion Detection System. There are two primary methods of detection: misuse detection (signature-based) and anomaly detection.

II. PROBLEM DEFINATION

Machine Learning (ML) algorithms present a promising avenue for addressing this challenge by providing the capability to discern patterns and anomalies within network data. The specific problem to be addressed in this study is the design, implementation, and evaluation of an Intrusion Detection System leveraging various ML algorithms.

This includes Supervised Learning models such as Support Vector Machines, Random Forests, Decision Trees, and Neural Networks, as well as Unsupervised Learning techniques like k-means and hierarchical clustering

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134127

III. SCOPE OF THE PROJECT

In order to systematically study and evaluate ML application intrusion detection domain algorithms, particularly in terms of their efficacy versus rule-based approaches. The activity will include assessing a variety of ML algorithms, including supervised learning models contain Support Vector Machines, Random Forests , Decision Trees and Neural Networks hierarchical clustering unsupervised learning techniques shall also be investigated the research can incorporate comprehensive comparative analysis using important measure such as accuracy precision recall and F1 score to quantify the efficiency with which each algorithm In general, the research work is more focused in finding out of what is best algorithm that can more clearly detect any mis adjustment in the network. One of the factors that contribute to formulating a good solution to solve network security issues is selecting an appropriate algorithm for detecting anomaly in the network.

The ML algorithms that are chosen for this study include SVM, Decision Tree, Logistic Regression, Naive Bayes and Random Forest. Here, the NSL-KDD dataset is used for training and testing of topic 6 ML model. The dataset is prepared and the features are extracted from it. These features are subsequently used in the ML algorithms and an analysis of these algorithm performances over different intrusions such as Probe, DoS, R2L U2R is made. The study provides a preliminary understanding of the ML algorithm that would be best utilized by Intrusion Detection System, which can detect how deviations in networks create chaos with respect to decision-making

IV. PROBLEM STATEMENT

Real-time detection enables intrusion detection systems to proactively identify and respond to threats as they emerge. By continuously analyzing network data in real-time, these systems can detect suspicious activities and potential security breaches before they escalate. This proactive approach is essential for maintaining a strong cyber security defense.

V. LITERATURE REVIEW

As of late, the increasing complexity and diversity of cyber threats have sparked a surge in interest in utilizing machine learning methods to improve intrusion detection systems (IDS). This literature review delves into the current state of research in this area, with a focus on the intersection of intrusion detection and machine learning.

In recent years, the escalating complexity and diversity of cyber threats have spurred a growing interest in the application of machine learning techniques to enhance intrusion detection systems (IDS).This literature review aims to provide an in-depth exploration of the current state of research in the field, focusing on the intersection of intrusion detection Numerous Studies have been conducted to evaluate the efficacy of machine learning algorithms in detecting various forms of cyber intrusions, ranging from traditional signature-based methods to more advanced anomaly detection

VI. DESCRIPTION

That is, it moves towards increasing accuracy. intrusion- detection efficiency. moving beyond traditional rule-based approaches. The project involves the diversification of selection and implementation. Supervised Learning models and other ML algorithms.

Examples of Supervised Learning (such as Support Vector Machines, Random Forests, Decision Trees and Neural Networks) unsupervised learning. techniques (e.g., k-means, hierarchical clustering) A detailed comparative quantitative assessments will be conducted through these algorithms with measures such as accuracy, precision, recall and F1 score. The research focuses on the flexibility of ML-based IDS through

incremental learning for the system's ability to identify and respond to new. threats. The research is also applied as it focuses on the practical aspects, and evaluates computational efficiency, scalability, usability aspect of ML based IDS solutions. Essentially, the project seeks to Offer opinions and suggestions into the process of introducing advanced technologies. adaptive IDS frameworks that successfully adapting to the dynamism. landscape of cyber threats

a. OBJECTIVES

The main objective of this project is to predict which machine learning algorithms will give the highest accuracy for the particular website. Machine Learning algorithms can be used to train and test data.

b. PROPOSED DESIGN

In this project, we have suggested a model to suggest the Intrusion Detection System based on different machine learning

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

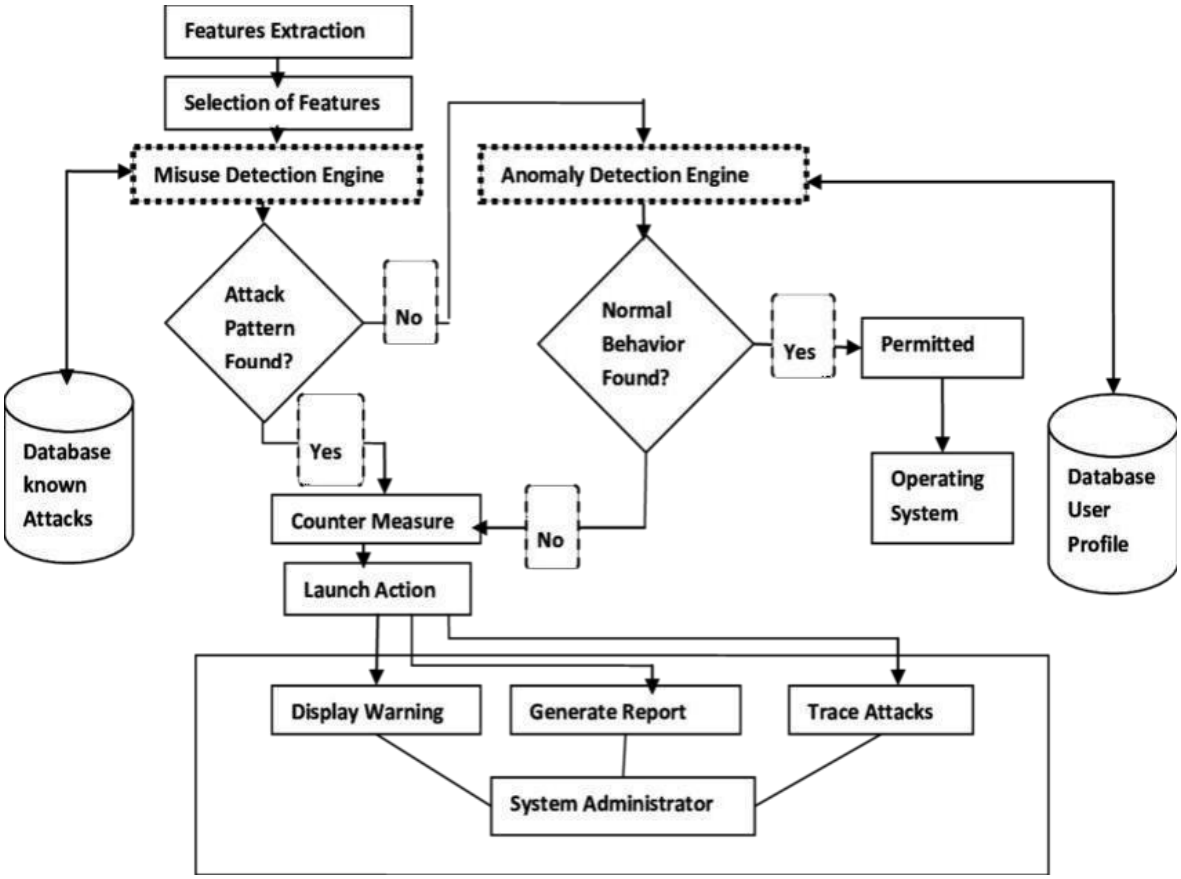
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134127

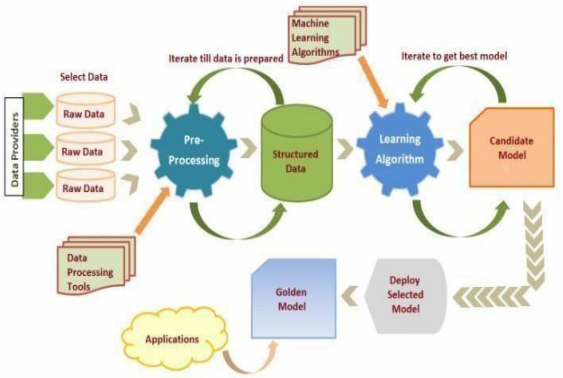
algorithms which will debug a website of any type and with which Intrusion Detection will be done with better accuracy.Previously, in old NIDS system NIDS If any new dynamic intrusion happens NIDS HIDS, which failed to discover the breaches. Therefore, this is the main weaknesses in the current system. An Intrusion Detection System (IDS) is a tracking system that detects atypical practices and alert when they are observed. On the basis of such alerts, a SOC analyst or an incident respond should act to research and resolve the problem while terminating the threat

c. GENERAL DESIGN



VII. METHODOLOGY

There are number of steps involved in performing a machine learning technique based on classification method using a data set. The basic steps in involved in performing a machine learning process are shown in figure



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

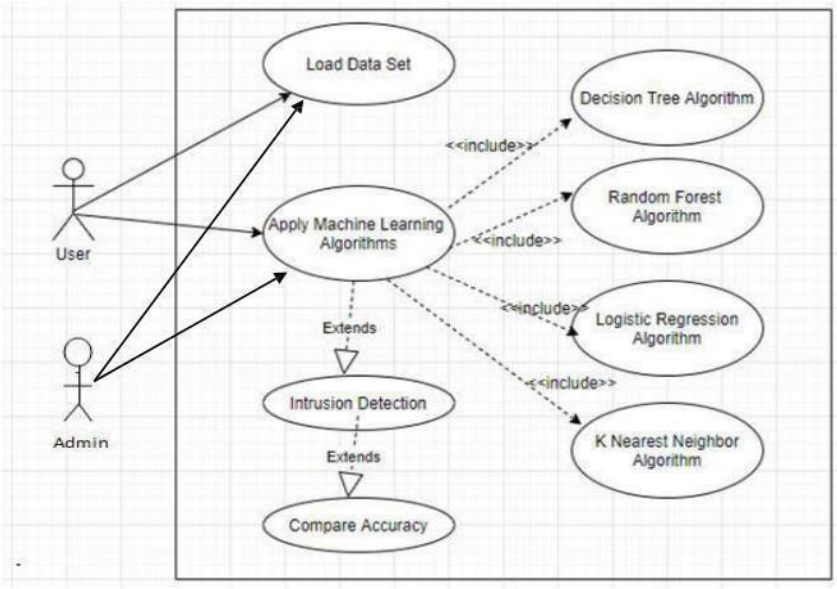
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134127

a. USE CASE DIAGRAM

Use Case Diagram captures the system's functionality and requirements by using actors and use cases. Use Cases model the services, tasks, function that asystem needs to perform. Use cases represent high-level functionalities and how a user will handle the system. Use-cases are the core concepts of Unified Modelling language modelling



NSL KDD Data set:

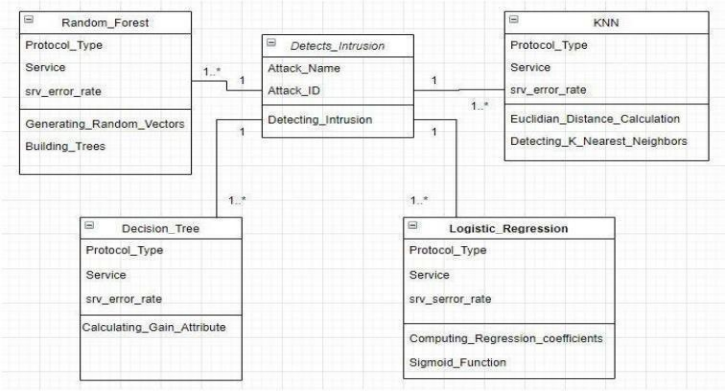
The NSL KDD data set is an improved version of the KDD Cup 99 data set, addressing some of its limitations. Itincludes various types of attacks and normal traffic, making it suitable for both binary and multi-class intrusion detection.

Support vector machine:

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. This is especially effective in scenarios where the data in the input feature space is not linearly separable. SVM works by finding the optimal hyper plane that best separates different classes in the feature space.Demonstrates some important concepts and features of support vector machines

b. CLASS DIAGRAM

Class diagram model class structure and contents using design elements such as classes, packages and objects. Class diagram describes 3 perspectives when designing a system Conceptual, Specification, Implementation. Classes are composed of three things: name, attributes and operations. Class diagrams also display relations such as containment, inheritance, associations etc. The association relationship is most common relationship in a class diagram. The association shows the relationship between instances of classes.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.134127

VIII. FUNCTIONAL REQUIREMENTS

Functional requirements for an intrusion detection system with machine learning involve specifying the capabilities and features that the system should possess to effectively detect and respond to network security threats. Here are some key functional requirements for such a system:

8.1 Hardware Requirements

Though our project is machine learning project which is success prediction of a Start up, doesn't require any hardware components but require some system configuration to run the project modules.

- RAM: 4 GB or 8 GB
- Space on Hard Disk: minimum 1

IX. DELIVERABLES

The deliverables for the proposed IDS with ML and comparison analysis project include a comprehensive set of outputs aimed at enhancing the understanding and application of advanced intrusion detection systems. Firstly, a detailed report outlining the project's scope, objectives, methodologies, and implementation strategies will be provided. This report will encompass a systematic review of ML algorithms, their selection criteria, and the rationale behind their Integration into the IDS framework. Additionally, the deliverables will include a well documented code base, ensuring transparency and Re producibility of the developed system. A set of pre processed and curated datasets, used for training and evaluating ML models, will be made available to facilitate further research and experimentation. The comparative analysis outcomes will be presented in a comprehensive format, detailing the performance metrics of each ML algorithm, thereby providing insights into their strengths and weaknesses. This information will be valuable for cyber security practitioners and researchers seeking to implement effective intrusion detection mechanisms. Furthermore, the deliverable will feature a user-friendly interface for system configuration and monitoring, enhancing the system's usability and practicality for security professionals. The project's success will be gauged through the effective integration of ML algorithms into the IDS, yielding improvements in intrusion detection accuracy and adaptability. Overall, these deliverable aim to contribute to the field of cyber security by offering both theoretical insights and practical tools for the development of resilient and adaptive intrusion detection systems

X. CONCLUSION

In conclusion, the application of machine learning in intrusion detection systems (IDS) represents a promising and evolving approach to enhance cyber security. The utilization of advanced algorithms enables the system to learn and adapt to emerging threats, offering a proactive defense mechanism against various cyber attacks. Machine learning- based IDS can effective detect anomalies and patterns in network traffic, providing a dynamic and intelligent defense against both known and unknown threats. However, it is essential to acknowledge the challenges associated with the implementation of machine learning in this context, such as the need for large and diverse datasets, potential false positives, and the requirement for continuous model updates to keep pace with evolving attack strategies. Moreover, a comprehensive comparison analysis of different machine learning techniques for intrusion detection is crucial for identifying the most suitable approach based on specific requirements and constraints. Ultimately, the integration of machine learning in intrusion detection systems holds great promise for fortifying cyber security postures, but ongoing research and refinement are necessary to address the ever-changing landscape of cyber threats

REFERENCES

[1]. Jie Wang, Jing Xu, Chengan Zhao, Yan Peng & Hongpeng Wang (2019) "An ensemble feature selection method for high-dimensional data based on sort aggregation", Systems Science & Control Engineering,
[2]. Yamada, A., Miyake, Y., Takemori, K., Studer, A., & Perrig, A. (2007, May). Intrusion detection For encrypted web accesses. In 21st International Conference on Advanced Information networking applicaation workshop
[3]. Goh, V. T., Zimmermann, J., & Looi, M. (2009, March). Toward intrusion detection for encrypted networks. In 2009 International Conference on Availability, Reliability and Security (pp. 540545). IEEE.
[4]. Venkata Ramani Varanasi, Shaik Razia,(2020), " A Comparative Evaluation of supervised and unsupervised algorithms for Intrusion Detection", International Journal of Advanced Trends in Computer Science and Engineering
[5]. Mc Gaughey, D., Semeniuk, T., Smith, R., and Knight, S. (2018, April). A systematic approach of feature selection for encrypted network traffic classification. In 2018 Annual IEEE International Systems Conference (SysCon) (pp. 1-8). IEEE.
[6]. Verma, V., & Kumar, R. (2014). A Unique approach to multimedia based dynamic symmetric key cryptography. International Journal of Computer Scienceand MobileComputing, 3(5), 1119-1128