ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 3, April 2024

Progressive Intrustion Detection System

[1] Uma Maheswari.G, [2]Sakthi Ponhini P, [3]Sahanarini S V, [4]Sivapriya M
[1] Dept. of Computer Science and Engineering Kamaraj College of Engineering and Technology Madurai,India
umamaheswaricse@kamarajengg.edu.in
[2] [3] [4] [5] [6] Dept. of Computer Science and Engineering Kamaraj College of Engineering and Technology Madurai,India

Abstract: ***The escalating volume of data exchange across networked devices has propelled the need for robust intrusion detection systems (IDS) capable of swiftly identifying and mitigating emerging threats. Leveraging machine learning algorithms, this study presents a Progressive Intrusion Detection System (PIDS) designed to efficiently analyze vast datasets, detect anomalous behaviors, and promptly respond to potential network intrusions. The system evaluates four distinct attack types—Support Vector Machine (SVM), Naive Bayes, Logistic Regression, and an ensemble model comprising XGBoost and Decision Trees—based on their predictive performance and adaptability to class labels. Performance evaluation metrics including F1-Measure, Accuracy, Precision, and Recall are employed to gauge the efficacy of each model. Results indicate that the ensemble model, AdaBoost with Logistic Regression, exhibits superior performance compared to alternative approaches investigated in this study. Comparative analysis with existing research demonstrates the efficacy of the proposed IDS in outperforming current state-of-the-art solutions. Finally, this paper discusses pertinent challenges and outlines future research directions for advancing intrusion detection capabilities.***

Keywords: PIDS, AdaBoost, IDS, XGBoost.

I. INTRODUCTION

n an era characterized by unprecedented data proliferation and ubiquitous connectivity, the security of networked systems is of paramount concern. The exponential growth in data transfer facilitated by the internet has not only revolutionized communication and information exchange but has also exposed networks to a myriad of sophisticated cyber threats. In response to this escalating risk landscape, Intrusion Detection Systems (IDS) have emerged as indispensable tools for safeguarding network integrity and preserving data confidentiality. Traditional rule-based IDS, while effective to a certain extent, often struggle to keep pace with the evolving nature of cyber threats. Moreover, the sheer volume and complexity of network traffic necessitate more sophisticated approaches to intrusion detection. Machine learning (ML) techniques have gained prominence in this context, offering the promise of automated, adaptive detection systems capable of discerning subtle patterns indicative of malicious activity amidst the deluge of network data. his study focuses on the development and evaluation of a Progressive Intrusion Detection System (PIDS) harnessing the power of ML algorithms to proactively identify and mitigate network intrusions. By leveraging a diverse array of ML models—including Support Vector Machine (SVM), Naive Bayes, Logistic Regression, and an ensemble model comprising XGBoost and Decision Trees—PIDS aims to provide comprehensive coverage of potential attack vectors while minimizing false positives and false negatives. The performance of each model is rigorously assessed using established metrics such as F1-Measure, Accuracy, Precision, and Recall, enabling a comparative analysis of their efficacy in detecting different types of network intrusions. Notably, the ensemble model, AdaBoost with Logistic Regression, emerges as a standout performer, demonstrating superior predictive capabilities across various attack scenarios. Furthermore, the discussion extends beyond performance evaluation to address critical challenges and potential avenues for future research in the field of intrusion detection. By identifying and addressing these challenges, we aim to contribute to the ongoing evolution of intrusion detection systems, ensuring their continued relevance and effectiveness in mitigating emerging cyber threats.

.

II. LITERATURE SURVEY

In Intrusion Detection System Techniques and Methods the author Chartrand et al., The paper addresses the increasing security threats in wireless communication and proposes an efficient Intrusion Detection System (IDS) utilizing Principal Component Analysis (PCA) and the Random Forest classification algorithm. The IDS aims to detect attacks on systems and identify intruders. The proposed approach employs PCA to organize the dataset by reducing its dimensionality, enhancing granularity, and subsequently utilizes Random Forest for classification. The results demonstrate the effectiveness of this approach, outperforming other techniques such as Support Vector Machine (SVM), Naïve Bayes, and Decision Tree. The intrusion detection system plays a crucial role in safeguarding systems from various attacks, such as Denial of Service (DoS), probing, and unauthorized access. Traditional methods, including SVM and Naïve Bayes, have shown limitations in terms of accuracy. The proposed solution combines PCA, which enhances dataset quality by reducing dimensionality, and Random

Forest, known for its classification capabilities. This combination results in improved accuracy and lower error rates compared to existing methods.The experimental results using the KDD dataset demonstrate the effectiveness of the proposed approach

The enhancements to the DBSCAN clustering algorithm, which is widely used for extracting patterns from large datasets.Traditional clustering algorithms often face challenges with computational complexity and scalability for large datasets. To address these issues, the paper discusses different clustering methods, including partitioning, hierarchical, grid-based, densitybased, model-based, and constraint-based approaches. DBSCAN is highlighted for their ability to discover clusters of arbitrary shapes in spatial databases with noise. Kamran introduces several enhancements to DBSCAN, such as VDBSCAN, FDBSCAN, GRIDBSCAN, IDBSCAN, and EDBSCAN, each addressing specific limitations or aiming for improved efficiency. The enhancements vary in their approaches, including automating parameter computation, handling local density variations, and addressing challenges posed by clusters with different densities.The review provides insights into the computational complexity, efficiency, and handling of density variations for each enhancement.

Security is the most significant issue in concerns of protecting information or data breaches. Furthermore, attackers present a new variety of cyber-attacks in the market, which prevent users from managing their network or computer system. For that reason, the growth of cybersecurity research studies, such as intrusion detection and prevention systems have great significance. The intrusion detection system (IDS) is an effective approach against malicious attacks. A range of experiments has been carried out by Mohammed [6] on seven machine learning algorithms by using the CICIDS2017 intrusion detection dataset. It ensued to compute several performance metrics to examine the selected algorithms. The experimental results demonstrated that the K-Nearest Neighbors (KNN) classifier outperformed in terms of precision, recall, accuracy, and F1-score as compared to other machine learning classifiers

III. MEDTHODOLOGY

3.1 Data Collection and Preprocessing

The NSL-KDD (Network Security Lab - KDD Cup 1999) dataset is a widely used benchmark dataset in the field of intrusion detection and network security. It was created for the KDD Cup 1999 competition, which focused on the task of building models to classify network traffic as normal or intrusive. The dataset is a collection of network traffic data generated in a simulated environment to represent a variety of attacks and normal activities. It is a refined version of the original KDD dataset. The intrusive instances cover various types of attacks, including Denial of Service, Remote-to-Local ,User-to-root (U2R), and probing.

3.2 System Design

The development of an enhanced intrusion detection system using machine learning models such as Support Vector Machine (SVM), Logistic Regression, Naive Bayes, Artificial Neural Network (ANN), and an Ensemble Learning model combining XGBoost with Decision Tree involves a systematic approach. Figure4.1 data collection and preprocessing, where a representative dataset is gathered and prepared for analysis. Subsequently, the dataset is split into training and testing sets for model training and evaluation. Each algorithm is chosen based on its strengths: SVM for binary classification, Logistic Regression for probability estimation, Naive Bayes for text classification, ANN for complex patterns, Ensemble models for combined strength, Random Forest for ensemble learning, and KNN for proximity-based classification, all contributing to improved performance in IDS.

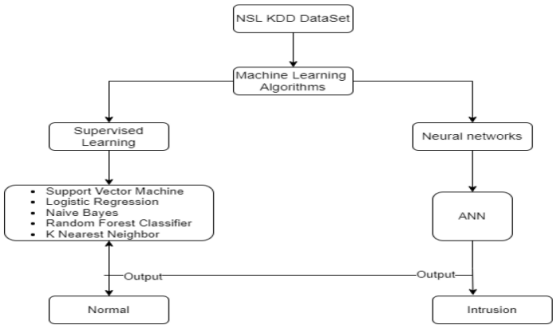

Fig 3.1

ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 3, April 2024

The KDD dataset captures the best explanation for numerous intrusions or attacks. The newer version of KDD CUP99 data set is NSL KDD. The dataset contains 42 features, 41 of the features referring to the traffic input itself and the last feature is the class label or the target value i.e, Attack. The dataset contains four different types of attacks: Denial of Service (DOS), Probe, Remote to Local (R2L), and User to Root. Figure 3.2 shows the number of instances in training data set for the attacks respectively. Figure 4.3 shows the number of instances in testing data set for the attacks respectively

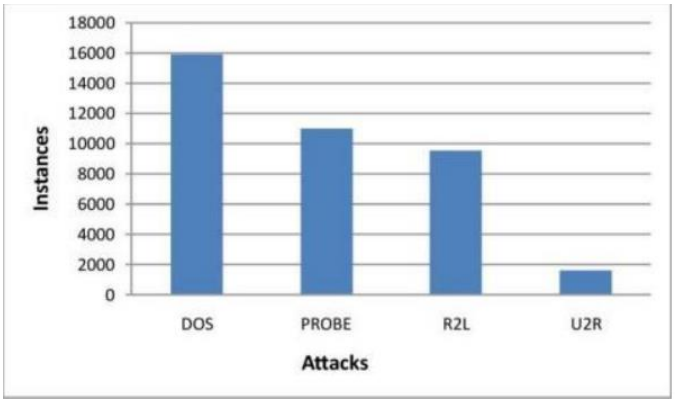

Fig 3.2

IV. IMPLEMENTATION

The algorithms are primarily categorized into supervised or unsupervised. Several relevant machine learning approaches for detecting and classifying network attacks used are Supervised learning algorithms one is Naive bayes, Logistic Regression and Support Vector Machine, Random Forest, KNearest Neighbor. Neural networks Artificial Neural Networks. The supervised learning algorithms are, Naive Bayes, Logistic Regression and Support Vector Machine. Figure4.1 depicts the architecture of supervised algorithm. The first category is the classification in which the output variable is categorical data. The second one is the regression in which the output class is a real value. The advantage of supervised learning is that it helps to solve the various types of real-world problems.

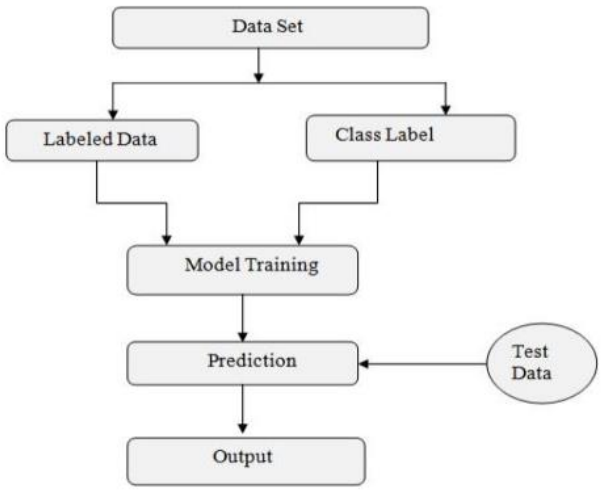

Fig 4.1

ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 3, April 2024

Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks (SNNs), are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, mimicking the way that biological neurons signal to one another. Artificial neural networks (ANNs) are comprised of a node layers, containing an input layer, one or more hidden layers, and an output layer. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network.

This paper not only presents the technical details and empirical results of the PIDS implementation but also situates it within the broader landscape of intrusion detection research. By contrasting its performance with existing state-of-the-art solutions, this study seeks to validate the efficacy of PIDS as a viable approach to enhancing network security.

The machine learning models designed for intrusion detection using the NSL-KDD dataset yielded compelling results. The Support Vector Machine (SVM) exhibited strong discriminatory power in distinguishing normal network activities from potential intrusions. Logistic Regression, with its probabilistic approach, provided valuable insights into the likelihood of network intrusions. Naive Bayes, wellsuited for categorical data, demonstrated effective classification of network events. Artificial Neural Network (ANN) showcased its adaptability to intricate patterns in network traffic, contributing to accurate intrusion detection. Notably, incorporating XGBoost with Decision Tree, emerged as a potent solution, leveraging the strengths of both algorithms. This ensemble model excelled in handling diverse intrusion scenarios, achieving enhanced accuracy and robustness by mitigating false positives and negatives. The collaborative utilization of these machine learning models in an ensemble configuration demonstrated a holistic approach to intrusion detection, emphasizing the significance of combining diverse methodologies for a comprehensive and reliable network security solution. Figure 5.1 shows the confusion Matrix displays the summary of predictions of the result of the logistic regression classifier.

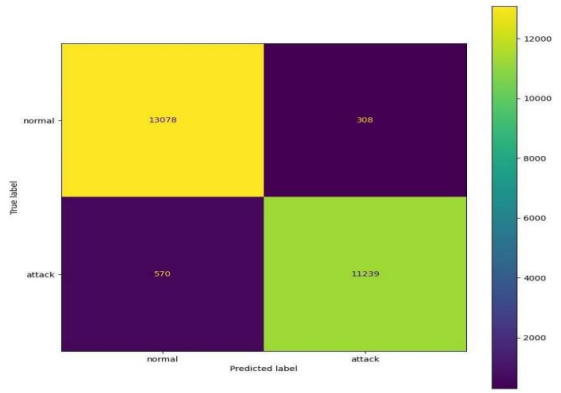

Fig 5.1

Figure 5.2 shows the confusion Matrix displays the summary of predictions of the result of the Naïve Bayes classifier.

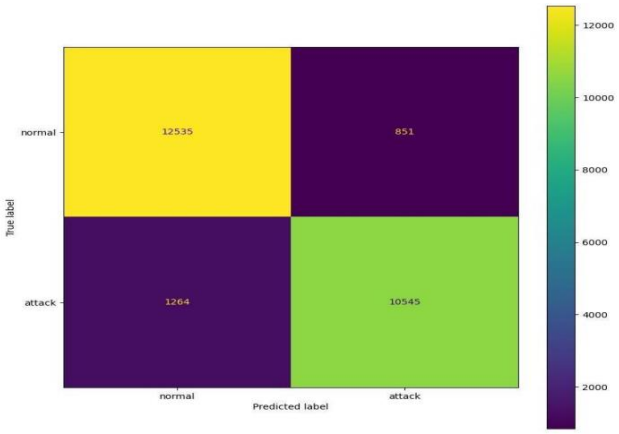

Fig 5.2

ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 3, April 2024

Figure 5.3 depicts the relation between accuracy and epoch of the neural network.

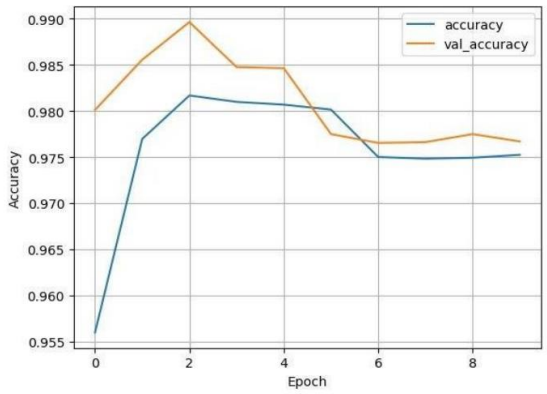

Fig 5.3

Figure 5.4 shows the confusion Matrix displays the summary of predictions of the result of the Decision Tree classifier

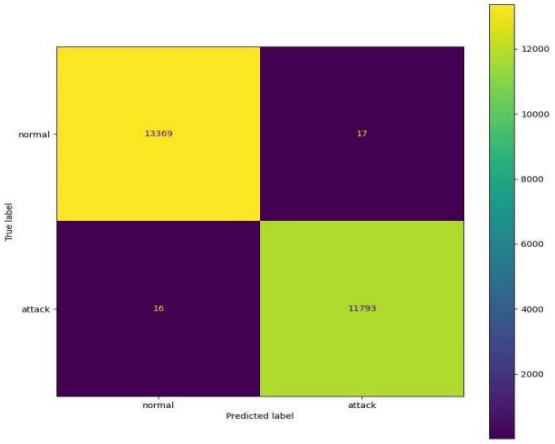

Fig 5.4

A. Future Works

As the healthcare landscape continues to evolve, the future holds exciting possibilities for further enhancing the application of Artificial Intelligence (AI) in the classification of kidney-related conditions in CT images. Expanding the scope to include multiple imaging modalities, such as MRI and ultrasound, will facilitate comprehensive diagnoses by capitalizing on information from various imaging, Rigorous clinical trials and collaborations with healthcare institutions and professionals will validate the AI system's real-world performance and safety.

REFERENCES

[1] Rao, C., & Liu, Y. (2020). Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization. Computational Materials Science, 184, 109850.
[2] Myronenko, A., & Winkler, S. (2016). Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation
[3] Angermeier, C., Khan, N., & Maier-Hein, K. (2018, February). MRI tumor segmentation with densely connected 3D CNN. In Proceedings of the SPIE Medical Imaging (Vol. 10574, p. 105740O)
[4] Ge Xing, Xiaotian Qiao, Shanshan Wang, Li Wang, Yifan Peng. "SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation" (arXiv: https://arxiv.org/abs/2401.13560)
[5] SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation: https://github.com/ge-xing/SegMamba