A Hybrid Classifier based on Machine Learning Algorithms for Intrusion Detection in Cloud Computing

Munish Saran[1], Rajan Kumar Yadav [2], Pranjal Maurya[3], Sangeeta Devi[4], Upendra Nath Tripathi[5]

[1,2,3,4,5] Department of Computer Science, Deen Dayal Upadhyaya Gorakhpur University, Gorakhpur, Uttar Pradesh, India

Abstract:

Cloud computing systems are susceptible to various types of attacks that can compromise their security. The system experiences a degradation in performance with respect to its integrity, confidentiality, and security. The introduction of an Intrusion Detection System has been proposed as a solution to address the aforementioned issues. This system is designed to identify various types of attacks, including but not limited to Normal Probe, DDoS, and R2L attacks. The present study proposes a hybrid classification approach for the development of an intrusion detection system. The hybrid clustering approach for intrusion detection has been proposed with the aim of effectively categorizing various types of attacks. This study introduces two innovative algorithms, namely the attack clustering algorithm and the machine learning and deep learning classifier, which are referred to as DBN (Deep Belief Networks), Random Forest, and Naïve Bayes. All of the models under consideration are deemed acceptable, and the research methodology employed yields an effective intrusion detection and prevention system for mitigating security breaches and intruder attacks

Keywords: Intrusion Detection System, Cloud Computing, Security Attack, Deep Learning, Machine Learning.

DOI: 10.24297/j.cims.2023.7.8

1. Introduction

Cloud Computing is a modern technology that enables the efficient utilization of computing infrastructures, as well as a commercial model for the sale of computing facilities and services. However, complicated and distributed frameworks present an appealing objective for potential intruders. Cloud computing has the potential to enhance productivity and decrease expenses; however, it also presents various new security hazards. The utilization of Intrusion Detection Systems (IDS) has been prevalent in the identification of malevolent activities in both network communication and servers [1] [2]. The term "computer network system" refers to a framework

designed to gather data pertaining to various crucial aspects, and subsequently scrutinize said data to identify any instances of non-compliance with network security policies or indications of a security breach [3].

Currently, numerous entities are transitioning their computational resources to the Cloud. This enhances the accessibility of users to the computer processing. Nonetheless, it introduces novel security risks and circumstances concerning the assurance and dependability of the system [4]. Cloud computing can be described as a complex system of interconnected networks that operate over the internet. As a result of this interconnectedness, there is an increased risk of security breaches due to the sophisticated nature of intruder attacks. Due to its distributed nature, cloud computing presents a higher likelihood of intrusion [5]. The examination of diverse methods for detecting and preventing intrusions is crucial. Various Intrusion Detection System (IDS) methodologies are employed to mitigate malevolent assaults in conventional networks.

2. Intrusion Detection System

An Intrusion Detection System (IDS) refers to a tool or software programme that is designed to oversee system or network operations with the aim of detecting malicious behavior or violations of policies. The system generates reports that are intended for management purposes [6].
The typical operational procedure of an intrusion detection system may be broken down into four distinct techniques, which are as follows:

1.1.1 Collection of Data

The process entails the acquisition of network traffic through specialised software, thereby facilitating the retrieval of pertinent details regarding the traffic, such as packet types, devices, and protocol specifications.

1.1.2 Selection of Features

As a result of the substantial amount of network traffic, the collected data is extensive. Consequently, a feature element is produced that solely comprises crucial information. The Internet Protocol (IP) header data, comprising of pertinent details such as both origin and destination IP addresses, packet type, and other identifying information, can be effectively employed in the detection of network-based intrusions.

1.1.3 Analyzing the Features

In this step, the data that was collected is analyzed so that it can be determined whether or not the information is abnormal. In general, the study effort makes use of a variety of techniques for the purpose of identifying intrusions [7].

1.1.4 System Response

The IDS will notify the system manager that an attack has occurred and will provide information on the scope of the attack. IDS also take part in the control of the harms by either blocking access to the network port or terminating the processes. Because of this, the manager of the system is able to quickly and simply determine whether or not the system has been compromised by the attacks.

3. Security in Cloud Computing

Cloud computing makes use of three main delivery models, each of which facilitates the delivery of a certain category of service to the end user. Software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS) are the three delivery models that supply the user with infrastructure resources, application platforms, and software respectively [8]. Additionally, each of these service models imposes a unique level of security requirement on the underlying cloud environment. The Infrastructure as a Service (IaaS) model serves as the basis upon which the Platform as a Service (PaaS) model and the Software as a Service (SaaS) model are constructed. In the same way that skills are passed down from generation to generation, so too are the problems and threats to information security. There are substantial compromises to be made between the integrated features of each model, the relative levels of complexity and flexibility, and the level of security [9]. The customers take on a greater share of the responsibility for the implementation and management of the security capacities if the cloud service provider is solely accountable for the security at the base of the security architecture. Attacks on the clouds are given-

Denial of Service:

The distributed structure of cloud environments renders them susceptible to Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks, which are inevitable. Denial of Service (DoS) attacks in cloud computing occur when attackers inundate virtual machines with a high volume of access requests, thereby impeding the availability of resources to legitimate users. Unauthorized individuals can gain access to a system by exploiting vulnerabilities and obtaining

the login and password details of an authorized user. This can result in unrestricted access to the entire system. The present attack is in contravention of the integrity of the cloud environment.

Attacks on Virtual Machine:

One type of cyber-attack compromises the hypervisor, allowing the attacker full control of the virtual computer. Using zero-day exploits, attackers can quickly gain access to the virtual computer. An attacker using a virtual machine has the potential to get access to the entire network. All potential forms of assault that can be carried out from within the system are classified as "inside attacks," which are carried out by this type of person in order to damage or distort information about the system [10].

Malware Injection Attack:

The adversary can infiltrate the system with malicious code, insert bogus commands, or swap metadata in such a way that it runs normally while still giving the impression that it is legitimate. Now a hacker has the ability to transform the data, alter the operation of the system, and cause the genuine user to have to wait for the completion of a task that was not requested by the actual user [11].

4. Problem Statement:

Because of its adaptability and scalability, cloud computing is quickly becoming the platform of choice for businesses of all sizes. However, due to the open and decentralized nature of its design, there is significant reason for worry about privacy and security. Cloud computing offers users a high level of service and safeguards their confidential information at the same time. Intruders may readily target the private data of end users, as well as the capacity for storage, bandwidth, and computational ability of the cloud network since cloud computing systems are spread. The network's integrity, confidentiality, and availability are all compromised when an intruder attack such as a probe, R2L, U2R, flooding, or DDoS occurs. The network's permission, authentication, and security are all put at risk as a result of these assaults. Traditional methods of network security, such as firewalls, are effective at preventing many assaults from the outside, but they are not designed to detect sophisticated attacks from inside the network or from the outside. In order to solve these issues, an intrusion detection system has been implemented to monitor the network and identify any malicious activity. It has the ability to give extra protection methods in a cloud environment that is dispersed.

Intrusion detection systems are capable of identifying malicious activity within a system through keeping track of changes in internet traffic, end user actions, and system setting. In the event of detecting a malicious activity within the system, an alert message is dispatched to either an individual or a monitoring console, prompting the initiation of preventative measures against such attacks. Our study involves the implementation of a hybrid classification-based approach for intrusion detection and prevention within a cloud environment. The present study implemented hybrid algorithms, namely packet scrutinization and classification model referred to as NK-RNN algorithm. The latter is a fusion of the nearest k-means algorithm and Recurrent Neural Network, and was employed to identify instances of intrusion in the system.

5. Literature Review:

Recent years have witnessed a significant amount of scholarly investigation into Intrusion Detection Systems (IDS) utilising machine learning techniques. Sandip Sonawane (2015) employed machine learning algorithm for the purpose of identifying irregularities in the KDD dataset.

Erxue Min et al. (2018) proposed the utilisation of decision trees and random forest (RF) in the context of anomaly detection. The empirical investigation of two distinct datasets provides evidence for the efficacy of the proposed model in generating dependable outcomes. In comparison to alternative models, this particular approach confers numerous advantages in relation to Accuracy (ACC), False Alarm Rate (FAR) and Precision.

In their study, Chiba z. et al. (2019) utilised artificial neural networks based on deep learning techniques to develop intrusion detection system models for anomaly detection on a given dataset. The outcomes of the model exhibited a high level of accuracy in detecting intrusions with a small false alarm rate, thus indicating its superiority over other contemporary methods.
T. Arvind (2020) proposes the integration of various machine learning techniques into a unified hybrid approach. The findings demonstrate that the hybrid techniques exhibit superior performance compared to the standalone models. Individuals with a desire to acquire further knowledge regarding machine learning techniques for Intrusion Detection Systems (IDS). In contemporary times, scholars have put forth various machine learning methodologies for Intrusion Detection Systems (IDS) that tackle one or more of the aforementioned challenges, thereby rendering machine learning algorithms advantageous for IDS.

In their study, Devarakonda A. et al. (2022) utilised a classification strategy consisting of four layers to discern four discrete types of attacks within the KDD dataset. The specified method yielded relatively low values for both the overall error and the misclassification error. The authors suggested reducing the number of characteristics in the original dataset as a means of enhancing accuracy and decreasing complexity, thereby simplifying the method. The authors omitted any mention of labelling errors that occurred, whereby a particular type of attack was erroneously classified as a different type of attack.

Attou, H. et al. (2023) employed the KDD method for the purpose of attack type classification, and the results indicated a minimal misclassification error. Notwithstanding, these models may encounter difficulties in contemporary multi-cloud environments that exhibit dynamic and closely associated attacks. The KDDcup99 dataset's relevance may be limited due to its age, potentially hindering its ability to accurately reflect contemporary network usage. Support Vector Machine (SVM) is a data mining technique utilised for the purpose of extracting anticipated data. The KDDCUP '99 IDS database was utilised by the author to perform classification using neural networks. The training data set yielded an accuracy rate of 90%, while a 10-fold cross validation experiment was conducted on the test set, resulting in an accuracy rate of 80%.

The scholarly discourse pertaining to Intrusion Detection Systems (IDS) encompasses a variety of classifiers and clustering methodologies, which encompasses unsupervised cluster analysis techniques. The imprecise categorization of certain assaults resulted in a decrease in the overall efficacy of attack identification methodologies. It is recommended that Deep Learning models based on machine learning be utilised to address the aforementioned concerns, with the aim of enhancing the accuracy of intrusion detection systems.

6. Methodology

This section presents the experimental validation of the proposed study. This part of the report displays graphical representations of performance measures and comparative analyses. This result shows that the proposed technique is efficient in detecting intrusions. When evaluating the suggested work experimentally, it outperforms the state-of-the-art with respect to a number of key performance measures. By using a range of performance metrics, such as F-score, false alarm rate, precision, and accuracy, to evaluate the recommended method.

3.1 Dataset Collection

NSL-KDD cup 99 is used as a dataset for Intrusion Detection System. There are a total of 591 records in the dataset, and each of those records has a total of 41 characteristics. Based on the IDs of the entities in the collection, packets were collected from those entities [18]. In order to better understand the security flaws in the system, this dataset has been prepared. There are four types of assaults included in the dataset: DoD, R2L, U2R, and probing. There is a risk of inaccurate predictions due to inadequate, noisy, or duplicate data in the dataset. That is why pre-processing is done to provide data before analysing the dataset using equation (1):

$$Y^{ij} = \frac{Y^{ij} - A^{i}}{(max - min)\ feature\ of\ j} \quad (1)$$

Where Y^{ij} represents the j[th] column attribute value in i[th] row of dataset and A is showing the mean value.

3.2 Clustering using Hybrid K-Means:

Following pre-processing, the data that arrives is subsequently subjected to the clustering process. The computation of intruded data from a vast quantity of data is a highly intricate and lengthy process, resulting in suboptimal outcomes due to the sheer magnitude of the data [19]. In order to mitigate this problem, the data has been subjected to clustering in the present section. The clustering process is characterized by a reduced time consumption and a decreased occurrence of false alarms. The algorithm in question is capable of detecting anomalies of both low and high frequency within the system.

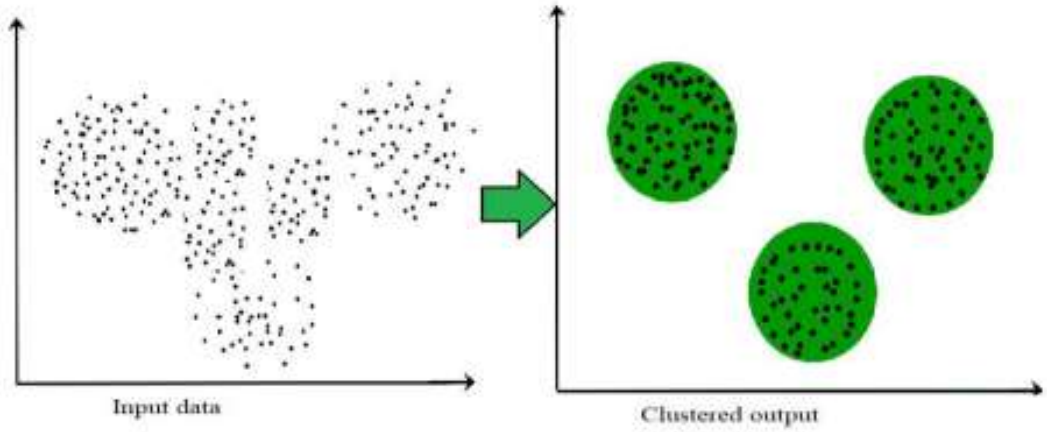

Figure 1: Clustering Output

The paper introduces a novel metric to address issues arising from the presence of mixed periodic and non-periodic attributes in clustering problems. Additionally, a Hybrid K-means algorithm is proposed as a viable solution for identifying clusters in such datasets. By utilizing

the distance measurement for periodic features and the Euclidean measurement for non-periodic features, it becomes feasible to establish a metric for quandaries that solely involve certain attributes that are periodic. Let be a subset of the set I= {1, 2,...,A}, representing the indexes of attributes that exhibit periodicity. Here, M denotes the total amount of features. The set D can be defined as the collection of periods of the periodic attributes, where D is represented as {Dj|j ∈ I}. The collection of periodic attributes, (0, D), will be represented by the variables x and y. The metric that denotes distances within this space can be expressed as –

$$D(x,y) = \sqrt{\sum_j^A D_j\,(x^2 - y^2)^2} \quad (2)$$

3.3. Working Methodology

The following is a description of how the functioning mechanism of work operates:

Input: Take NSL-KDD cup 99 data for process

Output: Retrieve the output score and forecasting for indications of both interacting and non-interacting nature.

Step-1: Collects NSL-KDD cup 99 and saves them in data objects.

Step-2: Convert data into numerical form.

Step-3: Initialize pre-processing and normalization process to deal with inadequate, noisy, or duplicate data.

Step-4: Hybrid K-means algorithm is proposed as a viable solution for identifying clusters in such datasets.

Step-5: Apply classification using DBN (Deep Belief Networks), Random Forest and Naïve Bayes.

Step -6: Perform classification and prediction.

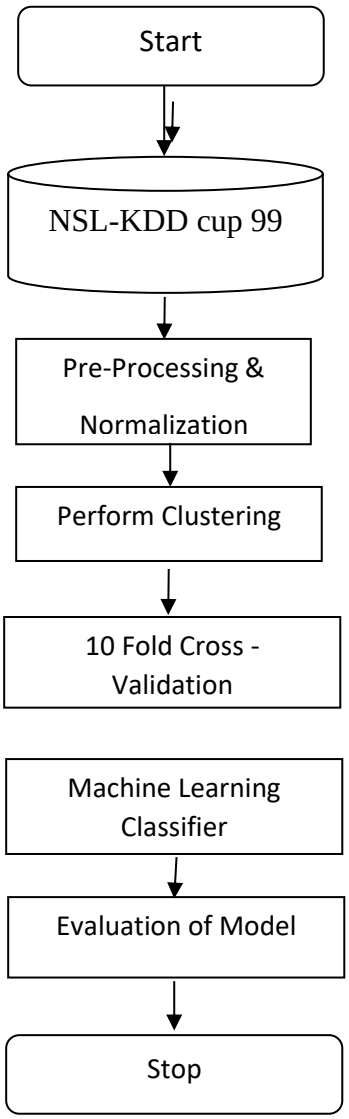

Figure 2: Proposed work flow diagram

7. Result and Discussion

The present study aims to identify various types of attacks, including Normal, DDoS, probe and R2L through the use of a hybrid classifier. This approach differs from previous research, which relied on Fuzzy-ANN methodology for detecting such attacks. The implementation of a Machine Learning classifier facilitates the classification of packets into two categories, namely normal and intruder packets. This classification process serves to mitigate the occurrence of false alarms within the system. This section presents an analysis of the experimental results obtained from the proposed Intrusion Detection System (IDS). The NSL KDD Cup 99 dataset was employed for experimental analysis. The evaluation of the suggested approach is assessed based on various metrics accuracy, precision, F-Score and False Alarm Rate. The outcomes of the intrusion detection system proposed in this study are illustrated in Figures 3-6 through simulation results.

The effectiveness of the suggested methodology is shown by comparing its performance to several alternative approaches. Clustering and optimum prediction are the two main considerations. Deep Belief Network (DBN), Random Forest, and Naive Bayes are used for the best prediction, while K-Means is used for clustering. The performance criteria listed below are used to evaluate the system under consideration. The performance indicators for the clustered dataset that were suggested in the preceding section are shown in Table 1.

Table 1: Comparative analysis of proposed system for (a) DBN (b) Random Forest and (c) Naïve Bayes

Model	Parameter (%)	Normal (%)	DDoS (%)	Probe (%)	R2L (%)
DBN	Accuracy	92.8	93.9	95.3	95.6
	Precision	95.2	95.8	97.6	96.3
	F-Score	90.1	91.6	92.1	93.5
	False Alarm Rate	0	0	0	2
Random Forest	Accuracy	88.0	89.3	91.2	92.8
	Precision	91.8	92.7	93.9	93.7
	F-Score	89.9	90.2	91.5	92.8
	False Alarm Rate	0	0	27	32
Naive Bayes	Accuracy	90.2	91.9	93.6	94.1
	Precision	93.3	94.1	95.6	95.3
	F-Score	90.0	90.6	91.3	92.1
	False Alarm Rate	0	0	5	9

In figure 3, the findings of the experiment indicate the classifier proposed in this study yields effective outcomes for both low-frequency and high-frequency attacks, achieving the highest level of accuracy. The identification of various types of attacks within the cloud environment enables the implementation of effective security measures against potential intrusions.

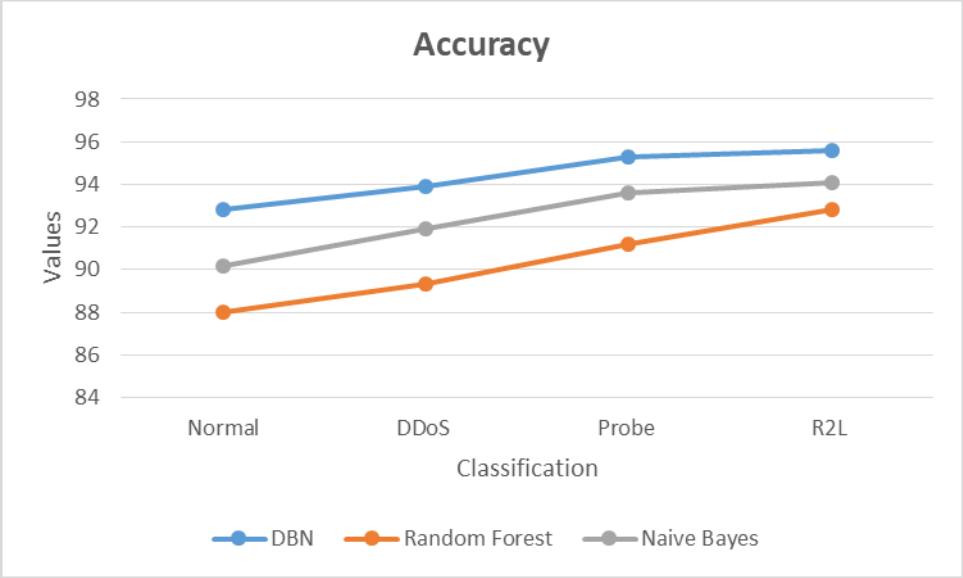

Figure 3: Graphical Representation of Accuracy

This paper presents a novel approach to designing an intrusion detection system by utilising deep learning techniques. Specifically, a high-precision intrusion detection system is proposed based on latent variables.

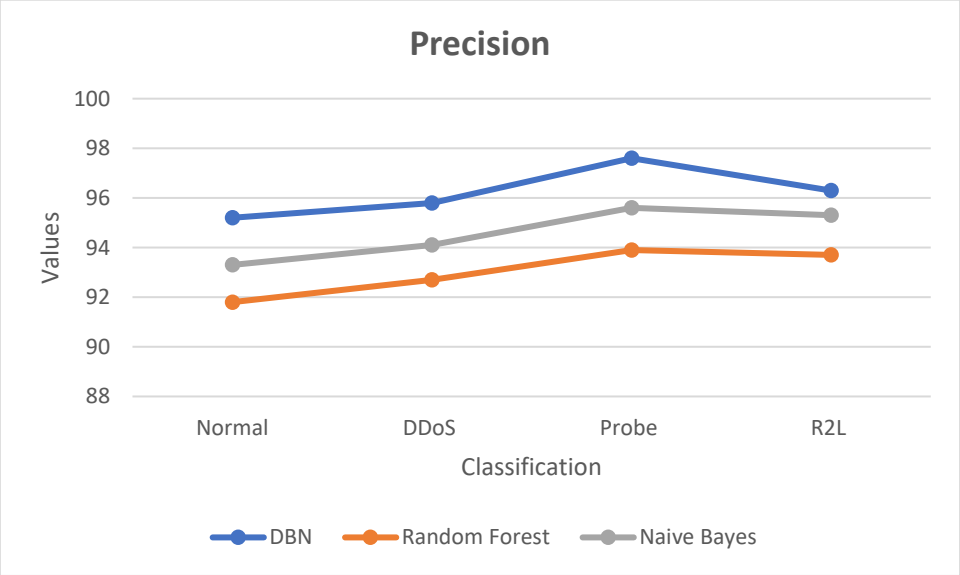

Figure 4: Graphical Representation of Accuracy

The comparison of F-Score determined by precision value between the prior research result and the proposed work result is illustrated in Figure 6. The graph illustrates that the proposed work exhibits a superior score in comparison to the prior research.

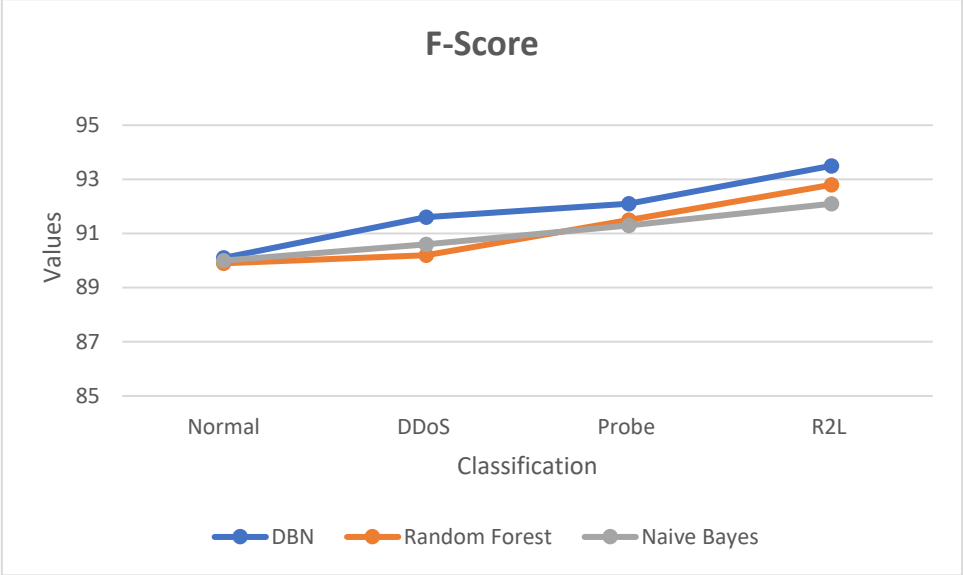

Figure 5: Graphical Representation of F-Score

This section presents an analysis of the false alarm rate in relation to various types of attacks, as well as the precise prediction of intruder packets within the context of cloud computing. A false negative event transpires when an intrusion detection system fails to detect a possible or authentic attack. Figure 6 is showing graphical representation of false alarm rate for various machine learning algorithm.

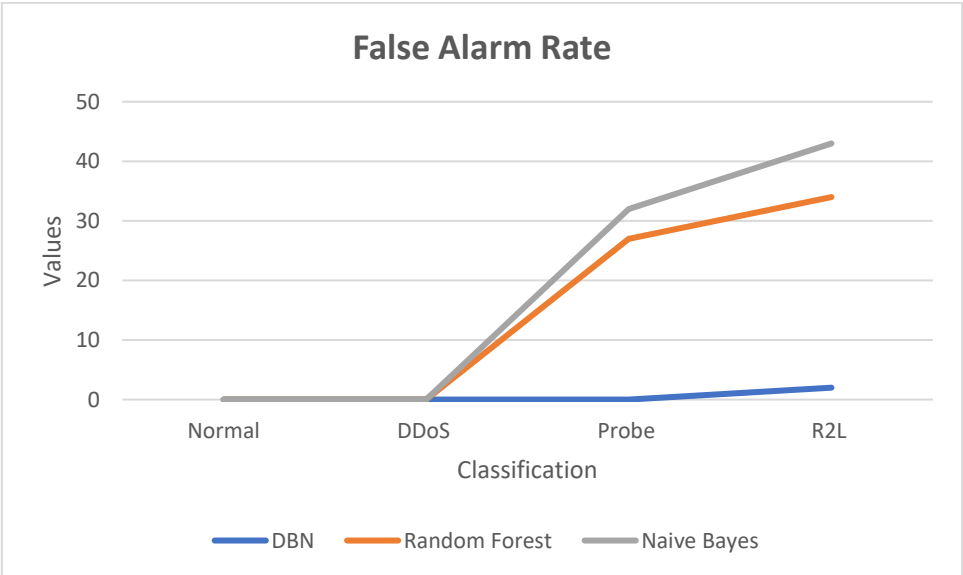

Figure 6: Graphical Representation of False Alarm Rate

8. Conclusion and Future Scope

The present study involved an analysis of the effectiveness of the hybrid algorithm based intrusion detection system that was proposed. The present study conducted a comparative analysis between the proposed anomaly detection systems. The proposed system's performance was evaluated by taking into account various performance metrics, including but not limited to Accuracy, Precision, F-Score and false alarm rate. The findings of the comparative analysis indicate that the system we have proposed exhibits greater efficacy in identifying intruder attacks within cloud systems

It was found the best result of the classifier is more than 95% for various machine learning and deep learning model. The observed disparity in accuracy between the classifier utilising all features and the aforementioned approach is negligible. The figures above depict the performance of the classifier. Due to the low rate of misclassification, this methodology has been deemed effective for the classification of intruder detection.

References

1. Balamurugan, E., Mehbodniya, A., Kariri, E., Yadav, K., Kumar, A., & Haq, M. A. (2022). Network optimization using defender system in cloud computing security based intrusion detection system withgame theory deep neural network (IDSGT-DNN). Pattern Recognition Letters, 156, 142-151.
2. Devi, B. T., Shitharth, S., & Jabbar, M. A. (2020, March). An Appraisal over Intrusion Detection systems in cloud computing security attacks. In 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 722-727). IEEE.
3. Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity, 2(1), 1-22.
4. Song Han, Miao Xie, Hsiao-hwa Chen, Yun Ling, Intrusion detection in cyber physical systems: techniques and challenges, Sys. J., IEEE 8 (2014) 1049–1059, https://doi.org/10.1109/JSYST.2013.2257594.
5. Santos, L., Rabadao, C., & Gonçalves, R. (2018, June). Intrusion detection systems in Internet of Things: A literature review. In 2018 13th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-7). IEEE.
6. Srivastava, D., Sharma, V., & Soni, D. (2019, April). Optimization of CSMA (Carrier Sense Multiple Access) over AODV, DSR & WRP routing protocol. In 2019 4th international conference on internet of things: Smart innovation and usages (IoT-SIU) (pp. 1-4). IEEE.

7. Srivastava, D., Soni, D., Sharma, V., Kumar, P., & Singh, A. K. (2022). An Artificial Intelligence Based Recommender System to analyze Drug Target Indication for Drug Repurposing using Linear Machine Learning Algorithm. Journal of Algebraic Statistics, 13(3), 790-797.
8. Soni, D., Srivastava, D., Bhatt, A., Aggarwal, A., Kumar, S., & Shah, M. A. (2022). An Empirical Client Cloud Environment to Secure Data Communication with Alert Protocol. Mathematical Problems in Engineering.
9. Kacha, L., & Zitouni, A. (2018). An overview on data security in cloud computing. Cybernetics Approaches in Intelligent Systems: Computational Methods in Systems and Software 2017, vol. 1, 250-261.
10. Namasudra, S. (2019). An improved attribute-based encryption technique towards the data security in cloud computing. Concurrency and Computation: Practice and Experience, 31(3), e4364.
11. Mishra, N., Sharma, T. K., Sharma, V., & Vimal, V. (2018). Secure framework for data security in cloud computing. In Soft Computing: Theories and Applications: Proceedings of SoCTA 2016, Volume 1 (pp. 61-71). Springer Singapore.
12. Sandip Sonawane, Rule based learning intrusion detection system using KDD and NSL KDD dataset, 04, Prestige International Journal of Management & IT - Sanchayan (2015) 135–145, https://doi.org/10.37922/PIJMIT.2015.V04i02.009.
13. Zouhair Chiba, Noreddine Abghour, Khalid Moussaid, Amina omri, Mohamed Rida, Intelligent approach to build A deep neural network based IDS for cloud environment using combination of machine learning algorithms, Computer Security 86 (2019), https://doi.org/10.1016/j.cose.2019.06.013.
14. Erxue Min, Jun Long, Qiang Liu, Jianjing Cui, Wei Chen, TR-IDS: anomaly-based intrusion detection through text-convolutional neural network and random forest, 2018, Secur. Commun. Network. (2018) 1–9, https://doi.org/10.1155/2018/4943509.
15. T. Arvind, A survey on building an effective intrusion detection system (IDS) using machine learning techniques, challenges and datasets, Int. J. Res. Appl. Sci. Eng. Technol. 8 (2020) 1473–1478, https://doi.org/10.22214/ijraset.2020.30598.
16. Davuluri, S. K., Srivastava, D., Aeri, M., Arora, M., Keshta, I., & Rivera, R. (2023, April). Support Vector Machine based Multi-Class Classification for Oriented Instance Selection. In 2023 International Conference on Inventive Computation Technologies (ICICT) (pp. 112-117). IEEE.

17. A Sharma, V., Kumar, L., & Srivastava, D. (2023). Machine Learning-Based Prediction of Users' Involvement on Social Media. In Advanced Applications of NLP and Deep Learning in Social Media Data (pp. 151-170). IGI Global.
18. Srivastava, D., Chui, K. T., Arya, V., Peñalvo, F. J., Kumar, P., & Singh, A. K. (2022). Analysis of Protein Structure for Drug Repurposing Using Computational Intelligence and ML Algorithm. International Journal of Software Science and Computational Intelligence (IJSSCI), 14(1), 1-11. http://doi.org/10.4018/IJSSCI.312562
19. Maheswaran, N., Bose, S., Logeswari, G., & Anitha, T. (2023). Hybrid Intrusion Detection System Using Machine Learning Algorithm. In Proceedings of Data Analytics and Management: ICDAM 2022 (pp. 333-346). Singapore: Springer Nature Singapore.