WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Machine Learning as a Tool for Intrusion Detection and Data Security (2024)

Dr. C Nandini[1], Mrs. Nethra H L[2], Aman Chaudhary[3], Aman Kumar[4], Ashutosh Ranjan[5], Ayush Kumar[6]

[1,2]Professor, Department of CSE, Dayananda Sagar Academy of Technology and Management
[3,4,5,6]Student, Department of CSE, Dayananda Sagar Academy of Technology and Management
DOI : https://doi.org/10.55248/gengpi.6.0125.0632

ABSTRACT—

Since the cyber threats are developing with the increase in the complexity and visibility, the protection and integrity of data have become vastly significant in the current day cybersecurity. Intrusion Detection Systems (IDS) are the systems that are necessary for protecting against intruder or unusual events in networks.

This paper aims the discussion on how best to apply AI to improve IDS especially in real-world scenarios. This is because instead of using traditional IDS, the proposed approach utilizes novel AI approaches such as machine learning and deep learning, which therefore enhance anomalous traffic detection reliability that does not produce high numbers of false alarms that would overwhelm the network. The paper aims at investigating the incorporation of AI feature with IDS for maintaining data authenticity and adequate security of the data across various-network domains.

In addition, this paper explains such functionalities that posit a failure for the actual world implementation including; scalability, adaptability to dynamic threats, and resource limitation. An evaluation of the experimental results shows that the IDS developed using AI yields a significant improvement in detecting advanced attacks while at the same time ensuring system stability. In this, the authors outline future research avenues for AI such as the use of compound models and improved techniques of real-time threat handling for strengthening cybersecurity architectures.

KEYWORDS—Artificial Intelligence, Intrusion Detection System, Cybersecurity, Real-World Deployment, Data integrity, Data security.

I. INTRODUCTION

The reliability and protection of data require and at times prefer a scientist developed within a framework of confidence in their voice and therefore, I must come

to terms with the clear definition of boundaries of that power.

A measure where one can determine if "it is the reliability and trustworthiness of a piece of data all through its life span determining its validity or non-validity" measure data integrity. Examples are the process of data integrity checks and validation.

The KPMG statistical data release, however, shows that the top senior executives, which comprised 2,190 individuals surveyed by the firm, were “skeptical about the use of AI to protect the integrity of data and data in general when employing AI by the organizations to make critical company decisions binding and dependent on data and its analytics.” 35% of employees say data analytics and collection work well for their organizations. Most (92%) are concerned a great deal about the potential for data and analytics to damage a company's reputation[5].

Only 35% said they have a lot of trust in the use of analytics and data by their companies from the feedback drawn from the 2190 senior executives in KPMG. A high level of anxiety was associated with the business continuity involving data, analytics, and AI, and approximately two- thirds of them are something between reluctant-against- to outright skeptical about their data and analytics.

Of course, before getting into the other practical applications of AI, one would have to be conversant with the basic principles that govern this science[6].

II. RELATED WORK

Abhishek Divekar et al (A. Divekar, 2018) used classification algorithms such as Naïve Bayes, SVM, Neural network, RF, KMeans, and Decision tree, and compared their performances to alternatives that arose much later than KDD'99. The study exhibited that UNSW-NB15 is a much better and shiny

alternative to KDD'99. The results confirmed that classifiers trained in terms of f1-Scores were far better than classifiers trained with KDD'99 and NSLKDD.

The authors of (Srivastava, 2018), made an attempt at the performance and effectiveness of NIDS. Two feature reduction methods, LDA and CCA, were used. Seven classifiers were applied with various measurement parameters: FPR, training time, accuracy, and ROC area. Algorithms used include random tree, naive Bayes, REPTree, RF, random committee, randomizable bagging, and filters. The combination of LDA and random tree on UNSW-NB15 was declared best.

In (2018), M. Belouch provides information regarding experimental studies, classifying commonly used classifiers-- NB, SVM, Decision tree, and RF- on the Apache Spark big data environment. They produced timing measures for detection, building, and prediction purposes in a network intrusion detection system. In their performance evaluation, the UNSW- NB15 data set was forwarded for their work, and they contend that the RF technique was the best regarding specificity, accuracy, and sensitivity, along with execution time compared to any of the other tested algorithms.

Authors in (Slay N.M., 2015) propose a hybrid feature reduction approach consisting of CP containing attribute value features followed by an ARM. First, the dataset was partitioned equally, reducing processing time; output from the CP technique was then fed into the ARM algorithm to reduce the number of features in their model. In EM clustering, logistic regression, and the naive Bayes algorithms were used for comparison and evaluation of the results for NIDS. With respect to this, no additional details were included in the document. Their claim is that the model could improve the accuracy obtained for the false alarm rate and, at the same time, the time needed for processing. The datasets were NSL- KDD and UNSW-NB15.

Many researchers in IDS have proposed fusion with machine learning. Some are dedicated to unknown attacks [1], while others are oriented towards known attacks [2]. Recently, deep-learning-based techniques have also been exploited in IDSs and demonstrated successful detection of intrusive behaviors [3][4].

A deep neural network (DNN) is used for network IDS in [12], which later provides an explanation-based framework to enhance a DL model's transparency. The authors implemented and validated several XAI techniques, including SHAP, the contrastive explanation method, LIME, and ProtoDash, using the NSL-KDD dataset. Another SHAP-based framework was similarly proposed to improve the transparency of IDS in [11]. The framework's performance was also tested on the NSL- KDD dataset.

An adversarial attack is handled through the integration of XAI with an ML-based IDS in [10]: a random forest classifier for detecting network intrusions is first built, and then the SHAP approach is employed to explain and interpret the outputs of the random-forest-based model. The performance of this scheme is evaluated using the CICIDS dataset.

An LRP method was developed in [13] to evaluate the relevance of input features and then provide online and offline feedback to users as to which features were most influential in the predictions made by the IDS.

For addressing the incorrect classifications made by ML/ DL-based IDS, an explanation approach is advances in [14]. This approach analyzes what needs to be changed to rectify the misclassification of a given dataset sample. These alterations further help to recognize the most salient features behind the erroneous classification. The proposed approach was evaluated using the NSLKDD dataset.

Even the above works have deployed XAI to explain ML/DL-based IDS. However, some of them limited themselves only to traditional ML algorithms, less complex and easier to interpret than DL algorithms [10]. Moreover, most designed a general XAI framework for whatever the targeted ML/DL-based IDS [9, 13]. This may, in reality, not be the best solution, since every ML/DL-based IDS had its impressive and specific input characteristics and performance; an XAI framework should take into account such characteristics as input if it is to explain the decisions of an ML/DL-based IDS.

III. EXISTING SYSTEM

Artificial intelligence represents an advanced computer simulation of human cognition and awareness. Within the field of AI application, two most dominant aspects are seen: logical reasoning and natural language comprehension[7].

However, there are some other important issues concerning computer network security, like data risks within network security nodes, network intrusion, and network security management. Based on this issue, this article will review research regarding computer network security technologies based on AI[8].

A. Working Model of Exisitng System

The objective of this investigate is to accomplish the inquire about objective by providing arrangements to data security problems discussed in different universal themes and to illuminate the problems more successfully and effectively than considering and solving them in the past. In this think about, the unique KDD Cup 99 dataset is supplanted with the Arrange Security Laboratory Database Information Disclosure (NSL-KDD) dataset, which provides a superior understanding of the get to behavior[15].

The strategies utilized in this think about incorporate six strategies: data collection, preparatory information, highlight assessment, feature selection, plan and confirmation. The malware assault gather is called "0=Normal, 1=DoS, 2=R2L, 3=U2R, 4=Packet Sniffing". Each assault dataset goes through all different classification calculations some time recently creating the result. This data is based on distinctive strategies like "Choice

Tree (Decision tree), KNearest Neighbor (KNN), Bolster Vector Machine (SVM), Convolutional Neural Arrange (CNN) classifiers". This is also done to assess whether the truth is way better compared to testing and preparing all the strategies to classify the material.

The author's commitment highlights to distinguish the fourth type of cyber-attack namedas bundle sniffing assault discussed in the technique and result area by implementing Convolutional Neural Arrange (CNN) with the exactness of 98.4% and early discovery of malware assault which is enough for securing database framework for an organization or person[16]

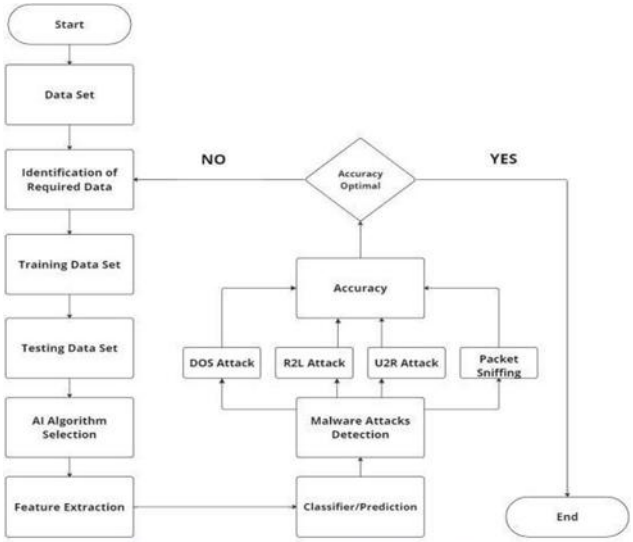


Figure 1. Proposed System for Malware Attacks detection

B. Results of the Attacks

The application of the machine learning methods is covered in this section. It also describes the widely used assessment metrics for machine learning techniques for intrusion detection systems. Table 1 displays the general confusion matrix, which is used to show how well our supervised learning algorithms work.

Table 1. Confusion Matrix

Actual Class	**Predicted Class**	
	Attack	Normal
Attack	True_Positive	False_Negative
Normal	False_Positive	True_Negative

B.1 DoS Attack

Tables 2 and 3 show the outcome with the application of classifier using Decision tree, SVM, KNN, CNN algorithm on our DoS Attack Data Set. Table 3 shows extra metrics which have been looked by classificators.Whereas Table 2 depicts the confusion matrix for DoS attacks that were classified based on the four aforementioned classifiers algorithms. Metrics such as F-measure, accuracy, recall, and precision.

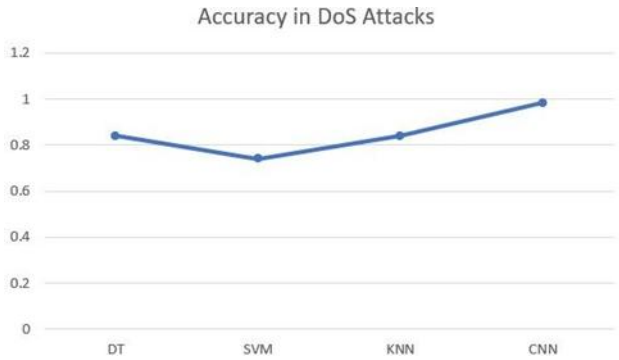
Table 2. Confusion matrix for four classifiers on DoS attack

DoS Attack		**Predicted Attacks**	**Classifier**
Actual Attacks	0	1	DT
0	9602	109	
1	2625	485	
	0	1	SVM
0	9677	34	
1	3578	3882	
	0	1	KNN
0	9653	58	
1	2645	4815	
	0	1	CNN

Table 3. Evaluation metrics for four classifiers on DoS attack

Metrics	Precision	Recall	F-Measure	Support	Classifiers
0	0.79	0.99	0.88	9711	DT
1	0.98	0.65	0.78	7460	
Accuracy	-	-	0.84	17171	
Macro avg	0.88	0.82	0.83	17171	
Weighted avg	0.87	0.84	0.83	17171	
0	0.73	1	0.84	9711	SVM
1	0.99	0.52	0.68	7460	
Accuracy	-	-	0.79	17171	
Macro avg	0.86	0.76	0.76	17171	
Weighted avg	0.84	0.79	0.77	17171	
0	0.78	0.99	0.88	9711	KNN
1	0.99	0.65	0.78	7460	
Accuracy	-	-	0.84	17171	
Macro avg	0.89	0.82	0.83	17171	
Weighted avg	0.87	0.84	0.83	17171	
0	0.77	0.99	0.87	9711	CNN
1	0.99	0.64	0.77	7460	
Accuracy	-	-	0.984	17171	
Macro avg	0.88	0.83	0.89	17171	
Weighted avg	0.86	0.85	0.89	17171	

Out of 12,821 applications, 9,602 were classified as adversaries. Out of 3,110 models included in the evaluation, the vendor is estimated to have only 485 models. Table 2 shows this. This study shows that decentralized decision trees provide better performance to network operations. The accuracy of this method is 0.84 but could be better. CNN and reconnaissance mission will prove it. After some training using decisions on DoS_attack_dataset, SVM classifier uses metric to compute classification algorithm accuracy and the accuracy is 0.79. A basic technique known as KNN classifier acquires patterns and classifies them based on degree of similarity measures like distance function. In contrast to decision trees, product confusion matrices enable prevention of network behavior and strong DoS attacks against predictions. The precision of this classification is similar in case of decision tree where it is 0.84. The CNN classifier gives the DoS events with an accuracy of 0.984. Figure 2 is a graph comparing four CNN models. A full measure of DoS is shown in Figure 2. This may be due to the development of deep neural networks in the CNN model, unlike other machine learning algorithms that only send the dataset to the category once.



B.2 R2L Attack

This is when a remote user tries to send a packet to in the R2L attack, the attacker tries to access unauthorized. As shown in table 4, confusion matrix displays the effectiveness of the classification of these R2L by decision tree attacks in the context of our study can be understood to mean effectiveness of the decision tree model in detection and blocking access points by giving bad predictions.

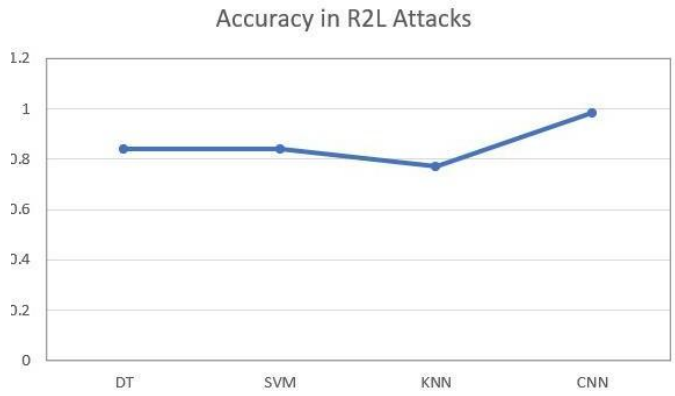
Table 4. Confusion matrix for four classifiers on R2L attack

R2L Attack		Predicted Attacks	Classifier
Actual Attacks	0	3	DT
0	9649	62	
1	2560	325	
	0		SVM
0	9711	0	
3	2885	0	
	0	3	KNN
0	9710	0	
3	2885	1	
	0	1	CNN
0	9635	72	
3	3091	445	

Table 5 Test Results for the Decision Tree and SVM methods is 84%. Since this is not a high accuracy in We will use a number of classifications for network security: For this reason, models with the highest accuracy were chosen, as these accuracy levels are not enough for cybersecurity purposes. KNN and further classification is performed to determine the accuracy. On the accuracy aspect of the classification. The algorithm crosses 77%, still, it cannot be ensured that overall security of the network. So far, we have demonstrated that machine learning methods can not attain R2L even when good predictions. The prediction of the CNN classifier is 0.984. This fact is helpful for predictive models of cybersecurity. Figure 3 it shows the graphical forms of different distributions. Traditionally, applied in defining R2L profiles.

Table 5. Evaluation metrics for four classifiers on R2L attack

Metrics	Precision	Recall	F-Measure	Support	Classifiers
0	0.79	0.99	0.88	9711	DT
1	0.84	0.11	0.2	7460	
Accuracy	-	-	0.84	17171	
Macro avg	0.82	0.55	0.54	17171	
Weighted avg	0.8	0.79	0.72	17171	
0	0.79	0.99	0.88	9711	SVM
1	0.84	0.11	0.2	7460	
Accuracy	-	-	0.84	17171	
Macro avg	0.82	0.55	0.54	17171	
Weighted avg	0.8	0.79	0.72	17171	
0	0.77	1	0.87	9711	KNN
1	0	0	0	7460	
Accuracy	-	-	0.77	17171	
Macro avg	0.39	0.5	0.44	17171	
Weighted avg	0.59	0.77	0.67	17171	
0	0.78	0.96	0.89	9711	CNN
1	0.98	0.66	0.78	7460	
Accuracy	-	-	0.984	17171	
Macro avg	0.88	0.83	0.88	17171	



B.3 U2R Attack

A User-to-root (U2R) attack happens when a local user who is access to the primary network but not to the backend. A network is provided with access to a root cause. Table 6 illustrates actual value, Incorrect value, Incorrect value, and incorrect values for the four tests in this experiment. More indicators are found in Table 7. This classification test has an output precision of 98.4%. This makes it very sensitive and suitable for the prediction of cyberattacks. Figure 4 Accuracy rate of 91% for the U2R attack with an SVM classifier. That is to say, the SVM therefore, the classifier may predict

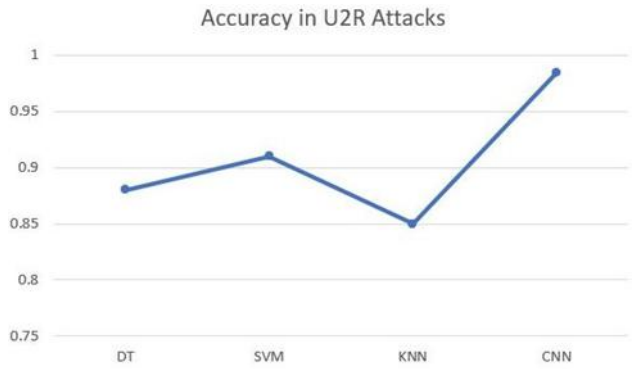
future U2R attacks of the network. This KNN product can also only give an accuracy of 85%, and further improvement is required for accessibilities in terms of searches even better. Estimated delivery rate of CNN classification results are 98.4%.

Table 6. Confusion matrix for four classifiers on U2R attacks

U2R Attack		Predicted Attacks	Classifier
Actual Attacks	0	5	DT
0	9706	5	
4	52	15	
	0	4	SVM
0	9711	0	
4	67	0	
	0	4	KNN
0	9709	2	
4	60	7	
	0	6	CNN
0	9788	8	
4	71	11	

Table 7. Evaluation metrics for four classifiers on U2R attack

Metrics	Precision	Recall	F-Measure	Support	Classifiers
0	0.99	1	1	9711	DT
1	0.75	0.22	0.34	7460	
Accuracy	-	-	0.88	17171	
Macro avg	0.87	0.61	0.67	17171	
Weighted avg	0.99	0.99	0.99	17171	
0	0.99	1	1	9711	SVM
1	0	0	0	7460	
Accuracy	-	-	0.91	17171	
Macro avg	0.5	0.5	0.5	17171	
Weighted avg	0.9	0.99	0.99	17171	



B.4 Packet Sniffing Attack

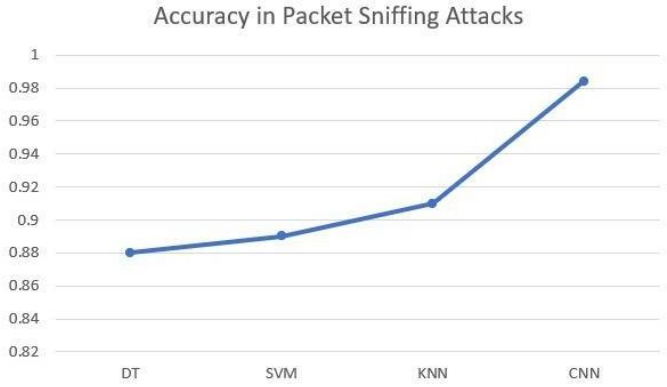
These cyber-attacks capture and examine the network packets to gathering of sensitive information. Table 8 aggregates the findings from confusion matrix after processing the data according to CNN SVM, KNN, and Classification with Decision Trees. It results in classification accuracy in determining the attacks. The results indicate the classification accuracy of Decision tree, SVM, KNN and according to CNN, 88%, 89%, 91%, and 98.4% the highest respective. CNN is therefore the correct classifier that can identify the attacks. Table 9 demonstrates the correctness of the classification algorithm which resulted in 98.4%.

Table 8. Confusion matrix for four classifiers on Packet Sniffing attacks

Packet Sniffing Attack		Predicted Attacks	Classifier
Actual Attacks	0	7	DT
0	9808	7	
4	65	14	
	0	6	SVM
0	9813	3	
4	77	3	
	0	5	KNN
0	9829	3	
4	80	8	
	0	12	CNN
0	9889	17	
4	81	23	

Table 9. Evaluation metrics for four classifiers on Packet Sniffing attacks

Metrics	Precision	Recall	F-Measure	Support	Classifiers
0	0.99	1	1	9711	DT
1	0.85	0.22	0.34	7460	
Accuracy	-	-	0.88	17171	
Macro avg	0.89	0.61	0.67	17171	
Weighted avg	0.99	0.99	0.99	17171	
0	0.98	1	1	9711	SVM
1	0	0	0	7460	
Accuracy	-	-	0.89	17171	
Macro avg	0.6	0.5	0.5	17171	
Weighted avg	0.8	0.99	0.99	17171	
0	0.99	1	1	9711	KNN
1	0.77	0.1	0.18	7460	
Accuracy	-	-	0.91	17171	
Macro avg	0.90	0.55	0.59	17171	
Weighted avg	0.98	0.99	0.99	17171	
0	0.99	1	1	9711	CNN
1	0.97	0.1	0.18	7460	
Accuracy	-	-	0.984	17171	
Macro avg	0.9	0.66	0.69	17171	
Weighted avg	0.93	0.99	0.99	17171	



Many tools have been integrated together to make a cluster. The classifier above it is supposed to grade it to make sure that it detects the attack. Packet sniffing attacks are used for all distributions. All the output by the this makes cluster as high as 98.4%. From the results shown in Table 9 and Figure 5, and can see very clearly that the CNN classifier produces the attack and the most likely forecast of the attack was measurement, the CNN classifier outputs results very close to there are several machine learning classifications or combinations: methods. 3.5. Early Detection The first major issue of this research is results of Decision tree, SVM to measure precision and time required at the time of search KNN and CNN would be addressed. It is demonstrated that CNN. Thus, classifier can detect network attacks earlier and better short detection time as compared to other algorithms, as shown in Figure 6.

C. Comparison of Algorithms

Showcase inquire about Science investigate Generation management, all data assembled can be connected in other utilizations, among others.Perhaps the most imperative run the show of machine learning is classification calculations[18]. They are utilized to classify unlabeled data classify into a few categories. There are the calculations that were utilized in the work:

Support Vector Machine (SVM): Compared to the other end algorithms, SVM is one of the most solid classifications Algorithms in machine learning since they are quick and straightforward prediction strategy. It produces a hyperplane which separates the classification all the information into the doled out classes of data points that bolster the vectors in a source of information.

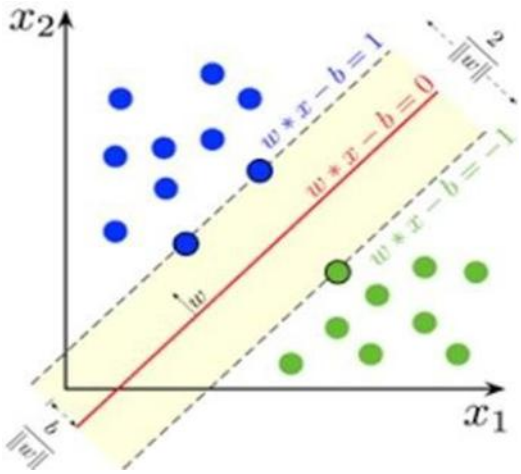


Figure 7. SVM

K-Nearest Neighbor (KNN): another trusted classifier such as autoencoders, and Isolation Forests for example algorithm used in classifying data into classes. One of its are used to scan network traffic for abnormal traffic promising feature is that it can be used for both classification patterns. For example, Random Forests, Gradient ends regression analysis.

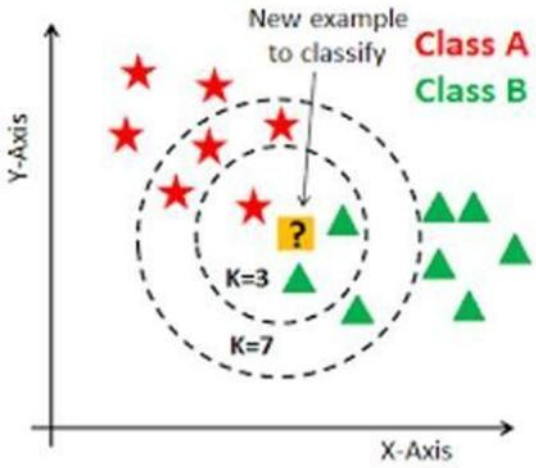


Figure 8. NSL-KDD details

Naïve Bayes (NB): They are able to predict the probability that whether a model given fits to a particular it is a Bayesian formulation. It is based on the hypothesis that, for example of a given class, the attribute value is independent of the values of the attributes. This theory It is called Class Conditional Independence.

$$P(H/X) = P(X/H).P(H)/P(X) \quad (1)$$

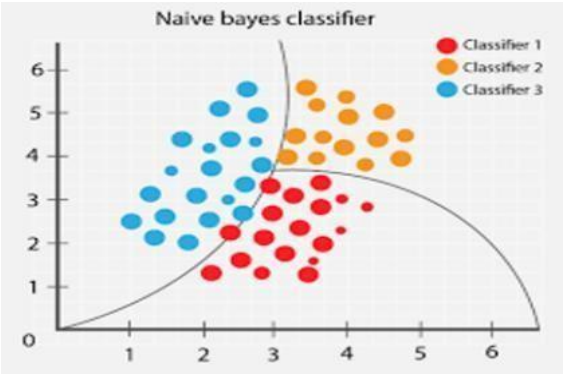


Figure 9. Naïve Bayes

IV. PROPOSED METHODOLOGY

The methodology that is to be used for the development of the multi model machine learning network intrusion detection system (NIDS) is of hybrid nature where in both supervised models and models based on anomaly detection are utilized to improve the detection accuracy. To begin, relevant network increased and scaled traffic data is sourced from temporal sources, preprocessed, and feature vectors are created using tools such as Zeek or Wireshark. The network anomaly detection models such as autoencoders, and Isolation Forests for example are used to scan network traffic for abnormal traffic patterns. For example, Random Forests, Gradient Boosted Trees, and other supervised models are used to identify the various attack types using datasets of labeled information such as CICIDS2017 or NSL-KDD. An ensemble method constructs these models using stacking or weighted voting. Each of these models will then use an ensemble method to enable them to vote, thus using weighted voting or stacking. The device is set up in a real-world scenario using frameworks such as TensorFlow Serving or Apache Kafka for constant observation and the reporting system sends notifications for the processed activities that are flagged. Updating and re-training is done continuously ensuring that the system is always ready for new and evolving threats.

V. OBJECTIVE OF THE PROPOSED SYSTEM

concerningperformance scalability, and preprocessing capabilities, leveraging Spark's distributed processing ability

A. **ENHANCED THREAT DETECTION:** THE PURPOSE IS TO to handle large amounts adequately on network data. Four DETECT USUAL AND NEW OR ZERO DAY ATTACKS WITH THE types of attacks are was detected in this research such as DoS HELP OF ENHANCED MACHINE LEARNING (ML) AND DEEP attack, R2L attack, U2R. Classification - four Classifiers learning (DL) tools. Determine a lot of attack types such as anomaly, DoS, malware and insiders.

Decision tree Attack and Packet Sniffing Attack SVM, KNN and CNN. The CNN must be right. Its detection accuracy rate is 98.4%, and it can detect within 10 seconds. It took longer

B. **Improved Accuracy:** REDUCE FALSE POSITIVE AND time than other machine learning classifiers to give an FALSE-NEGATIVE RATES WITH THE HELP OF THE AI-BASED environment of better database security policies[19]. MODELS THAT HAD BEEN TRAINED WITH A RANGE OF BIG Improved security structure and recognition of malware

AND BALANCED DATA SETS. IMPROVE THE PERFORMANCES attacks. Based on state-of-the-art machine learning algorithms.

OF IDS BY DETECTING RELATIVELY NEGLIGIBLE In this search, the first database was covered by three

VARIATIONS OF TRAFFIC FLOW CHARACTERISTICS.

C. **REAL-TIME INTRUSION DETECTION:** FACILITATE

D. algorithms SVM NB and KNN with neighborhood of 3.

Thus, KNN is discarded immediately after obtaining the below result:. If done, then the other two operate using the second

PROMPT AND SCRIPT-BASED REACTION TO DEVIATIONS IN database algorithms. SVM has performed very well, regardless of

THE NETWORK TO REDUCE HARM.CAPTURES AND what the database-size or type of attacks- included was this

EXAMINES NETWORK TRAFFIC FLOW IN REAL-TIME WITH contents of model, to be optimized in subsequent works

MINIMAL DELAY. according to processing time and also we'll work on its implementation from behind a firewall and test in real time along

E. **DATA INTEGRITY AND SECURITY:** PROTECT NETWORK with the continuously improving AI technology, its application in DATA, DEPRIVING THE INTRUDER OR THE UNAUTHORIZED Technology Security Risk Identification is becoming PERSON OF ACCESS TO IT OR INTERFERING WITH IT. increasingly common[20].

SECURES INFORMATION OR DATA AGAINST CYBER CRIMINALS OR THREATS

Using the technology of AI it facilitates the collection and analysis of large amounts of data it improves the quality and accuracy of risk identification.The application of AI

F. **COST-EFFECTIVENESS:** LOWER OPERATIONAL technology in technology security risk it may make risk

EXPENSES THROUGH THE APPLICATION OF AI AIDED IDS identification more efficient and dependable identification,

INTO PRESENT SECURITY SYSTEMS.DECREASE THE DEMAND reduction in labor-intensive processes, and this further makes

FOR THE CONFIGURATION AND SUPERVISION OF NUMEROUS technology security risk identification even intelligently.

MANUAL RULES.

Results And Discussions

This study aims to improve database intrusion detection systems (DIDS) based on Machine Learning Methods. The effectiveness of various methods with data compression network connectivity analysis and assessment shall be done to improve database intrusion detection systems (DIDS). In a comparative test, the algorithm and simulator of Apache Spark were evaluated to improve data link connectivity, connectivity error handling, and security intercepting. This this is designed to manage updates in networks by controlling them through sending fast ones questions. This paper covers the practice and evaluation of a multi-layered secure relational database an AI-based management system which access search system. Its content is merged with a safe deductivedatabase management system. It encompasses machine learning techniques that are important for efficiency, optimization, and preventing through database intrusion detection by Spark system (DIDS))[17].

It has been shown in this paper that application machine learning approaches to optimize database intrusions. DIDS or Detection Systems: Installation and Evaluation of a secure deductive database management system deduced reasoning was incorporated with a multilevel secure relational database management system architectural issues raised and a sample implementation. Machine learning methods for conducting IDS appears to work pretty well.

However, the algorithm of random forest still is imperfect, though data in the provided paper can be free from a extremely large applicability domain, and would embrace then using a better algorithm in subsequent research[21].

Acknowledgment

All of us wish to extend our warmest appreciation to all who assisted in the completion of this research on AI-based Intrusion Detection Systems.

Finally, above all, thanks go to our advisors and mentors for Useful guidance and constructive criticism and constant supporting their investigating process. Their quality is significantly contributed by experience and knowledge of this book.

Thanks to my institution that sponsored the research work with access to resources and equipments necessary for it, as well as the funding for our work.

This work is for the greater research community aspiring to seek greater opportunities in cyber security through the integration of artificial intelligence.

References

[1] R. Jain and H. Shah. “An anomaly detection in smart cities modeled as wireless sensor network”. In International Conference on Signal and Information Processing (IConSIP), pages 1–5, Oct 2016.E.

[2] C. Ioannou, V. Vassiliou, and C. Sergiou. “An intrusion detection system for wireless sensor networks”. In 24th International Conference on Telecommunications (ICT), , pages 1–5, May 2017.

[3] Ahmad Javaid, Quamar Niyaz, Weiqing Sun, and Mansoor Alam. “A deep learning approach for network intrusion detection system”. In Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (Formerly BIONETICS), pages 21–26, 2016.

[4] *C. Yin, Y. Zhu, J. Fei, and X. He. A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access, 5:21954–21961, 2017.*

[5] Safa Otoum, Burak Kantarci and Hussein Mouftah “A Comparative Study of AI-based Intrusion Detection Techniques in Critical Infrastructures”. In ACM Transactions on Internet Technology (TOIT), Volume 21, Issue 4 Article No.: 81,

[6] Pages 1 - 22.

[7] Rachid Tahri, Youssef Balouki, Abdessamad Jarrar, and Abdellatif Lasbahani “Intrusion Detection System Using machine learning Algorithms”. In ITM Web of Conferences 46, 2022.

[8] Rafeeq Ahmad, Humayun Salahuddin, Attique Ur Rehman, Abdul Rehman, Muhammad Umar Shafiq, M Asif Tahir, and Muhammad Sohail Afzal “Enhancing Database Security through AI-Based Intrusion Detection System”. In Journal of Computing & Biomedical Informatics Volume 07 Issue 02, 2024.12.8.

[9] BO-Xiang Wang, Jiann-Liang Chen and Chiao-Lin Yu “An AI- Powered Network Threat Detection System”. In IEEE Access, 2022.

[10] Zakaria Abou El Houda, Bouziane Brik, and Sidi-Mohammed Senouci “A Novel IoT-Based Explainable Deep Learning Framework for Intrusion Detection System”. In IEEE Xplore, 2022.

[11] S. Mane and D. Rao, “Explaining Network Intrusion Detection System Using Explainable AI Framework,”2021https://www.researchgate.net/publication/350061199_ExplainiNetwork_Intrusion_Detection_System_Using_Explainable_AI_F ramework.

[12] M. Wang et al., "An Explainable Machine Learning Framework for Intrusion Detection Systems," IEEE Access, vol. 8, 2020, pp. 73,127–41.

[13] S. Wali and I. Khan, "Explainable AI and Random Forest Based Reliable Intrusion Detection System Detection System," Dec. 2021. DOI:10.36227/ techrxiv.17169080.v1.

[14] K. Amarasinghe and M. Manic, "Improving User Trust on Deep Neural Networks Based Intrusion Detection Systems," Proc. IECON 2018 — 44th Annual Conf. IEEE Industrial Electronics Society, 2018, pp. 3262–68.

[15] D. L. Marino et al., "An Adversarial Approach for Explainable Ai in Intrusion Detection Systems," 2018. DOI: 10.1109/IECON.2018.8591457.

[16] Mohammed Mahmoud "THE RISKS AND VULNERABILITIES OF ARTIFICIAL INTELLIGENCE USAGE IN

[17] INFORMATION SECURITY". In International Conference on Computational Science and Computational Intelligence (CSCI),2023.

[18] Nandini, C., and Shiva Sumanth Reddy. "Detection of Communicable and NonCommunicable Disease Using Lenet- Bi-Lstm Model in Pathology Images." International Journal of System Assurance Engineering and Management, springer India- 2022, doi:10.1007/s13198-022-01702-5. (Q3 journal).

[19] Kumar, P. R., Meenakshi, S., Shalini, S., Devi, S. R., & Boopathi, S. (2023). Soil Quality Prediction in Context Learning Approaches Using Deep Learning and Blockchain for Smart Agriculture. In R. Kumar, A. Abdul Hamid, & D.

[20] Binti Ya'akub (Eds.), Effective AI, Blockchain, and E- Governance Applications for Knowledge Discovery and Management (pp. 1-26). IGI Global Scientific Publishing.

[21] Shantakumar Patil, Nagaraj M Lutimath, D Jogish, Premjyoti, Bhargav S Patil, "Prediction of Heart Disease Using Hybrid Naïve Bayes Technique", IEEE 22nd International Symposium on Communications and Information Technologies (ISCIT), Sydney, Australia, 16th -18th Oct 2023, pp. 257-261.

[22] N. Kumar, P. Nandihal, M. R. B, P. K. Pareek, N. T and S. S. R, "A Novel Machine Learning-Based Artificial Voice Box," 2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), Bangalore, India, 2022, pp. 1-7, doi: 10.1109/ICATIECE56365.2022.10046967.

[23] Decentralized Malware Attacks Detection using Blockchain S. Sheela, S. Shalini, D. Harsha, V.T. Chandrashekar, Ayush Goyal ITM Web Conf. 53 03002 (2023)DOI:10.1051/itmconf/20235303002.