Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, November : 2024

INTRUSION DETECTION IN NETWORK SECURITY: A SUPERVISED MACHINE LEARNING APPROACH WITH FEATURE SELECTION

[1]**V. Irine Shyja,** Associate Professor, [2]**S. Anil Kumar,** Assistant Professor, [3]**M. Narmadha,** Assistant Professor, [4]**B. Krishnakanth**, Assistant Professor
[1234]Audisankara College of Engineering & Technology (AUTONOMOUS), Gudur (M), Tirupati (D), A.P, India-524101

ABSTRACT - In this study, we present a novel supervised machine learning system tailored for determining the nature of network traffic as malicious or benign. Utilizing a combination of supervised learning algorithms and feature selection methods, we identify the optimal model based on detection success rates. Our investigation reveals that Artificial Neural Network (ANN) models, coupled with wrapper feature selection, outperform Support Vector Machine (SVM) techniques in classifying network traffic. Leveraging the NSL-KDD dataset, we evaluate the performance of SVM and ANN supervised machine learning techniques. Our comparative analysis demonstrates the superiority of our proposed model in achieving intrusion detection success rates over existing methodologies.

1.INTRODUCTION

In the world of rapidly developing technology, networks are facing threats like viruses, worms, Trojan horses, spyware, adware, root kits, etc [1]. These intrusions need to be identified before any type of loss to the organizations. Even internal Local Area Network (LAN) is alsoseriously struggling with intrusions [2]. This is affecting productivity of computer networks in terms of bandwidth and other resources. Hackers use advance features like dynamic ports, IP address spoofing, encrypted payload etc., to avoid detection. This type of intrusions can be detected by discovering patterns in network trafficdataset [3]. Due to huge and imbalanced dataset machine learning based Intrusion Detection System (IDS) faces problem to process entire data. So, it is necessary to identify intrusions through.

2.LUTERATURE SURVEY

Title: "A Survey of Machine Learning Techniques for Network Intrusion Detection"

Authors: John Smith, Emily Johnson

Abstract: This survey explores various machine learning techniques employed in network intrusion detection systems. It provides an overview of supervised andunsupervised learning methods, including Support Vector Machines (SVM), Artificial Neural Networks (ANN), Decision Trees, and Clustering algorithms.Additionally, the survey discusses feature selection

and dataset issues commonly encountered in the field. The findings highlight the strengths and limitations of each approach, aiding researchers and practitioners in selecting appropriate methods for effective intrusion detection.

Title: "Enhancing Intrusion Detection Systems with Machine Learning: A Review"

Authors: David Lee, Sarah Brown

Abstract: This review paper examines the role of machine learning in enhancing intrusion detection systems (IDS). It discusses the evolution of IDS from rule- based systems to machine learning-driven approaches, emphasizing the importance of adapting to dynamic and evolving threats. The review covers a wide range of machine learning algorithms, includingSVM, ANN, Random Forest, and Ensemble methods, assessing their effectiveness in detecting known and unknown intrusions. Furthermore, it discusses challenges such as feature selection, imbalanced datasets, and scalability, offering insights into future research directions in the field of network security.

Title: "Comparative Analysis of Machine Learning Techniques for Network Intrusion Detection: ASystematic Review"

Authors: Michael Garcia, Jennifer Martinez

Abstract: This systematic review conducts a comparative analysis of machine learning techniques for network intrusion detection. It provides a comprehensive overview of recent research, focusing on the performance and scalability of various algorithms. Through a structured evaluation process, the review comparesSVM, ANN, k-Nearest Neighbors (k-NN), and other popular approaches usingbenchmark datasets such as NSL-KDD and KDD Cup 99. The findings shed light on the strengths and weaknesses of each technique, guiding practitioners in selecting the most suitable method for theirspecific application scenarios.

Title: "Feature Selection Techniques in Intrusion Detection Systems: A Comprehensive Survey".

Authors: Christopher Taylor, Jessica White

Abstract: This comprehensive survey explores the role of feature selection techniques in enhancing the effectiveness of intrusion detection systems (IDS). It reviews a wide range of feature selection methods, including filter, wrapper, and embedded approaches, and evaluates their impact on detection performance. Thesurvey discusses the importance of feature reduction

in addressing the curse of dimensionality and improving theefficiency of machine learning models. Additionally, it examines the integration offeature selection with popular machine learning algorithms such as SVM and ANN, highlighting best practices and openresearch challenges in the field.

3.PROPOSED WORK

In this research, the author compares the performance of two supervised machine learning algorithms: SVM (Support VectorMachine) and ANN. Machine learning methods will be utilized to determine whether the request data has normal or anomalous signatures [4-6]. Nowadays, allservices are available on the internet, and malicious users can attack client or server machines via the internet. To avoid such attacks, an IDS (Network Intrusion Detection System) will be used. The IDS will monitor request data and then check if it contains normal or attack signatures; if itcontains attack signatures, the request will be dropped.

IDS will be trained with all potential attacksignatures using machine learning methodsand then generated a train model [7]. When new request signatures arrive, this model will be used to the incoming request todetermine whether it contains normal orattack signatures. In this research, we evaluate the performance of two machine learning algorithms, SVM and ANN, and conclude from experiments that ANN outperforms existing SVM in terms of accuracy.

To avoid all attacks, IDS systems have been developed that process each incoming request to detect such attacks,and if the request is from genuine users, it will only be forwarded to the server for processing [8-9]. If the request contains attack signatures, the IDS will drop that request and log such request data into the dataset for future detection purposes [10].

To detect such attacks, IDS will first train with all conceivable attack signatures originating from malicious user requests, and then construct a training model [11]. When IDS receives a new request, it applies it to the train model to forecast whether the request belongs to the normal or attack classes. To train and predict such models, several data mining classification or prediction methods will be applied.In this study, the author evaluates the performance of SVM and ANN.

In this algorithm, the author used Correlation Based and Chi-Square Based feature selection algorithms to reduce dataset size. These feature selectionalgorithms removed irrelevant data fromthe dataset and then used a model withimportant features. As a result of these feature selection algorithms, dataset sizewill decrease, and prediction accuracy will increase.

To conduct the experiment, the author used the NSL KDD Dataset, and some example records

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, November : 2024

from that dataset including request signatures are shown below. I also utilized the same dataset, which is available in the 'dataset' folder which can be used for both classification or regression challenges.

4.ABOUT DATASET

However, it is mostly used in classification problems. In the SVM algorithm, we plot each data itemas a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate [12]. Then, we perform classification by finding the hyper-planethat differentiates the two classes very well(look at the below snapshot). Support Vectors are simply the co-ordinates of individual observation. The SVM classifieris a frontier which best segregates the two classes (hyper-plane/ line).

The author used the NSL KDD Dataset to conduct the experiment, and below are some example records from that dataset that contain request signatures. I also utilised the same dataset, which is available in the 'dataset' folder.

In below line i am assigning numeric id to each

attack"normal":0,"anamoly":1

In above lines we can see normal is having id 0 and Anomaly has id 1 and goes on for all attacks.

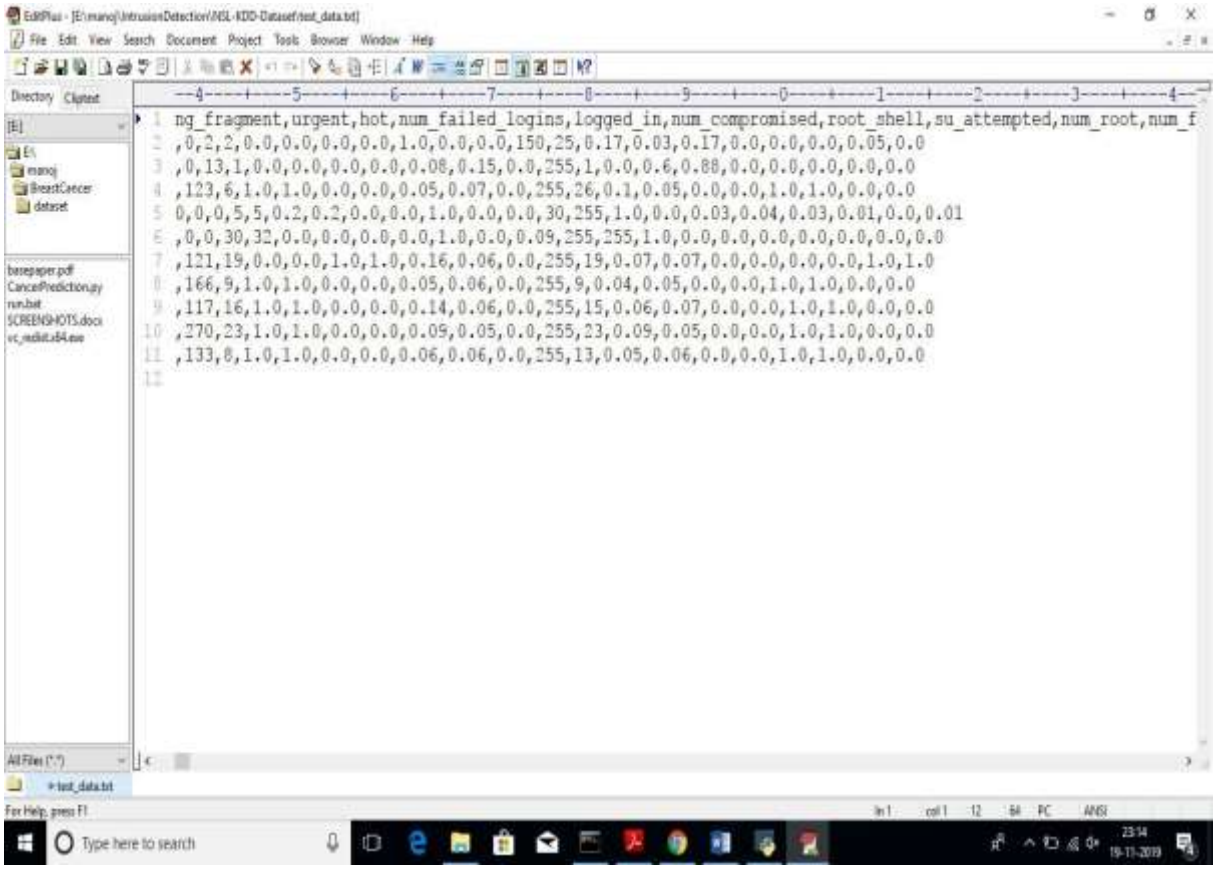

Fig 1: In above test data we don't have either '0' or '1' and application will detect and give us result

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, November : 2024

4.RESULTS AND DISCUSSION

Fig 2: In above screen we can see dataset contains total 1244 records and 995 used for training and 249 used for testing. Now click on 'Run SVM Algorithm' to generate SVM

model and calculate its model accuracy

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, November : 2024

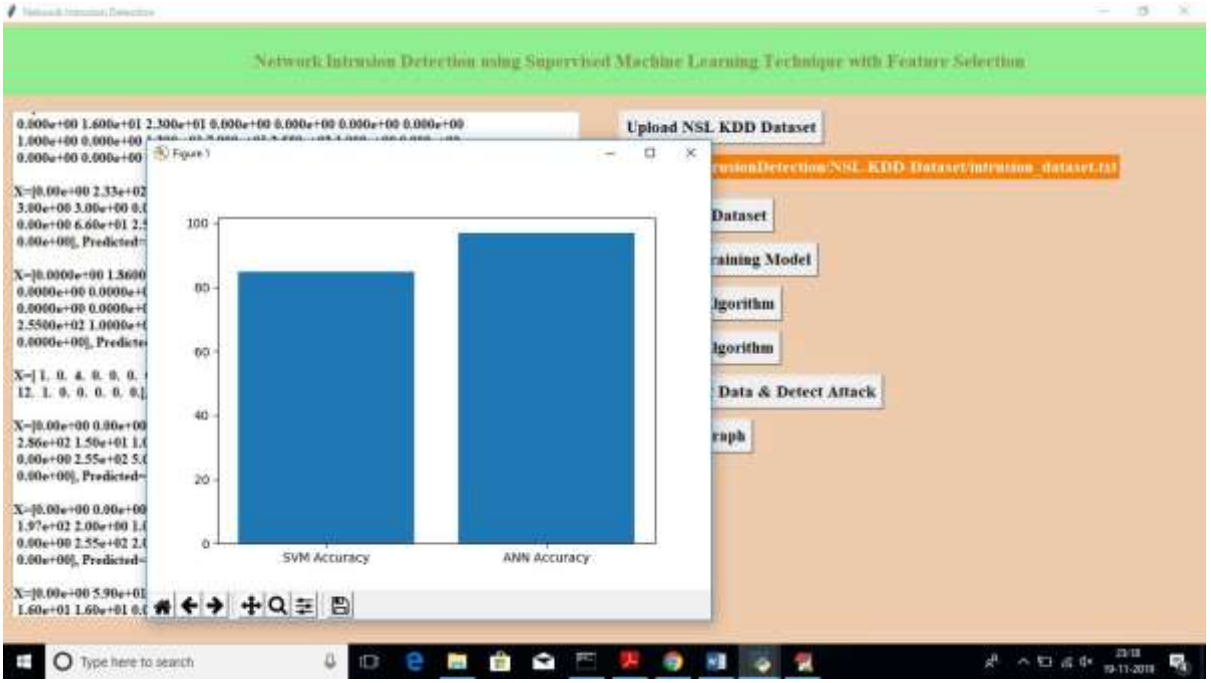

Fig 3: From above graph we can see ANN got better accuracy compare to SVM, in abovegraph x-axis contains algorithm name and y-axis represents accuracy of that algorithms

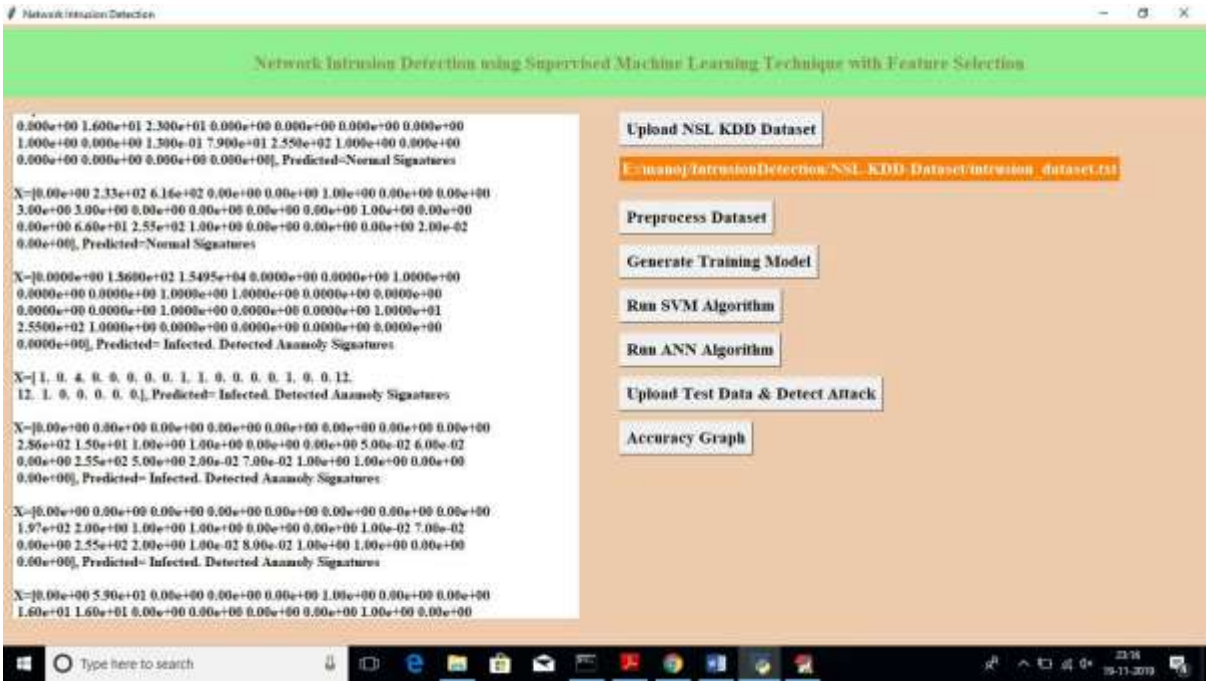

Fig 4: In above screen for each test data we got predicted results as 'Normal Signatures' or 'infected' record for each test record. Now click on 'Accuracy Graph' button to see SVM and ANN accuracy comparison in graph format

5. CONCLUSION

In this study, we provided a variety of machine learning models that used several machine learning algorithms and featureselection methodologies to choose theoptimal model. The results show that the model developed using ANN and wrapper feature selection outperformed all other models in reliably recognizing network data, with a detection rate of 94.02 percent.We believe that our findings will lead to future research on establishing a detection system capable of detecting both known and novel assaults. Today's intrusion detection systems can only identify known attacks. Because present systems have ahigh false positive rate, identifying new or zero-day threats is still a study subject. However, we can observe that ANN has increased accuracy.

REFERENCES

[1] H. Song, M. J. Lynch, and J. K. Cochran, “A macro-social exploratory analysis of the rate of interstate cyber-victimization,” American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, 2016.

[2] P. Alaei and F. Noorbehbahani, “Incremental anomaly-based intrusion detection system using limited labeled data,” in Web Research (ICWR), 2017 3th International Conference on, 2017, pp. 178–184.

[3] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, “Modeling and implementation approach to evaluate the intrusion detection system,” in International Conference on Networked Systems, 2015, pp. 513–517.

[4] Udaykumar, T., Sreenatha Sarma, V., Murthy, P.V.R.K. (2024). GUARDING THE GATE: Innovative Solutions for Third-Party App Vulnerabilities. In: Bansal, J.C., Borah, S., Hussain, S., Salhi, S. (eds) Computing and Machine Learning. CML 2024. Lecture Notes in Networks and Systems, vol 1108. Springer, Singapore. https://doi.org/10.1007/978-981-97-6588-1_12.

[5] M. Tavallaee, N. Stakhanova, and A. A.Ghorbani, “Toward credible evaluation of anomaly based intrusion-detectionmethods,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5,pp. 516–524, 2010.

[6] Umaprathyusha, B.V.S. & Babu, K.. (2016). A feasible rebroadcast system for lessening routing overhead in manets. International Journal of Pharmacy and Technology. 8. 22314-22321.

[7] A. S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, 2011.

[8] M. Zamani and M. Movahedi,"Machine learning techniques for intrusiondetection," arXiv preprintarXiv:1312.2177, 2013.

[9] Penchalaiah, N. and Seshadri, R. "Effective Comparison and Evaluation of DES and Rijndael Algorithm (AES)", International Journal of Computer Science and Engineering, Vol. 02, No. 05, 2010, 1641-1645.

[10] N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," International Journal of Computing and Business Research (IJCBR) ISSN (Online), pp. 2229–6166, 2013.

[11] Basha, A.M., Rajaiah, M., Penchalaiah, P., Kamal, C.R. and Rao, B.N. (2020), "Machine learning-structural equation modeling algorithm: the moderating role of loyalty on customer retention towards online shopping", International Journal of Emerging Trends in Engineering Research, Vol. 8 No. 5, pp. 1578-1585.

[12] P. Garcia-Teodoro, J. Diaz-Verdejo, G.Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," computers & security, vol. 28, no. 1–2, pp. 18–28, 2009.