International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

A Comparative Study for Improved Network Intrusion Detection

M.Anitha[1],E.Nagaraju[2],T.Soniya[3]

#1 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.
#2 Assistant Professor in the Department of MCA,SRK Institute of Technology, Vijayawada.
#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

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 also seriously 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 traffic dataset [3]. Due to huge and imbalanced dataset machine learning based Intrusion Detection System (IDS) faces problem to processentire 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 and unsupervised 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, including SVM, 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: A Systematic 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 compares SVM, ANN, k-Nearest Neighbors (k-NN), and other popular approaches using benchmark 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 their specific 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. The survey discusses the importance of feature reduction in addressing the curse of dimensionality and improving the efficiency of machine learning models. Additionally, it examines the integration of feature selection with popular machine learning algorithms such as SVM and ANN, highlighting best practices and open research challenges in the field.

3.PROPOSED WORK

In this research, the author compares the performance of two supervised machine learning algorithms: SVM (Support Vector Machine) and ANN. Machine learning methods will be utilized to determine whether the request data has normal or anomalous signatures. Nowadays, all services 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 it contains attack signatures, the request will be dropped.

IDS will be trained with all potential attack signatures using machine learning methods and then generated a train model. When new request signatures arrive, this model will be used to the incoming request to determine whether it contains normal or attack 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. If the request contains attack signatures, the IDS will drop that request and log such request data into the dataset for future detection purposes.

To detect such attacks, IDS will first train with all conceivable attack signatures originating from malicious user requests, and then construct a training model. 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 selection algorithms removed irrelevant data from the dataset and then used a model with important features. As a result of these feature selection algorithms, dataset size will decrease and prediction accuracy will increase.

To conduct the experiment, the author used the NSL KDD Dataset, and some example records from that dataset including request signatures are shown below. I also utilized the same dataset, which is available in the 'dataset' folder.

3.1 IMPLEMENTATION

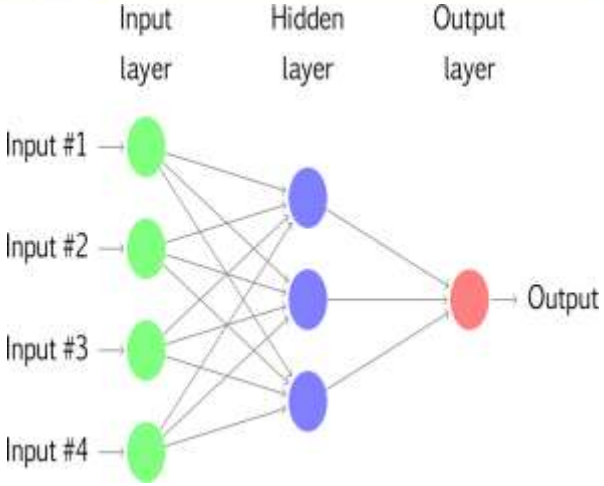
3.1.1 ANN

To demonstrate how to construct an ANN neural network-based image classifier, we will construct a six-layer neural network that will detect and differentiate one image from another. This network that we will construct is a very modest network that can also be run on a CPU. Traditional neural networks that excel at picture classification have many more parameters and take a long time to train on a standard CPU. However, our goal is to demonstrate how to use TENSORFLOW to construct a real-world convolutional neural network.

Neural Networks are mathematical models that are used to tackle optimization problems. They are built up of neurons, which are the basic computation units of neural networks. A neuron receives an input (say x), does some computation on it (say, multiplying it by w and adding another variable b), and produces a value (say, z= wx + b). This value is transferred to a non-linear function called activation function (f) to generate the neuron's final output (activation). There are numerous types of activation functions. Sigmoid is a prominent activation function. The neuron that uses the sigmoid function as an activation function is referred to as a sigmoid neuron. Neurons are named based on their activation functions, and there are many different types of them, such as RELU and TanH.

If you stack neurons in a single line, it’s called a layer; which is the next building block of neural networks. See below image with layers

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362



To predict image class multiple layers operate on each other to get best match layer and this process continues till no more improvement left.

3.1.2 SVM ALGORITHM

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as 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. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot).Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes (hyper-plane/ line).

4.ABOUT DATASET

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.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

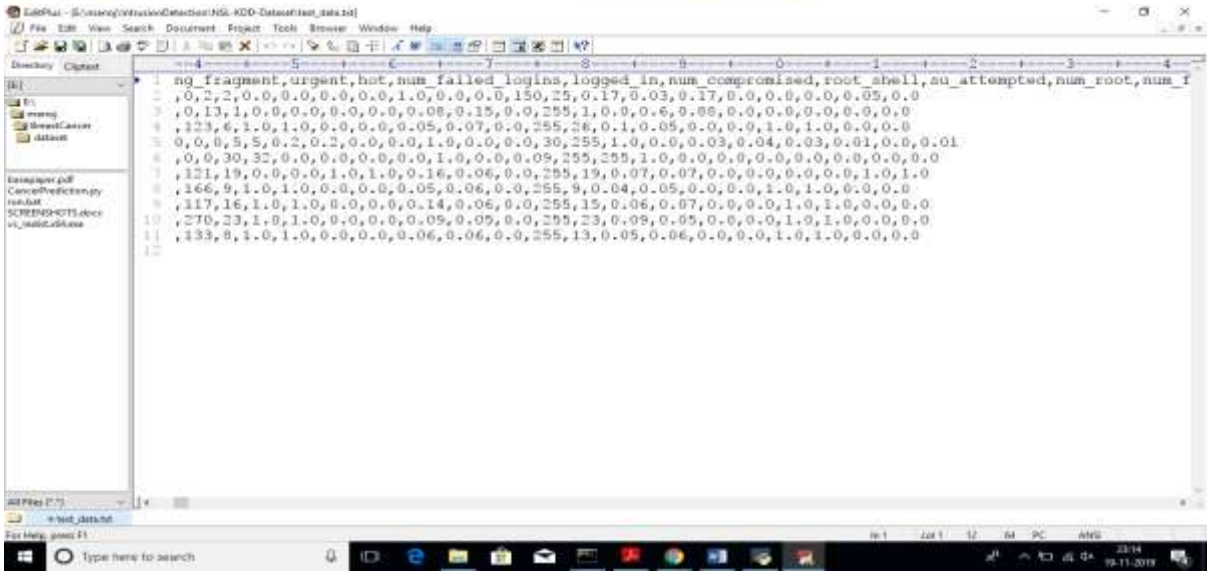


Fig 1:In above test data we don't have either '0' or '1' and application will detect and give us result

4.RESULTS AND DISCUSSION

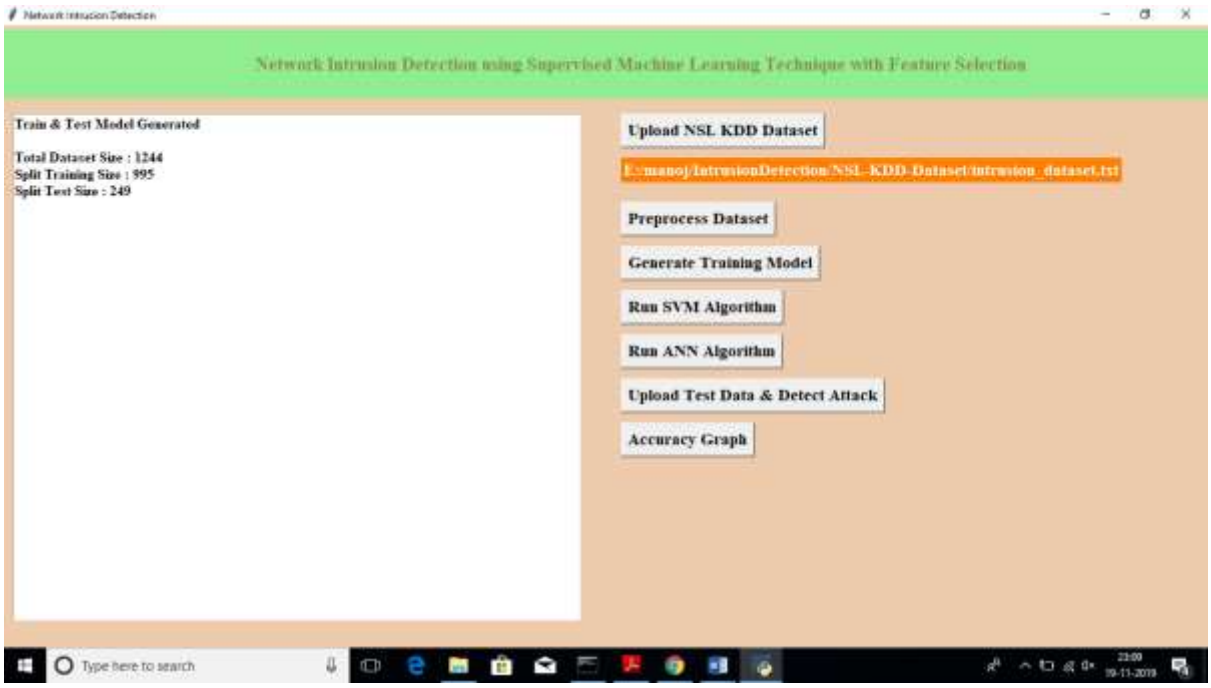


Fig 2:In above screen we can see dataset contains total 1244 records and 995 used for training and 249 used for testing. Now click on 'Run SVM Algorithm' to generate SVM model and calculate its model accuracy

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

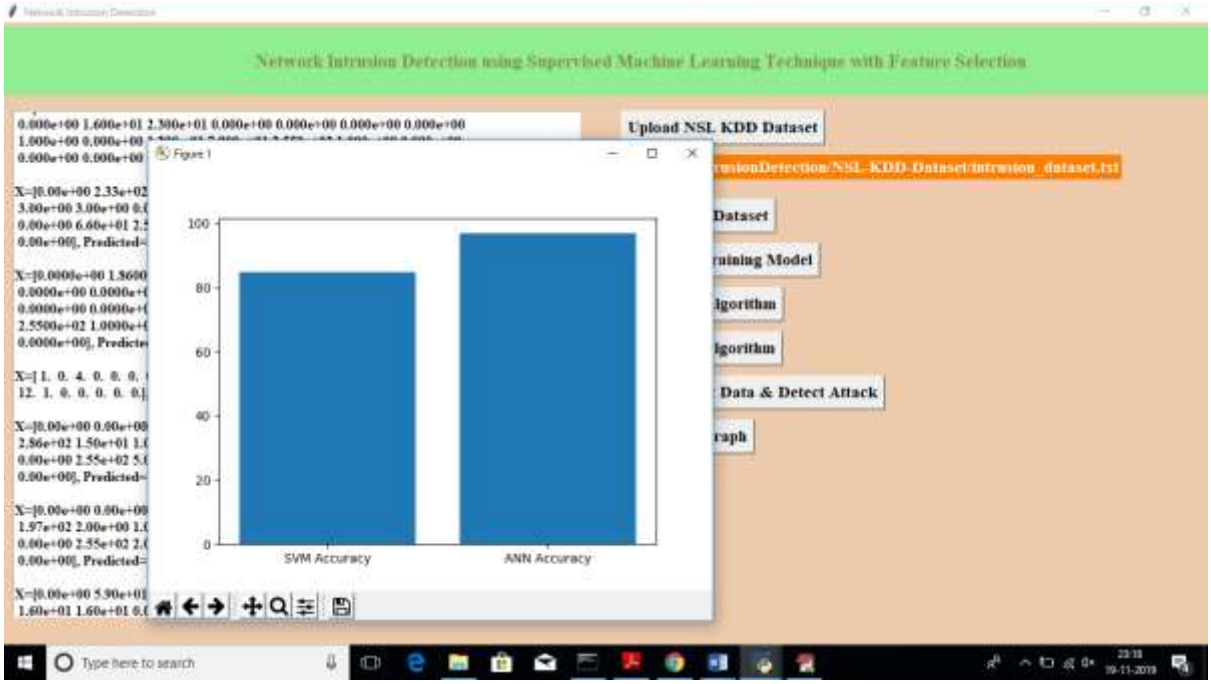


Fig 3:From above graph we can see ANN got better accuracy compare to SVM, in above graph x-axis contains algorithm name and y-axis represents accuracy of that algorithms

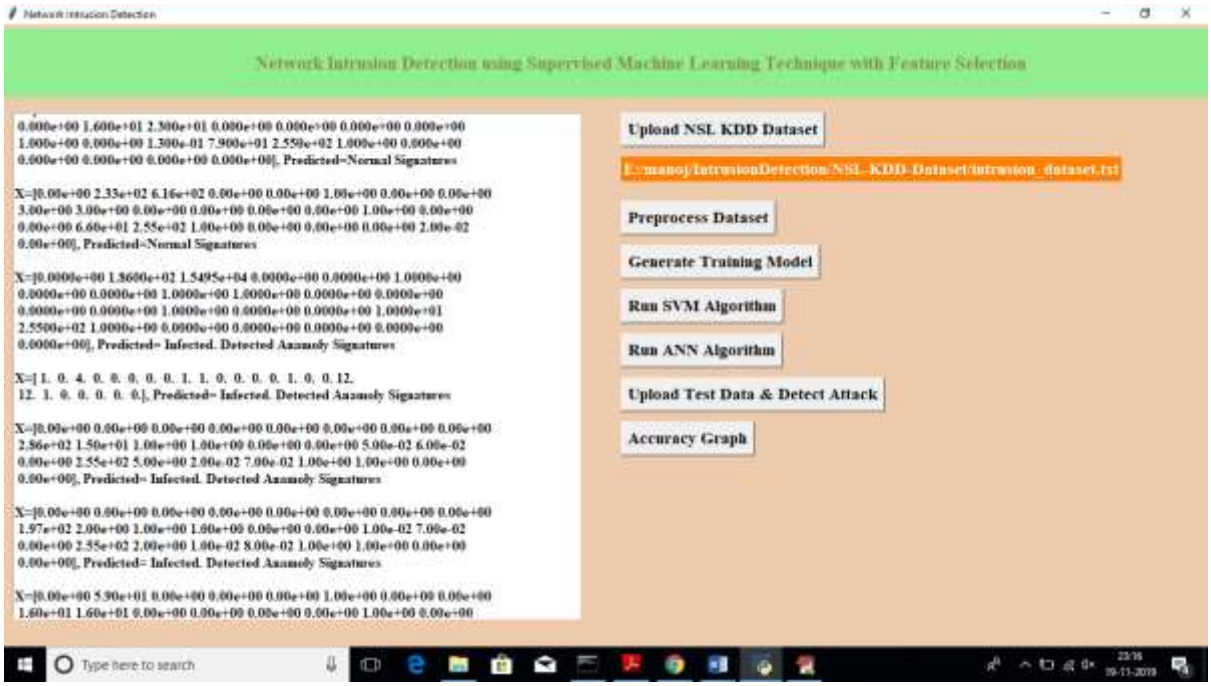


Fig 4:In above screen for each test data we got predicted results as 'Normal Signatures' or 'infected' record for each test record. Now click on 'Accuracy Graph' button to see SVM and ANN accuracy comparison in graph format

5.CONCLUSION

In this study, we provided a variety of machine learning models that used several machine learning algorithms and feature selection methodologies to choose the optimal 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 a high 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]M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomalybased intrusion-detection methods," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516–524, 2010.

[5]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.

[6]M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," arXiv preprint arXiv:1312.2177, 2013.

[7]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.

[8] 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

Author's Profiles

Ms.M.Anitha Working as Assistant Professor & Head of Department of MCA ,in SRK Institute of technology in Vijayawada. She done with B .tech, MCA ,M. Tech in Computer Science .She has 14 years of Teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Machine Learning with Python and DBMS.

Mr.E.Nagaraju completed his Masters of Computer Applications. He has published A Paper Published on ICT Tools for Hybrid Inquisitive Experiential Learning in Online Teaching-a case study Journal of Engineering Education Transformations, Month 2021, ISSN 2349- 2473, eISSN 2394-1707. Currently working has an Assistant professor in the department of MCA at SRK Institute of Technology, Enikepadu, NTR (DT). His areas of interest include Artificial Intelligence and Machine Learning.

Ms.Tadaka.Soniya is an MCA Student in the Department of Computer Application at SRK Institute Of Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.Sc(MPCs) from Gayatri college of science and management , Munasbpeta, Srikakulam, Srikakulam District. Her area of interest are DBMS and Python.