International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 13th Apr 2022. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=ISSUE-04

DOI: 10.48047/IJIEMR/V11/I04/30

Title **Network Intrusion Detection Using Supervised Machine Learning Techniques with Feature Selection**

Volume 11, Issue 04, Pages: 206-212

Paper Authors
N.Chandbi, K.Likhitha, M.Supriya, N.Deepika, Dr. N .Rajeswari

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Network Intrusion Detection Using Supervised Machine Learning Techniques with Feature Selection

N.Chandbi[1], K.Likhitha[2], M.Supriya[3], N.Deepika[4], Dr. N .Rajeswari[5]

[1,2,3,4]Under Graduate Student, [5]Associate Professor Computer Science and Engineering,
Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, India.
[1]chandbi.nandanapu@gmail.com ,[2]likhithakunchala27@gmail.com
[3]supriyamiriyala9@gmail.com, [4]deepikanagudu945@gmail.com
[5]rajeswari.gec@gmail.com

Abstract:

To classify community visitors as harmful or benign, a novel supervised computer researching machine is built. A combination of supervised studying algorithm and characteristic determination strategy was employed to find the high-quality mannequin thinking about detecting success rate. This study discovered that when it comes to recognising network traffic, ANN-based totally laptop mastering with wrapper function resolution surpasses the guide vector laptop (SVM) method. To assess performance, the NSL-KDD dataset is utilised to identify community visitors using supervised laptop learning techniques such as SVM and ANN. In terms of intrusion detection success rate, comparative research shows that the proposed mannequin is more environmentally friendly than other contemporary trends.

Keywords: Intrusion, feature selection, SVM, ANN

1. Introduction

As a result of the rapid expansion of information technology, computer networks are widely employed by industry, business, and a variety of facets of human life[1].As a result, IT managers must work hard to build a trustworthy network. Furthermore, information technology is advancing at a breakneck pace has generated significant challenges in constructing a dependable network. This is a very difficult assignment. A range of attacks are putting the availability, integrity, and privacy of computer networks at jeopardy. One of the most popular types of attacks is a denial of service attack (DOS). The goal of a distributed denial-of-service attack is to temporarily disable a number of end-user services. Network resources are typically depleted, and unwanted request systems are overworked. As a result, DOS has come to be used as a catch-all term for all types of attacks intended at depleting computer and network resources. As a result, identifying all forms of attacks is extremely difficult; as a result, intrusion detection systems (IDS) have become an important part of network security.

It's used the keep track on network traffic and send out alerts if any attacks are identified. IDSs can monitor network traffic on a single device (host intrusion detection system) or the entire network (network intrusion detection system), which is the most frequent type. In general, there are two types of intrusion detection systems (IDS). [2] Attacks are detected using an anomaly-based intrusion detection system based on previously recorded normal activities. As a result, we compare real-time traffic to typical traffic data from the past. This form of intrusion detection system is often used since it can detect new types of invasions. Keep a record of the most important values of a false positive alarm, nevertheless. As a result, a number of legitimate packages have been labelled as attack packages. An exploit intrusion detection system, on the other hand, uses attack signature stores to detect attacks. There are no false alarms, but Alerts can be successfully triggered by new sorts of attacks (new signatures). Because traditional intrusion detection systems rely on costly human input to evaluate whether the target is a normal or an attack packet, machine learning algorithms[2] can be employed as a substitute. Machine learning approaches to building intrusion detection models have recently attracted a lot of attention. Detecting normal and aberrant intrusive activity using machine learning methods, Intrusion detection is addressed as a classification task in this model. An effective categorization strategy can be used to build an accurate intrusion detection model. As a result, the suggested technique in this paper evaluates and develops an intrusion detection system model based on a Knowledge Discovery in Databases (KDD) dataset using machine learning classifiers such as C4.5 Decision Tree and Naive Bayes classifiers.

II Related Works

2.1 "Evaluation of machine learning algorithms for intrusion detection system," 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, 2017, pp. 000277- 000282. M. Almseidin, M. Alzubi, S. Kovacs, and M. Alkasassbeh, "Evaluation of machine learning algorithms for intrusion detection system," 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, 2017.

One of the used options against harmful attacks is an intrusion detection device (IDS). In addition, attackers are continually changing their equipment and strategies. Imposing a generic IDS device, on the other hand, is a challenging undertaking. A number of experiments have been carried out and assessed in this research in order to determine a number of computing device learning classifiers most of which were based on the KDD intrusion dataset. In order to consider the classifiers that were chosen, it was able to generate a number of overall performance metrics. To increase the detection charge of the intrusion detection system, the focus was first on false good and false pleasant overall performance measurements. Experiments

revealed that the choice desk classifier had the lowest rate of false alarms, while the random woodland classifier had the best overall accuracy rate

III Machine Learning (ML)

This study's author compares the results of two supervised machine learning approaches. Two examples are SVM (Support Vector Machine) and ANN (Automatic Neural Network) (Artificial Neural Networks). To identify whether the request data has typical or atypical signatures, machine learning algorithms will be used. An IDS (Network Intrusion Detection System) will be used to monitor request data and identify whether it contains normal or attack signatures; if it does, the request will be dropped.

Using machine learning techniques, the IDS will be educated with all conceivable attack signatures, and a train model will be used to assess whether new requests include normal or attack signatures. In this paper, we compare the accuracy of two machine learning algorithms, SVM and ANN, and conclude that SVM outperforms ANN. IDS systems were developed to detect all attacks and only send valid user requests to the server for processing; if the request contains attack signs, the IDS will drop the request and log the data into a dataset for future detection.

To detect such attacks, an IDS will be first trained with all conceivable attack signatures generated by malicious user requests, and then a training model will be developed. To assess whether the train model belongs to the normal or attack class, IDS will send a new request to it. Several data mining categorization or prediction methodologies will be employed to train such models and forecasts.

In this paper, the author compares the performance of SVM and ANN.

To reduce the size of the dataset, the author created a feature selection strategy based on Correlation Based and Chi-Square. This feature selection approach removed irrelevant data from the dataset prior to applying the key characteristic. As a result of these feature selection techniques, the dataset size will be lowered and prediction accuracy will improve.

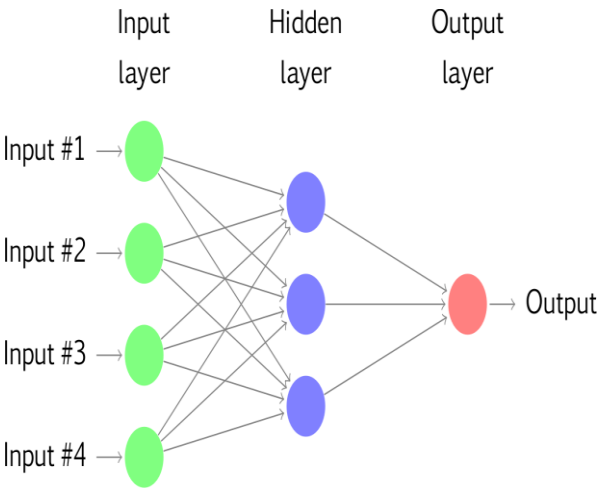
3.1 MATERIALS AND METHODS

We'll build a six-layer neural network to recognise and differentiate one image from another to show how to build an ANN neural network-based image classifier. This network that we will build is simple one that can be run on a computer. Traditional neural networks for image classification have a lot more parameters and take a long time to train on a normal CPU. Our goal, however, is to show how to use TENSORFLOW to build a real-world convolutional neural network.

Mathematical models known as neural networks are utilised to solve optimization problems. Neurons, the basic computation units of neural networks, are used to construct them. A neuron takes an input

(say x), performs some calculations on it (say, multiplying it by w and adding another variable b), and then outputs a value (say, z= wx + b). This value is routed via a non-linear function known as activation function (f) to form the neuron's final output (activation). Activation functions exist in a variety of shapes and sizes. A well-known activator is the sigmoid function. The neuron that the sigmoid function activates. Neurons are classified according to their activation roles, and there are a variety of them, including RELU and TanH.

A layer is the next building component of neural networks, and it is formed by stacking neurons in a single line. Layers can be seen in the image below.



To forecast image class, numerous layers interact with one another to find the best match layer, and this process is repeated until no more improvement is possible.

3.2 SVM ALGORITHM

The "Support Vector Machine" (SVM) is a supervised machine learning technology that can be used to solve issues such as classification and regression. However, it is mostly used to tackle categorization problems. Each data item is represented as a point in n-dimensional space (where n is the number of features you have), with the value of each feature being the SVM algorithm's value of a specific coordinate. Then we locate the hyper-plane that best distinguishes the two classes to complete classification. Support Vectors are made up of individual observation co-ordinates. The SVM classifier is a frontier that best distinguishes between the two classes (hyperplane /line).

IV Experimental Results

The author conducted the experiment using NSL KDD Dataset, and example of request signatures from that dataset are shown below. I also used the same dataset from 'dataset' folder.

I'm assigning numeric ids to each attack in the line below: "normal":0,"anamoly":1

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Normal has id 0 and Anomaly has id 1 in the above lines, and so on for all attacks.

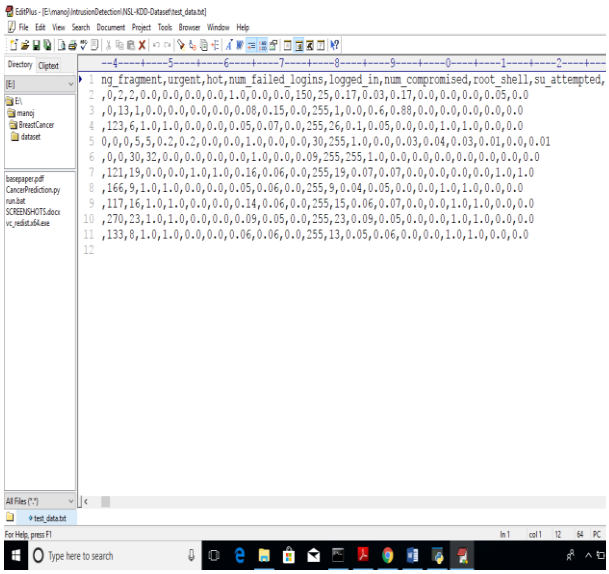


Fig 1:we don't have either a 0 or 1 in the given test data ,thus the application will identify this and return a response.

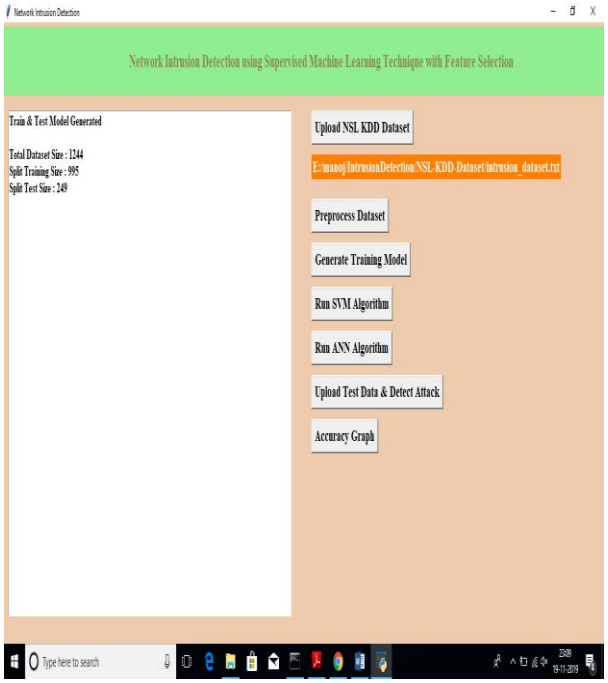


Fig 2: The above dataset has 1244 elements, with 995 would be used the training and 249 using it for testing. To create an SVM model and verify its accuracy, select 'Run SVM Algorithm.'

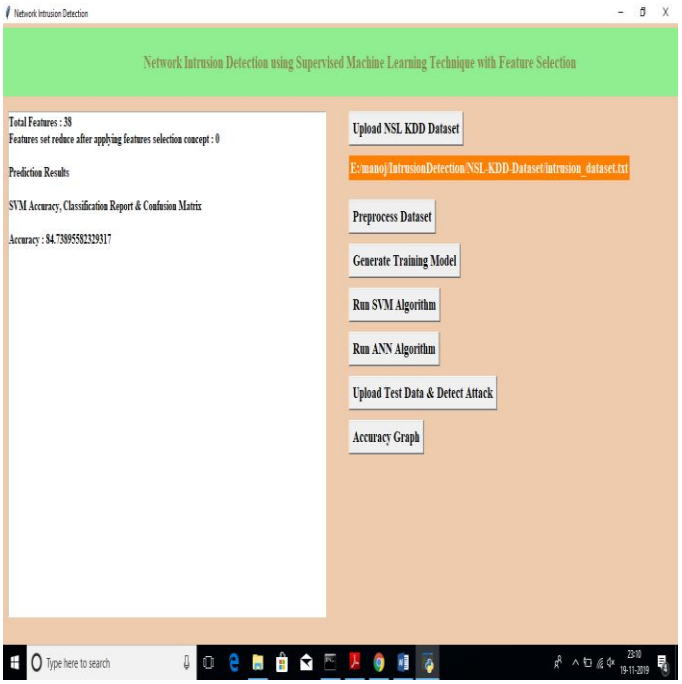


Fig 3: We can observe SVM in the image above. With SVM, we achieved an accuracy of 84.73 percent; now, click on 'Run the ANN Algorithm' to obtain the ANN accuracy.

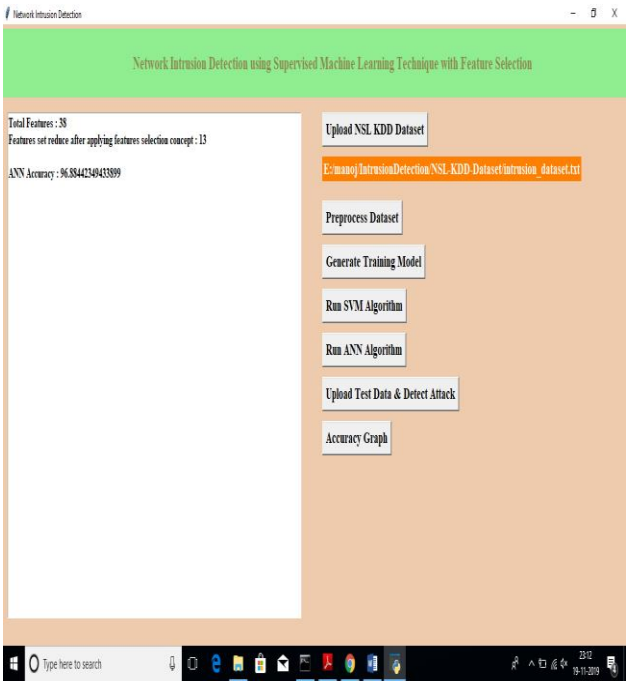


Fig 4:We acquired 96.88 percent accuracy in the previous screen; now we'll upload test data

and forecast whether it's normal or contains an attack by clicking the 'Upload Test Data & Detect Attack' button. All test data is unclassified as 0 or 1, and the application forecasts and gives us a result. Some test data records are listed below.

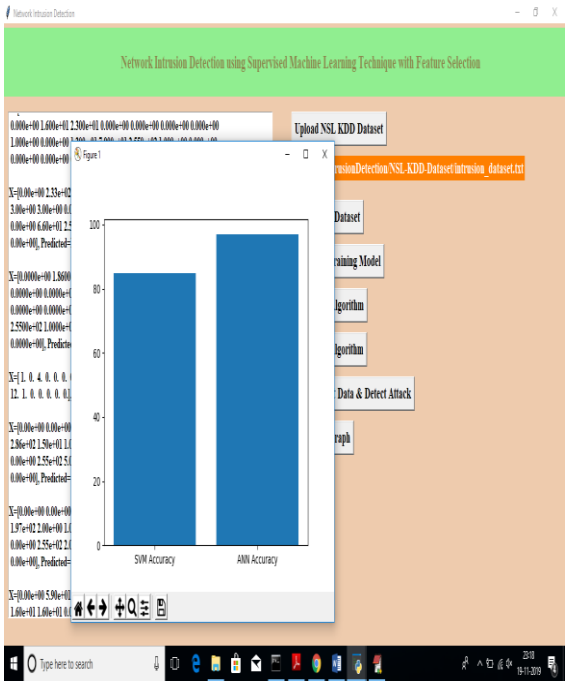


Fig 5: We can see that ANN is more accurate than SVM in the graph above; the x-axis shows the algorithm name, while the y-axis reflects the method's accuracy.

V. Conclusion

We presented a number of machine learning models in this paper, each of which used a different machine learning algorithm and feature selection method to arrive at the best model. With a detection rate of 94.02 percent, the model created with ANN and wrapper feature selection outperformed all other models in successfully detecting network data. Our findings are expected to benefit future research in the development of a detection system capable of identifying both known and novel assaults. Intrusion detection systems can now only detect the known intrusions. Detecting new attacks or zero-day assaults remains as research area because to the huge probability of detection of existing techniques. On either hand, ANN looks to be more accurate.

References

[1]American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, 2016. "A macro-social exploratory analysis of the rate of interstate cyber-victimization."

[2]"Incremental anomaly-based intrusion detection system with limited labelled data," in Web Research (ICWR), 2017 3rd International Conference on, pp. 178–184. P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system with limited labelled data," in Web Research (ICWR), 2017 3rd International Conference on, pp. 178–184.

[3]In International Conference on Networked Systems, 2015, pp. 513–517, "Modeling and implementation technique to evaluate the intrusion detection system."

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

[5] "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 strategies for intrusion detection," arXiv preprint arXiv:1312.2177, 2013. [7] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," arXiv preprint arXiv:1312.2177, 2013.

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