Contents

1. Abstract…..1

2. Acknowledgement… 2

3. Keywords 3

4. List of figures.....4

5. List of tables.....5

6. Introduction….....6-7

7. Literature Survey……..8-9

8. Related Work……10

9. Existing System…….11

10. Proposed System……12-13

11. Dataset…….14-15

12. Preprocessing……16

13. Algorithm…….17-19

14. Result……20

15. Conclusion .....….21

16. References..... 22

1. Abstract

Given the increasing complexities of today's network environments, more and more hosts are becoming vulnerable to attacks and hence it is important to look at systematic, efficient and automated approaches for Intrusion Detection. Network Security is to protect computer network against hacking, misuse, unauthorized changes to the system and securing a computer network infrastructure. Network attack is the intrusion or threat can be defined as any deliberate action that attempts unauthorized access, information manipulation, or rendering the system unstable by exploiting the existing vulnerabilities in the system. Intrusion Detection system (IDS) / Intrusion Prevention System (IPS) has become a prerequisite in computer networks. IDS/IPS is a device or software application that monitors network or system activities for malicious activities. These type of IDS/IPS used in the network is known as Network based IDS/IPS.

The goal of a Network Intrusion Detection System (NIDS) is to alert a system administrator each time an intruder tries to penetrate the network. A misuse NIDS defines attacks via a table of malicious signatures: if an ongoing activity matches a signature in the table, an alarm is raised. Thousands of organizations depend upon such systems because they are simple to understand, enable administrators to customize the signature database, and provide concrete information about the events that have occurred.

In this project we compared the use of various machine learning techniques, to the use of various deep learning techniques for detecting intrusions in networks. We utilized various supervised machine learning techniques and deep learning techniques to train and build multiple classification models that can classify attack type of network traffic versus normal type of network traffic. We compared the test accuracies of these multiple classification models to identify the best model for performing network intrusion detection. We concluded that using a combination of Autoencoder for feature selection, followed by fully connected deep learning model for classification gave better accuracy compared to using machine learning models.

3. Keywords

Keywords	Descriptions
Auto-Encoder	An autoencoder is a type of Artificial neural network used to learn efficient codings of unlabled data.
DNN	DNNs are typically feedforward Networks in which data flows From the input layer to the output Layer without looping back.
DOS	DOS attack is a cyber-attack in which the perpetrator seeks to make a machine or network resource unavailable to its intended users by temporarily or indefinitely disrupting services of a host connected to the internet.

4.List of Figures

Fig 1: System Architecture and Flow

Fig 2: Model Structure

Fig 3: Flow-Diagram of Proposed Algorithm

Fig 4: Accuracies for the ML classification technique

Fig 5: Plot for the training and testing accuracy of the DNN that is using AutoEncoder (Blue; Test accuracy, Red: Train accuracy)

5. List of Tables

Table 1: Comparison between Misuse and Anomaly Detection

Table 2: Architecture of the DNN using AutoEncoder

Table 3: Details about the Dataset

Table 4: Various files that are available for this dataset and their description

Table 5: Details of the NSL-KDD training Dataset

Table 6: Details of the NSL-KDD testing Dataset

Table 7: Machine Learning Methods used

Table 8: Accuracies for the ML classification techniques

Table 9: Architecture of the AutoEncoder

Table 10: Architecture of the DNN using AutoEncoder

Table 11: Accuracies achieved by DNN that used AutoEncoder

Table 12: Comparison of accuracies achieved by various model

6. Introduction

Introduction to Network Security

Network security consists of the provisions and policies adopted by a network administrator to prevent and monitor unauthorized access, misuse, modification, or denial of a computer network and network-accessible resources. Network security involves the authorization of access to data in a network, which is controlled by the network administrator. Users choose or are assigned an ID and password or other authenticating information that allows them access to information and programs within their authority. In the past, hackers were highly skilled programmers who understood the details of computer communications and how to exploit vulnerabilities. Today almost anyone can become a hacker by downloading tools from the Internet. These complicated attack tools and generally open networks have generated an increased need for network security and dynamic security policies. The easiest way to protect a network from an outside attack is to close it off completely from the outside world. A closed network provides connectivity only to trusted known parties and sites; a closed network does not allow a connection to public networks.

Network Attack and its Types

Networks attacks are subject to attacks from malicious sources. Network attack is the intrusion or threat can be defined as any deliberate action that attempts unauthorized access of Information manipulation and by exploiting the existing vulnerabilities in the system. A Network attack is deliberate exploitation of computer systems, technology-dependent enterprises and networks. Network attacks use malicious code to alter computer code, logic or data, resulting in disruptive consequences that can compromise data and lead to cybercrimes, such as information and identity theft.

Types of Attacks

Passive Attacks- An network intruder intercepts data travelling through the network (e.g) Wiretapping, port and idle Scanner

Active Attacks - An intruder initiates commands to disrupt the network's normal operation. (e.g) DOS, Spoofing, SQL Injection, Cross-site Scripting

Intrusion :

A deliberate attempt to enter a network and break the security of the network and thus breaking the confidentiality of the information present in the systems of the network . The person who tries to attempt such an action is called as an Intruder and the action can be termed as Network Intrusion. It is any set of activities that attempt to compromise the integrity, confidentiality or availability of a resource.

Intrusion Detection System (IDS)

An IDS can be a piece of installed software or a physical appliance that monitors network traffic in order to detect unwanted activity and events such as illegal and malicious traffic, traffic that violates security policy, and traffic that violates acceptable use policies. Many IDS tools will also store a detected event in a log to be reviewed at a later date or will combine events with other data to make decisions regarding policies or damage control. An IPS is a type of IDS that can prevent or stop unwanted traffic. The IPS usually logs such events and related information. An Intrusion Prevention System (IPS) goes one step further and not only detects attacks but attempts to prevent them as well.

Advantages of Intrusion Detection Systems :

1. The network or computer is constantly monitored for any invasion or attack.

2. The system can be modified and changed according to needs of specific client and can help outside as well as inner threats to the system and network.

3. It effectively prevents any damage to the network.

4. It provides user friendly interface which allows easy security management systems.

5. Any alterations to files and directories on the system can be easily detected and reported

7. Literature Survey

Stated by (Kazienko & Dorosz, 2003), an Intrusion Detection System is a defence mechanism, which detects hostile activities in a network. System will be compromise if the intrusion is not detected and possible prevented. One of the major benefits of intrusion detection system is it provides an overview of any unusual unscrupulous activities. According to (Amoroso, 1999), intrusion detection is “a process of identifying and responding to malicious activity targeted at computing and networking resources".

Even though there are firewall and antivirus programs installed to protect their computer from any unwanted access, it can still be vulnerable to any unauthorised user. With the inclusion of network intrusion detection and prevention system, there will be another protection layer against potential hackers.

Intrusion detection and prevention systems are much more secure than common firewall technology. Although considered to be an expansion of the original intrusion detection system, they are actually more a way of controlling who has access to a computer network. They not only control access, but also detect entry to the network, so the two systems are closely linked.

There are 4 types of detection system. One of the systems is network-based detection system where it is mostly used on virtual private servers, remote access servers, and routers by analyzing various network protocols (Sturmer, 2013). Wireless intrusion detection system works much like network-based system only that it applies on wireless networks (Adams). Access point misuse is one of the illegal activities that are monitored by the system. In host based system, works on an individual computers. Any changes on the file system, abnormal network traffic and odd application process (Sturmer). Whereas for network behaviour analysis, it detects any irregularity in the system and the also the amount of traffic flow of the network (Seehorn).

. Statistics from (Magalhaes, 2003):

- Almost 90% of interconnected networks that uses Intrusion Detection System detected computer security breach in the last 12 months even though there are several firewalls installed.
- Computer Security Institute, 4/7/02 reported that 80% reported financial losses in excess of $455M was caused by intrusion and malicious acts thereafter. • Millions of jobs have been affected because of intrusion.
- Only 0.1% of companies are spending the appropriate budget on Intrusion Detection System.

- Intrusion Detection System are mostly mistaken as a firewall or its substitute.
- By using Intrusion Detection System will act as an additional barrier on top of an antivirus. Most organizations using antivirus software do not use IDS.

8. Related Work

Chunjie Zhou, Shuang Huang, et al.[1] in this paper they have used an anomaly detection based on multimodel has proposed and intelligent detection algorithms are designed. Classifier based on an intelligent hidden Markov model. A novel multimodel-based anomaly intrusion detection system with embedded intelligence and resilient coordination for the field control system in industrial process automation is designed. In this system, an anomaly detection based on multimodel has proposed, and the corresponding intelligent detection algorithms are designed. Furthermore, to overcome the disadvantages of anomaly detection, a classifier based on an intelligent hidden Markov model, have designed to differentiate the actual attacks from faults. The unique feature of the proposed intelligent intrusion detection system has that it uses complete multiple models of PCS built through the integration of multi domain knowledge to detect system anomaly, and employs HMM models to identify attacks from the sequential anomaly alerts. So in conclusion, the proposed intelligent intrusion detection system can detect the attack from both spatial and temporal aspects. In addition, since our intrusion detection system developed for PCSs takes into account the system knowledge instead of attack signatures, unknown type and new type of attack can also be detected by the proposed IDS. This paper is based on anomaly intrusion detection without consideration of attack knowledge. For the future, a comprehensive intrusion detection system for PCSs, which integrates system knowledge and attack knowledge, will be researched to optimize resources and time.

Al-Jarrah, O. ; Dept. of Comput. Eng., et al.[2] in this paper they have used an intelligent system to maximize the recognition rate of network attacks by embedding the temporal behavior of the attacks into a TDNN neural network structure. The proposed system consists of five modules: packet capture engine, preprocessor, pattern recognition, classification, and monitoring and alert module. This system captures packets in real time using a packet capture engine that presents the packets to a preprocessing stage using two pipes. The preprocessing stage extracts the relevant features for port scan and host sweep attacks, stores the features in a tapped line of a TDNN, and produces outputs that represent possible attack behaviors in a pre-specified number of packets. These outputs are used by the pattern recognition neural networks to recognize the 23 attacks, which are classified, by the classifier network to generate attack alerts. DARPA data sets are used to evaluate the systems in terms of recognition capability and throughput. Test results show that this system detects all types of attacks much faster than rule based systems such as SNORT.

9. Existing System

Depending on HOW the intrusion detection is performed there are 2 types of IDS:

Misuse Detection:

Misuse detection is driven by KNOWN attacks. Known attacks are used to define patterns of malicious network activities.

Anomaly Detection:

Anomaly detection is suitable for detection UNKNOWN attacks, based on profiles defining normal and anomalous behaviors.

feature	IDS detection techniques	
	Misuse detection	***Anomaly detection***
Attacks detected	Known attacks only	Any type
Attack background data required	Yes	No
False alarm rate	Low	High
Need update	Yes	No
Attack type	Defined	Cannot be defined
Protection tool identification	Yes	No

Table 1: Comparison between Misuse and Anomaly Detection

10. Proposed System

- In the proposed system, initially we used machine learning techniques to pre-process the dataset.
- On this pre-processed dataset, we applied Auto-Encoder to further perform the feature selection.
- Then we trained a model using a fully connected deep neural network. In this fully connected deep neural network we used the Auto-Encoder which we constructed in the previous step, in it's hidden layers.
- This proposed system gave high accuracies while detecting the network intrusions compared to the existing systems.

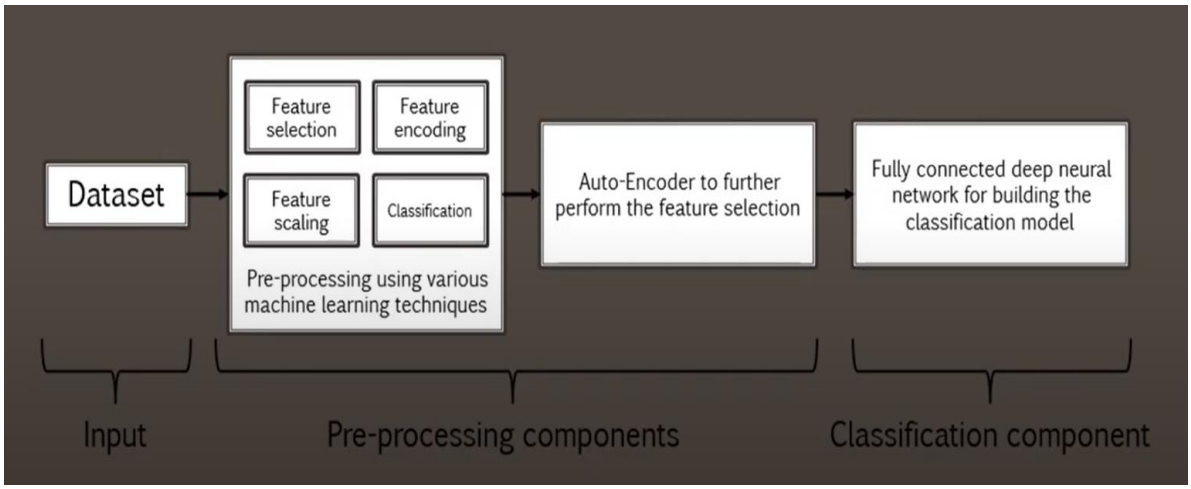

Fig 1: System Architecture and Flow

System Architecture and Flow:

Activation functions:

- Activation function is used to determine the output of each node in the neural network.
- ReLU is the most used activation function while working with fully connected deep neural networks. The function returns 0 if it receives any negative input, but for any positive value x it returns that value back.
- ReLU overcomes the vanishing gradient problem, allowing models to learn faster and perform better.
- Sigmoid is used as an activation function in the output layer. Sigmoid takes a real value as input and outputs another value between 0 and 1. It's good for a classifier.

Optimizer function:

- Optimizers are algorithms or methods used to change the attributes of neural network such as weights and learning rate in order to reduce the losses and to provide the most accurate results.
- Adam is the most used optimizer while working with fully connected deep neural networks.
- Adam is fast and converges rapidly, when compared to other optimizers

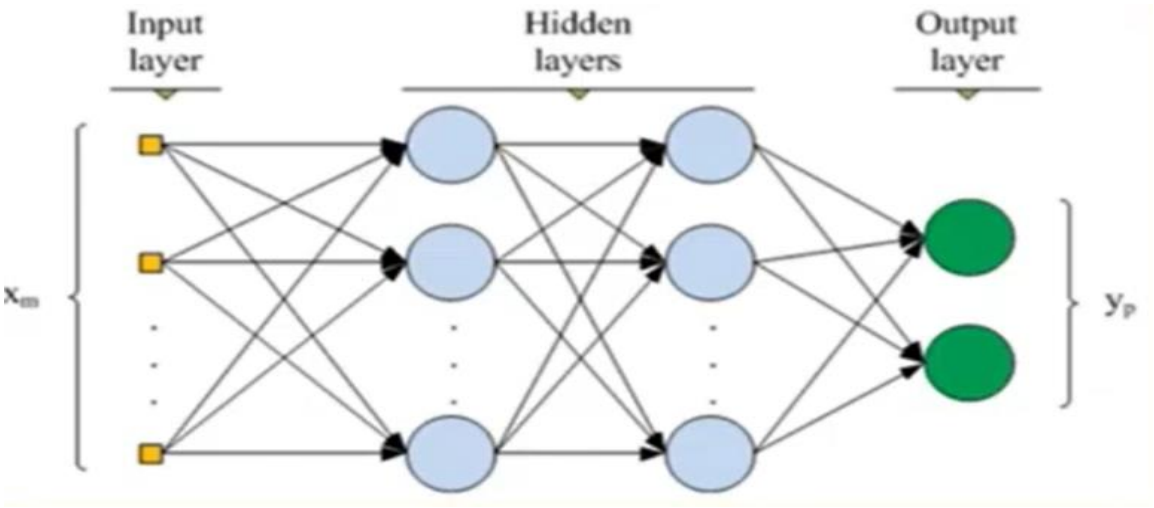

Fig 2: Model Structure

Architecture of the DNN using AutoEncoder	
Number of layers	9
Activation function in hidden layers	ReLU
Activation function in output layers	Sigmoid
Optimizer	Adam
Dropout rate	0.30
Number of epochs	100
Batch size	16

Table 2: Architecture of the DNN using AutoEncoder

11. Dataset

- Name: NSL-KDD
- Advantages of the NSL-KDD dataset:
- No redundant records in train and test datasets

Total number of features in the dataset	41
Number of labels in the train dataset	22
Number of numerical features in the dataset	38
Number of categorical features in the dataset	3

Table 3: Details about the Dataset

Name	Description
KDDTrain+.TXT	Complete NSL-KDD dataset for training, including attack-type labels and difficulty level
KDDTrain+_20Percent.TXT	20% subset of the KDDTrain+.TXT
KDDTest+.TXT	Complete NSL-KDD dataset for testing, including attack-type labels and difficulty level
KDDTest-21.TXT	This is a subset of the KDDTest+.txt file

Table 4: Various files that are available for this dataset and their description

- More details about the training and testing dataset:

Details of the training dataset	Value	
Number of rows and columns	(25191, 42)	
Total number of features	41	
Number of categorical features	3	
Number of numerical features	38	
Are there any missing values	No	
Are there any duplicate records	No	
Number of different values for label	22	
Label distribution	normal	13448
	attack	11743

Table 5: Details of the NSL-KDD training Dataset

Details of the training dataset	Value	
Number of rows and columns	(11850, 42)	
Total number of features	41	
Number of categorical features	3	
Number of numerical features	38	
Are there any missing values	No	
Are there any duplicate records	No	
Number of different values for label	38	
Label distribution	normal	2152
	attack	9698

Table 6: Details of the NSL-KDD testing Dataset

12. Preprocessing

Below is the list of pre-processing steps that we performed:

- Feature selection
- Feature encoding
- Feature scaling

Feature selection:

We performed feature selection to identify the most important features that contribute towards the label, so that we can eliminate the least contributing features, thereby improving the computational efficiency. Below is the list of feature selection techniques we used.

- Chi-Squared Test
- Random Forest Classifier
- Extra Trees Classifier

Feature encoding:

We performed feature encoding to encode the categorical features in the dataset, as most of the machine learning algorithms work well with noncategorical data. Below is the list of feature encoding techniques we used,

- OneHotEncoder
- LabelEncoder
- BinaryEncoder

Feature scaling:

We performed feature scaling to scale features in the dataset, so that all the features are in the same range and that the model that we build is not biased on some of the features just because they have values at a higher range. Below is the list of feature encoding techniques we used,

- Min-Max
- Standardization
- Binarizing.

13. Algorithm

Below is the procedure followed in our project:

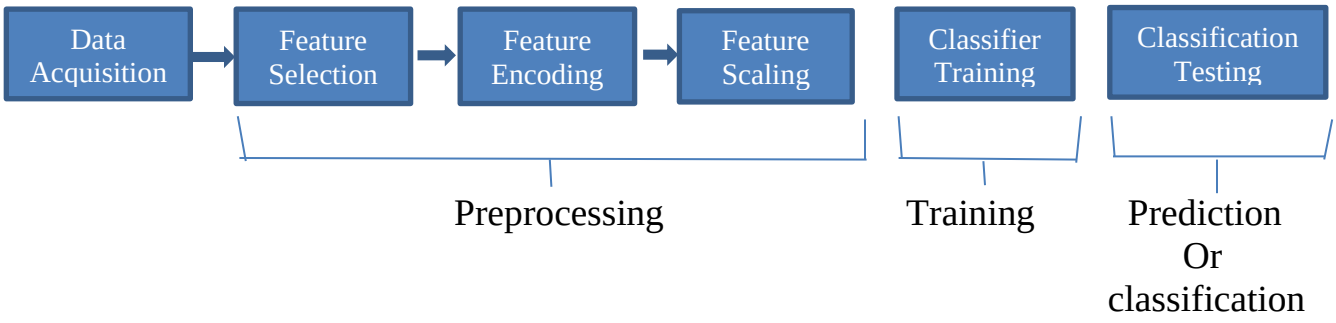

Fig 3: Flow-Diagram of Proposed Algorithm

Machine learning methods: We used the below list of machine learning methods:

Machine learning classifiers	Description
Decision Tree	Decision tree uses the tree representation to solve the problem. In decision tree each leaf node corresponds to a class label and attributes are represented on the internal node of the tree.
Random Forest	Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction.
Extra Trees	Extra tree, is very similar to a Random forest classifier and only differs from it in the manner of construction of the decision trees in the forest.
KNN	KNN is a model that classifies data points based on the points that are most similar.

Table 7: Machine Learning Methods used

- We used these machine learning classification methods and built models for classification of normal type of network request versus network intrusion.

- Accuracies measured for the four ML classification techniques we used:

ML model name	Accuracy	
	Training	Testing
ExtraTreesClassifier_OneHotEncoder_ Standardization_DecisonTree	0.99875	0.8164
ExtraTreesClassifier_OneHotEncoder_ Standardization_RandomForestClassifier	0.99875	0.8210
ExtraTreesClassifier_OneHotEncoder_ Standardization_ExtraTreesClassifier	0.99875	0.8204
ExtraTreesClassifier_OneHotEncoder_ Standardization_KNN	0.99865	0.8164

Table 8: Accuracies for the ML classification techniques

- Accuracies measured for the four ML classification techniques we used:

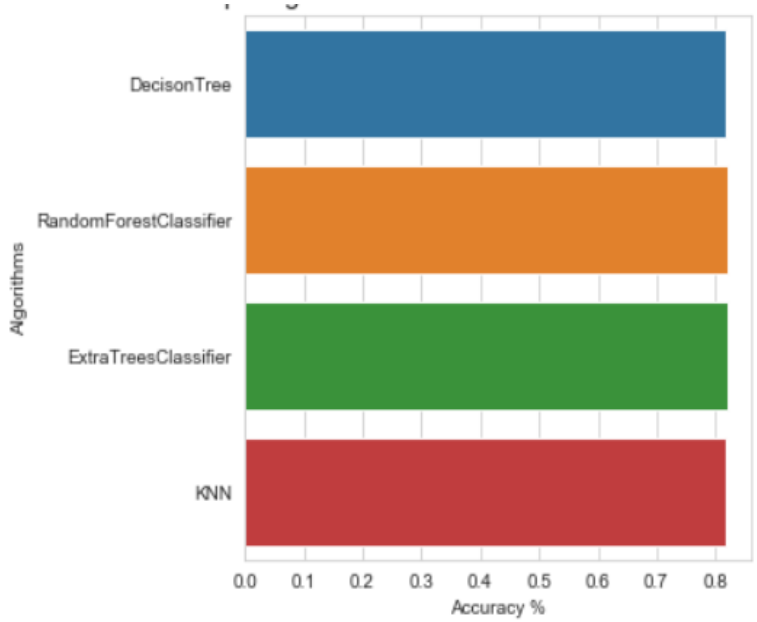

Fig 4: Accuracies for the ML classification techniques

- After building the various machine learning models and measuring their accuracies, we built an Autoencoder to further perform feature selection on the preprocessed dataset. We used 8 layers for the AutoEncoder and used RMSProp as an optimizer function.

Architecture of the AutoEncoder	
Number of layers	8
Activation function in hidden layers	ReLU
Optimizer	RMSprop
Number of epochs	100
Batch size	16

Table 9: Architecture of the AutoEncoder

- As a next step, we built a fully connected deep neural network, using the AutoEncoder in the hidden layers if this fully connection deep neural network.

Architecture of the DNN using AutoEncoder	
Number of layers	9
Activation function in hidden layers	ReLU
Activation function in output layers	Sigmoid
Optimizer	Adam
Dropout rate	0.30
Number of epochs	100
Batch size	16

Table 10: Architecture of the DNN using AutoEncoder

- Accuracies measured for the fully connected deep neural network:

Deep neural network that uses AutoEncoder layers in hidden layers	Accuracy	
	Training	Testing
	0.9860	0.98585

Table 11: Accuracies achieved by DNN that used AutoEncoder

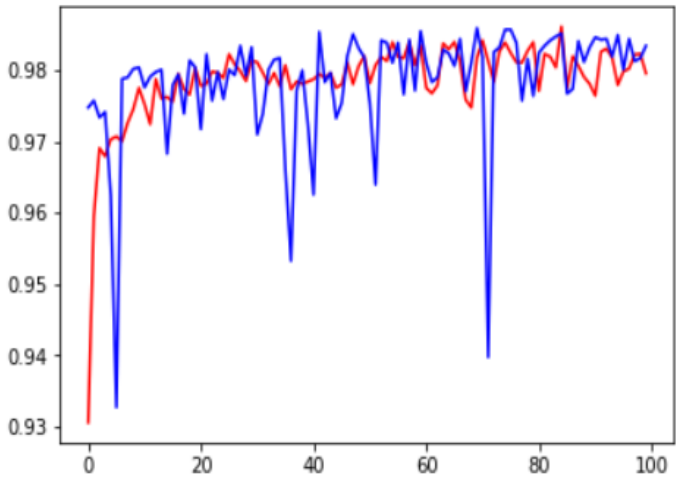

Fig 5: Plot for the training and testing accuracy of the DNN that is using AutoEncoder (Blue; Test accuracy, Red: Train accuracy)

14. Results

- We evaluated the various machine learning techniques and deep learning techniques against the NSL-KDD dataset. Here is the comparison between the accuracies noted from the various machine learning models and the deep learning using layers from AutoEncoder.

- We clearly identified that the accuracies achieved by the deep neural network that is using layers from Autoencoder in its hidden layers, showed better accuracies compared to models built using the classic machine learning algorithms.

- On the training dataset,the accuracies achieved by the classic machine learning classification algorithms is 0.99875, whereas the accuracy achieved by the DNN using Autoencoder is 0.9860.

- On testing dataset ,the accuracies achieved by the classic machine learning classification algorithms is 0.8210, whereas the accuracy achieved by the DNN using Autoencoder is 0.9858.

Accuracies achieved by various models		
	Training	Testing
Machine learning mode: (ExtraTreesClassifier_OneHotEncoder_ Standardization_RandomForestClassifier)	0.99875	0.8210
Deep neural network using AutoEncoder layers as hidden layers	0.9860	0.9858

Table 12: Comparison of accuracies achieved by various models

15. Conclusion

We observed that the model created using Deep neural network using the AutoEncoder layers as hidden layers has shown better results when compared to the results from model created using machine learning algorithms.

We want to further experiment using Generative Adversarial networks, type of neural networks and also perform experiments towards building an intrusion prevention system versus an intrusion detection system.

16. References

1. Chunjie Zhou, Shuang Huang, Naixue Xiong, Senior Member, IEEE, Shuang-Hua Yang, Senior Member, IEEE, Huiyun Li, Yuanqing Qin, and Xuan Li., Design and Analysis of Multimodel-Based Anomaly Intrusion Detection Systems in Industrial Process Automation, IEEE Transactions On Systems, Man, And Cybernetics: Systems. Year: 2015, Volume: PP, Issue: 99 , DOI: 10.1109/TSMC.2015.2415763

2. Al-Jarrah, O. ; Dept. of Comput. Eng., Jordan Univ. of Sci. & Technol., Irbid, Jordan ; Arafat, A. , Network Intrusion Detection System Using Attack Behavior Classification, 5th International Conference on Information and Communication Systems (ICICS), 2014.

3. B. Mukherjee, L. T. Heberlein and K. N. Levitt, "Network intrusion detection", IEEE Netw., vol. 8, pp. 26-41, May 1994.

4. D Larson, "Distributed denial of service attacks–holding back the flood", Netw. Secur., vol. 2016, no. 3, pp. 5-7, 2016.

5. R. C. Staudemeyer, "Applying long short-term memory recurrent neural networks to intrusion detection", South Afr. Comput. J., vol. 56, no. 1, pp. 136-154, 2015.

6. S. Venkatraman and M. Alazab, "Use of data visualisation for zero-day Malware detection", Secur. Commun. Netw., vol. 2018, Dec. 2018