IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Network Intrusion Detection Using Supervised Machine Learning

Marelly Shivaram[1], K . Sravanthi[2]

[1]Pg student , Department of Computer Science and Engineering , Kakatiya University College Of Engineering and Technology Warangal -506009, (TS).

[2]Assistant Professor, Department of Computer Science and Engineering , Kakatiya University College Of Engineering and Technology Warangal -506009, (TS).

Abstract: Contrasted with the past, improvements in PC and correspondence innovations have given broad and propelled changes. The use of new innovations give incredible advantages to people, organizations, and governments, be that as it may, messes some up against them. For instance, the protection of significant data, security of put away information stages, accessibility of information and so forth. Contingent upon these issues, digital fear based oppression is one of the most significant issues in this day and age. Digital fear, which made a great deal of issues people and establishments, has arrived at a level that could undermine open and nation security by different gatherings, for example, criminal association, proficient people and digital activists. Along these lines, Intrusion Detection Systems (IDS) has been created to maintain a strategic distance from digital assaults. Right now, learning the bolster support vector machine (SVM) calculations were utilized to recognize port sweep endeavors dependent on the new CICIDS2017 dataset with 97.80%, 69.79% precision rates were accomplished individually. Rather than SVM we can introduce some other algorithms like random forest, Decision Trees, Logistic Regression where these algorithms can acquire accuracies more accuracy.

INTRODUCTION

IDSs are security solutions that, like antivirus software, firewalls, and access control schemes, are designed to make information and communication systems more secure. IDS arose as a result of the inadequacy of traditional security methods. The following subsections discuss the network security, firewalls and IDSs, respectively. According to Cisco , network security involves any action that is tailored to ensure that there is usefulness and reliable integrity of the user's network and data. This activity incorporates both tangible and intangible innovations to computer systems. Accessing the network is usually under the control of active network security. It can detect and prevent a variety of threats from getting into or proliferating throughout the user's network at any given time. The majority of security threats are purposefully created by malicious people seeking a benefit, gaining publicity, or harming someone. Network security issues can be loosely classified into five interconnected areas, 1. Confidentiality: The contents of the transmitted communication should only be understood by the sender and the intended receiver because the message could be intercepted by eavesdroppers. Encryption is used to accomplish this. 2. Message integrity assures that the delivered message's content isn't tampered with, either intentionally or accidentally. Checksum and hash functions are used to accomplish this. 3. Verification: the party sending, and the one receiving the information, ought to have a way of verifying their identity. Each party should be able to verify the identity of the other. 4. No repudiation deals with the possibility of someone denying sending a message or carrying out an action. It is achieved through digital signatures. 5. Operational security: this is a security process used to prevent important materials of a company or an institution from being accessed by unauthorized individuals. Nearly all institutions, including banks and higher learning institutions, among others, possess or use a network that happens to be linked to the public Internet. At some point, the networks can easily be tampered with without the owner's consent. Malicious people can introduce worms into the network's host, access the institution's confidential documents, change the organization's network configuration, and launch disk operating system attacks. For this reason, firewalls and IDSs are put into use to counter attacks that may arise against a company's network. Networks of companies or institutions are organized into two categories:

internal networks and demilitarized zone . The internal network of the company or institution can only be accessed by the network administrators or the workers within the company. The demilitarized zone (DMZ) can be accessed by anyone. Having a demilitarized zone within any organization plays a very crucial role. It adds an extra layer of security to the company's internal network because the hosts that are the most susceptible to attacks are the ones that provide services to users who are not within the internal network, for instance, electronic mail, website, and domain name system servers. Due to the high number of organizations that are facing attacks, the organizations are placed within a sub network to protect the rest of the network within the organization from receiving attacks. Only the information exposed in the DMZ within an organization can be accessed by an external host. The rest of the organization's network cannot be accessed by any means from an external host. Nevertheless, having a separation of the organization's network while not developing tactics that can control network traffic doesn't make any sense. Consequently, a common mechanism of security is the addition of a firewall. As we saw in the previous section, a packet filter (firewall) inspects packets such as ICMP, TCP, IP, and UDP header fields when determining whether to allow them past the firewall. However, Deep Packet Inspection (DPI) is required to detect many attack types, particularly those that the packet filter cannot detect. A device that not only analyses the headers of all packets traveling through it (unlike a packet filter) but also does deep packet inspections has a place in intrusion prevention. An Intrusion Prevention System acts when a device detects a suspect packet or a suspicious series of packets and drops them to prevent them from accessing the organization's network. Intrusion Detection Systems are used when a device can let packets pass by it on their way to the corporate network but sends an alarm to the network administrator or logs the packets. In this section, we'll look at intrusion detection in further depth. Intrusion Detection Systems are computer based security and defense systems that monitor, identify, and analyze harmful activity on hosts or networks. The purpose of an intrusion detection system is to ensure that the security of a computer system or network based on integrity, confidentiality, and availability is maintained. The Intrusion Detection System, upon detecting that an intrusion has occurred and that the firewall failed to mitigate or stop the attack or intrusion . The firewall is the first protection against intrusion. At the same time, using the Intrusion Detection System is based upon the certainty that an attack will occur that the firewall cannot eliminate or mitigate. The Intrusion Detection System can be classified in different ways, based on the monitored platform or the technique they employ to identify anomalous activity.

PROBLEM STATEMENT:

The task is to build a network intrusion detector, a predictive model capable of distinguishing between bad connections, called intrusions or attacks, and good normal connections. Providing security to the industrial networks using IT solutions may not be a reasonable approach because of the different functionalities that these networks have. Hence, to effectively protect the ICS network from the increasing number of intrusions and reduce their impact, an efficient Intrusion Detection Systems (IDS) which can minimize the effects of the attacks is vital. However, existing IDSs have shown inefficiency in detecting zero-day attacks. They also suffer from false positives (unnecessary alarm) and false negatives (which impact the security), which affect the performance and accuracy of the ICS. When designing an efficient IDS framework, the problem that struggles developers is to intertwine various components to reduce these drawbacks.

EXISTING SYSTEM

Blameless Bayes and Principal Component Analysis (PCA) were been used with the KDD99 dataset by Almansob and Lomte .Similarly, PCA, SVM, and KDD99 were used Chithik and Rabbani for IDS . In Aljawarneh et al's. Paper, their assessment and examinations were conveyed reliant on the NSL-KDD dataset for their IDS model Composing inspects show that KDD99 dataset is continually used for IDS .There are 41 highlights in KDD99 and it was created in 1999. Consequently, KDD99 is old and doesn't give any data about cutting edge new assault types, example, multi day misuses and so forth. In this manner we utilized a cutting-edge and new CICIDS2017 dataset in our investigation.

However, there are several issues with the accessible existing public dataset, like uneven information, or out-of-date content and therefore the likes are similar. These issues have mostly restricted the event of analysis during this explicit space.

PROPOSED SYSTEM

Every approach for implementing an intrusion detection system has its own pros and cons, a degree apparent from the discussion conducted for comparisons among the varied ways. Thus, it's tough to settle on a specific methodology to implement an intrusion detection system over the others. Datasets for network intrusion detection are important resources for coaching and testing systems. The Machine Learning method doesn't work while not the representative information, and getting such a dataset is tough and long. Network info updates in no time that brings to the ML model coaching with larger problem. The Model has to be retrained semi-permanent/long-termed and quickly. Thus progressive learning and long learning are the long run focus within the study of this field within the future.

Important steps of the algorithm are given in below. 1) Normalization of every dataset. 2) Convert that dataset into the testing and training. 3) Form IDS models with the help of using random forest, Decision Trees, Logistic Regression algorithms. 4) Evaluate every model's performance.

IMPLEMENTATION

In this section, the methodology of the research is discussed. According to the literature studies, there is a critical need for the creation of effective machine learning and deep learning models for identifying attacks in datasets. The dataset NSL-KDD was analyzed and trained using four Machine Learning algorithms Random Forest (RF), Decision Tree, Logistics Regression and Support Vector Machine (SVM). The general layout of the methodology.

a)Dataset Collection:

NSL-KDD is a condensed version of the original KDD dataset that was acquired from the Canadian Institute for Cyber security [21]. It has the same features as KDD. Each record has 41 features and one class attribute. Each connection is classified as either an attack or a normal connection. NSL-KDD has a total of 39 attacks, each of which is classified into one of four categories: DOS, R2L, U2R, and Probing. For building our models, we used 25,192 instances as training. Next, these trained models were evaluated and tested using 11851 instances. Finally, the rest of the dataset was used for validation.

DOS: denial-of-service, which means preventing authorized users' access to a service, such as syn flooding.

R2L: This refers to breaking into a remote machine to get access to the victim's machine, such as guessing passwords.

U2R: When a normal account is used to log into a victim machine and tries to gain root privilege, using a technique such as a buffer overflow.

Probing: examining and scanning the victim's machine for vulnerabilities in terms of learning more about it, such as port scanning.

Data pre-processing

Pre-processing the data is a very important step in preparing the data to be fed into the algorithm. The goal of data preparation is to eliminate ambiguity in the dataset and provide IDS with accurate data. It unifies feature selection and normalization. Many symbolic attributes in the dataset, such as flags and protocol types, have nominal values. These values must be converted to numeric values for the dataset to perform better. Multi-class classification problems (4 attack classes and normal classes) and binary classification problems (normal or attack) have been transformed using discredited datasets in bin 10.

b) Feature selection

Feature Selection produces more enhanced and efficient subsets by eliminating redundant and unrelated features. Correlation is a popular and successful strategy for identifying the most closely linked characteristics in any dataset; it defines the strength of the relationship between features, based on the assumption that features are conditionally independent given the class. A good feature subset contains

features that are highly correlated (predictive of) the class yet uncorrelated and not predictive of one another. The table shows the result of CFS Sub Set Eval-Best First was chosen for feature selection used in WEKA.

Split and discretization

The main objective of discretization is to improve the overall classification performance while reducing storage space because discretized data takes up less space. An important step before classification is considered using several classifiers employing discrete data and classifiers using discrete data discretization. Discretization is numeric attributes that were discretized by use of a discretization filter using unsupervised 10 bin discretization on Weka. Also, one of the most important steps for building any machine learning model is splitting the dataset into training and testing modules. In this study, the dataset was split into two, 80% of data for training, 20% for testing, and the rest for validation, which is 1% of the data. Then, we renamed every attack label for binary and multi-classifications as normal traffic or attacks and determined the type of main categories of attacks on the datasets DOS, Probe, R2L, and U2R.

a) Classification process

For the supervised machine learning algorithms used to evaluate the performance of NIDS over the NSL-KDD dataset in this study, we used Support Vector Machines (SVM), Random Forest (RF), Decision Tree and Logistic Regression algorithms for each type of feature selection method. In general, every process of classification in machine learning is divided into five steps:

- **Logistic Regression Algorithm**

  It is a SML model that is very commonly or widely used for the classification. Performance of LR model for linearly separable classes is very well and even easy to implement. Specially, in industry it is most commonly used. In general LR is used for binary classification as it is a linear model but using technique OvR it may be used for classification of multi class [9]. LR is applied on dataset by considering three different train test ratio (80:20, 60:40, and 70:30) to predict whether the bank currency is forge or genuine. For train test ratio 80:20 ROC curve and learning curves are drawn. Accuracy of LR is observed around 98% .

- **Decision Tree Algorithm:**

  It is a classification model having a structure like a tree. DT is incrementally developed by breaking down the data set into smaller subsets. DT results are having two types of nodes Decision nodes and leaf nodes. For an example consider a decision node i.e., Outlook and it have branches as Rainy, Overcast and Sunny representing values of the tested feature. Hours Played i.e., a leaf node it gives the decision on numerical targeted value. DT can handle both numerical as well as categorical data [8]. DT is applied on dataset by considering three different train test ratio (80:20, 60:40, and 70:30) to predict whether the bank currency is forge or genuine. For train test ratio 80:20 ROC curve and learning curves are drawn. Accuracy of DT has been observed around 99%.

- **Random Forest Algorithm**

  Random Forest is that the prevalent supervised technique. it's useful for mainly doing classification challenges and also regression challenges. RF is one amongst the classifiers which holds multiple decision trees in each subset of an assumed data set and computes the everyday value that enhances prediction accurateness for the dataset. The random forest doesn't depend upon decision trees. Instead, it gets a prediction from every tree so forecasts the last result which is made upon polls of prevalence estimations. The more trees within the forest, the upper the accuracy and avoid over fitting problems. it's supported the ensemble technique concept, which mixes multiple classifiers to unravel a thorny problem and improves model performance.

- **Support Vector Machine (SVM)**

  The SVM is already known as the best learning algorithm for binary classification. The SVM, originally a type of pattern classifier based on a statistical learning technique for classification and regression with a variety of kernel functions, has been successfully applied to a number of pattern recognition applications. Recently, it has also been applied to information security for intrusion detection. Support Vector Machine has become one of the popular techniques for anomaly intrusion detection due to their good generalization nature and the ability to overcome the curse of dimensionality .Another positive aspect of SVM is that it is useful for finding a global minimum of the actual risk using structural risk minimization, since it can generalize well with kernel tricks even in high-dimensional spaces under little training sample conditions. The SVM can select appropriate setup parameters because it does not depend on traditional empirical risk such as neural networks. One of the main advantage of using SVM for IDS is its speed, as the capability of detecting intrusions in real-time is very important. SVMs can learn a larger set of patterns and be able to scale better, because the classification complexity

does not depend on the dimensionality of the feature space. SVMs also have the ability to update the training patterns dynamically whenever there is a new pattern during classification.

b) Evaluation metrics

The evaluation of the produced classification models is an important phase. It's also done through the use of a variety of evaluation metrics. The following are used on evaluation metrics: • True Positives (TP) the total number of malicious packets correctly classified. • True Negatives (TN) the total number of correctly classified as normal. • False Positives (FP) the total number of malicious packets incorrectly classified as attacks. • False Negatives (FN) the total number of malicious packets incorrectly classified as normal. Classification accuracy is the most commonly used statistic for evaluating a model, however, it is not a reliable predictor of its performance.

Accuracy:-The appropriate classification ratio is the proportion of correctly classified samples to the total number of input samples. It is calculated using the following formula:

Accuracy = (TP+TN)/(TP+FP+FN+TN)

Precision: It's the number of successfully classified positive samples divided by the number of samples that the classifier predicted as positive (i.e. the proportion of positive samples correctly classified to the all predicted as positive). Its formula is as follows:

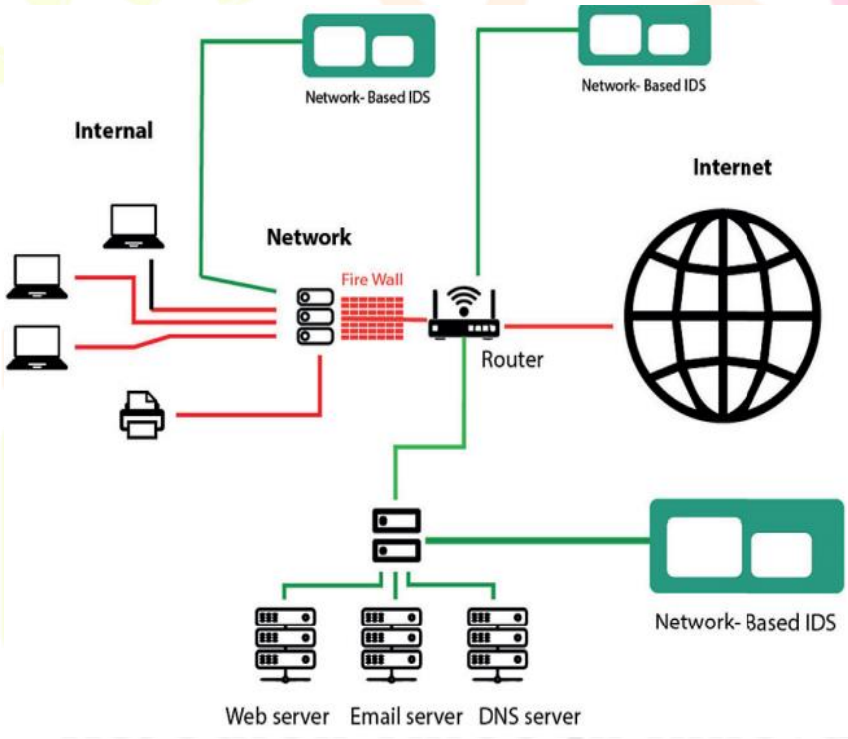
Precision = TP/(TP+FP)

Recall: It is calculated by dividing the number of correctly classified positive samples by the total number of positive samples passed.

Recall = TP/(TP+FN) (

Mathews Correlation Coefficient (MCC): It represents the relative correlation between observed and predicted binary classifications.

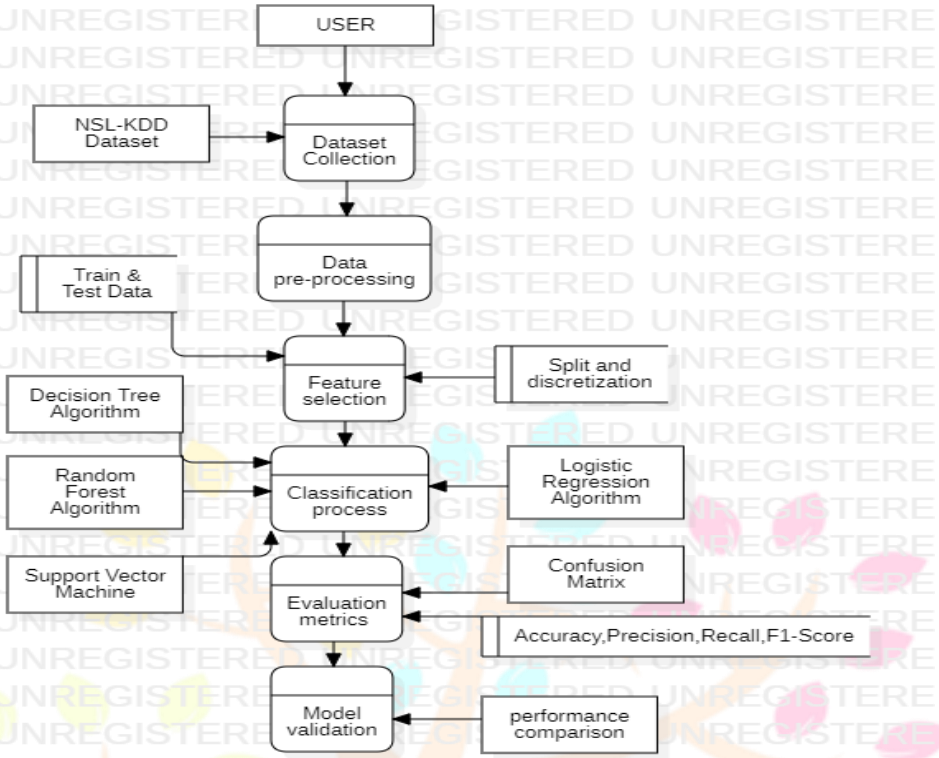
MCC = (TP*TN – FP*FN). / / sqrt[(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)]



Model validation

In the final step, the model will be implemented and trained based on the decisions made in the previous processes, and then validated to see if it meets all of the preconditions and to see how accurate it is at predicting with new data. The model's flaws and limitations are recognized as a result of these assessments, allowing the required measures to be taken to address them. In comparison to other algorithms, the experiment shows that RF has the highest accuracy, followed by the ML algorithm shows that selecting 13 features for each algorithm provide high accuracy in the binary class. The models have a closer accuracy of 98.92% and F-measure 98.9%, respectively. It shows that the model is the best for detecting DOS attacks. It shows the same results for multi-class, with slight changes in accuracy, which was high in model RF.

SYSTEM DESIGN AND ARCHITECTURE

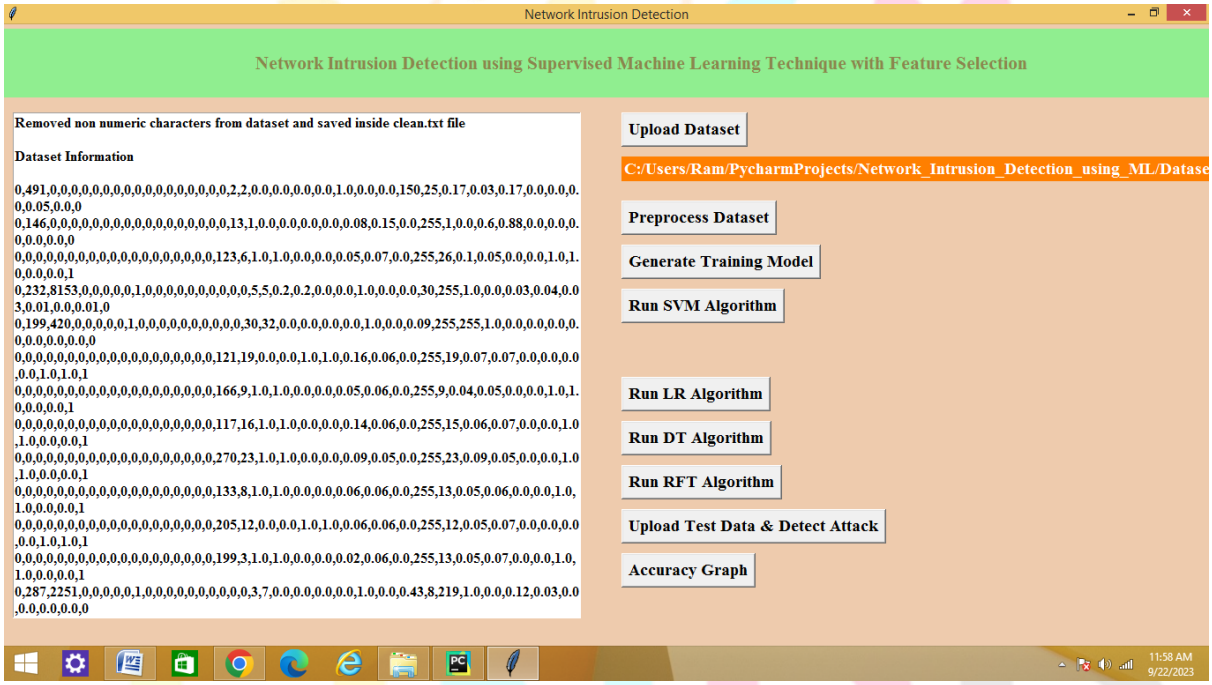


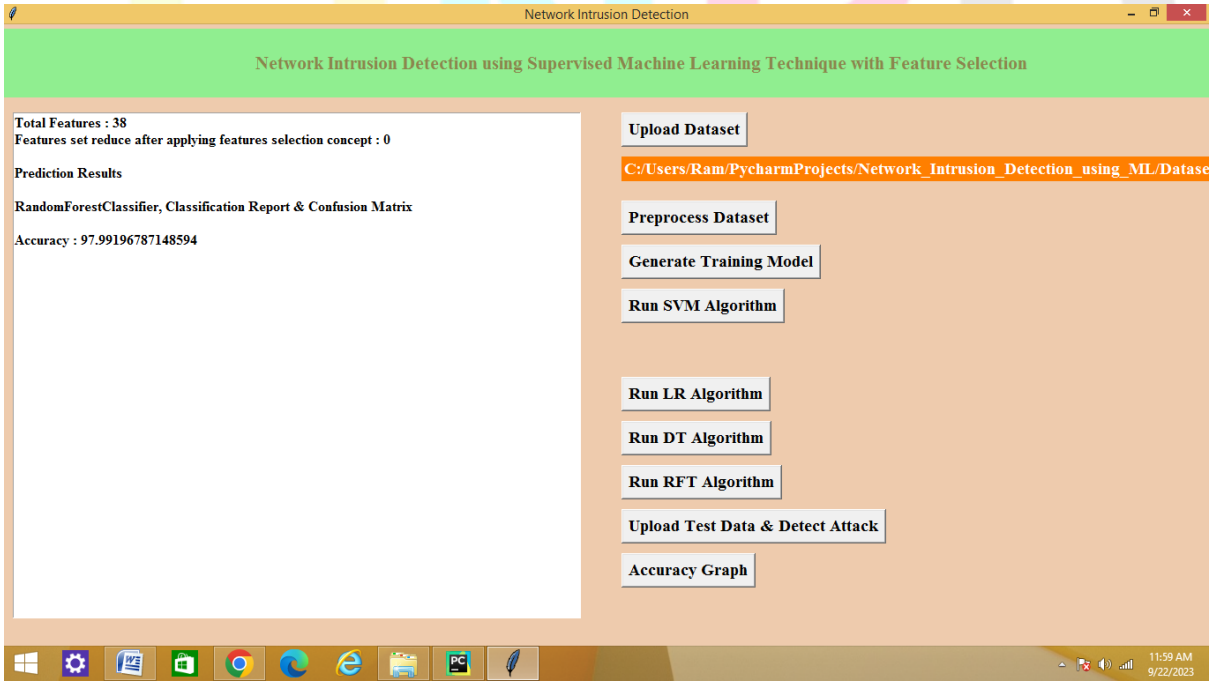
DATASET EXAMPLE:

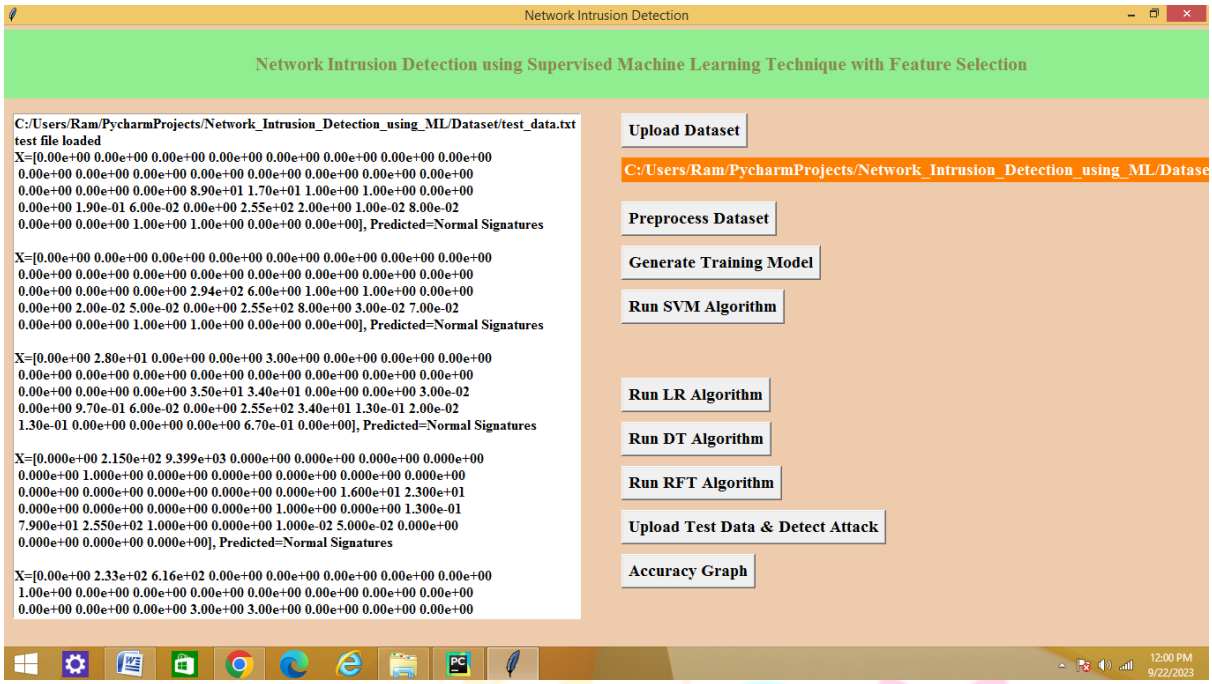
duration,protocol_type,service,flag,src_bytes,dst_bytes,land,wrong_fragment,urgent,hot,num_failed_logins,logged_in,num_compromised,root_shell,su_attempted,num_root,num_file_creations,num_shells,num_access_files,num_outbound_cmds,is_host_login,is_guest_login,count,srv_count,serror_rate,srv_serror_rate,rerror_rate,srv_rerror_rate,same_srv_rate,diff_srv_rate,srv_diff_host_rate,dst_host_count,dst_host_srv_count,dst_host_same_srv_rate,dst_host_diff_srv_rate,dst_host_same_src_port_rate,dst_host_srv_diff_host_rate,dst_host_serror_rate,dst_host_srv_serror_rate,dst_host_rerror_rate,dst_host_srv_rerror_rate,label

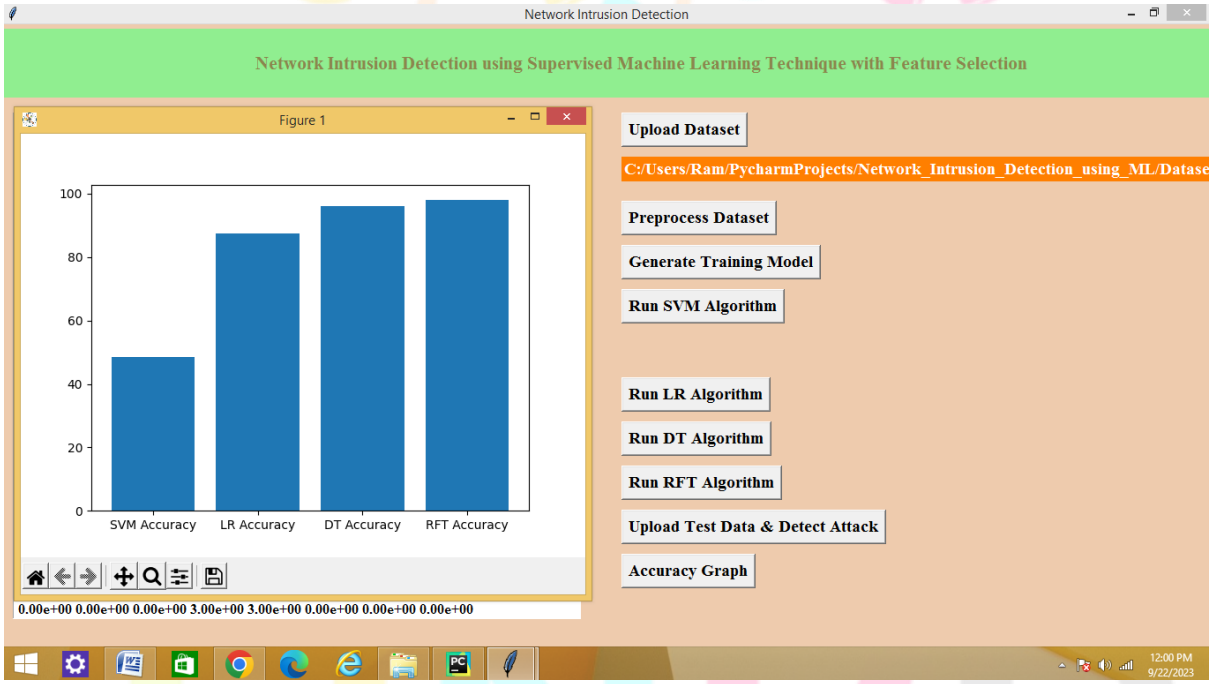
RESULTS


Network Intrusion Detection
Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection
Dataset loaded
Upload Dataset
C:/Users/Ram/PycharmProjects/Network_Intrusion_Detection_using_ML/Datase
Preprocess Dataset
Generate Training Model
Run SVM Algorithm
Run LR Algorithm
Run DT Algorithm
Run RFT Algorithm
Upload Test Data & Detect Attack
Accuracy Graph
11:58 AM
9/22/2023


Network Intrusion Detection
Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection
Removed non numeric characters from dataset and saved inside clean.txt file
Dataset Information
0,491,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,0.0,0.0,0.0,0.0,1.0,0.0,0.0,150,25,0.17,0.03,0.17,0.0,0.0,0.0,0.0,0.05,0.0,0
0,146,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,13,1,0.0,0.0,0.0,0.0,0.08,0.15,0.0,255,1,0.0,0.6,0.88,0.0,0.0,0.0,0.0,0.0,0,0
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,123,6,1.0,1.0,0.0,0.0,0.05,0.07,0.0,255,26,0.1,0.05,0.0,0.0,1.0,1.0,0.0,0.0,1
0,232,8153,0,0,0,0,1,0,0,0,0,0,0,0,0,0,5,5,0.2,0.2,0.0,0.0,1.0,0.0,0.0,30,255,1.0,0.0,0.0,0.03,0.04,0.03,0.01,0.0,0.01,0
0,199,420,0,0,0,0,1,0,0,0,0,0,0,0,0,0,30,32,0.0,0.0,0.0,0.0,1.0,0.0,0.0,09,255,255,1.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0,0
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,121,19,0.0,0.0,1.0,1.0,0.16,0.06,0.0,255,19,0.07,0.07,0.0,0.0,0.0,0.0,1.0,1.0,1
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,166,9,1.0,1.0,0.0,0.0,0.05,0.06,0.0,255,9,0.04,0.05,0.0,0.0,1.0,1.0,0.0,0.0,1
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,117,16,1.0,1.0,0.0,0.0,0.14,0.06,0.0,255,15,0.06,0.07,0.0,0.0,1.0,1.0,0.0,0.0,1
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,270,23,1.0,1.0,0.0,0.0,0.09,0.05,0.0,255,23,0.09,0.05,0.0,0.0,1.0,1.0,0.0,0.0,1
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,133,8,1.0,1.0,0.0,0.0,0.06,0.06,0.0,255,13,0.05,0.06,0.0,0.0,1.0,1.0,0.0,0.0,1
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,205,12,0.0,0.0,1.0,1.0,0.06,0.06,0.0,255,12,0.05,0.07,0.0,0.0,0.0,0.0,1.0,1.0,1
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,199,3,1.0,1.0,0.0,0.0,0.02,0.06,0.0,255,13,0.05,0.07,0.0,0.0,1.0,1.0,0.0,0.0,1
0,287,2251,0,0,0,0,1,0,0,0,0,0,0,0,0,0,3,7,0.0,0.0,0.0,0.0,0.1,0.0,0.0,0.43,8,219,1.0,0.0,0.0,0.12,0.03,0.0,0.0,0.0,0.0,0
Upload Dataset
C:/Users/Ram/PycharmProjects/Network_Intrusion_Detection_using_ML/Datase
Preprocess Dataset
Generate Training Model
Run SVM Algorithm
Run LR Algorithm
Run DT Algorithm
Run RFT Algorithm
Upload Test Data & Detect Attack
Accuracy Graph
11:58 AM
9/22/2023


Network Intrusion Detection
Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection
Total Features : 38
Features set reduce after applying features selection concept : 0
Prediction Results
RandomForestClassifier, Classification Report & Confusion Matrix
Accuracy : 97.99196787148594
Upload Dataset
C:/Users/Ram/PycharmProjects/Network_Intrusion_Detection_using_ML/Datase
Preprocess Dataset
Generate Training Model
Run SVM Algorithm
Run LR Algorithm
Run DT Algorithm
Run RFT Algorithm
Upload Test Data & Detect Attack
Accuracy Graph
11:59 AM
9/22/2023





REFERENCES

1. Cisco Annual Internet Report (2018–2023) White Paper. (2022, January 23). Cisco. https://www. cisco.com/c/en/us/solutions/collateral/executiveperspectives/annual-internet-report/white-paperc11–741490.html

2. Dyn Analysis Summary of Friday October 21 Attack (2022, February 20). https://web.archive.org/ web/20200620203923

3. Dartigue, C., Jang, H.I., Zeng, W. A new data-mining based approach for network intrusion detection. In Seventh Annual Communication Networks and Services Research Conference. 2009; 372–377.

4. García-Teodoro, P., Díaz-Verdejo, J., MaciáFernández, G., Vázquez, E. Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers & Security. 2009; 28(1–2); 18–28.

5. Cisco. What Is Network Security? (2022, February,8). Cisco. https://www.cisco.com/c/en/us/products/security/what-is-network-security.html

6. Kurose, J.F., Ross, K.W. Computer Networking: A Top-Down Approach (6th Edition). Pearson, 2012.

7. Tanenbaum, A., Wetherall, D. Computer Networks (5th Edition). Pearson, 2010.

8. Fernandes, G., Rodrigues, J.J.P.C., Carvalho, L.F., Al-Muhtadi, J.F., Proença, M.L. A comprehensive survey on network anomaly detection. Telecommunication Systems. 2018; 70(3): 447–489.

9. Othman, S.M. Alsohybe, N.T., Ba-Alwi, F.M., Zahary, A.T. Survey on intrusion detection system types. 2018; 7(4): 444–463.

10. Pal Singh, A., Deep Singh, M. Analysis of HostBased and Network-Based Intrusion Detection System. International Journal of Computer Network and Information Security, 2014; 6(8): 41–47.

11. Ferrag, M.A. Maglaras, L. Moschoyiannis, S., Janicke, H. Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. Journal of Information Security and Applications. 2020; 50.

12. Boutaba, R. Salahuddin, M.A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., Caicedo, O.M. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. Journal of Internet Services and Applications. 2018; 9(1).

13. Buczak, A.L., Guven, E.A Survey of Data Mining and Machine Learning Methods for Cyber Se curity Intrusion Detection. IEEE Communications Surveys & Tutorials. 2016; 18(2): 1153–1176.

14. Chaabouni, N., Mosbah, M., Zemmari, A., Sauvignac, C., Faruki, P. Network Intrusion Detection for IoT Security Based on Learning Techniques. IEEE Communications Surveys & Tutorials. 2019; 21(3): 2671–2701.

15. Berman, D., Buczak, A., Chavis, J., Corbett, C. A Survey of Deep Learning Methods for Cyber Security. Information. 2019; 10(4): 122.

16. Mahdavifar, S., Ghorbani, A.A. Application of deep learning to cybersecurity: A survey. Neurocomputing. 2019; 347: 149–176.

17. Ahmed, M., Naser Mahmood, A., Hu, J. A survey of network anomaly detection techniques. Journal of Network and Computer Applications. 2016; 60: 19–31.

18. Ring, M., Wunderlich, S., Scheuring, D., Landes, D., Hotho, A. A survey of network-based intrusion detection data sets. Computers & Security. 2019; 86: 147–167.

19. Bhuyan, M.H., Bhattacharyya, D.K., Kalita, J.K. Network Anomaly Detection: Methods, Systems and Tools. IEEE Communications Surveys & Tutorials. 2014; 16(1): 303–336.

20. Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H., Wang, C. Machine Learning and Deep Learning Methods for Cybersecurity. IEEE Access. 2018; 6: 35365–35381.

21. UNB (2021, Novamber 15). https://www.unb.ca/ cic/datasets/nsl.html

22. Chumachenko, K. Machine learning methods for malware detection and classification., 2017.

23. Zou, J., Han, Y., So, S.S. Overview of artificial neural networks. Methods in molecular biology (Clifton, N.J.). 2008; 458: 15–23.

24. Dong, B., Wang, X. Comparison deep learning method to traditional methods using for network intrusion detection. In 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN). 2016; 581–585.

25. Mahesh, B. Machine Learning Algorithms – A Review. International Journal of Science and Research (IJSR). 2020; 381–386.

26. Farnaaz, N., Jabbar, M.A. Random forest modeling for network intrusion detection system. Procedia Computer Science. 2016; 89: 213–217.

27. Bhumgara, A., Pitale, A. Detection of Network Intrusions using Hybrid Intelligent Systems. 1st International Conference on Advances in Information Technology (ICAIT). 2019; 500–506.

28. Kumar, K., Batth, J.S. Network Intrusion Detection with Feature Selection Techniques using Machine-Learning Algorithms. International Journal of Computer Applications. 2016; 150(12): 1–13.

29. Dhanabal, L., Shantharajah, S.P. A study on NSLKDD dataset for intrusion detection system based on classification algorithms. International journal of advanced research in computer and communication engineering. 2015; 4(6): 446–452.