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

Machine Learning Techniques for Detecting Cyber Attacks

Ms.M.ANITHA[1], Mr. EJJIVARAPU NAGARAJU [2], Mr. S. Sai Ram [3]

#1 Assistant professor in the Master of Computer Applications in the SRK Institute of Technology, Enikepadu, Vijayawada, NTR District

#2 Assistant professor in the Master of Computer Applications SRK Institute of Technology, Enikepadu, Vijayawada, NTR District

#3 MCA student in the Master of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District.

Abstract_ In contrast to the past, advancements in computer and correspondence technology have brought about significant and rapid changes. Although some people, organizations, and governments benefit greatly from the application of new innovations, others work against them. For instance, the safety of important data, the security of stored information platforms, information accessibility, and so on. Dependent upon these issues, computerized dread based persecution is quite possibly of the main issue nowadays. Digital apprehension, which has caused numerous problems for individuals and organizations, has reached a point where other groups, such as criminal organizations, knowledgeable individuals, and digital activists, could compromise open and national security. Thusly, Interruption Location Frameworks (IDS) has been made to avoid advanced attacks.

Keywords: data security, accessibility of information, digital fear, Intrusion Detection Systems.

1.INTRODUCTION

Today, political and commercial entities are increasingly engaging in sophisticated cyberwarfare to damage, disrupt, or censor information content in computer networks. In designing network protocols, there is a need to ensure reliability against intrusions of powerful attackers that can even control a fraction of parties in the network. The controlled parties can launch both passive (e.g., eavesdropping, nonparticipation) and active attacks (e.g., jamming, message dropping, corruption, and forging). Intrusion detection is the process of dynamically monitoring events occurring in a computer system or network, analysing them for signs of possible incidents and often interdicting the unauthorized access. This is typically accomplished by automatically collecting information from a variety of systems and network sources, and then analysing the information for possible security problems. Traditional intrusion detection and prevention techniques, like firewalls, access control mechanisms, and encryptions, have several limitations in fully protecting networks and systems from increasingly sophisticated attacks like denial of service. Moreover, most systems built based on such techniques suffer from high false positive and false negative detection rates and the lack of continuously adapting to changing malicious behaviours. In the past decade, however, several Machine Learning (ML) techniques have been applied to the problem of intrusion detection with the hope of improving detection rates and adaptability. These techniques are often used to keep the attack knowledge bases up-to-date and comprehensive. In recent days, cyber-security and protection against numerous cyber-attacks are becoming a burning question. The main reason behind that is the tremendous growth of computer networks and the vast number of relevant applications used by individuals or groups for either personal or commercial use, especially after the acceptance of the Internet of Things (IoT). The cyber-attacks cause severe damage and severe financial losses in large-scale networks. The existing solutions like hardware and software firewalls, user's authentication, and data encryption methods are not sufficient to meet the challenge of upcoming demand, and unfortunately, not able to protect the computer network's several cyber-threats. These conventional security structures are not sufficient as safeguard due to the faster rigorous evolution of intrusion systems. Firewall only controls every access from network to network, which means prevent access between networks. But it does not provide any signal in case of an internal attack. So, it is obvious to develop accurate defense techniques such as machine learning-based intrusion detection system (IDS) for the system's security In general, an intrusion detection system (IDS) is a system or software that detects infectious activities and violations of policy in a network or system. An IDS identifies the inconsistencies and abnormal behavior on a network during the functioning of daily activities in a network or system used to detect risks or attacks related to network security, like denial-ofservice (Dos). An intrusion detection system also helps to locate, decide, and control unauthorized system behaviour such as unauthorized access, or modification and destruction. There are different types of intrusion detection systems based on the user perspective. For instance, they are hostbased and network-based IDS

2.LITERATURE SURVEY

2.1 R. Christopher, "Port scanning methods and the protection towards them," SANS Institute, 2001.

Port Scanning is one of the most famous strategies attackers use to find out offerings that they can take advantage of to smash into systems. All structures that are linked to a LAN or the Internet by means of a modem run offerings that hear to established and now not so time-honored ports. By port scanning, the attacker can locate the following records about the focused systems: what offerings are running, what customers very own these services, whether or not nameless logins are supported, and whether or not sure community offerings require authentication. Port scanning is carried out via sending a message to every port, one at a time. The variety of response obtained suggests whether or not the port is used and can be probed for similarly weaknesses. Port scanners are necessary to community safety technicians due to the fact they can expose viable protection vulnerabilities on the centered system. Just as port scans can be ran in opposition to your systems, port scans can be detected and the quantity of data about open offerings can be confined utilising the applicable tools. Every publicly handy device has ports that are open and on hand for use. The object is to restrict the publicity of open ports to approved customers and to deny get admission to to the closed ports.

2.2 M. C. Raja and M. M. A. Rabbani, "Combined evaluation of guide vector computer and precept element

evaluation for ids," in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.

Compared to the previous protection of networked structures has grow to be a integral normal trouble that influences individuals, firms and governments. The fee of assaults towards networked structures has improved melodramatically, and the techniques used through the attackers are persevering with to evolve. For example, the privateness of vital information, safety of saved records platforms, availability of information etc. Depending on these problems, cyber terrorism is one of the most necessary problems in today's world. Cyber terror, which triggered a lot of issues to men and women and institutions, has reached a degree that ought to threaten public and united states of america safety via a range of agencies such as crook organizations, expert individuals and cyber activists. Intrusion detection is one of the options towards these attacks. A free and high quality strategy for designing Intrusion Detection Systems (IDS) is Machine Learning. In this study, deep getting to know and help vector laptop (SVM) algorithms have been used to observe port scan tries based totally on the new CICIDS2017 dataset Introduction Network Intrusion Detection System (IDS) is a software-based software or a hardware gadget that is used to perceive malicious conduct in the community [1,2]. Based on the detection technique, intrusion detection is categorized into anomaly-based and signature-based.

2.3 S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection machine thru function choice evaluation and constructing hybrid environment friendly model," Journal of Computational Science, vol. 25, pp. 152–160, 2018.

community security, intrusion detection performs an essential role. Feature subsets acquired with the aid of one-of-a-kind function resolution techniques will lead to unique accuracy of intrusion detection. Using man or woman characteristic resolution technique can be unstable in special intrusion detection scenarios. In this paper, the concept of ensemble is utilized to characteristic determination to modify characteristic subsets. Feature choice is transformed into a two-category problem, and strange variety of characteristic determination techniques is used for balloting technique to figure out whether or not a characteristic is required or discarded. In real operation, imply limit impurity, random wooded area classifier, balance selection, recursive function removing and chi-square take a look at are used. Feature subsets got from them will be adjusted via our proposed approach to get ensemble function subsets. To take a look at the performance, help vector machine, selection tree, knn and multi-layer understanding are used to study and evaluate the classification accuracy with ensemble characteristic subsets. Three intrusion detection records sets, which include kddcup99, cidds-001 and unsw_nb15 are used in our experiments. The pleasant end result is mirrored on cidds-001 with a 99.40% classification accuracy.

3.PROPOSED SYSTEM

The training and detection of a cyberattack can be accomplished with the help of machine learning algorithms. An email notification can be sent to users or security engineers as soon as the attack is detected. It is possible to use any classification algorithm to determine whether it is a DoS/DDoS attack or not. Support Vector Machine (SVM), a supervised learning method that analyzes data and recognizes patterns, is an example of a classification algorithm. At the same time, we use decision tree and random forest algorithms for effective detection. Since we cannot control when, where, or how an attack may occur, and absolute prevention against these cannot yet be guaranteed, our best shot for the time being is early detection, which will help mitigate the risk of such incidents causing irreparable damage. Associations can utilize existing arrangements or fabricate their own to identify digital assaults at a beginning phase to limit the effect. Any framework that requires insignificant human mediation would be great

3.1 IMPLEMETATION

3.1.1 Gathering the datasets: We gather all the r data from the kaggale website and upload to the proposed model

3.1.2 Generate Train & Test Model: We have to preprocess the gathered data and then we have to split the data into two parts training data with 80% and test data with 20%

3.1.3 Run Algorithms: For prediction apply the machine learning models on the dataset by splitting the datasets in to 70 to 80 % of training with these models and 30 t0 20 % of testing for predicting

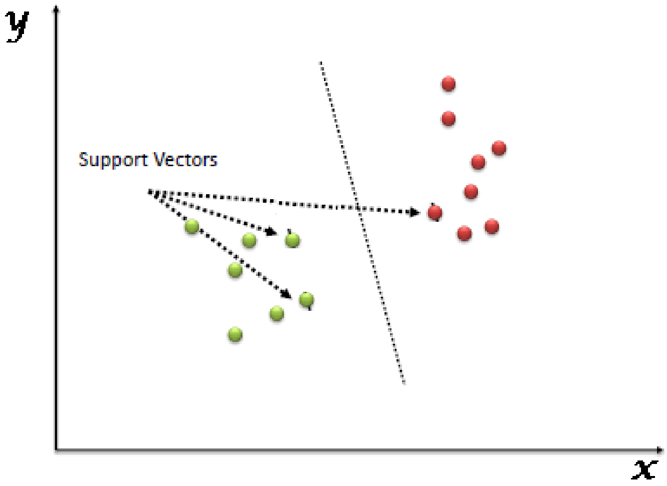
3.1.4 Obtain the accuracy: In this module we will get accuracies

3.1.5 Predict output: in this module we will output in graph

3.2 About Algorithms

3.2.1 **Support Vector Machine**

"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).

3.2.2 Decision Tree Classification Algorithm

○ o Decision Tree is a method of supervised learning that can be used to solve classification and regression problems, but it is typically used to solve classification problems. It is a classifier with a tree structure, with internal nodes representing a dataset's features, branches representing the decision rules, and each leaf node representing the result.

○ o The Decision Node and the Leaf Node are the two nodes in a Decision tree. Leaf nodes are the results of decisions and do not contain any additional branches, whereas Decision nodes are used to make any decision and have multiple branches.

○ o The features of the given dataset are used to make decisions or conduct tests.

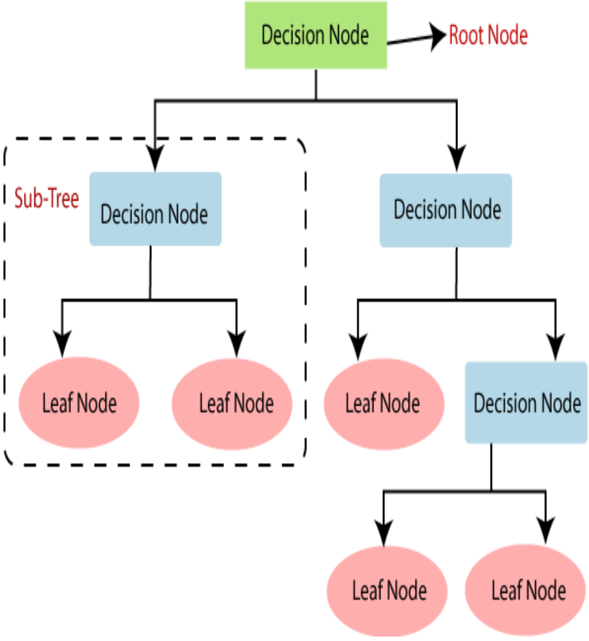
○ o It is a graphical representation that is used to get all of the possible solutions to a problem or decision based on the conditions that have been given.

○ o It is referred to as a decision tree because, like a tree, it begins at the root node and grows into a structure similar to a tree.

○ o The CART algorithm, which stands for Classification and Regression Tree algorithm, is used to construct a tree.

○ o A decision tree simply asks a question and divides the tree into subtrees based on the answer (yes/no).

○ o Below graph makes sense of the general design of a choice tree:



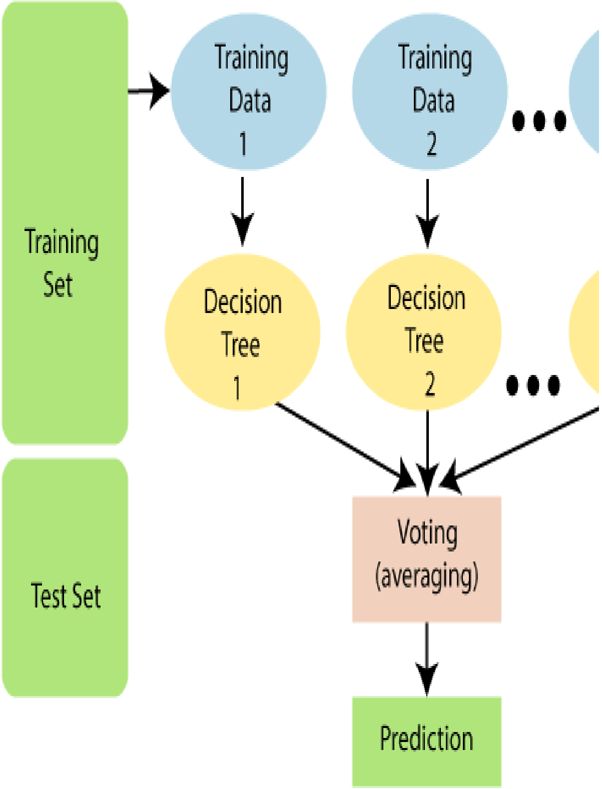
3.2.3 Random Forest

The supervised learning method includes the well-known random forest machine learning algorithm. In ML, it can be utilized for both regression and classification issues. It is based on the idea of ensemble learning, in which multiple classifiers are combined to solve a complex problem and boost the model's performance.

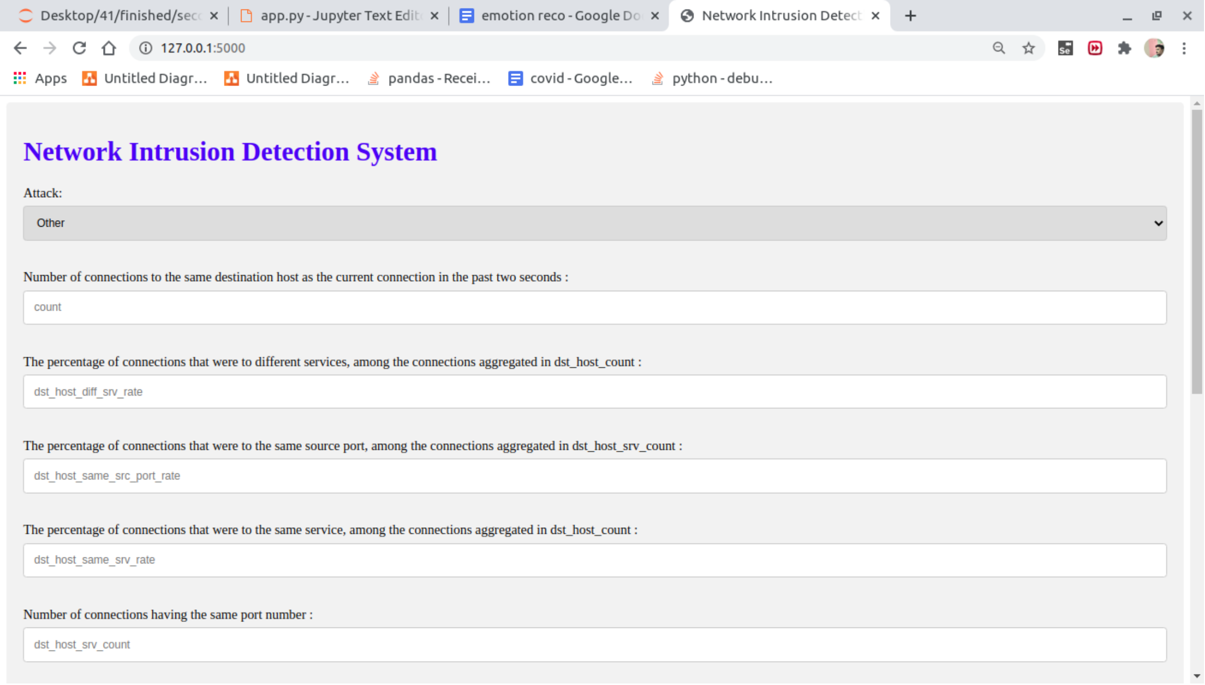
"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset," as the name suggests, "Random Forest is a classifier." Rather than depending on one choice tree, the irregular timberland takes the forecast from each tree and in light of the greater part votes of expectations, and it predicts the last result.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

The below diagram explains the working of the Random Forest algorithm:



4.RESULTS AND DISCUSSIONS



Network Intrusion Detection System

Attack:

Other

Number of connections to the same destination host as the current connection in the past two seconds :

count

The percentage of connections that were to different services, among the connections aggregated in dst_host_count :

dst_host_diff_srv_rate

The percentage of connections that were to the same source port, among the connections aggregated in dst_host_srv_count :

dst_host_same_src_port_rate

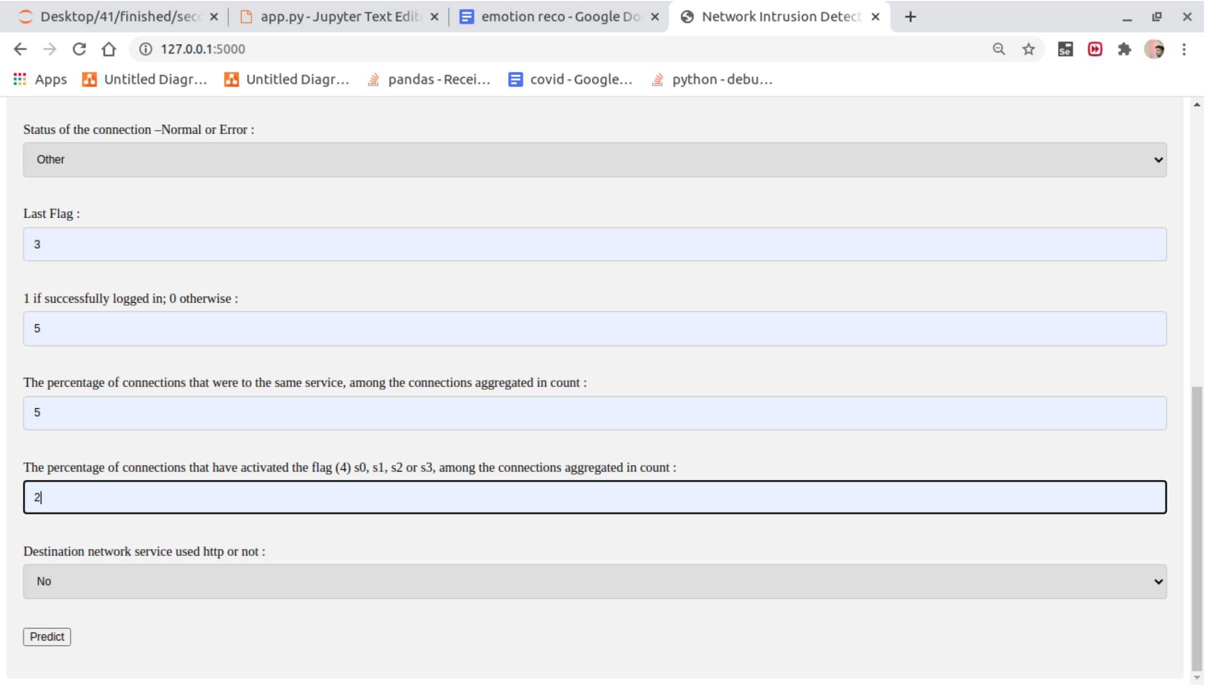
The percentage of connections that were to the same service, among the connections aggregated in dst_host_count :

dst_host_same_srv_rate

Number of connections having the same port number :

dst_host_srv_count

Fig 1:Home Screen



Status of the connection –Normal or Error : Other

Last Flag : 3

1 if successfully logged in; 0 otherwise : 5

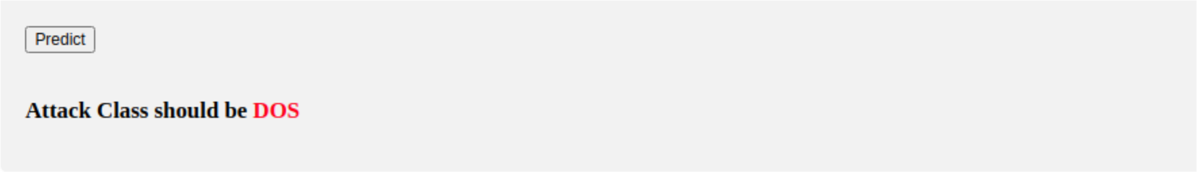
The percentage of connections that were to the same service, among the connections aggregated in count : 5

The percentage of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in count : 2

Destination network service used http or not : No

Fig 2:Enter the input

Predict attack -



Attack Class should be DOS

Fig 3:Predict output

5.CONCLUSION

Currently, evaluations of assistance with supporting vector machine, Choice tree, Arbitrary Woods, and massive learning assessments based on current dataset have been introduced gradually. Significant learning estimation beat SVM, decision tree, and RF in general, according to the data. Port scope efforts, like other forms of attacks, will be used with AI and extensive learning computations. Each of these estimates assists us in locating the digital attack in the network. It happens in the same way that there may have been numerous assaults in the past. When these assaults are detected, the location of the assaults will be recorded in some datasets. Therefore, by utilizing these datasets, we will anticipate whether the digital assault has ended. These forecasts should be possible using four different calculations, such as Decision tree and random forest. This paper helps identify which calculation predicts the best precision rates, which helps predict the best outcomes to determine whether or not digital attacks occurred.

FUTURE SCOPE

In enhancement we will add some ML Algorithms to increase accuracy

REFERENCES

[1] K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.

[2] R. Christopher, “Port scanning techniques and the defense against them,” SANS Institute, 2001.

[3] M. Baykara, R. Das¸, and I. Karado ˘gan, “Bilgi g ¨uvenli ˘gi sistemlerinde kullanilan arac¸larin incelenmesi,” in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.

[4] S. Staniford, J. A. Hoagland, and J. M. McAlerney, “Practical automated detection of stealthy portscans,” Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

[5] S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, “Surveillance detection in high bandwidth environments,” in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1. IEEE, 2003, pp. 130–138.

[6] K. Ibrahimi and M. Ouaddane, “Management of intrusion detection systems based-kdd99: Analysis with lda and pca,” in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1–6.

[7] N. Moustafa and J. Slay, “The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems,” in Building Analysis Datasets and Gathering

Experience Returns for Security (BADGERS), 2015 4th International Workshop on. IEEE, 2015, pp. 25–31.

[8] L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, “Detection and classification of malicious patterns in network traffic using benford’s law,” in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.

[9] S. M. Almansob and S. S. Lomte, “Addressing challenges for intrusion detection system using naive bayes and pca algorithm,” in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.

[10] M. C. Raja and M. M. A. Rabbani, “Combined analysis of support vector machine and principle component analysis for ids,” in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.

[11] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, “Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model,” Journal of Computational Science, vol. 25, pp. 152–160, 2018.

[12] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward generating a new intrusion detection dataset and intrusion

traffic characterization." in ICISSP, 2018, pp. 108–116.

[13] D. Aksu, S. Ustebay, M. A. Aydin, and T. Atmaca, "Intrusion detection with comparative analysis of supervised learning techniques and fisher score feature selection algorithm," in International Symposium on Computer and Information Sciences. Springer, 2018, pp. 141–149.

[14] N. Marir, H. Wang, G. Feng, B. Li, and M. Jia, "Distributed abnormal behavior detection approach based on deep belief network and ensemble svm using spark," IEEE Access, 2018.

[15] P. A. A. Resende and A. C. Drummond, "Adaptive anomaly-based intrusion detection system using genetic algorithm and profiling," Security and Privacy, vol. 1, no. 4, p. e36, 2018.

[16] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[17] R. Shouval, O. Bondi, H. Mishan, A. Shimoni, R. Unger, and A. Nagler, "Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in sct," Bone marrow transplantation, vol. 49, no. 3, p. 332, 2014.

Author's Profiles

AUTHOR PROFILES

Ms.M.ANITHA completed her Master of Computer Applications and Masters of Technology. Currently working as an Assistant professor in the Department of Masters of Computer Applications in the SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Machine Learning with Python and DBMS.

Mr. EJJIVARAPU 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.

Mr. S. Sai Ram is an MCA student in the Master of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. He has Completed Degree in B.Sc (chemistry) from K B N Degree College(Affiliated to Krishna University),Vijayawada. His areas of interest are DBMS, JavaScript, Blockchain, Machine Learning with Python, HTML, CSS and Bootstrap.