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

CYBER ATTACK DETECTION MODEL BASED MACHINE LEARNING APPROACH

Kuncham Ramya (1), Godi Sravanthi(2), Lakkepogu Vivek (3), Gajjala Ramacharan Reddy(4), Chalamakurthi Hemanth Kumar(5), Shaik Chand Basha6)

[1] Asst.Professor,CSE(Artificial Intelligence) Department,ABRCET,Kanigiri, Andhra Pradesh, India.

[2,3,4,5,6] B.Tech Student, CSE(Artificial Intelligence) Department, ABRCET, Kanigiri, Andhra Pradesh, India.

Abstract

Stood out from the past, enhancements in PC and correspondence advancements have given expansive and moved changes. The utilization of new developments give inconceivable benefits to individuals, associations, and governments, nevertheless, some against them. For example, the assurance of critical information, security of set aside data stages, availability of data, etc. Dependent upon these issues, advanced anxiety based abuse is perhaps the main issues nowadays. Computerized fear, which made a lot of issues individuals and foundations, has shown up at a level that could subvert open and country security by various social occasions, for instance, criminal affiliation, capable individuals and advanced activists. Thusly, Intrusion Detection Systems (IDS) has been made to keep an essential separation from advanced attacks.

INTRODUCTION

Lately, the world has seen a critical evolution in the various spaces of associated innovations like brilliant matrices, the Internet of vehicles, long haul advancement, and 5G correspondence. By 2022, it is normal that the quantity of IP-associated gadgets will be multiple times bigger than the worldwide populace, delivering 4.8 ZB of IP traffic yearly, as revealed by Cisco [1]. This sped up development raises overpowering security worries because of the trading of enormous measures of sensitive data through asset compelled gadgets and over the untrusted "Internet" utilizing heterogeneous advances and correspondence conventions. To keep up feasible and secure the internet, progressed security controls and flexibility investigation ought to be applied in the prior stages before sending.

The applied security controls are answerable for forestalling, identifying, and reacting to assaults. For location purposes an interruption recognition framework (IDS) is a generally utilized procedure for identifying interior and outer interruptions that objective a system, just as irregularities that show likely interruptions and dubious exercises. An IDS includes a bunch of instruments and mech anisms for observing the PC framework and the organization traffic, as well as breaking down exercises with the point of detecting potential interruptions focusing on the framework. An IDS can be executed as signature-based, inconsistency based, or mixture IDS. In signature-based IDS, interruptions are identified by contrasting observed practices and pre-characterized interruption designs, while oddity put together IDS centers with respect to knowing typical conduct in or der to distinguish any deviation [2]. Various strategies are utilized to recognize oddities, for example, factual based, information based, and AI procedures; as of late, profound learning techniques have been researched.

LITERATURE SURVEY

This segment presents different late achievements around here. It ought to be noticed that we just examine the work that have utilized the NSL-KDD dataset for their perfor mance benchmarking. Subsequently, any dataset alluded from here on out ought to be considered as NSL-KDD. This methodology permits a more exact examination of work with other found in the writing. Another restriction is the utilization of preparing information for both preparing and testing by most work. At long last, we examine a couple of profound learning based methodologies that have been attempted so far for comparable sort of work.

One of the most punctual work found in writing utilized ANN with improved strong back-spread for the plan of such an IDS [6]. This work utilized just the preparation dataset for preparing (70%), approval (15%) and testing (15%). As expected, utilization of unlabelled information for testing brought about a reduction of execution. A later work utilized J48 choice tree classifier with 10-overlay cross-approval for testing on the preparation dataset [4]. This work utilized a decreased list of capabilities of 22 highlights rather than the full arrangement of 41 highlights. A comparable work assessed different well known regulated tree-based classifiers and tracked down that Random Tree model performed best with the most extensive level of exactness alongside a decreased bogus alert rate [5].

Numerous 2-level characterization approaches have likewise been master presented. One such work utilized Discriminative Multinomial Naive Bayes (DMNB) as a base classifier and Nominal-to Binary directed separating at the second level alongside 10-crease cross approval for testing [9]. This work was hide the reached out to utilize Ensembles of Balanced Nested Dichotomies (END) at the main level and Random Forest at the second level [10]. True to form, this upgrade resulted in an improved location rate and a lower bogus positive rate. Another 2-level execution utilized head segment examination (PCA) for the list of capabilities decrease and afterward SVM (utilizing Radial Basis Function) for last classification, brought about a high recognition precision with just the preparation dataset and full 41 highlights set. A decrease in features set to 23 came about in far better location exactness in a portion of the assault classes, however the general execution was diminished [11]. The creators improved their work by utilizing data gain to rank the highlights and afterward a conduct based element determination to lessen the list of capabilities to 20. This brought about an improvement in detailed precision utilizing the preparation dataset [12].

Existing system :

Within the ever-growing and quickly increasing field of cyber security, it is nearly impossible to quantify or justify the explanations why cyber security has such an outsized impact. Permitting malicious threats to run any place, at any time or in any context is a long way from being acceptable, and may cause forceful injury. It particularly applies to the Byzantine web of consumers and using the net and company information that cyber security groups are finding it hard to shield and contain. Cyber security may be a necessary thought for people and families alike, also for businesses, governments, and academic establishments that operate inside the compass of the world network or net. With the facility of Machine Learning, we will advance the cyber security landscape. Today's high-tech infrastructure, that has network and cyber security systems, is gathering tremendous amounts of data and analytics on almost all the key aspects of mission-critical systems. Whereas people still give the key operational oversight and intelligent insights into today's infrastructure. Most intrusion detection systems are focused on the perimeter attack surface threats, starting with your firewall. That offers protection of your network's northsouth traffic, but what it doesn't take into account is the lateral spread (east-west) that many network threats today take advantage of as they infiltrate your organization's network and remain there unseen. We know this is true because research has shown that only 20% of discovered threats come from northsouth monitoring. When an IDS (Intrusion Detection System) detects suspicious activity, the violation is typically reported to a security information and event management (SIEM) system where real threats are ultimately determined amid benign traffic abnormalities or other false alarms. However, the longer it takes to distinguish a threat, the more damage can be done. An IDS is immensely helpful for monitoring the network, but their usefulness all depends on what you do with the information that they give you. Because detection tools don't block or resolve potential issues, they are ineffective at adding a layer of security unless you have the right personnel and policy to administer them and act on any threats.

- An IDS cannot see into encrypted packets, so intruders can use them to slip into the network.
- An IDS will not register these intrusions until they are deeper into the network, which leaves your systems vulnerable until the intrusion is discovered. This is a huge concern as encryption is becoming more prevalent to keep our data secure.
- significant issue with an IDS is that they regularly alert you to false positives. In many cases false positives are more frequent than actual threats.

Proposed System :

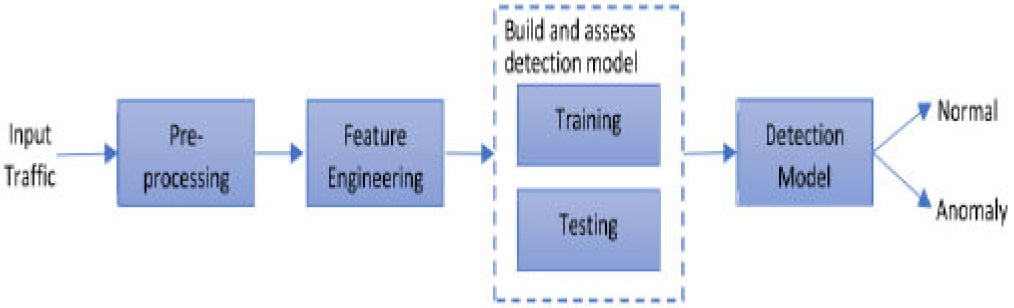
At the present time, assessments of help vector machine, ANN, CNN, Random Forest and significant learning estimations reliant upon current CICIDS2017 dataset were presented moderately. Results show that the significant learning estimation performed generally best results over SVM, ANN, RF and CNN. We will use port scope attempts just as other attack types with AI and significant learning computations, apache Hadoop and shimmer advancements together ward on this dataset later on. So by utilizing these datasets we will anticipate if digital assault is finished. These forecasts should be possible by four calculations like SVM, ANN, RF, CNN this paper assists with distinguishing which calculation predicts the best precision rates which assists with foreseeing best outcomes to recognize the digital assaults occurred or not.

Advantages:

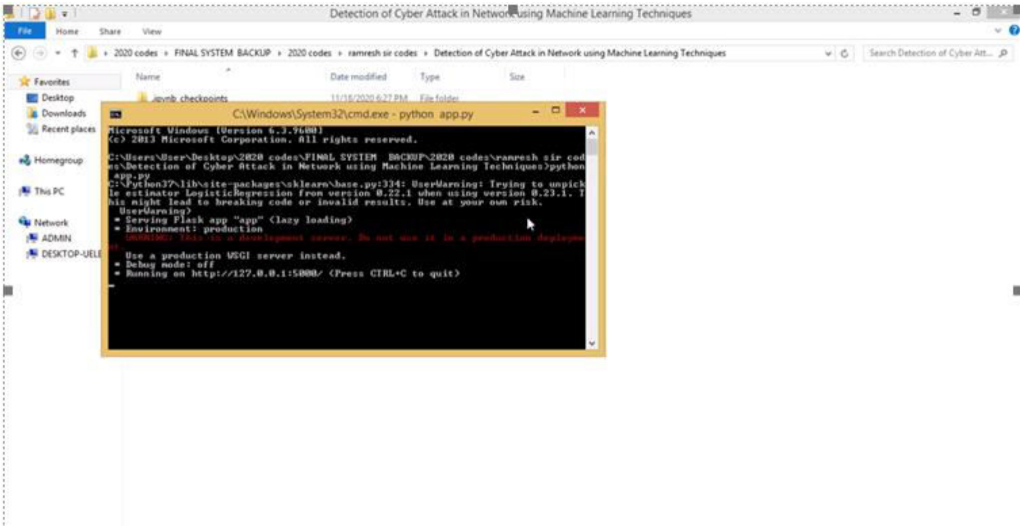
Advantages of the proposed systems are follows:

- □ Protection from malicious attacks on your network.
- □ Deletion and/or guaranteeing malicious elements within a preexisting network.
- □ Prevents users from unauthorized access to the network.
- □ Deny's programs from certain resources that could be infected.
- □ Securing confidential information

System architecture :

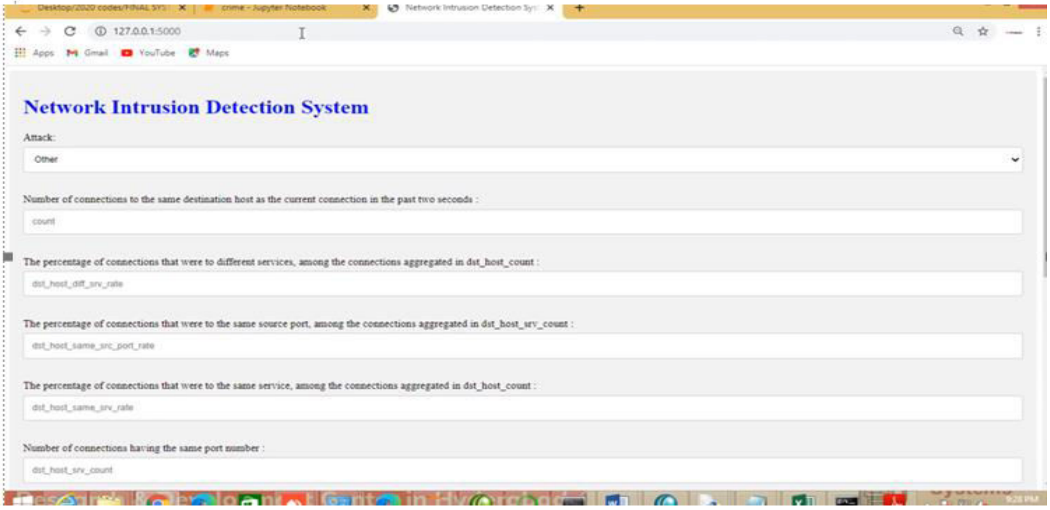


OUTPUT SCREENS


Detection of Cyber Attack in Network using Machine Learning Techniques
File
Home
Share
View
2020 codes › FINAL SYSTEM BACKUP › 2020 codes › ramresh sir codes › Detection of Cyber Attack in Network using Machine Learning Techniques
Search Detection of Cyber Att...
Favorites
Desktop
Downloads
Recent places
Homegroup
This PC
Network
ADMIN
DESKTOP-UELE
Name
Date modified
Type
Size
.ipynb_checkpoints
11/18/2020 6:27 PM
File folder
C:\Windows\System32\cmd.exe - python app.py

```
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\Users\User\Desktop\2020 codes\FINAL SYSTEM  BACKUP\2020 codes\ramresh sir cod
es\Detection of Cyber Attack in Network using Machine Learning Techniques>python
 app.py
C:\Python37\lib\site-packages\sklearn\base.py:334: UserWarning: Trying to unpick
le estimator LogisticRegression from version 0.22.1 when using version 0.23.1. T
his might lead to breaking code or invalid results. Use at your own risk.
  UserWarning)
 * Serving Flask app "app" (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployme
nt.
   Use a production WSGI server instead.
 * Debug mode: off
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
_
```



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 :

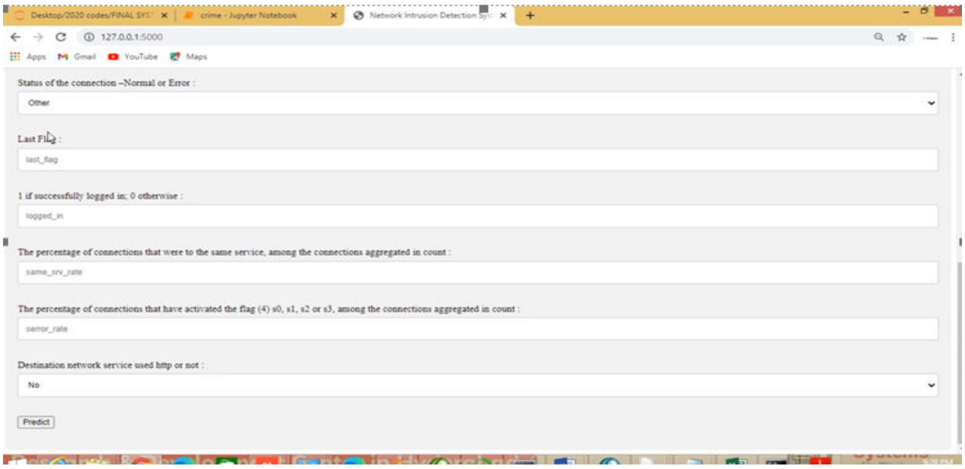
dst_host_same_src_port_rate

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



Status of the connection –Normal or Error :
Other

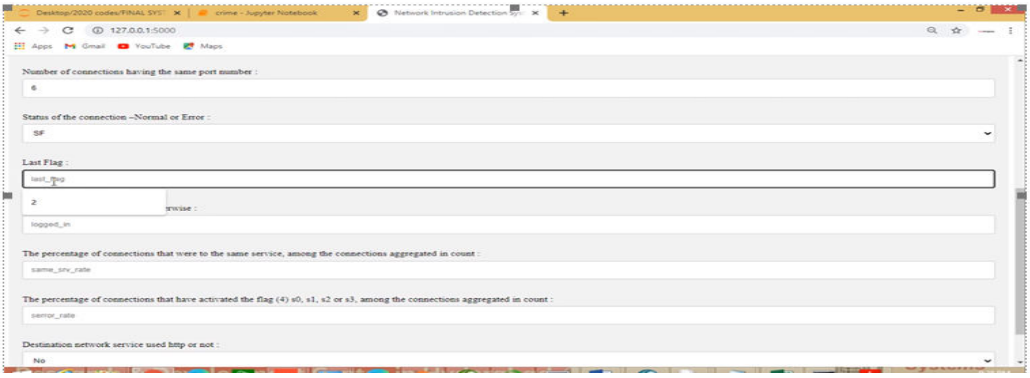
Last Flag :
last_flag

1 if successfully logged in; 0 otherwise :
logged_in

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

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

Destination network service used http or not :
No


Desktop/2020 codes/FINAL SYST
crime - Jupyter Notebook
Network Intrusion Detection Sys
127.0.0.1:5000
Apps
Gmail
YouTube
Maps

Number of connections having the same port number :
6

Status of the connection –Normal or Error :
SF

Last Flag :
last_flag
2

rwise :
logged_in

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

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

Destination network service used http or not :
No

CONCLUSION

At the present time, assessments of help vector machine, ANN, CNN, Random Forest and significant learning estimations reliant upon current CICIDS2017 dataset were presented moderately. Results show that the significant learning estimation performed generally best results over SVM, ANN, RF and CNN. We will use port scope attempts just as other attack types with AI and significant learning computations, apache Hadoop and shimmer advancements together ward on this dataset later on. Every one of these estimation assists us with recognizing the digital assault in network. It occurs in the manner that when we think about long back a long time there might be such countless assaults occurred so when these assaults are perceived then the highlights at which esteems these assaults are going on will be put away in some datasets. So by utilizing these datasets we will anticipate if digital assault is finished. These forecasts should be possible by four calculations like SVM, ANN, RF, CNN this paper assists with distinguishing which calculation predicts the best precision rates which assists with foreseeing best outcomes to recognize the digital assaults occurred or not.

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] Rashmi T V. "Predicting the System Failures Using Machine Learning Algorithms". International Journal of Advanced Scientific Innovation, vol. 1, no. 1, Dec. 2020, doi:10.5281/zenodo.4641686.

[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] Girish L, Rao SKN (2020) "Quantifying sensitivity and performance degradation of virtual machines using machine learning.", Journal of Computational and Theoretical Nanoscience, Volume 17, Numbers 9-10, September/October 2020, pp. 4055-4060(6) https://doi.org/10.1166/jctn.2020.9019

.

[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.