International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Cyber Hacking Breaches Prediction Using Cat Boost Machine Learning

Gurram Kesavakrishna[1], N. Anand Reddy[2]
M Tech Student[1] , Assistant Professor[2]
Department of Computer Science and Engineering, Siddhartha Educational Academy Groups of Institutions, Tirupathi, Andhra Pradesh, India

Article Info

Publication Issue :
Volume 8, Issue 6
November-December-2022

Page Number : 156-164

Article History
Accepted: 05 Nov 2022
Published: 15 Nov 2022

ABSTRACT

Cyber-physical systems (cps) have made significant progress in many dynamic applications due to the integration between physical processes, computational resources, and communication capabilities. However, cyber-attacks are a major threat to these systems. Unlike faults that occurs by accidents cyber-physical systems, cyber-attacks occur intelligently and stealthy. Some of these attacks which are called deception attacks, inject false data from sensors or controllers, and also by compromising with some cyber components, corrupt data, or enter misinformation into the system. If the system is unaware of the existence of these attacks, it won't be able to detect them, and performance may be disrupted or disabled altogether. Therefore, it is necessary to adapt algorithms to identify these types of attacks in these systems. It should be noted that the data generated in these systems is produced in very large number, with so much variety, and high speed, so it is important to use machine learning algorithms to facilitate the analysis and evaluation of data and to identify hidden patterns. In this research, the CPS is modeled as a network of agents that move in union with each other, and one agent is considered as a leader, and the other agents are commanded by the leader. The proposed method in this study is to use the structure of deep neural networks for the detection phase, which should inform the system of the existence of the attack in the initial moments of the attack. The use of resilient control algorithms in the network to isolate the misbehave agent in the leader-follower mechanism has been investigated. In the presented control method, after the attack detection phase with the use of a deep neural network, the control system uses the reputation algorithm to isolate the misbehave agent. Experimental analysis shows us that deep learning algorithms can detect attacks with higher performance that usual methods and can make cyber security simpler, more proactive, less expensive and far more effective.

Keywords : Decision Tree, Random Forest Classifier, Support Vector Machine, Cat Boost

I. INTRODUCTION

Recent advances in technology have led to the introduction of cyber-physical systems, which due to their better computational and communicational ability and integration between physical and cyber-components, has led to significant advances in many dynamic applications. But this improvement comes at the cost of being vulnerable to cyber- hacking. Cyber-physical systems are made up of logical elements and embedded computers, which communicate with communication channels such as the Internet of Things (IoT). More specifically, these systems include digital or cyber components, analog components, physical devices and humans that designed to operate between physical and cyber parts. In other words, a cyber-physical system is any system that includes cyber and physical components and humans, and has the ability to trade between the physical and cyber parts. In cyber-physical systems, the security of these types of systems becomes more important due to the addition of the physical part.

Physical components including sensors, which receive data from the physical environment, maybe attacked and be injected incorrect data into the system. One of the most important challenges of a cyber-physical system, in its physical part is the presence of a large number of sensors in the environment, which collect so much data, with so much variety, and at high speed. Also, the connection between the sensors and the necessary calculations and the analysis of the obtained data will be among the main challenges. Therefore, one of the most important features of a cyber-physical system is to communicate between these sensors, compute and control the system

The security of cyber-physical systems to detect cyber-attacks is an important issue in these systems . It should be noted that cyber-attacks occur in irregular ways, and it is not possible to describe these attacks in a regular and orderly manner. In general, cyber attacks in cyber-physical systems are divided into two main types: denial of service(Dos) and deception attacks. In denial of service, the attacker prevents communication between network nodes and communication channels. However, in the deception attacks that inject false data to system, which are carried out by abusing system components , such as sensors or controllers and it can corrupt data or enter incorrect information into the system and cause misbehaving.

These attacks can be detected by system monitoring in the system. But if the attacker can plan a high-level attack to prevent himself from being identified, these attacks are called stealthy deception attacks, and other common methods of counteracting such attacks will not work. Therefore, it is important to be aware of the attacks that occur in order to respond in a timely manner to attackers. In other words, the security system must be aware of the attack, otherwise it will not be able to detect and control the attack. Cyber defense can be improved by using security analytic to search for hidden patterns and how to deceive.

II. RELATED WORKS

S.]Kwon, Cheolhyeon, Weiyi Liu, and Inseok Hwang. "Security analysis for cyber-physical systems against stealthy deception attacks." In 2013 American control conference, IEEE (2013): 3344-3349 The security issue in the state estimation problem is investigated for a networked control system (NCS). The communication channels between the sensors and the remote estimator in the NCS are vulnerable to attacks from malicious adversaries. The false data injection attacks are considered. The aim of this paper to find the so-called insecurity conditions under which the estimation system is insecure in the sense that there exist malicious attacks that can bypass the anomaly detector but still lead to unbounded estimation errors. In particular, a new necessary and sufficient condition for the insecurity is derived in the case that all communication channels are compromised by the adversary. Moreover, a specific algorithm is proposed for generating attacks with which the estimation system is insecure. Furthermore, for the insecure system, a system protection scheme through which only a few (rather than all) communication channels require protection against false data injection attacks is proposed. A simulation example is utilized to demonstrate the effectiveness of the proposed conditions/algorithms in the secure estimation problem for a flight vehicle.

Pajic, Miroslav, James Weimer, Nicola Bezzo, Oleg Sokolsky, George J. Pappas, and Insup Lee. ”Design and implementation of attack-resilient cyberphysical systems: With a focus on attack-resilient state estimators.” IEEE Control Systems Magazine 37, no. 2 (2017): 66-81. Recent years have witnessed a significant increase in the number of security-related incidents in control systems. These include high-profile attacks in a wide range of application domains, from attacks on critical infrastructure, as in the case of the Maroochy Water breach [1], and industrial systems (such as the StuxNet virus attack on an industrial supervisory control and data acquisition system [2], [3] and the German Steel Mill cyberattack [4], [5]), to attacks on modern vehicles [6]-[8]. Even high-assurance military systems were shown to be vulnerable to attacks, as illustrated in the highly publicized downing of the RQ-170 Sentinel U.S. drone [9]-[11]. These incidents have greatly raised awareness of the need for security in cyberphysical systems (CPSs), which feature tight coupling of computation and communication substrates with sensing and actuation components. However, the complexity and heterogeneity of this next generation of safety-critical, networked, and embedded control systems have challenged the existing design methods in which security is usually consider as an afterthought.

Sheng, Long, Ya-Jun Pan, and Xiang Gong. ”Consensus formation control for a class of networked multiple mobile robot systems.” Journal of Control Science and Engineering 2012 (2012). Embedded computational resources in autonomous robotic vehicles are becoming more abundant and have enabled improved operational effectiveness of cooperative robotic systems in civilian and military applications. Compared to autonomous robotic vehicles that operate single tasks, cooperative teamwork has greater efficiency and operational capability. Multirobotic vehicle systems have many potential applications, such as platooning of vehicles in urban transportation, the operation of the multiple robots, autonomous underwater vehicles, and formation of aircrafts in military affairs [1–3]. The study of group behaviors for multirobot systems is the main objective of the work. Group cooperative behavior signifies that individuals in the group share a common objective and action according to the interest of the whole group. Group cooperation can be efficient if individuals in the group coordinate their actions well. Each individual can coordinate with other individuals in the group to facilitate group cooperative behavior in two ways, named local coordination and global coordination. For the local coordination, individuals react only to other individuals that are close, such as fish engaged in a school.

Zeng, Wente, and Mo-Yuen Chow. ”Resilient distributed control in the presence of misbehaving agents in networked control systems.” IEEE transactions on cybernetics 44, no. 11 (2014): 2038-2049. In this paper, we study the problem of reaching a consensus among all the agents in the networked control systems (NCS) in the presence of misbehaving agents. A reputation-based resilient distributed control algorithm is first proposed for the leader-follower consensus network. The proposed algorithm embeds a resilience mechanism that includes four phases (detection, mitigation, identification, and update), into the control process in a distributed manner. At each phase, every agent only uses local and one-hop neighbors' information to identify and isolate the misbehaving agents, and even compensate their effect on the system. We then extend the proposed algorithm to the leaderless consensus network by introducing and adding two recovery schemes (rollback and excitation recovery) into the current framework to guarantee the accurate convergence of the well-behaving agents in NCS. The effectiveness of the proposed method is demonstrated through case studies in multirobot formation control and wireless sensor networks.

Sun, Hongtao, Chen Peng, Taicheng Yang, Hao Zhang, and Wangli He. "Resilient control of networked control systems with stochastic denial of service attacks." Neurocomputing 270 (2017): 170-177: This paper focuses on resilient control of networked control systems (NCSs) under the denial of service (DoS) attacks which is characterized by a Markov process. Firstly, the packets dropout are modeled as Markov process according to the game between attack strategies and defense strategies. Then, an NCS under such game results is modeled as a Markovian jump linear system and four theorems are proved for the system stability analysis and controller design. Finally, a numerical example is used to illustrative the application of these theorems. Networked control systems (NCSs) have received an increasing attention in the past decades. Now, NCSs have been widely applied in industrial processes, electric power networks, and intelligent transportation and so on. With the growing of the NCSs, network, as a critical element in an NCS, is vulnerable to cyber-threats which can menace the control systems.

III. Methodology

Proposed system:

Proposed several machine learning models to classify whether there will be a cyber- hacking or not, but none have adequately addressed this misdiagnosis problem. Also, similar studies that have proposed models for evaluation of such performance classification mostly do not consider the heterogeneity and the size of the data Therefore, we propose a Support Vector, Decision Tree, Random forest and Cat Boost Classifier classification techniques.

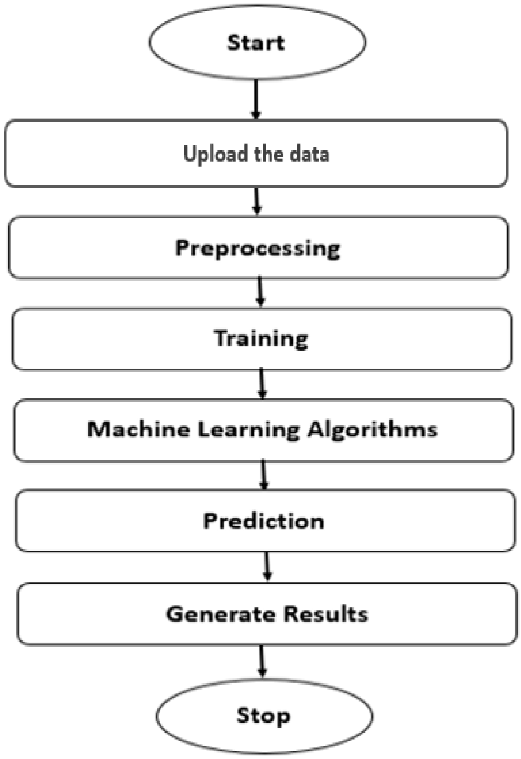


Figure 1: Block diagram

ADVANTAGES:

- High accuracy.
- Time Saving.
- Low complexities.
- High reliability.

EXISTING METHOD

In the existing system, implementation of machine learning algorithms is bit complex to build due to the lack of information about the data visualization. Mathematical calculations are used in existing system for SVM model building this may takes the lot of time and complexity. To overcome all this, we use machine learning packages available in the scikit-learn library.

DISADVANTAGES:

- Low accuracy.
- Time consuming.
- High complexities.

IV. Implementation:

The project has implemented by using below listed s

DECISION TREE:

Decision tree is a flowchart-like tree structure where an internal node represents feature(or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees are easy to understand and interpret.

The basic idea behind any decision tree algorithm is as follows:

1. Select the best attribute using Attribute Selection Measures (ASM) to split the records.

2 .Make that attribute a decision node and breaks the dataset into smaller subsets.

3 .Starts tree building by repeating this process recursively for each child until one of the conditions will match:

•All the tuples belong to the same attribute value.

•There are no more remaining attributes.

•There are no more instances.

Random Forest Classifier:

A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems.

A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.

A random forest eradicates the limitations of a decision tree algorithm. It reduces the over fitting of datasets and increases precision. It generates predictions without requiring many configurations in packages (like Scikit-learn).

Features of a Random Forest Algorithm:

•It's more accurate than the decision tree algorithm.

•It provides an effective way of handling missing data.

•It can produce a reasonable prediction without hyper-parameter tuning.

•It solves the issue of over fitting in decision trees.

•In every random forest tree, a subset of features is selected randomly at the node's splitting point.

Decision trees are the building blocks of a random forest algorithm. A decision tree is a decision support technique that forms a tree-like structure. An overview of decision trees will help us understand how random forest algorithms work.

A decision tree consists of three components: decision nodes, leaf nodes, and a root node. A decision tree algorithm divides a training dataset into branches, which further segregate into other branches. This sequence continues until a leaf node is attained. The leaf node cannot be segregated further.

The nodes in the decision tree represent attributes that are used for predicting the outcome. Decision nodes provide a link to the leaves. The following diagram shows the three types of nodes in a decision tree.

Support Vector Machine:

A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for each category, they're able to categorize new text.

So you're working on a text classification problem. You're refining your training data, and maybe you've even tried stuff out using Naive Bayes. But now you're feeling confident in your dataset, and want to take it one step further. Enter Support Vector Machines (SVM): a fast and dependable classification algorithm that performs very well with a limited amount of data to analyze.

Perhaps you have dug a bit deeper, and ran into terms like linearly separable, kernel trick and kernel functions. But fear not! The idea behind the SVM algorithm is simple, and applying it to natural language classification doesn't require most of the complicated stuff.

Steps for implementation:

- Import the dataset.
- Explore the data to figure out what they look like.
- Pre-process the data.
- Split the data into attributes and labels.

Cat Boost:

CatBoost is a high-performance open source library for gradient boosting on decision trees. CatBoost is an algorithm for gradient boosting on decision trees. It is developed by Yandex researchers and engineers, and is used for search, recommendation systems, personal assistant, self-driving cars, weather prediction and many other tasks at Yandex and in other companies, including CERN, Cloudflare and Careem taxi. It is in open-source and can be used by anyone. Catboost, the new kid on the block, has been around for a little more than a year now, and it is already threatening XGBoost, Light GBM.

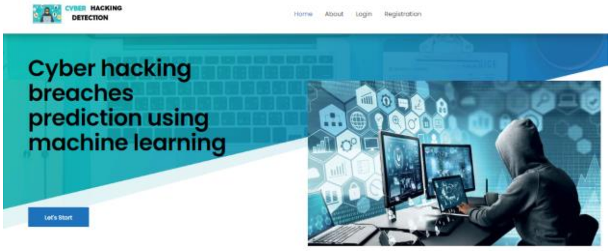
Catboost achieves the best results on the benchmark, and that's great. Though, when you look at datasets where categorical features play a large role, this improvement becomes significant and undeniable.

While training time can take up longer than other GBDT implementations, prediction time is 13–16 times faster than the other libraries according to the Yandex benchmark. Catboost's default parameters are a better starting point than in other GBDT algorithms and it is good news for beginners who want a plug and play model to start experience tree ensembles or Kaggle competitions. Some more noteworthy advancements by Catboost are the features interactions, object importance and the snapshot support. In addition to classification and regression, Catboost supports ranking out of the box.

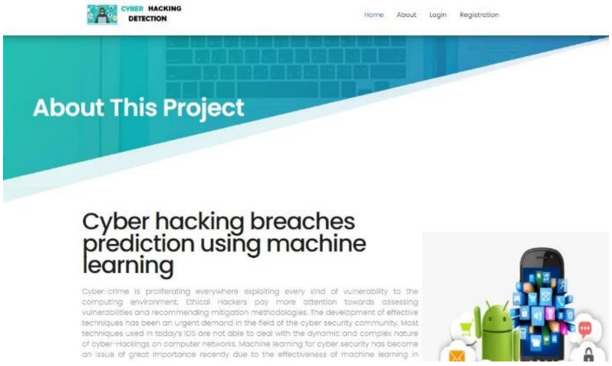
V. Results and Discussion

The following screenshots are depicted the flow and working process of project.

Home Page: This is the home page of Cyber hacking breaches prediction using machine learning



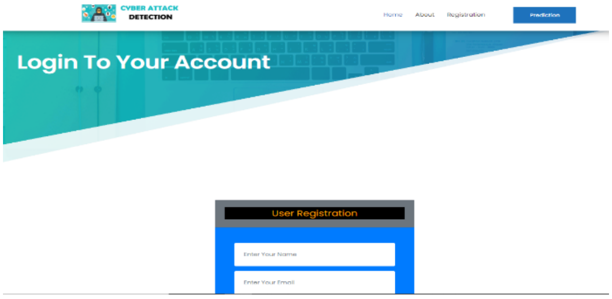
About Us:



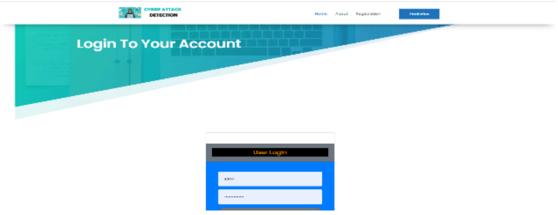
Cyber hacking breaches prediction using machine learning

Cyber crime is proliferating everywhere exploiting every kind of vulnerability to the computing environment. Ethical Hackers pay more attention towards assessing vulnerabilities and recommending mitigation methodologies. The development of effective techniques has been an urgent demand in the field of the cyber security community. Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-Hackings on computer networks. Machine learning for cyber security has become an issue of great importance recently due to the effectiveness of machine learning in

Register Page:

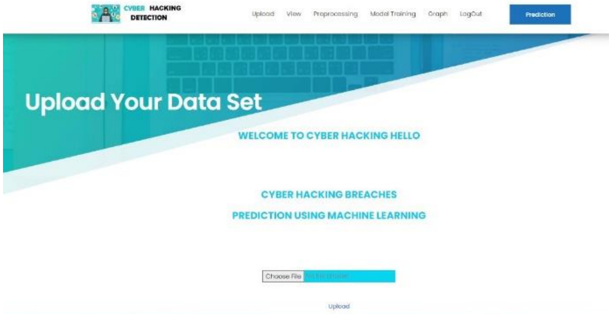


Login Page:

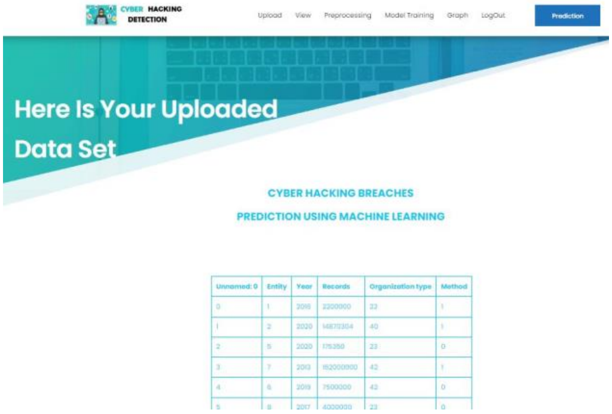


Load:

In the load page, users can load the cyber dataset.

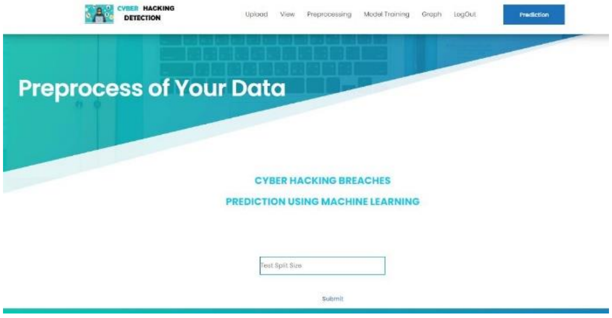


View Data: Here we can see the uploaded data set..

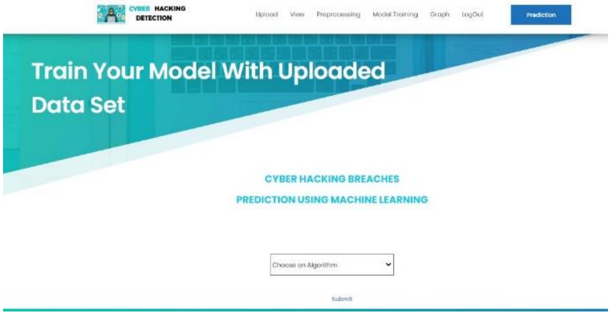


Unnamed: 0	Entity	Year	Records	Organization type	Method
0	1	2016	2200000	22	1
1	2	2020	14870304	40	1
2	5	2020	175350	23	0
3	7	2013	152000000	42	1
4	6	2019	7500000	42	0
5	8	2017	4000000	23	0

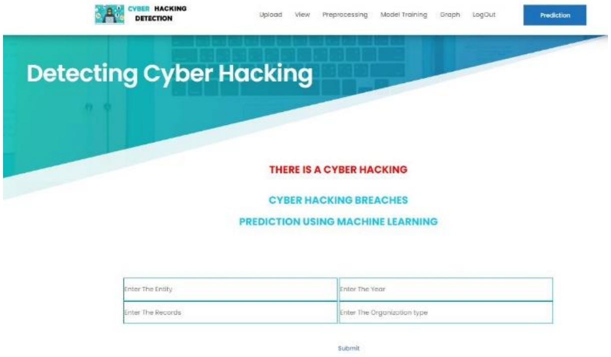
Pre-process: Here we can pre-process and split our data into train and test



Model: Here we train our data with different ML algorithms.



Prediction: This page show the detection result of the Cyber hacking breaches prediction using machine learning data



Detecting Cyber Hacking

THERE IS A CYBER HACKING

CYBER HACKING BREACHES

PREDICTION USING MACHINE LEARNING

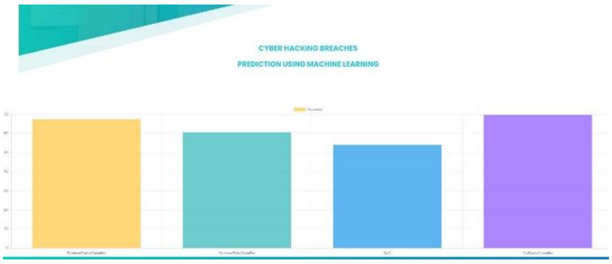
Enter The Entity

Enter The Year

Enter The Records

Enter The Organization type

Graph:



CYBER HACKING BREACHES

PREDICTION USING MACHINE LEARNING

VI. Conclusion

In this study, an attempt was made to use the resilient control consensus method in complex discrete cyber-physical networks with a number of local Cyber hacking breaches off. By applying this control method, it was observed that even in the presence of Cyber hacking breaches, the system can remain stable and isolate the Cyber hacking node and the performance of the system is not weakened. Using the neural network used in this study, it was observed that with a deep neural network, with 7 hidden layers, the system shows better performance. Also in a recurrent neural network integrated with a deep neural network, a deep layer network with a linear function performs better. Therefore, it can be said that the system has less complexity. So With deep learning method, systems can analyse patterns and learn from them to help prevent similar attacks and respond to changing behaviour. In short, machine learning can make cyber security simpler, more proactive, less expensive and far more effective. After observing the state of the system reported by the neural network, the control system makes decisions based on it and, if there is an Cyber hacking, detects it and isolates it, so as not to have a detrimental effect on the behaviour of other agents. In future research, more attacks on agents can be considered, also data mining and other machine learning methods, such as support vector machine (SVM) algorithms or other types as recurrent CatBoost to evaluate system performance improvemsssssents.

VII. FUTURE SCOPE

There are quite a few things that can be polished or be added in the future work. • We have opted to use two data mining classifies in this project namely the ID3 and Naive Bayes classifier. There are more classieres such as the Bayesian network classifier, Neural Network classifier and C4.5 classifier. Such classifiers were not included in this paper and could be counted in future to give a more data to be compared with.

VIII. REFERENCES

[1]. Kwon, Cheolhyeon, Weiyi Liu, and Inseok Hwang. ”Security analysis for cyber-physical systems against stealthy deception attacks.” In 2013 American control conference, IEEE (2013): 3344-3349.

[2]. Pajic, Miroslav, James Weimer, Nicola Bezzo, Oleg Sokolsky, George J. Pappas, and Insup Lee. ”Design and implementation of attack-resilient cyberphysical systems: With a focus on attack-resilient state estimators.” IEEE Control Systems Magazine 37, no. 2 (2017): 66-81.

[3]. Sheng, Long, Ya-Jun Pan, and Xiang Gong. ”Consensus formation control for a class of networked multiple mobile robot systems.” Journal of Control Science and Engineering 2012 (2012).

[4]. Zeng, Wente, and Mo-Yuen Chow. ”Resilient distributed control in the presence of misbehaving agents in networked control systems.” IEEE transactions on cybernetics 44, no. 11 (2014): 2038-2049.

[5]. Sun, Hongtao, Chen Peng, Taicheng Yang, Hao Zhang, and Wangli He. ”Resilient control of networked control systems with stochastic denial of service attacks.” Neurocomputing 270 (2017): 170-177.

[6]. Zhang, Haotian, and Shreyas Sundaram. ”Robustness of information diffusion algorithms to locally bounded adversaries.” In 2012 American Control Conference (ACC), IEEE (2012): 5855-5861.

[7]. Fu, Weiming, Jiahu Qin, Yang Shi, Wei Xing Zheng, and Yu Kang. ”Resilient Consensus of Discrete-Time Complex Cyber-Physical Networks under Deception Attacks.” IEEE Transactions on Industrial Informatics (2019).

[8]. Ozay, Mete, Inaki Esnaola, Fatos Tunay Yarman Vural, Sanjeev R. Kulkarni, and H. Vincent Poor. ”Machine learning methods for attack detection in the smart grid.” IEEE transactions on neural networks and learning systems 27, no. 8 (2015): 1773-1786.

[9]. Tianfield, Huaglory. ”Data mining based cyber-attack detection.” System simulation technology 13, no. 2 (2017): 90-104.

[10]. Pasqualetti, Fabio, Florian Dorfler, and Francesco Bullo. ”Attack detection and ¨ identification in cyber-physical systems.” IEEE Transactions on Automatic Control 58, no. 11 (2013): 2715-2729.

Cite this article as :

Gurram Kesavakrishna, N. Anand Reddy, "Cyber Hacking Breaches Prediction Using Cat Boost Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 156-164, November-December 2022.

Journal URL : https://ijsrcseit.com/CSEIT228616