Shodhshauryam, International Scientific Refereed Research Journal
Available online at : www.shisrrj.com
OPEN ACCESS
© 2024 SHISRRJ | Volume 7 | Issue 2
ISSN : 2581-6306
doi : https://doi.org/10.32628/SHISRRJ

Cyber Hacking Breaches Prediction Using Machine Learning

T. Rajasekhar[1], Mukku Keerthna[2]

[1]Assistant Professor, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India

[2]Post Graduate, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India

Article Info

Publication Issue :
March-April-2024
Volume 7, Issue 2

Page Number : 173-179

Article History
Received : 15 March 2024
Published : 30 March 2024

ABSTRACT

The combination of physical processes, computational resources, and communication capabilities has driven major advancements in many dynamic applications of cyber-physical systems (cps). Cyberattacks, however, pose a serious risk to these systems. Cyber-attacks are smart and covert, in contrast to cyber-physical system defects that arise accidentally. Certain attacks, referred to as deception attacks, introduce erroneous data into the system by manipulating sensors or controllers, or by breaching cyber components and contaminating or introducing false information. The system may experience performance issues or become completely disabled if it is not aware that these attacks are occurring. Consequently, in order to recognize these kinds of assaults in these systems, algorithms must be modified. These systems generate large amounts of different, rapidly created data, so that's essential to use machine learning techniques to identify hidden trends and facilitate data analysis and review. This study models the CPS as a network of moving agents that work in union with among themselves. The model recognizes a leader agent and gives commands to the other agents in the network. The study's suggested approach makes use of deep neural network architecture for the detection stage, which should alert the system to the attack's presence in the early stages. Researchers have investigated isolating the misbehaving agent in the leader-follower system utilizing robust control methods within the network. In the proposed control strategy, the phase of assault detection is executed by a deep neural network, after The control system employs a reputation algorithm to identify and separate the agent exhibiting misconduct. Through experimental study, we can see that deep learning algorithms are able to detect assaults at a better performance level.

Keywords: Decision Tree, Random Forest, CatBoost, Adaboost, Logistic Regression, KNN, SVC

I. INTRODUCTION

The emergence of the The digital age has presented opportunities never before possible. for creativity and connectedness. Cyber hacking breaches, on the other hand, are becoming an increasingly dangerous threat as a result of it. Cyberattacks have grown more dangerous and sophisticated in recent years, endangering people, companies, and even entire countries.

This project aims to comprehensively investigate and analyze cyber hacking breaches that occurred in the past year. We will delve into the methods, motivations, and impacts of these breaches to gain a deeper understanding of the evolving landscape of cyber threats. By examining a range of high-profile cases, we intend to identify common vulnerabilities and attack vectors.

Our objectives include mapping the tactics employed by hackers, assessing the effectiveness of security measures, and evaluating the financial and reputational costs incurred by victims. Additionally, we will explore the ethical, legal, and regulatory aspects surrounding cyberattacks and data breaches.

This research not only serves as a valuable resource for cybersecurity professionals but also contributes to raising awareness among individuals and organizations about the importance of robust digital security. By shedding light on the ever-evolving world of cyber hacking breaches, we aim to empower stakeholders to fortify their defenses and safeguard their digital assets in an increasingly interconnected world.

II. PROPOSED FRAMEWORK

Numerous machine learning models have been suggested for determining the likelihood of a cyber hack, yet none have sufficiently tackled the issue of misdiagnosis. Furthermore, similar studies focusing on evaluating classification performance often overlook the complexities of data heterogeneity and size, failing to provide comprehensive solutions Consequently, we recommend the following classification techniques: SVM, Random Forest, Decision Tree, and CatBoost.

Advantages:

1. Early Detection

2. Adaptive Analysis

3. Pattern recognition

4. Improved Incident Response

5. Enhanced Risk Mitigation

III. LITERATURE SURVEY

[1] **Kwon, Cheolhyeon , Weiyi Liu, and Inseok Hwang. "Security analysis forcyber-physical systems against stealthy deception attacks."**

The difficulty of state estimation in a networked control system (NCS) raises security concerns that are investigated. A hostile attacker may attempt to compromise the pathways via which data is sent from the NCS's sensors to the remote estimator.

[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."**

Events pertaining to the safeguarding control systems have become more frequent in recent years. These include well-known attacks from a variety of application domains.

[3] Sheng, Long, Ya-Jun Pan, and Xiang Gong. "Consensus formation control for aclass of networked multiple mobile robot systems."

Due to the increasing availability of embedded computational resources in autonomous robotic vehicles, cooperative robotic systems are becoming more operationally effective in both civilian and military environments.

[4] Zeng, Wente, and Mo-Yuen Chow. "Resilient distributed control in the presence ofmisbehaving agents in networked control systems."

This research examines the challenge of achieving agreement among all agents inside networked control systems (NCS) when there are agents that exhibit misbehavior.

[5] Sun,Hongtao, Chen Peng, Taicheng Yang, Hao Zhang, and Wangli He. "Resilient control of networked control systems with stochastic denial of service attacks."

Here, we study robust the management of networked control systems (NCSs).against DoS attacks, with a Markov process serving as a distinguishing characteristic.

IV. METHODOLOGY

4.1 Model selection: Gather a comprehensive dataset containing information about past cyber hacking incidents. This dataset should include features such as time of attack, type of attack (e.g., phishing, malware, DDoS), target system or network, duration of attack, methods used, and any other relevant information.

4.2 Data preprocessing: When data is preprocessed, it is cleaned up by eliminating missing values and transformed into numerical representations for categorical variables using methods like label or one-hot encoding.

4.3 Feature engineering: Discover the critical features that play a vital role in forecasting hacking breaches by utilizing various methods, including correlation analysis, assessing feature importance through tree-based models, or leveraging domain expertise to cherry-pick or engineer novel features.

4.4 Model selection and training: Select the right machine learning algorithms for your classification jobs based on the challenge at hand and the characteristics of your data. to optimize hyperparameters, ensuring improved model performance.

4.5 Model evalution and selection: Examine the performance of multiple models using evaluation metrics and opt for the one exhibiting the highest performance on the validation set.

Subsequently, validate the chosen model on the testing set to evaluate its capacity for generalization and verify its accuracy in predicting outcomes on new, unseen data.

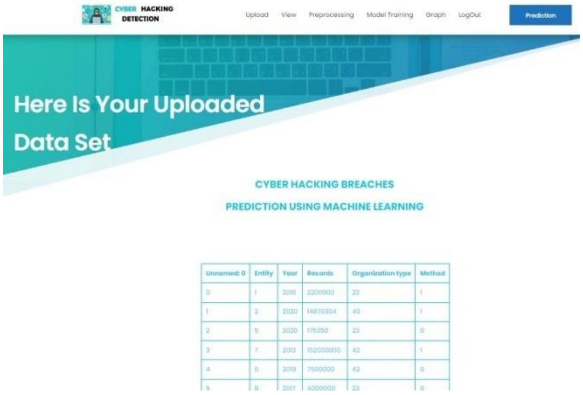
4.6 datasets used

Researchers and practitioners utilize a variety of datasets to forecast cyber hacking breaches; in this case, I'm using the datasets listed below.

4.6.1 Data breaches(2004-20021):

This dataset includes various features related to cyber hacking incidents, such as attack type, target system, time of attack, duration, impact, attack vector, data breach details, regulatory compliance, industry sector, attack source, attack tools.

It contain attributes like entity, organization type, records, year, method.

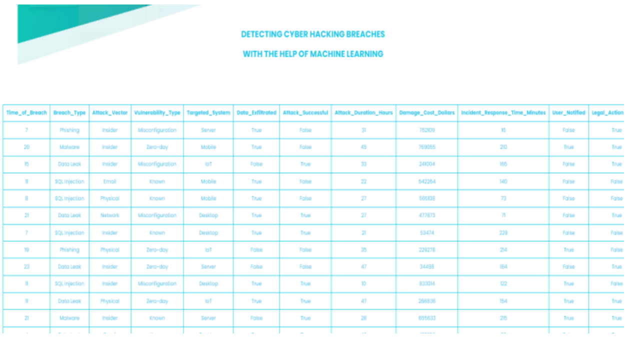


Unnamed: 0	Entity	Year	Records	Organization type	Method
0	1	2018	2200000	23	1
1	2	2020	14870304	40	1
2	5	2020	175350	23	0
3	7	2013	152000000	42	1
4	6	2019	7500000	42	0
5	8	2017	4000000	23	0

4.6.2 Cyber hacking breaches prediction Dataset:

Cyber hacking breaches prediction dataset plays an important role for breaches. Here the explanation of cyber hacking breaches prediction dataset and significant features.

Cyber hacking breaches dataset contains Time_of_breach,Breach_type,Vulnarability-type,Targeted_system,Data_exfiltrated,Attac_sucssfull. Prediction of cyber hacking breaches is very helpful for most of the people, business etc.


DETECTING CYBER HACKING BREACHES
WITH THE HELP OF MACHINE LEARNING

Time_of_Breach	Breach_Type	Attack_Vector	Vulnerability_Type	Targeted_System	Data_Exfiltrated	Attack_Successful	Attack_Duration_Hours	Damage_Cost_Dollars	Incident_Response_Time_Minutes	User_Notified	Legal_Action
7	Phishing	Insider	Misconfiguration	Server	True	False	31	762109	15	False	True
20	Malware	Insider	Zero-day	Mobile	True	False	45	769065	210	True	True
15	Data Leak	Insider	Misconfiguration	IoT	False	True	33	241004	165	False	True
11	SQL Injection	Email	Known	Mobile	True	False	22	542264	140	False	False
8	SQL Injection	Physical	Known	Mobile	True	False	27	565138	73	False	False
21	Data Leak	Network	Misconfiguration	Desktop	True	True	27	477873	71	False	True
7	SQL Injection	Insider	Known	Desktop	True	True	21	53474	229	False	False
19	Phishing	Physical	Zero-day	IoT	False	False	35	229278	214	True	False
23	Data Leak	Insider	Zero-day	Server	False	False	47	34498	184	False	True
11	SQL Injection	Insider	Misconfiguration	Desktop	True	True	10	833014	122	True	False
11	Data Leak	Physical	Zero-day	IoT	True	True	47	266836	154	True	True
21	Malware	Insider	Known	Server	False	True	28	655633	215	True	True

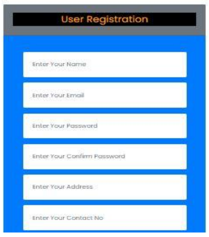
V. EXPERIMENTAL SETUP

The experimental setup for predicting cyber hacking breaches using machine learning involves several key steps.

Clearly define the problem statement and objectives. Determine what constitutes a hacking breach and the scope of the prediction. Gathering a comprehensive dataset containing historical information about cyber hacking incidents .Gather relevant datasets containing features related to system logs, network traffic, user activities, and other relevant data sources. Ensure the datasets are representative of the problem domain and include sufficient examples of both normal behavior and hacking breaches. It is essential to maintain data privacy and security, comply with relevant regulations, and adhere to ethical guidelines regarding the use of sensitive information.

VI. IMPLEMENTATION

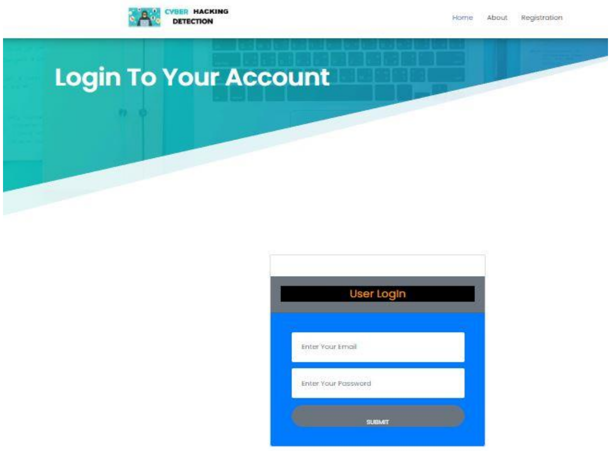
It requires some steps to be followed.
Register: user will register with the user details
Login: User will login using email and password.
Load: user will load the dataset
Preprocess: The user will handle data preprocessing and divide it into training and
testing sets.
Model: Here, we use several ML techniques to train our data.
Prediction: It presents the result of the detection.
Logout: Finally logout from the application.


User Registration

Enter Your Name
Enter Your Email
Enter Your Password
Enter Your Confirm Password
Enter Your Address
Enter Your Contact No

A. Fig 1

First, the user need to register by entering the following details like username, Email, Password, at last they need to confirm the password.



Login To Your Account

User Login

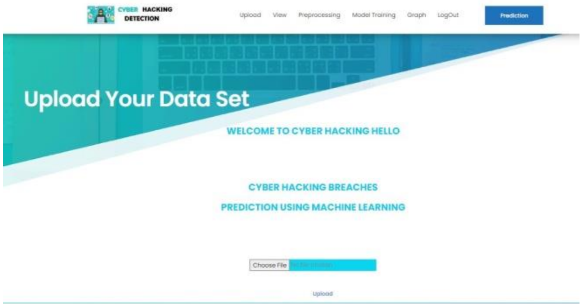
Enter Your Email

Enter Your Password

SUBMIT

B. Fig 2

After registration the user needs to login with respective email id or user name and password .After login in to the website the following image will be displayed.



Upload Your Data Set

WELCOME TO CYBER HACKING HELLO

CYBER HACKING BREACHES

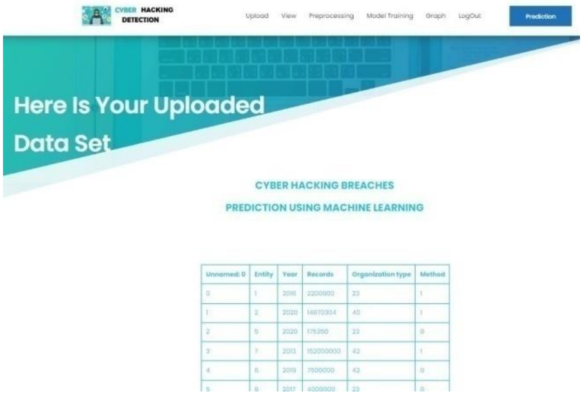
PREDICTION USING MACHINE LEARNING

Choose File

Upload

C. Fig 3

Here the user will upload the Cyber dataset.



Here Is Your Uploaded Data Set

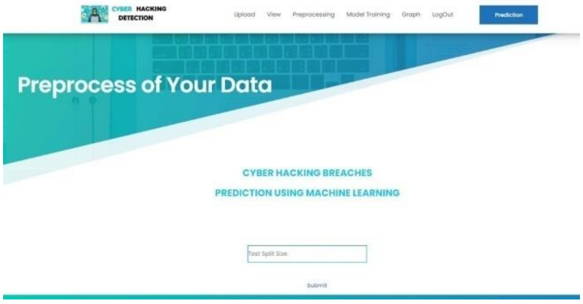
CYBER HACKING BREACHES

PREDICTION USING MACHINE LEARNING

Unnamed: 0	Entity	Year	Records	Organization type	Method
0	1	2018	2200000	23	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

D. Fig 4

Here the user will see the upload dataset and what are the attributes that are present in the Dataset .



Preprocess of Your Data

CYBER HACKING BREACHES

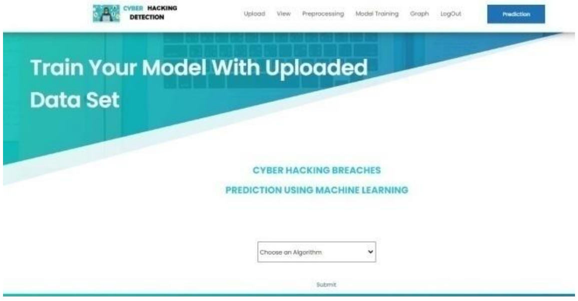
PREDICTION USING MACHINE LEARNING

Test Split Size

Submit

E. Fig 5

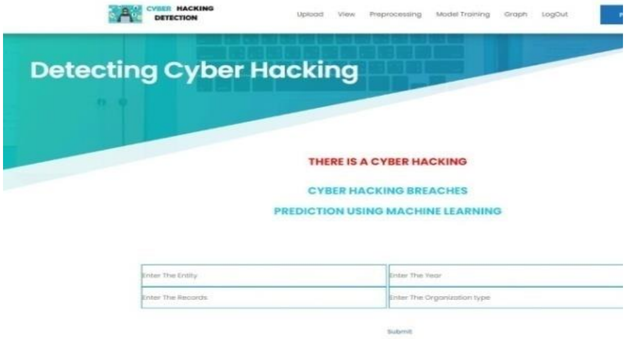
Here, the user preprocesses the data by dividing it into train and test sets. Preprocessing in this context entails removing duplicates and preparing the data in executable format.



Choose an Algorithm

F. Fig 6

Here the user will train the model with uploaded data set, By selecting the algorithm the data will be trained.



THERE IS A CYBER HACKING

Enter The Entity

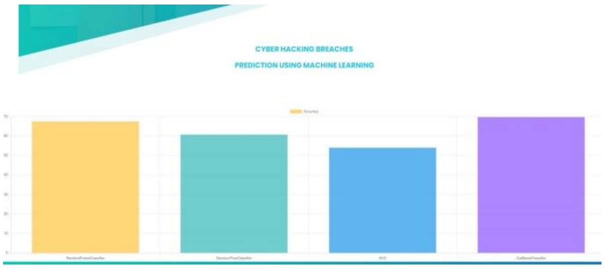
Enter The Year

Enter The Records

Enter The Organization type

G. Fig 7

Here the system will predict whether there is a cyber hack or not.



H. Fig 8

The above graph shows the accuracy level using various machine learning algorithms.

VII. CONCLUSION

The conclusion of the proposed approach for cyber hacking breaches prediction lies in its innovative utilization of machine learning algorithms to proactively identify and forecast potential cyber threats. Unlike traditional methods, this framework leverages advanced data analytics to discern subtle patterns and anomalies within vast datasets, enabling early detection of potential hacking breaches. The integration of machine learning models empowers the system to adapt and evolve with emerging cyber threats, enhancing its predictive capabilities. This novel approach not only contributes to the field of cybersecurity but also establishes a proactive and dynamic paradigm for anticipating and mitigating cyber risks, thus fortifying the resilience of digital ecosystems.

VIII. FUTURE ENHANCEMENT

When you are part of an organization, whether you are an employee or a manager, one thing you may already have realized is that Time is the most essential part of the organization. From coming and leaving office on time to delivering projects or finishing tasks on time, the clock plays a very important role in our lives. Essentially that is the reason why almost all of the organizations have implemented attendance marking systems.

Back in the days marking attendance was straight forward. You sign on a register every day when you enter the office premises. In larger organizations, it was an easy job to manipulate the register to falsify your actual working hours and entry time.

Fast forward a couple of years and we can see some technological advancements in attendance marking. Biometric devices like fingerprint scanner was a huge step forward. It made attendance marking almost foolproof preventing buddy punching and other hacks.

IX. 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. “Resilientdistributed 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.” Neuro computing 270 (2017): 170-177.