VISVESVARAYA TECHNOLOGICAL UNIVERSITY
BELAGAVI

A Project Report

On

CYBER HACKING BREACHES PREDICTION USING MACHINE LEARNING

Submitted in Partial fulfillment of requirement for the
Bachelor of Engineering
in

Computer Science & Engineering

By

N NAVEEN UPADHYAYA 3VC19CS091

Under the guidance of
Mr. Sampath Kumar R
Assistant Professor,
Dept of CSE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
ACCREDITED BY NATIONAL BOARD OF ACCREDITATION
RAO BAHADUR Y MAHABALESHWARAPPA ENGINEERING COLLEGE
ACCREDITED BY NAAC WITH B++
CANTONMENT, BALLARI-583104, KARNATAKA
2022 – 2023

VEERASHAIVA VIDYAVARDAHKA SANGHA'S

RAO BAHADUR Y MAHABALESHWARAPPA ENGINEERING COLLEGE

(AFFILIATED TO VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELGAUM, APPROVED BY AICTE, NEW DELHI & ACCREDITED BY NAAC WITH B++)
BELLARY – 583104, KARNATAKA

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

ACCREDITED BY NATIONAL BOARD OF ACCREDATATION

CERTIFICATE

This is to certify that project work entitled **"CYBER HACKING BREACHES PREDICTION USING MACHINE LEARNING"** is bonafied work carried out by **N NAVEEN UPADHYAYA (3VC19CS091)** of 8th Semester in Partial fulfillment for the award of degree of Bachelor of Engineering in Computer Science and Engineering of **Visvesvaraya Technological University, Belgaum** during the year **2022-2023**. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect project work prescribed for the said **Bachelor of Engineering Degree.**

Signature of Guide

Mr. Sampath Kumar R
Assistant Professor,
Dept of CSE, RYMEC.

Signature of HOD

Dr. H. GIRISHA
HOD, Dept of CSE,
RYMEC.

Signature of Principal

Dr. T. HANUMANTHA REDDY
RYMEC, BALLARI

Name of Examiners: **Signature with Date**

1)

2)

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crowned our effort with success.

We express our sincere gratitude to a Principal **Dr. T. Hanumantha Reddy** for giving us an opportunity to carry out our academic project

We wish to place on record our grateful thank to **Dr. H. Girisha Head** of the Department, Computer Science and Engineering RYMEC, Ballari for providing encouragement and guidance.

We hereby like to thank **Mr. Sampath Kumar R. Assistant Professor**, Department of Computer Science and Engineering, on their periodic inspection time to time evaluation of the project and for the support, coordination, valuable suggestions and guidance given to us in completion of the project Also we thank the members of the faculty of Computer Science and Engineering Department whose suggestions enable us to surpass many of these seemingly impossible hurdles. We also thank out guides and lastly we thank everybody who has directly or indirectly helped us in the course of this Project.

PROJECT ASSOCIATES

KRISHNA V R	**3VC19CS063**
N NAVEEN UPADHYAYA	**3VC19CS091**
N ASHOK REDDY	**3VC19CS090**
ABDUL MUIZ ALI	**3VC19CS189**

ABSTRACT

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. 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. 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.

CO-PO MAPPING

Course name: PROJECT WORK **Course Code:18CSP83**

CO NO.	Course Outcomes
C413.1	Understand and analyse the requirements of the end user considering economic, social and environmental factors for providing a feasible solution with strong basics of Computer Science subjects.
C413.2	Design, develop and demonstrate feasible solution for the identified problem with good literature survey using modern tools.
C413.3	Prepare well-structured report of the project with plagiarism check and communicate the same in different phases / journals/ conferences / project exhibitions.
C413.4	Coordinate and execute the assigned task and evaluate with the team members within in specified time in concern with the project guide.

CO-PO MAPPING														
CO No.	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO1	PSO2	PSO3
C413.1	3	3				3	3				3	2			
C413.2			3	3	3							2	3	3	3
C413.3								3				2			
C413.4									3	3		2			

CONTENTS

APPENDICES

APPENDIX A: CERTIFICATES OF ONLINE COURSES ATTENDED

APPENDIX B: CERTIFICATE OF PROJECT EXHIBITION

APPENDIX C: PAPER PUBLISHED

APPENDIX D: FINAL PLAGIARISM CHECK REPORT

APPENDIX E: INSTALLATION AND EXECUTION STEPS

LIST OF FIGURES

CHAPTER 1

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 are 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, may be attacked and 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.

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, deception attacks that inject false data into the systemem, which are carried out by abusing system components, such as sensors or controllers, and 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. Cyberdefense can be improved by using security analytics to search for hidden patterns and how to deceive.

1.1 PROBLEM STATEMENT

- The problem area of cyber hacking breach prediction is the difficulty in accurately predicting cyber-attacks and identifying vulnerabilities in real time.
- Machine learning has emerged as a promising solution for predicting cyber-attacks, but there are several challenges that need to be addressed. One of the major challenges is the lack of high-quality labeled datasets, which are required for training machine learning models.
- Another challenge is the complexity and dynamic nature of cyber-attacks. Cyber attackers are constantly developing new techniques and strategies, making it difficult to develop models that can keep up with the evolving threat landscape.
- Moreover, cyber-attacks can have significant financial, legal, and reputational consequences for organizations, making it critical to accurately predict and prevent them
- Overall, the problem area of cyber hacking breaches prediction is the need to develop accurate, robust, and scalable models that can predict cyber-attacks in real-time, despite the constantly evolving threat landscape, the lack of high-quality labeled datasets, and the complexity and dynamic nature of cyber-attacks.

1.2LITERATURE SURVEY

[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

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. False data injection attacks are considered. The aim of this paper is 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.

[2] Pajic, Miroslav, James Weimer, Nicola Bezzo, Oleg Sokolsky, George J. Pappas, and Insup Lee. "Design and implementation of attack-resilient cyber-physical 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 and industrial systems (such as the StuxNet virus attack on an industrial supervisory control and data acquisition system and the German Steel Mill cyberattack), to attacks on modern vehicles. 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. These incidents have greatly raised awareness of the need for security in cyber-physical 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.

[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).

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. The study of group behaviors for multirobot systems is the main objective of the work. Group cooperative behavior signifies that individual 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.

[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.

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.

[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.

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 is 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, 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.

1.3 JUSTIFICATION OF THE PROJECT

- The purpose of cyber hacking breaches prediction is to identify potential cyber threats and vulnerabilities before they can be exploited by attackers. Predictive cybersecurity measures use machine learning algorithms and other techniques to analyze large amounts of data, including network traffic, system logs, and user behavior, in order to detect anomalous activity that may indicate a cyber-attack is imminent or underway.
- By predicting cyber-attacks, organizations can take proactive measures to prevent or mitigate the impact of an attack. This may include implementing additional security controls, conducting security awareness training for employees, and updating security policies and procedures. Early detection and response to cyber-attacks can also help minimize financial losses, prevent data theft or loss, and protect the organization's reputation.
- In addition, cyber hacking breaches prediction can help organizations stay compliant with regulations and standards related to cybersecurity. Many regulatory frameworks, such as the General Data Protection Regulation (GDPR) and the Payment Card Industry Data Security Standard (PCI DSS), require organizations to have effective cybersecurity measures in place to protect sensitive data and prevent cyber-attacks.
- Overall, the purpose of cyber hacking breaches prediction is to proactively identify and mitigate cyber threats in order to protect the confidentiality, integrity, and availability of data, as well as the reputation and financial stability of organizations.

1.4 USES OF THE PROJECT:

- The use of cyber hacking breaches prediction is to improve an organization's cybersecurity posture by detecting and preventing cyber-attacks before they can cause harm. Predictive cybersecurity measures use machine learning algorithms and other techniques to analyze large amounts of data, including network traffic, system logs, and user behavior, in order to detect anomalous activity that may indicate a cyber-attack is imminent or underway.
- The main use of cyber hacking breaches prediction is to provide organizations with early warning and situational awareness of potential cyber threats. This allows organizations to

take proactive measures to prevent or mitigate the impact of an attack. By identifying vulnerabilities in their systems, organizations can take steps to fix them before they can be exploited by attackers.

- In addition, cyber hacking breaches prediction can help organizations comply with regulatory frameworks related to cybersecurity, such as the General Data Protection Regulation (GDPR) and the Payment Card Industry Data Security Standard (PCI DSS). Compliance with these regulations is critical to avoid legal and financial penalties.
- Another use of cyber hacking breaches prediction is to help organizations prioritize their cybersecurity efforts and allocate resources more effectively. By identifying the most critical vulnerabilities and potential attack vectors, organizations can focus their resources on protecting the most important assets.

CHAPTER 2

SYSTEM FUNCTIONAL SPECIFICATION

2.1 Function Performed

- Data Collection: The system should be able to collect data from various sources, such as network traffic, system logs, and user behavior. The data should be stored in a central repository for analysis.
- Data Preprocessing: The system should preprocess the collected data to remove noise, handle missing values, and normalize the data. This would involve various preprocessing techniques such as data cleaning, data transformation, and data reduction.
- Feature Extraction: The system should be able to extract relevant features from the preprocessed data. These features could be related to network traffic patterns, system log entries, or user behavior.
- Machine Learning Algorithms: The system should use various machine learning algorithms such as random forest, decision trees, neural networks, and support vector machines to analyze the extracted features and detect anomalous activity.
- Visualization: The system should be able to visualize the detected attacks and provide a dashboard to display real-time information on the cybersecurity posture of the organization.
- Reporting: The system should generate reports on the detected attacks and provide recommendations on how to improve the cybersecurity posture of the organization.
- Overall, the system functional specification of cyber hacking breaches prediction should include data collection, preprocessing, feature extraction, machine learning algorithms, visualization and reporting components to provide a comprehensive cybersecurity solution.

2.2 User Input Specification

The input design is link between the system and the user. The input specification mainly focuses on the data that is been uploaded by the user. By selecting the different algorithm, the user may get the accuracy of each algorithm based on the uploaded data.

2.3 User Output Specification

A quality output is one which meets the requirements of the user and the data that is been uploaded. In cyber hacking breach prediction depending upon the data that is uploaded the different algorithm is used to predict the accuracy.

2.4 External Restrictions

- Data Availability: Access to comprehensive and high-quality data is crucial for training accurate machine learning models. However, data on cyber hacking incidents may be limited, particularly for certain industries or types of attacks. Data availability issues can affect the model's performance and generalizability.
- Data Quality and Reliability: The quality and reliability of the data used for training can impact the effectiveness of the predictive models. Inaccurate or incomplete data may lead to biased or unreliable predictions.
- Evolving Tactics: Cyber attackers continuously evolve their tactics, techniques, and procedures (TTPs) to bypass security measures. Machine learning models trained on historical data may struggle to keep up with new and emerging attack methods, reducing their effectiveness in predicting novel breach scenarios.
- Lack of Labels: Obtaining labeled data where each instance is explicitly classified as a breach or non-breach can be challenging. Labeling data accurately requires expertise and may not always be feasible, leading to limitations in supervised learning approaches.

2.5 Internal Restrictions

- Model Complexity and Interpretability: Advanced machine learning models, such as deep learning neural networks, can be highly complex and often referred to as "black boxes." This lack of interpretability makes it difficult to understand and explain the rationale behind the model's predictions, which can be a challenge for regulatory compliance and trust.

- False Positives and Negatives: Predictive models may generate false positive or false negative predictions. False positives can lead to unnecessary alerts and increased workload for security teams, while false negatives can result in undetected breaches. Balancing the trade-off between these two types of errors is crucial but can be challenging.

- Adversarial Attacks: Machine learning models used in cyber hacking breach prediction can be susceptible to adversarial attacks. Adversaries may attempt to manipulate input data to deceive the model or exploit vulnerabilities in the learning algorithms themselves, compromising the accuracy and reliability of predictions.

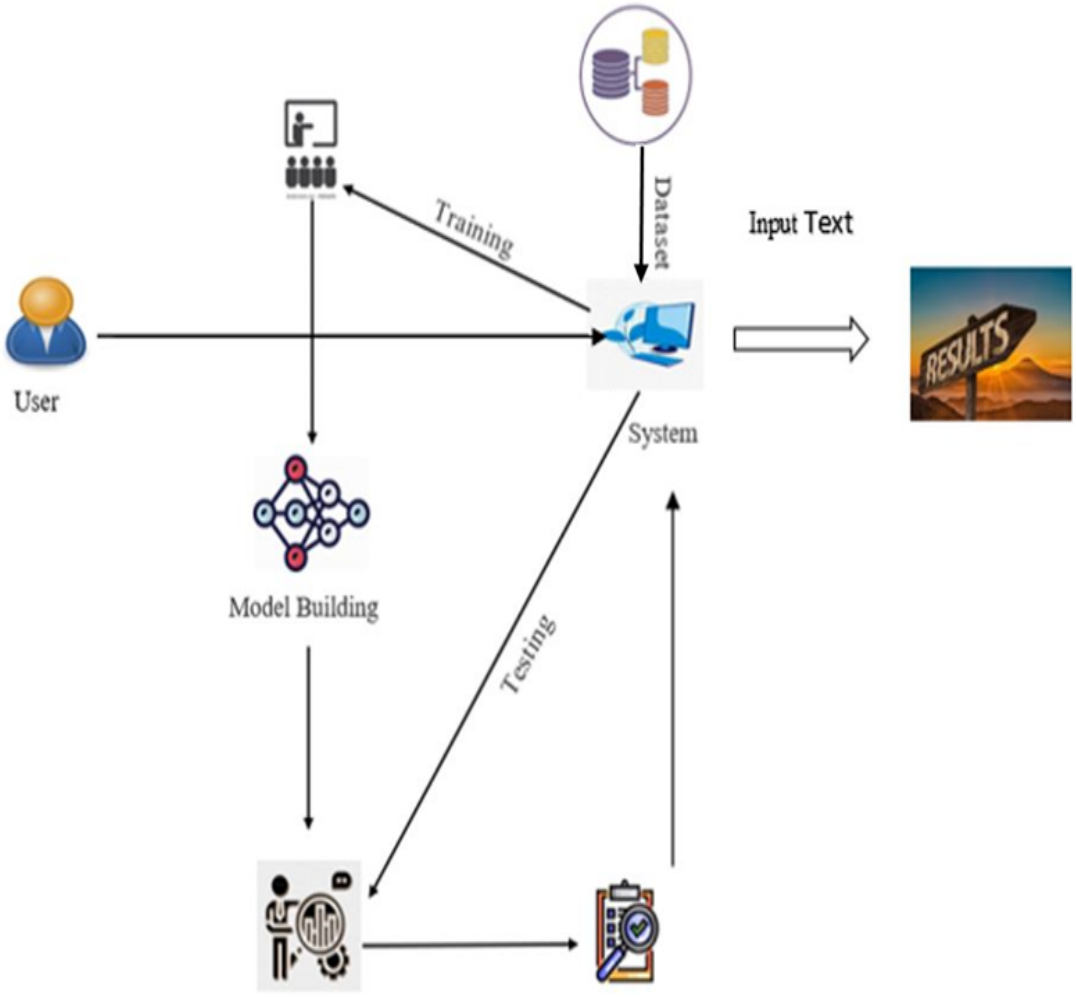
- Resource Requirements: Developing and maintaining an effective cyber hacking breach prediction system using machine learning requires significant computational resources, storage capacity, and expertise. Implementing and managing the necessary infrastructure and personnel can be costly and resource-intensive.

CHAPTER 3

SYSTEM DESIGN

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.

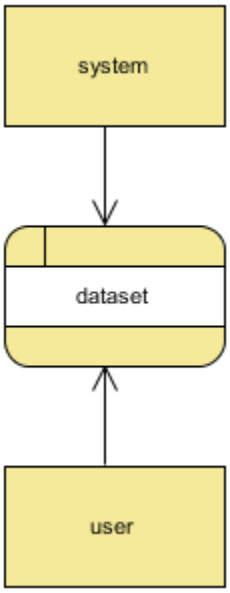
3.1 System Architecture



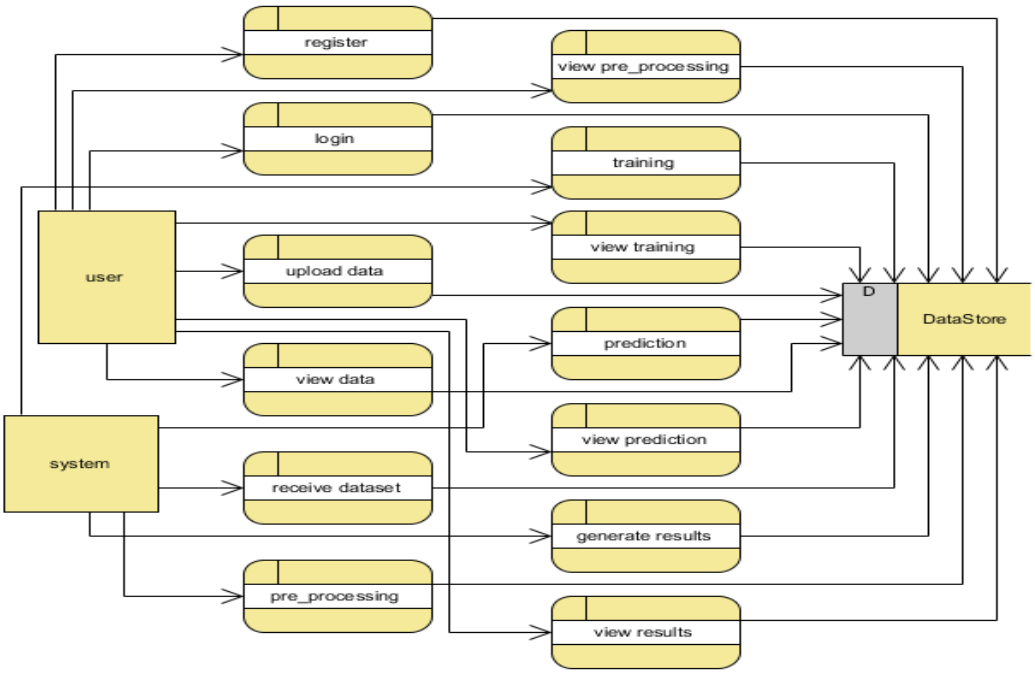
3.2 System Data Flow Diagram

A Data Flow Diagram (DFD) is a traditional way to visualize the information flows within a system. A neat and clear DFD can depict a good amount of the system requirements graphically. It can be manual, automated, or a combination of both. It shows how information enters and leaves the system, what changes the information and where information is stored. The purpose of a DFD is to show the scope and boundaries of a system as a whole. It may be used as a communications tool between a systems analyst and any person who plays a part in the system that acts as the starting point for redesigning a system.

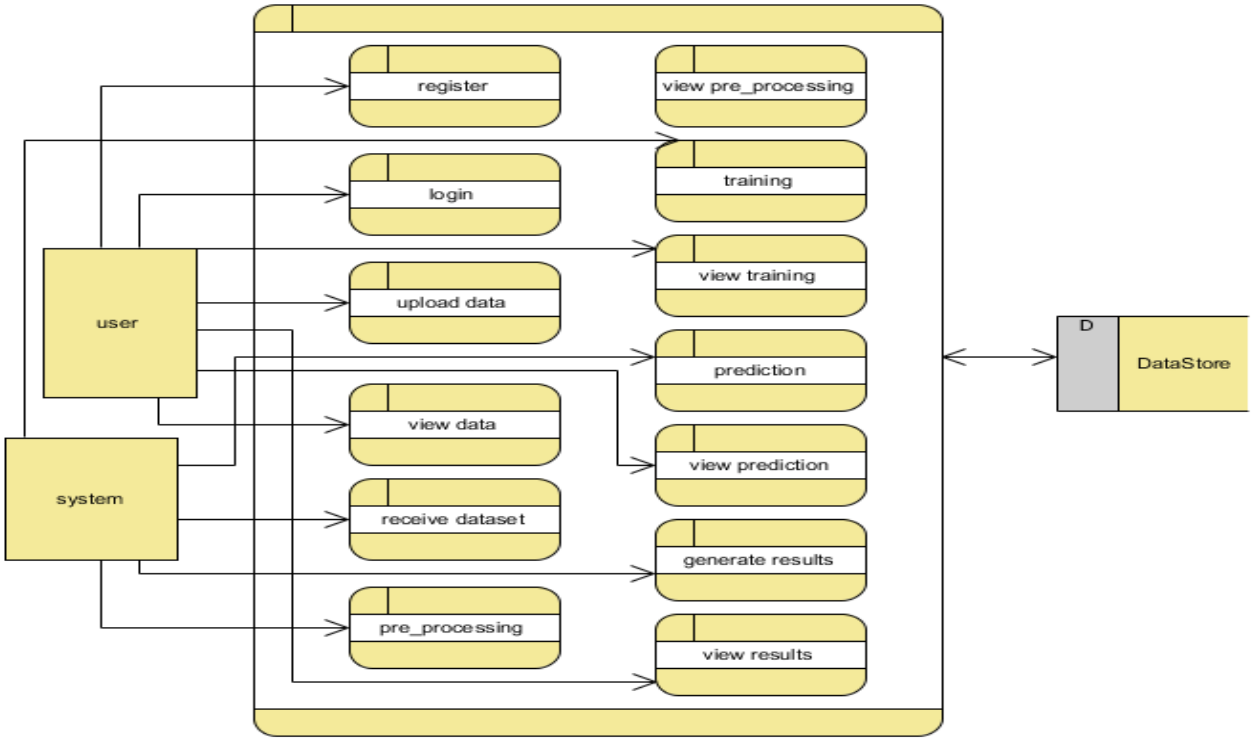
Contrast Level:



Level 1 Diagram:



Level 2 Diagram:



3.3 Description of System Operation

3.3.1 User:

1. **Register:**

   Users can register for the Cyber hacking breaches prediction using machine learning application here.

2. **Login:**

   After registering, the user can access his portal**.**

3. **View Home page:**

   Here user view the home page of the Cyber hacking breaches prediction using machine learning web application.

4. **View about page:**

   In the about page, users can learn more about the poverty classification.

5. **Input Model:**

   The user must provide input values for the certain fields in order to get results.

6. **View Results:**

   User view's the generated results from the model.

7. **View score:**

   Here user have ability to view the score in percentage.

3.3.2 System

1. **Working on dataset:**

System checks for data whether it is available or not and load the data in csv files.

2. **Pre-processing:**

Data need to be pre-processed according the models it helps to increase the accuracy of the model and better information about the data.

3. **Training the data:**

After pre-processing the data will split into two parts as train and test data before training with the given algorithms.

4. **Model Building**

To create a model that predicts the personality with better accuracy, this module will help user.

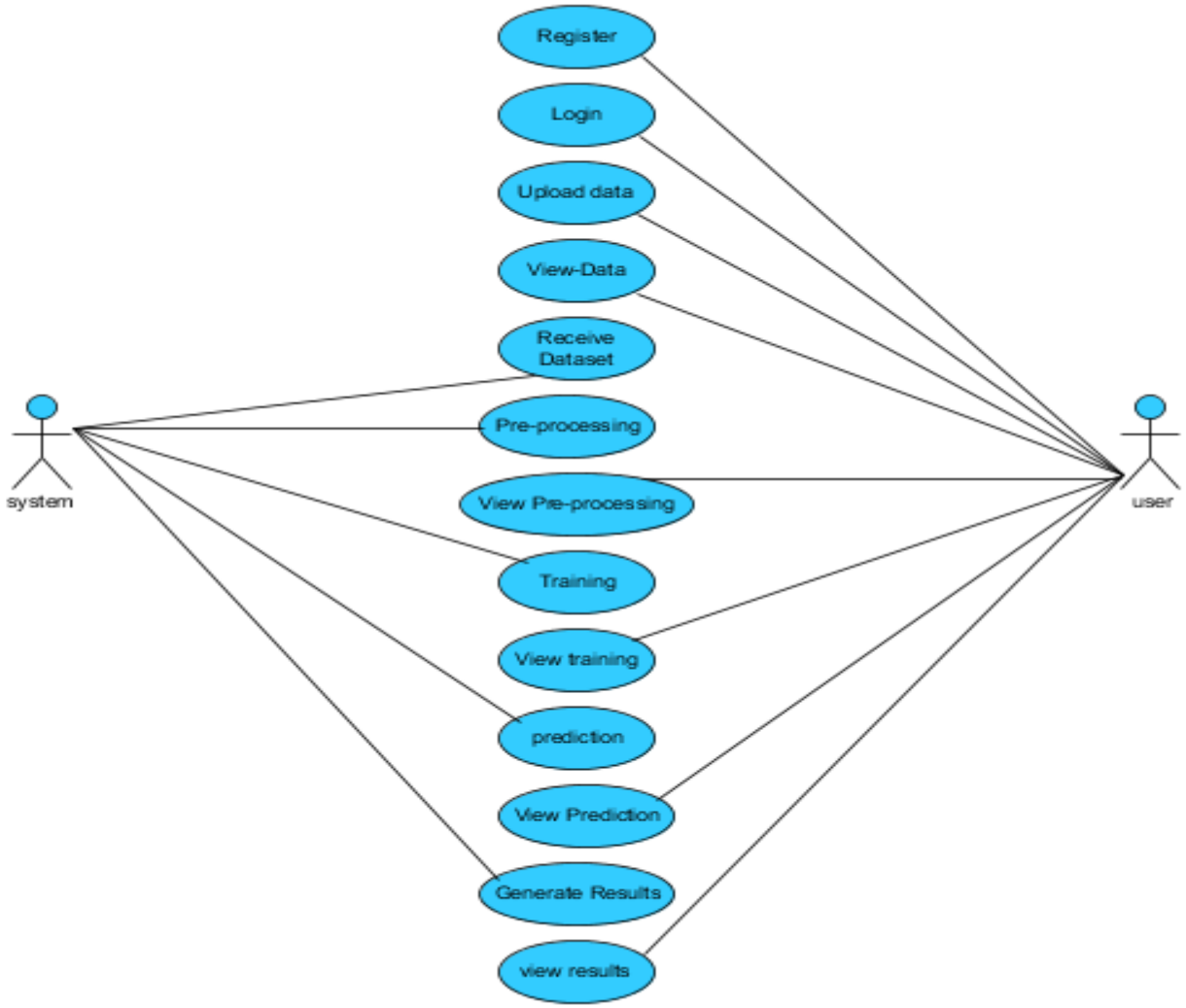
5. **Generated Score:**

Here user view the score in percentage.

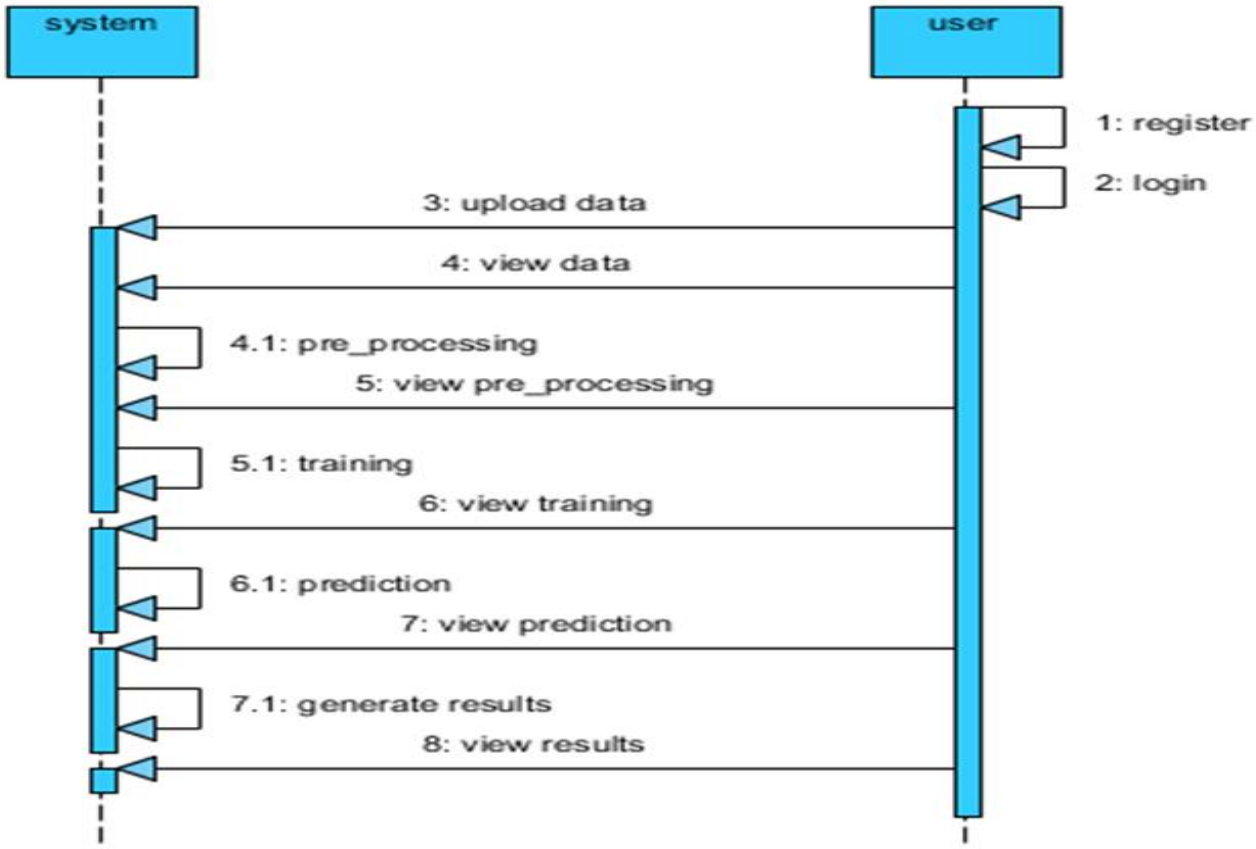
6. **Generate Results:**

We train the machine learning algorithm and predict the cyber-attack detection.

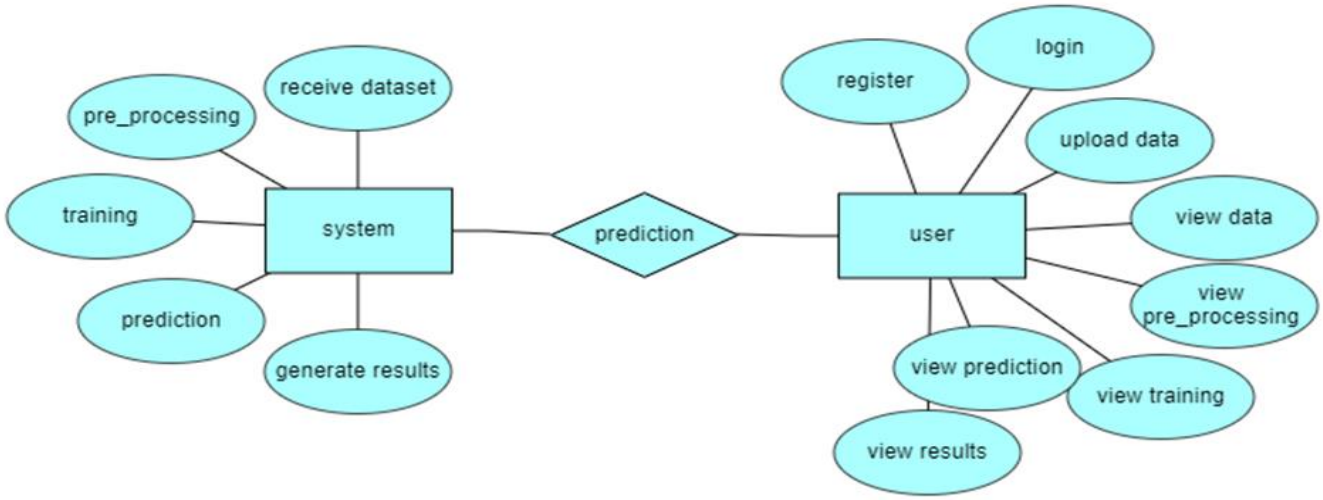
3.4 USE CASE DIAGRAMS



3.5 Sequence Diagram



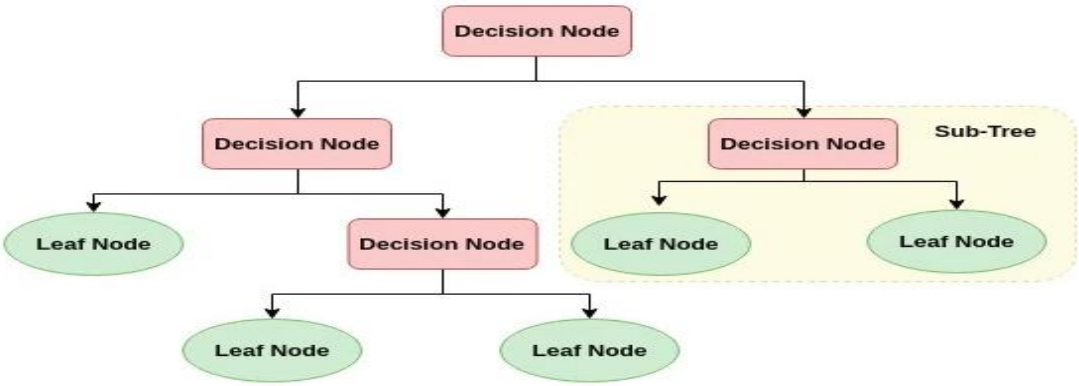
3.6 ER Diagram



3.7 Algorithm Specification

1. 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.

2. 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.

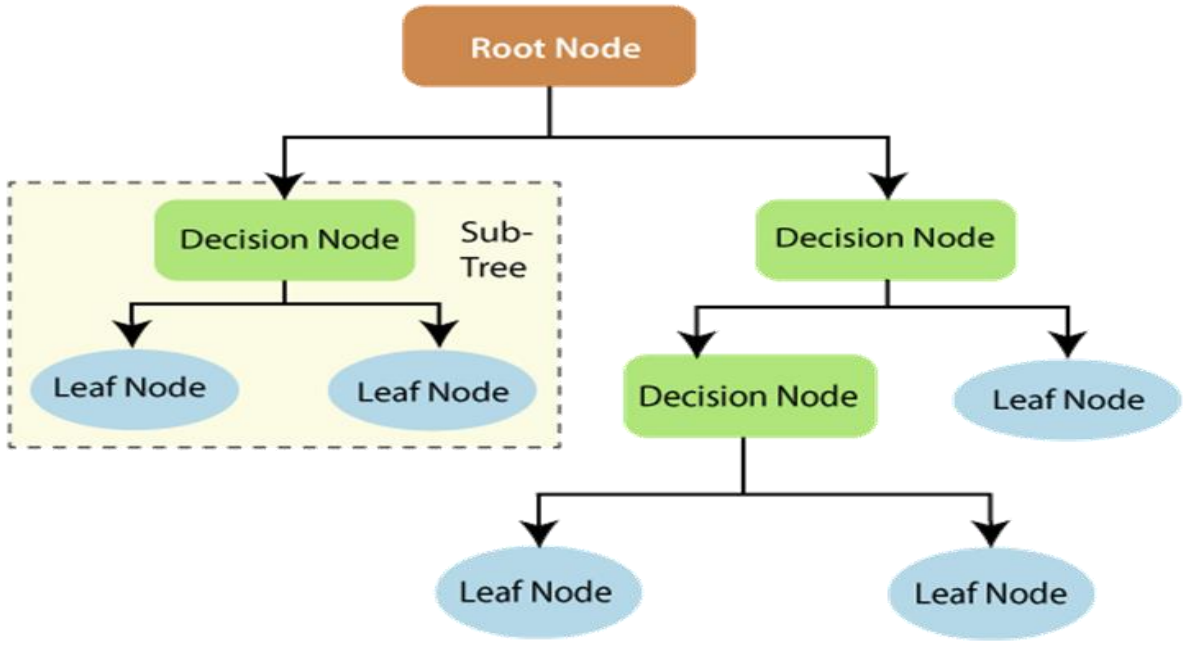
A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

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.



Support Vectors

Support vectors are data points that are closer to the hyper plane and influence the position and orientation of the hyper plane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyper plane. These are the points that help us build our SVM.

Large Margin Intuition

In logistic regression, we take the output of the linear function and squash the value within the range of [0,1] using the sigmoid function. If the squashed value is greater than a threshold value (0.5) we assign it a label 1, else we assign it a label 0. In SVM, we take the output of the linear function and if that output is greater than 1, we identify it with one class and if the output is -1, we identify is with another class. Since the threshold values are changed to 1 and -1 in SVM, we obtain this reinforcement range of values ([-1, 1]) which acts as margin.

Cost Function and Gradient Updates

In the SVM algorithm, we are looking to maximize the margin between the data points and the hyper plane. The loss function that helps maximize the margin is hinge loss.

$$c(x, y, f(x)) = \begin{cases} 0, & \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases}$$

$$c(x, y, f(x)) = (1 - y * f(x))_+$$

Hinge loss function (function on left can be represented as a function on the right)

The cost is 0 if the predicted value and the actual value are of the same sign. If they are not, we then calculate the loss value. We also add a regularization parameter the cost function. The objective of the regularization parameter is to balance the margin maximization and loss. After adding the regularization parameter, the cost functions look as below.

$$min_w \lambda \parallel w \parallel^2 + \sum_{i=1}^{n} (1 - y_i \langle x_i, w \rangle)_+$$

Loss function for SVM

Now that we have the loss function, we take partial derivatives with respect to the weights to find the gradients. Using the gradients, we can update our weights.

$$\frac{\delta}{\delta w_k} \lambda \parallel w \parallel^2 = 2\lambda w_k$$

$$\frac{\delta}{\delta w_k} \left(1 - y_i \langle x_i, w \rangle\right)_+ = \begin{cases} 0, & \text{if } y_i \langle x_i, w \rangle \geq 1 \\ -y_i x_{ik}, & \text{else} \end{cases}$$

Gradients

When there is no misclassification, i.e. our model correctly predicts the class of our data point, we only have to update the gradient from the regularization parameter.

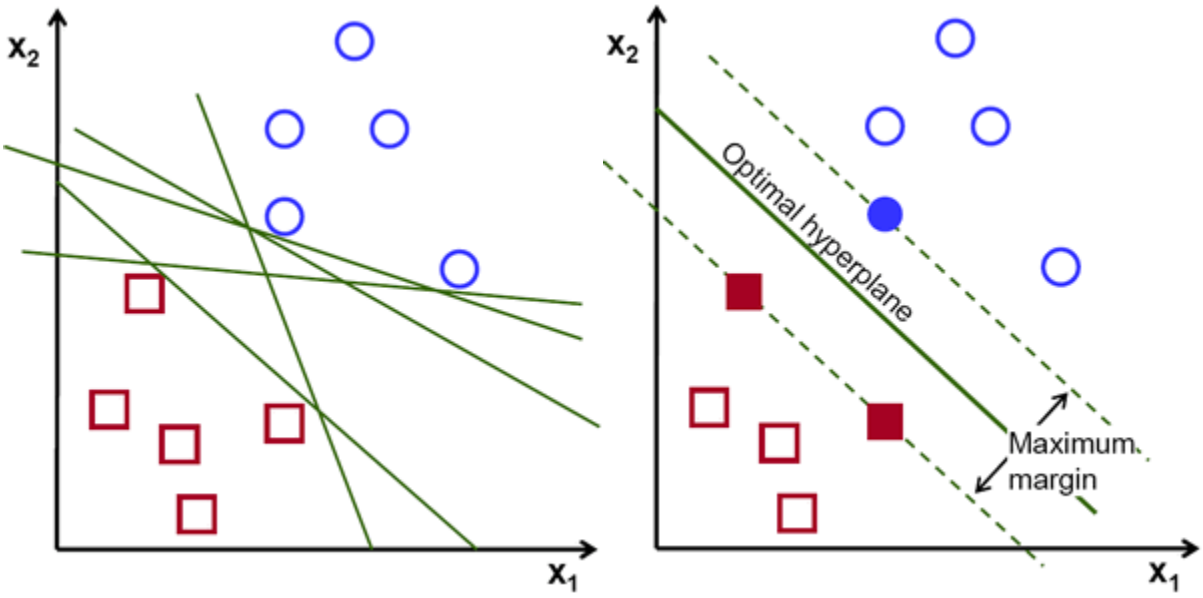
$$w = w - \alpha \cdot (2\lambda w)$$

Gradient Update — No misclassification

When there is a misclassification, i.e. our model make a mistake on the prediction of the class of our data point, we include the loss along with the regularization parameter to perform gradient update.

3. SUPPORT VECTOR MACHINES:

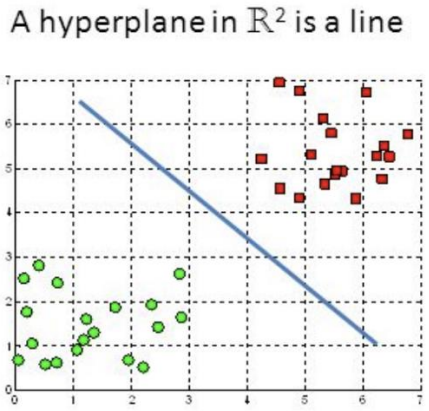
The objective of the support vector machine algorithm is to find a hyper plane in an N-dimensional space (N — the number of features) that distinctly classifies the data points.

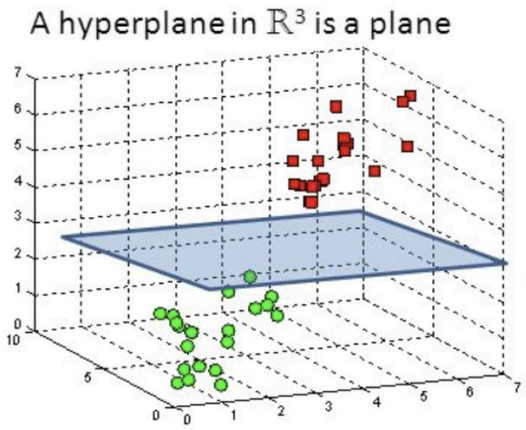


Possible hyper planes :

To separate the two classes of data points, there are many possible Hyper planes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e. the maximum distance between data points of both classes. Maximizing the margin distance provides some

reinforcement so that future data points can be classified with more confidence.





4.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, LightGBM.

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.

3.8 Equipment Configuration

Hardware:

Operating system : Windows 7 or 7+

RAM : 8 GB

Hard disc or SSD : More than 500 GB

Processor : Intel 3rd generation or high or Ryzen with 8 GB Ram

Software:

Software's : Python 3.6 or high version

IDE : PyCharm.

3.9 Implementation Languages

The language used for the implementation of this project is Python, it has become one of the most popular programming languages in the world recent years.

Why Python?

- It has a simple syntax that mimics natural language, so it's easier to read and understand. This makes is quicker to build projects, and faster to improve on them.
- It's versatile. Python can be used for many different tasks, from web development to machine learning.
- It's beginner friendly, making it popular for entry-level coders.
- It's open source, which means it's free to use and distribute, even for commercial purposes.
- Python's archive of modules and libraries-bundles of code that third- party users have created to expand Python's capabilities-is vast and growing
- Python has a large and active community that contributes to Python's pool of modules and libraries, and acts as a helpful resource for other programmers. The vast support community means that if coders run into a stumbling block, finding a solution is relatively easy; somebody is bound to have run into the same problem before.

CHAPTER 4

SYSTEM VERIFICATION

Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software testing can also provide an objective, independent view of the software to allow the business to appreciate and understand the risks of software implementation. Test techniques inched, but are not limited to, the process of executing a program or application with the intent of finding software bugs.

Software testing can be stated as the process of validating and verifying of a software program/application/product which meets the requirements that guided its design and development also works as expected and can be implemented with the same characteristics Software testing, depending on the testing method employed, can be implemented at any time in the development process. However, most of the test effort traditionally occurs after the requirements have been defined and the coding process has been completed having been shown that fixing a bug is less expensive when found earlier in the development process. Although in the Agile approaches most of the test effort is, conversely, on-going. As such, the methodology of the test is governed by the software development methodology adopted.

Different software development models will focus the test effort at different points in the development process. Newer development models, such us Agile, often employ test driven development and place un increased portion of the testing in the hands of the developer, before it reaches a formal team of testers.

In a more traditional model, most of the test execution occurs after the requirements have been defined and the coding process has been completed.

Based on various parameters there are different methods of testing. A few commonly used ones are as follows.

Functional and Non-functional testing

Functional testing refers to activities that verify a specific action or function of the code. These are usually found in the code requirements documentation, although some development methodologies work from use cases or user stories. Functional tests tend to answer the question of "can the user do this" or "does this particular feature work."

Non-functional testing refers to uspects of the software that may not be related to a specific function or user action, such as scalability or other performance, behavior under certain constraints, or security Testing will determine the flake point, the point at which extremes of scalability or performance leads to unstable execution Non-functional requirements tend to be those that reflect the quality of the product, particularly in the context of the suitability perspective of its users.

Compatibility testing

A common cause of software failure (real or perceived) is a lack of its compatibility with other application software, operating systems (or operating system versions, old or new), or target environments that differ greatly from the original (such as a terminal or GUI application intended to be run on the desktop now being required to become a web application, which must render in a web browser). For example, in the case of a lack of backward compatibility, this can occur because the programmers develop and test software only on the latest version of the target environment, which not all users may be running. This results in an unintended consequence that the latest work may not function on earlier versions of the target environment, or on older hardware that carlier versions of the target environment was capable of using. Sometimes such issues can be fixed by proactively abstracting operating system functionality into a separate program module or library.

Verification and Validation

The process of evaluating software to determine whether the products of a given development phase satisfy the conditions imposed at the start of that phase is called Verification.

The process of evaluating software during or at the end of the development process to determine whether it satisfies specified requirements is called Validation. Validation checks that the product design satisfies or fits the intended usage (high-level checking)-i.e., you built the right product. This is done through dynamic testing and other forms of review,

Validation ensures that the product actually meets the user's needs, and that the specifications were correct in the first place, while verification is ensuring that the product has been built according to the requirements and design specifications. Validation ensures that you built the right thing. Verification ensures that you built it right Validation confirms that the product, as provided, will fulfill its intended use.

4.1 FUNCTIONS TO BE TESTED

- Data Preprocessing
- Model Training
- Model Evaluation
- False Positive and False Negative Analysis
- Performance Comparison

4.2 TESTING METHODOLOGIES

Software testing methods are traditionally divided into white and black-box testing. These two approaches are used to describe the point of view that a test engineer takes when designing test cases.

- White-box testing is when the tester has access to the internal data structures and algorithms including the code that implements these.

- Black-box testing treats the software as a "black box-without anyknowledge of internal implementation.
- Grey-box testing involves having knowledge of internal data structures and algorithms for purposes of designing tests, while executing thosetests at the user, or black-box level.

4.3 TESTING LEVELS

Tests are frequently grouped by where they are added in the software development process, or by the level of specificity of the test. The main levels during the development process are unit, integration, and systems testing that are distinguished by the test target without implying a specific process model.

4.3.1 Unit Testing

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

4.3.2 Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g., components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

4.3.3 Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

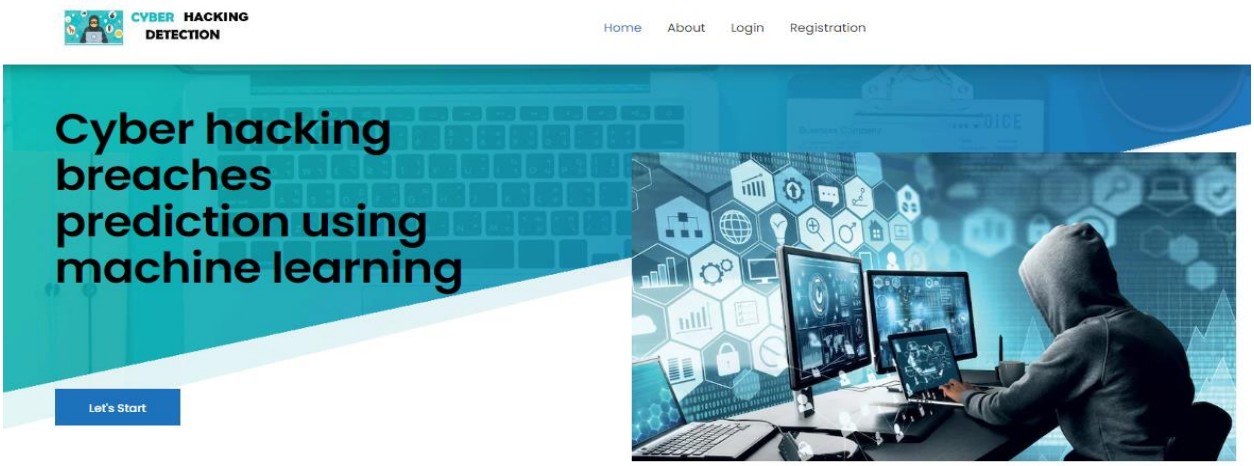
CHAPTER 5

RESULTS

5.1 USER SCREENS

5.1.1 Home Page:

Here user view the home page of Cyber hacking breaches prediction using machine learning web application.

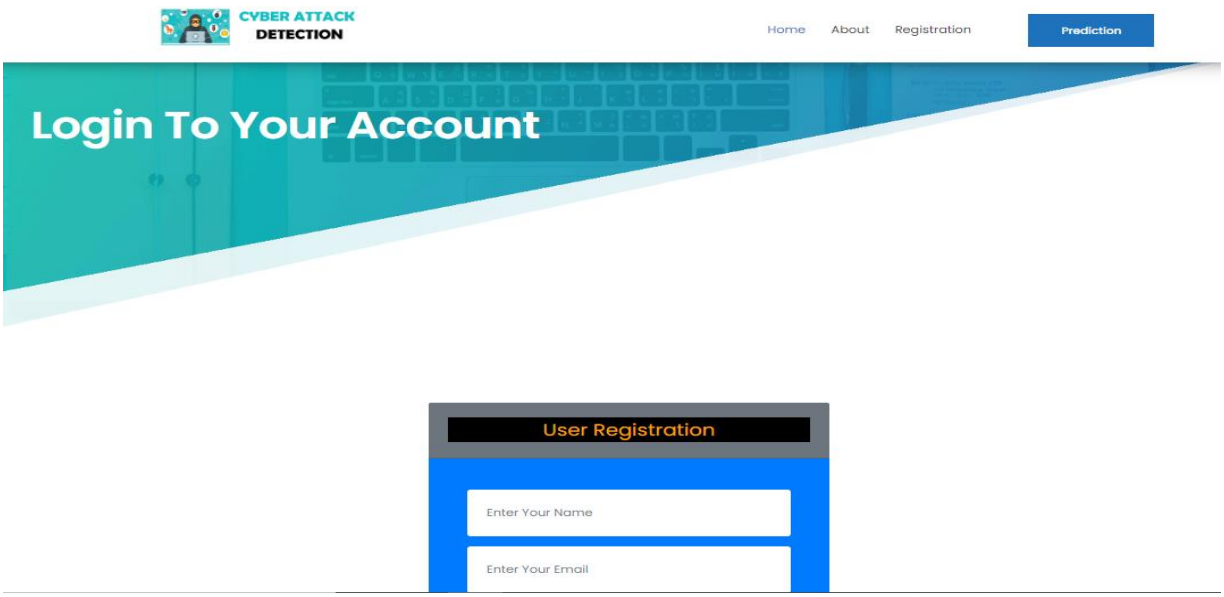


5.1.2 About:

Here we can read about our project.

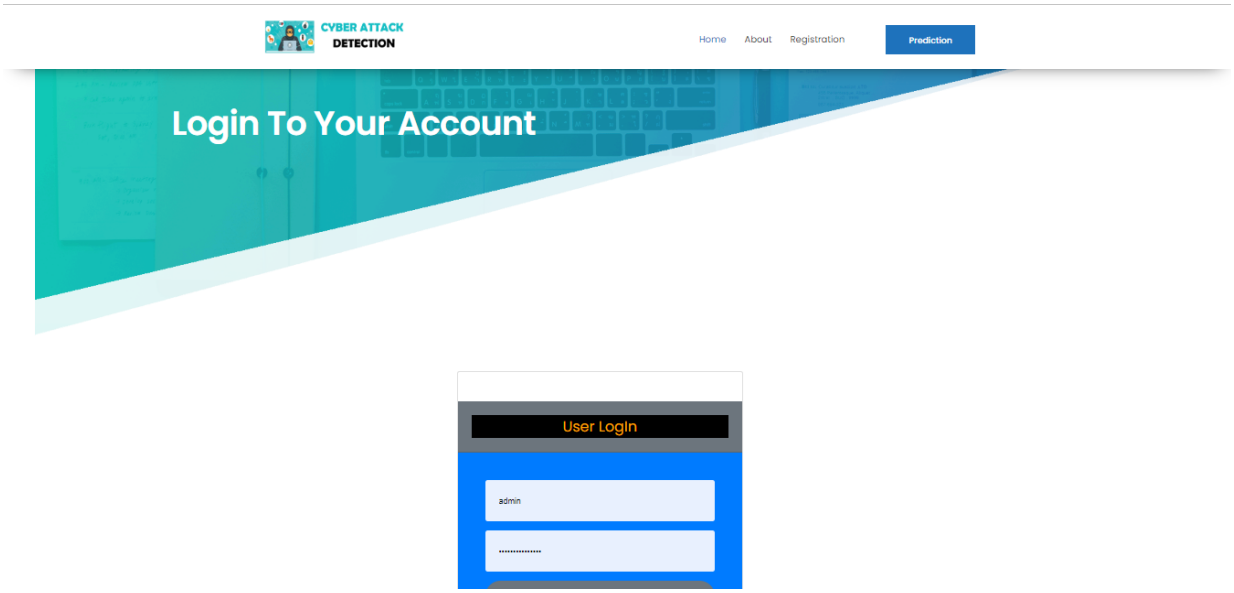
5.1.3 Register:

In the page, users need to register by entering his credentials.



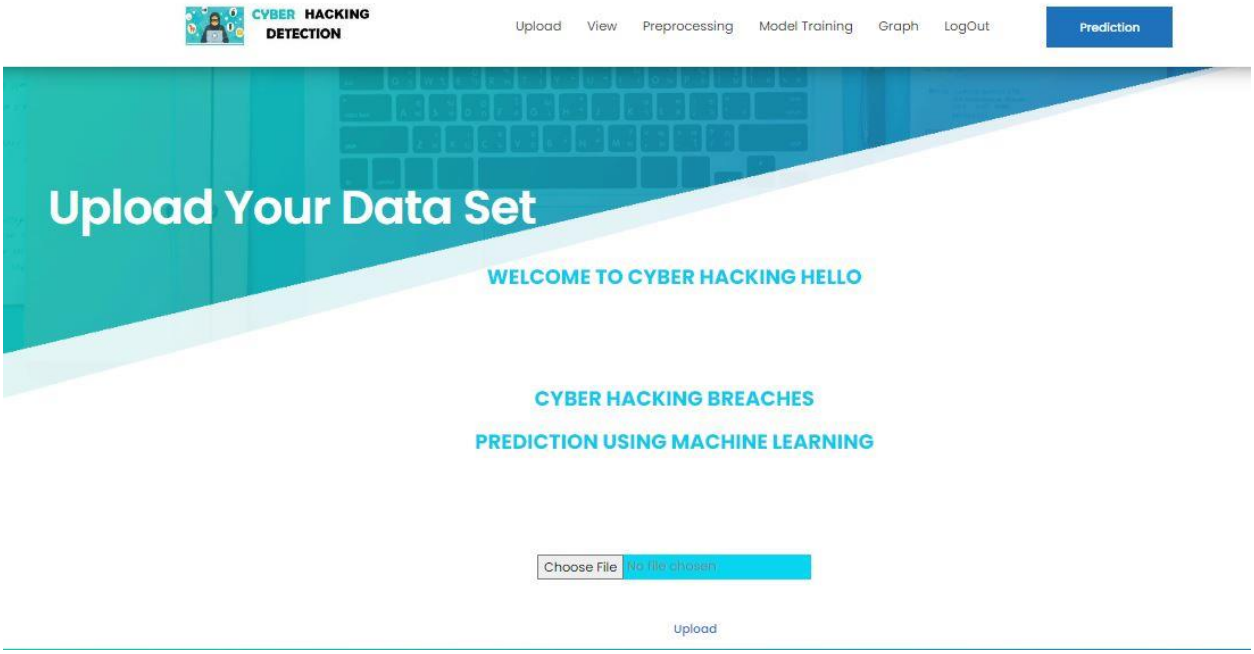
5.1.4 Log in:

In the page, users have to enter the credentials to enter into the Cyber hacking breaches prediction using machine learning.



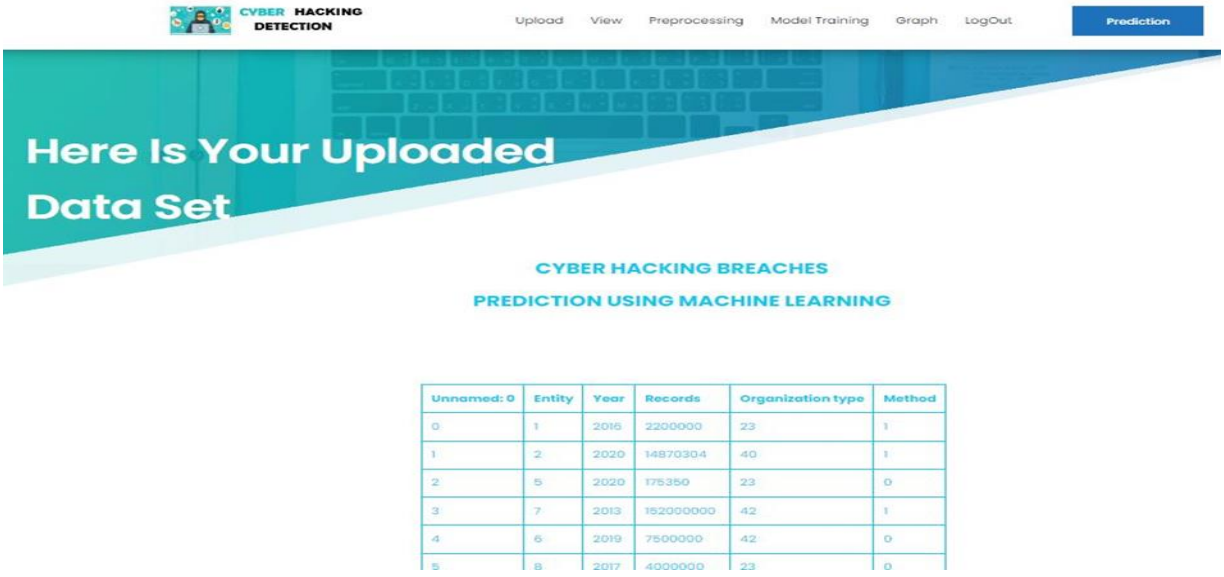
5.1.5 Load:

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



5.1.6 View:

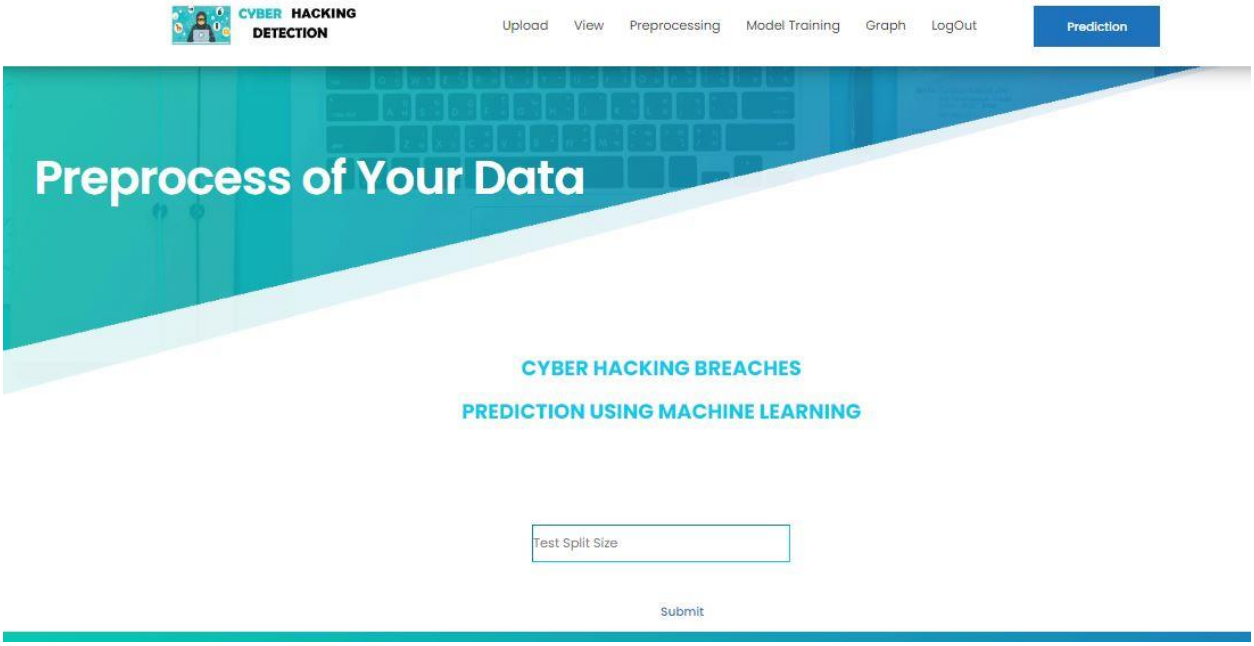
Here we can see the uploaded data set.



Unnamed: 0	Entity	Year	Records	Organization type	Method
0	1	2016	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

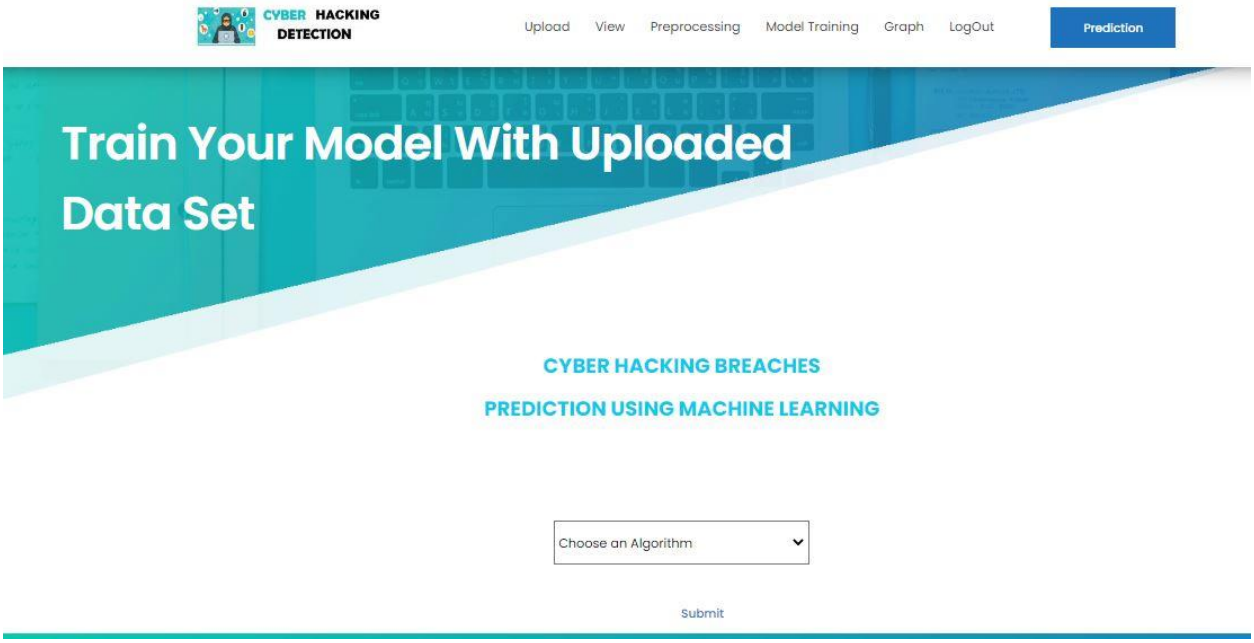
5.1.7 Pre-process:

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



5.1.8 Model:

Here we train our data with different ML algorithms.



5.1.9 Prediction:

This page shows the detection result of theCyber hacking breaches prediction using machine learning data.

5.1.10 Graph:

CHAPTER 6

CONCLUSION

6.1 SUMMARY

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 anCyber 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 improvements.

6.2 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.

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, InakiEsnaola, FatosTunayYarmanVural, 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.

Certificate of Achievement

Short Course: Cisco CCNA Security

This is to certify that

KRISHNA V R

has successfully completed the Short Course

Cisco CCNA Security

Grade: Distinction (85/100)

Lecturer: Matt Constable

Completed: January 8, 2023

CHale

Chantelle Hale
CEO, IT Masters
Adjunct Lecturer, CSU

IT Masters
itmasters.edu.au

Charles Sturt University

Certificate of Achievement

Short Course: Cisco CCNA Security

This is to certify that

N Naveen Upadhyaya

has successfully completed the Short Course

Cisco CCNA Security

Grade: High Distinction (85/100)

Lecturer: Matt Constable

Completed: January 12, 2023

CHale

Chantelle Hale
CEO, IT Masters
Adjunct Lecturer, CSU

IT Masters
itmasters.edu.au

Charles Sturt University

Certificate of Achievement

Short Course: Cisco CCNA Security

This is to certify that

N Ashok Reddy

has successfully completed the Short Course

Cisco CCNA Security

Grade: High Distinction (86/100)

Lecturer: Matt Constable

Completed: January 12, 2023

CHale

Chantelle Hale
CEO, IT Masters
Adjunct Lecturer, CSU

IT Masters
itmasters.edu.au

Charles Sturt University

Certificate of Achievement

Short Course: Cisco CCNA Security

This is to certify that

Abdul Muiz Ali

has successfully completed the Short Course

Cisco CCNA Security

Grade: High Distinction (87/100)

Lecturer: Matt Constable

Completed: January 12, 2023

CHale

Chantelle Hale
CEO, IT Masters
Adjunct Lecturer, CSU

IT Masters
itmasters.edu.au

Charles Sturt University

V. V. Sangha's

RAO BAHADUR Y. MAHABALESWARAPPA ENGINEERING COLLEGE

Cantonment, Ballari-583104, Karnataka.
08392-244809, Telefax: 08392-242148, Website: www.rymec.in, email: principalrymec@gmail.com
(Affiliated to VTU, Belagavi and Approved by AICTE, New Delhi)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

INSTITUTION'S INNOVATION COUNCIL

NBA NATIONAL BOARD OF ACCREDITATION

CERTIFICATE OF MERIT

This is to certify that Ms./Mr. KRISHNA V R of

has been Awarded the Certificate of Participation in State Level Project Exhibition 2023 Conducted by

Department of Computer Science and Engineering on 4th May 2023.

HOD CSE
Dr. H. Girisha

Principal
Dr. T. Hanumantha Reddy

RAO BAHADUR Y. MAHABALESWARAPPA ENGINEERING COLLEGE

Cantonment, Ballari-583104, Karnataka.
08392-244809, Telefax: 08392-242148, Website: www.rymec.in, email: principalrymec@gmail.com
(Affiliated to VTU, Belagavi and Approved by AICTE, New Delhi)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

INSTITUTION'S INNOVATION COUNCIL

NBA NATIONAL BOARD OF ACCREDITATION

CERTIFICATE OF MERIT

This is to certify that Ms./Mr. N. NAVEEN UPADHYAYA *of*

has been Awarded the Certificate of Participation in State Level Project Exhibition 2023 Conducted by

Department of Computer Science and Engineering on 4th May 2023.

HOD CSE
Dr. H. Girisha

Principal
Dr. T. Hanumantha Reddy

V. V. Sangha's

RAO BAHADUR Y. MAHABALESWARAPPA ENGINEERING COLLEGE

Cantonment, Ballari-583104, Karnataka.
08392-244809, Telefax: 08392-242148, Website: www.rymec.in, email: principalrymec@gmail.com
(Affiliated to VTU, Belagavi and Approved by AICTE, New Delhi)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

INSTITUTION'S INNOVATION COUNCIL

NBA NATIONAL BOARD of ACCREDITATION

CERTIFICATE OF MERIT

This is to certify that Ms./Mr. N. ASHOK REDDY of

has been Awarded the Certificate of Participation in State Level Project Exhibition 2023 Conducted by

Department of Computer Science and Engineering on 4th May 2023.

HOD CSE
Dr. H. Girisha

Principal
Dr. T. Hanumantha Reddy

V. V. Sangha's

RAO BAHADUR Y. MAHABALESWARAPPA ENGINEERING COLLEGE

Cantonment, Ballari-583104, Karnataka.
08392-244809, Telefax: 08392-242148, Website: www.rymec.in, email: principalrymec@gmail.com
(Affiliated to VTU, Belagavi and Approved by AICTE, New Delhi)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

INSTITUTION'S INNOVATION COUNCIL

NBA NATIONAL BOARD OF ACCREDITATION

CERTIFICATE OF MERIT

This is to certify that Ms./Mr. ABDUL MUIZ ALI *of*

has been Awarded the Certificate of Participation in State Level Project Exhibition 2023 Conducted by Department of Computer Science and Engineering on 4th May 2023.

HOD CSE
Dr. H. Girisha

Principal
Dr. T. Hanumantha Reddy

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

IJARSCT

CERTIFICATE OF PUBLICATION

INTERNATIONAL STANDARD SERIAL NUMBER
ISSN NO: 2581-9429

THIS IS TO CERTIFY THAT

Krishna V R

Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka, India

HAS PUBLISHED A RESEARCH PAPER ENTITLED

Cyber Hacking Breaches Prediction using Machine Learning

IN IJARSCT, VOLUME 3, ISSUE 7, APRIL 2023

Certificate No: 042023-A1916
www.ijarsct.co.in

Crossref
DOI: 10.48175/IJARSCT-9479
www.doi.org
www.crossref.org

Scientific Journal Impact Factor 7.301 Journal Impact Factor
www.sjifactor.com

Editor-in-Chief

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

IJARSCT

CERTIFICATE OF PUBLICATION

INTERNATIONAL STANDARD SERIAL NUMBER
ISSN NO: 2581-9429

THIS IS TO CERTIFY THAT

N Ashok Reddy

Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka, India

HAS PUBLISHED A RESEARCH PAPER ENTITLED

Cyber Hacking Breaches Prediction using Machine Learning

IN IJARSCT, VOLUME 3, ISSUE 7, APRIL 2023

Certificate No: 042023-A1917

www.ijarsct.co.in

Crossref
DOI: 10.48175/IJARSCT-9479
www.doi.org

www.crossref.org

7.301
Scientific Journal Impact Factor
Journal Impact Factor

www.sjifactor.com

ISSN 2581-9429 IJARSCT

Editor-in-Chief

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

IJARSCT

CERTIFICATE OF PUBLICATION

INTERNATIONAL STANDARD SERIAL NUMBER
ISSN NO: 2581-9429

THIS IS TO CERTIFY THAT

N Naveen Upadhyaya

Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka, India

HAS PUBLISHED A RESEARCH PAPER ENTITLED

Cyber Hacking Breaches Prediction using Machine Learning

IN IJARSCT, VOLUME 3, ISSUE 7, APRIL 2023

Certificate No: 042023-A1918

www.ijarsct.co.in

Crossref
DOI: 10.48175/IJARSCT-9479
www.doi.org

www.crossref.org

7.301
Scientific Journal Impact Factor
Journal Impact Factor

www.sjifactor.com

Editor-in-Chief

International Journal of Advanced Research In Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

IJARSCT

CERTIFICATE Of Publication

INTERNATIONAL STANDARD SERIAL NUMBER
ISSN NO: 2581-9429

THIS IS TO CERTIFY THAT

Abdul Muiz Ali

Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka, India

HAS PUBLISHED A RESEARCH PAPER ENTITLED

Cyber Hacking Breaches Prediction using Machine Learning

In IJARSCT, Volume 3, Issue 7, April 2023

Certificate No: 042023-A1919

www.ijarsct.co.in

Crossref
DOI: 10.48175/IJARSCT-9479
www.doi.org

www.crossref.org

7.301
Scientific Journal Impact Factor
Journal Impact Factor

www.sjifactor.com

Editor-in-Chief

IJARSCT

Impact Factor: 7.301

Scientific Journal Impact Factor

www.sjifactor.com

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

ISSN No. : 2581-9429

A Double Blind Peer-Reviewed Refereed Monthly Journal

Volume 3, Issue 7, April 2023

SCOPE DATABASE INDEXED

www.ijarsct.co.in

DOI: 10.48175/568

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

Cyber Hacking Breaches Prediction using Machine Learning

Sampath Kumar R[1], Krishna V R[2], N Ashok Reddy[2], N Naveen Upadhyaya[4], Abdul Muiz Ali[5]
Assistant Professor, Department of Computer Science[1]
Students, Department of Computer Science[2,3,4,5]
Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka, India
krishna.cse.rymec@gmail.com, ashokreddy.cse.rymec@gmail.com,
naveen.cse.rymec@gmail.com, muiz.cse.rymec@gmail.com

Abstract: *Cyber-attacks are a major threat to these systems. Unlike faults that occur by accidents in cyber-physical systems, cyber-attacks occur intelligently and stealthily. 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. 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. 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 misbehaving agent. Experimental analysis shows us that deep learning algorithms can detect attacks with higher performance than usual methods and can make cyber security simpler, more proactive, less expensive, and far more effective.*

Keywords: Cyber-attacks, Cyber Hacking, Cyber-hacking breaches

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 exposed tocyber-hacking. Cyber-physical systems are made up of different typesof logical elements and embedded system computers. Which will be communicating with communication channels such as the Internet of Things. More specifically, these systems include different digital or cyber components, analogy components, physical devices, and humans that are designed to operate between physical and cyber parts. In other words, we can say that a cyber-physical system is a type of system that includes cyber components and physical components and humans and can 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. The Physical component, which includes sensors that receive data from the physical environment, can be attacked and injected with 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 a large number of data, with so much variety, and at high speed. Also, the connection between the sensors, the necessary calculations, and the analysis of the obtained data will be one of the main challenges. Therefore, one of the most important features of a cyber-physical system is to communicate between these sensors and 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 different 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, deception attacks that administer false data to the system, are carried out by abusing system components, such as sensors or controllers, and they can corrupt data or enter incorrect information into the system and cause misbehaving.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

II. LITERATURE SURVEY

Here is a literature survey on the paper presents a security analysis of cyber-physical systems (CPS) against stealthy deception attacks, which are aimed at manipulating the behavior of the system without being detected model, Kwon, Cheolhyeon, Weiyi Liu, and Inseok Hwang "Security analysis for cyber-physical systems against stealthy deception attacks using machine learning model, in IEEE (2013). This paper proposed the design and implementation of attack-resilient cyber-physical systems with a focus on attack-resilient state estimators. The authors propose a framework for designing resilient systems that includes threat models, security requirements, attack detection and identification, and response strategies. Pajic, Miroslav, James Weimer, Nicola Bezzo, Oleg Sokolsky, George J. Pappas, and Insup Lee. "a Design and implementation of attack-resilient cyber-physical systems: With a focus on attack-resilient state estimators' model", IEEE Control Systems (2017). This paper presents a consensus formation control algorithm for a group of mobile robots with directed communication networks. The proposed algorithm ensures that the robots converge to a desired formation and maintain it despite disturbances and communication delays, 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). This paper proposes a distributed control approach for networked control systems that can tolerate the presence of misbehaving agents. The authors describe a method for designing a resilient controller that can mitigate the effects of malicious or faulty agents in a network, while still achieving satisfactory performance, Zeng, Wente, and Mo-Yuen Chow. "Resilient distributed control in the presence of misbehaving agents in networked control systems", IEEE Transactions on Cybernetics (2014).The authors propose a resilient control scheme that uses a state feedback controller and an observer-based controller to mitigate the effects of the DoS attacks model, 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).

III. PROPOSED SYSTEM

Many machine learning models have been proposed to determine whethera cyberattack is likely to occur, but it is not enough to solve this serious problem. In addition, similar studies that propose models for evaluating such activities often do not take into account the variability andsize of the data. Therefore, we propose support vectors, decision trees, random forests, and Cat Boost classifier techniques.

IV. METHODOLOGY

Using machine learning to predict network hacking is an active area of research and development. Here is an overview of the proposed system:

- Data collection: Collecting and analyzing historical data on cyber-attacks and security breach is the first step in creating a forecast. This information may include information about the type of attack, the target, the time of the attack, and other relevant factors.
- Preliminary Data: Collected data must be pre-processed before being fed into the machine learning algorithm. These steps include removing missing values, excluding items, and modifying the data to make it suitable for analysis.
- Feature extraction: The next step is to extract relevant features from previous data. These attributes may include the type of attack, its focus, the duration of the attack, and other relevant information.
- Model selection: After feature extraction, we need to choose the appropriate machine learning algorithm to predict network hacking. Some popular algorithms for this purpose include decision trees, random forests, and neural networks.
- Model Training: After choosing the machine learning algorithm, we must train and extract the model using the previous data. This includes dividing the data into training and validation and then showing the structure of the training process.
- Model Evaluation: After the model is trained, we need to evaluate its performance in the validation process. This helps us determine if the model is over or under-fitting the data and whether there is room for improvement.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

- Model Deployment: Finally, once we are satisfied with the model's performance, we can send it to production to predict future cyber-attacks. The system should be constantly monitored and updated as new information becomes available.

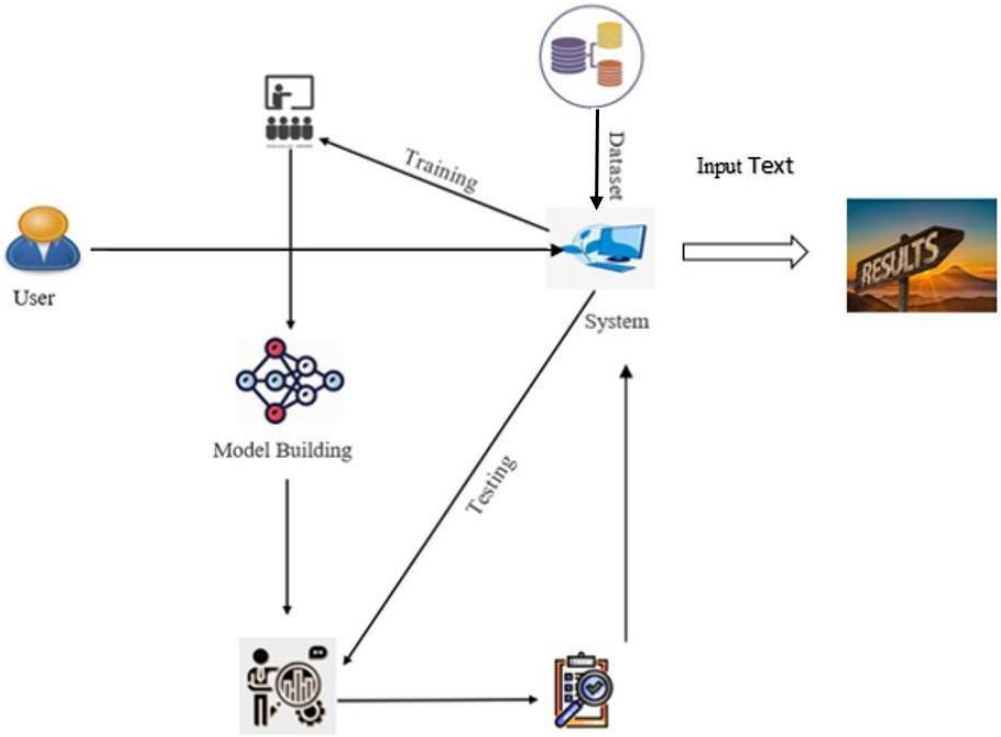
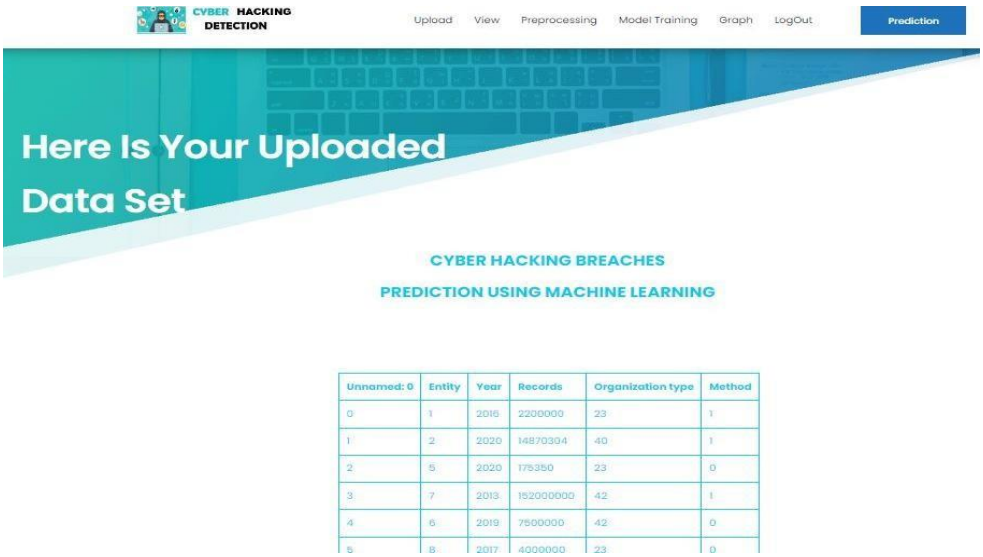


Figure 1: Representation of the architecture model.

V. EXPERIMENTAL RESULT

Figure 2: Representation of the main page.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023



Unnamed: 0	Entity	Year	Records	Organization type	Method
0	1	2016	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

Figure 3: Representation of the Uploaded data set.

VI. CONCLUSION

In this study, an attempt was made to use the resilient control consensus method in complex discrete cyber-physical networks in several local Cyber hacking breaches. 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. The system has less complexity. So,withthe deep learning method, systems can analyze patterns and learn from them to help prevent similar attacks and respond to changing behavior. In short, machine learning can make cybersecurity 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 Cyber hacking, detects it and isolates it, so as not to have a detrimental effect on the behavior of other agents.

REFERENCES

[1]. "Security analysis for cyber-physical systems against stealthy deception attacks". In 2013 American control conference. Available [online]
[2]. "Design and implementation of attack-resilient cyber-physical systems: With a focus on attack-resilient state estimators." IEEE Control Systems Magazine (2017). Available [online]
[3]. "Consensus formation control for a class of networked multiple mobile robot systems." Journal of Control Science and Engineering 2012 (2012). Available [online]
[4]. "Resilient distributed control in the presence of misbehaving agents in networked control systems."IEEE Transactions on Cybernetics (2014). Available [online]
[5]. "Resilient control of networked control systems with stochastic denial of service attacks." Neurocomputing (2017). Available [online]
[6]. "Robustness of information diffusion algorithms to locally bounded adversaries." In 2012 American Control Conference (ACC), IEEE (2012). Available [online]
[7]. "Resilient Consensus of Discrete-Time Complex Cyber-Physical Networks under Deception Attacks." IEEE Transactions on Industrial Informatics (2019). Available [online]
[8]. "Machine learning methods for attack detection in the smart grid." IEEE Transactions on neural networks and learning systems 27, (2015). Available [online]
[9]. "Data mining based cyber-attack detection." System simulation technology 13, (2017). Available [online]:
[10]. "Attack detection and identification in cyber-physical systems." IEEE Transactions on Automatic Control (2013). Available [online]

Lambert Publication's

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

ISSN No. : 2581-9429

Scope Database
Journal indexing & Citation Analysis

SJIFactor
Scientific Journal Impact Factor

Impact Factor: 7.301

Scientific Journal Impact Factor
www.sjifactor.com

Google Scholar

Crossref

Lambert Publication's

www.ijarsct.co.in

7972611953

doi

EduBirdie™

PLAGIARISM CHECKER BY EDUBIRDIE
CHECK YOUR ESSAY FOR FREE

For faster, more accurate results, please fill all the fields below

Select Your Type of Paper

- ◯ Essay (Any Type)
- ◯ Web Site Content
- ◯ Resume
- ◯ Other

Paste Your Title Here

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 are 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

The length of the text: **2829** (No spaces: **2829**) ↻ Check another text

100.0% The uniqueness of the text

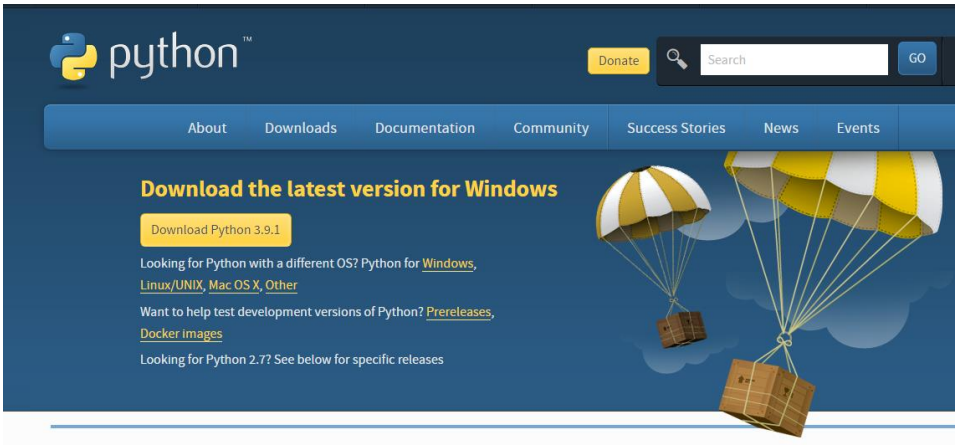
INSTALLATION AND EXECUTION STEPS:

The two main steps for installation and execution are:

1. Install Python IDE.
2. Download PyCharm and install it successfully.

Installing Python IDE:

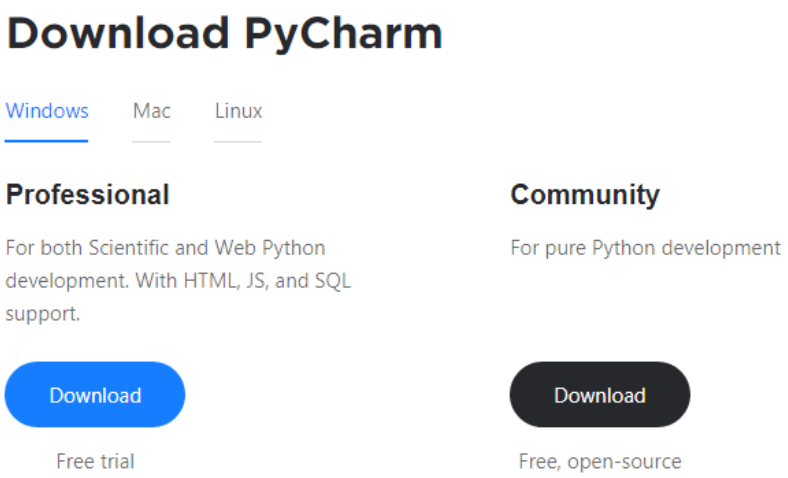
1. To download and install Python visit the official website of Python https://www.python.org/downloads/ and choose your version.



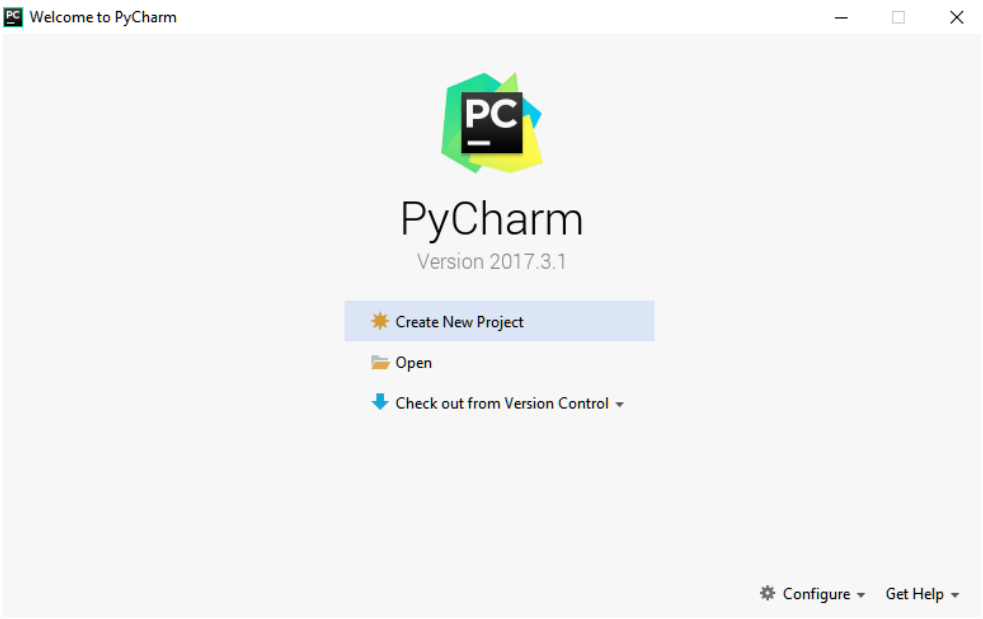
2. Once the download is complete, run the exe for install Python. Now click on Install Now.
3. You can see Python installing at this point.
4. When it finishes, you can see a screen that says the Setup was successful. Now click on "Close".

Installing PyCharm:

1. To download PyCharm visit the website https://www.jetbrains.com/pycharm/download/ and click the "DOWNLOAD" link under the Community Section.



2. Once the download is complete, run the exe for install PyCharm. The setup wizard should have started. Click “Next”.
3. On the next screen, Change the installation path if required. Click “Next”.
4. On the next screen, you can create a desktop shortcut if you want and click on “Next”.
5. Choose the start menu folder. Keep selected JetBrains and click on “Install”.
6. Wait for the installation to finish.
7. Once installation finished, you should receive a message screen that PyCharm is installed. If you want to go ahead and run it, click the “Run PyCharm Community Edition” box first and click “Finish”.
8. After you click on "Finish," the Following screen will appear.

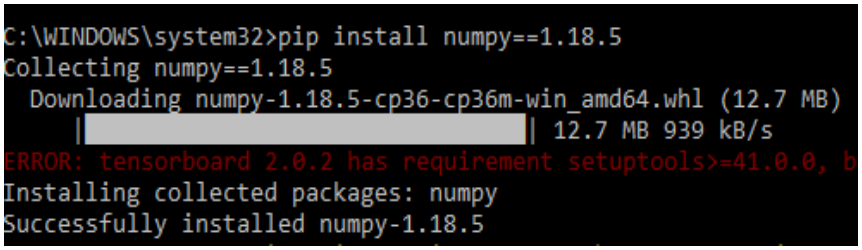


9. You need to install some packages to execute your project in a proper way.

10. Open the command prompt/ anaconda prompt or terminal as administrator.

11. The prompt will get open, with specified path, type “pip install package name” which you want to install (like NumPy, pandas, sea born, scikit-learn, Matplotlib, Pyplot)

Ex: Pip install NumPy



STUDENTS DETAILS

1. Name: Krishna V R
2. USN: 3VC19CS063
3. Phone NO: 7760298650
4.Email-ID: krishna.cse.rymec@gmail.com
5. Permanent Address: #220, Near Markendaya Swamy Temple, Molakalmuru-577535

1. Name: N Naveen Upadhyaya
2. USN: 3VC19CS091
3. Phone NO: 8105877089
4. Email-ID: naveenupadhyaya.cse.rymec@gmail.com
5. Permanent Address: SRI Balaji Nilaya House no15 Ward no 19 Gollanarasappa colony near Ganesh temple Snpet Ballari-583101

1. Name: N Ashok Reddy
2. USN: 3VC19CS090
3. Phone NO: 8884714574
4. Email-ID: ashokreddy.cse.rymec@gmail.com
5. Permanent Address: D3 , Vibhava appartment, Nandhi layout , sirguppa road,Ballari -583101

1.Name : Abdul Muiz Ali
2. USN: 3VC19CS189
3. Phone No: 9945250678
4. Email-ID: muiz.cse.rymec@gmail.com
5. Permanent Address: 182/31 Nandi Colony, 1st Cross , Cantonment, Ballari-583104