CYBER THREAT PREDICTIVE ANALYTICS FOR IMPROVING CYBER SUPPLY CHAIN SECURITY

CAPSTONE PROJECT REPORT

(Project Term June-July 2024)

Submitted by

1) Nadendla Sushmitha Chowdary 12200798
2) Mandala Pavani 12200890
3) Nadendla Tharun 12204843
4) Dwarapu Komal Harsha 12216924
5) K.V.B Vasanth Rayulu 12205149

Course Code: CAP769

Under the Guidance of

Asst. Prof. Anjana Sharma
UID: 30771

Lovely Professional University

School Of Computer Applications

CERTIFICATE

This is to certify that project entitled "**CYBER THREAT PREDICTIVE ANALYTICS FOR IMPROVING CYBER SUPPLY CHAIN SECURITY"** submitted in partial fulfillment of the requirement for the award of degree of Master of Computer Application in the discipline of School of Computer Application, is a Bonafide project work carried out by Nadendla Sushmitha Chowdary[12200798], Mandala Pavani [12200890] , Nadendla Tharun[12204843], Dwarapu Komal Harsha[12216924], K.V.B.Vasanth Rayulu[12205149], under our supervision and that no part of this project has been submitted for any other degree of diploma.

(Signature of Supervisor)

Asst. Prof. Anjana Sharma
UID: 30771

School of Computer Application
Lovely Professional University
Phagwara, Punjab.

DECLARATION OF STUDENT

We, Nadendla Sushmitha Chowdary [12200798], Mandala Pavani [12200890], Nadendla Tharun[12204843], Dwarapu Komal Harsha[12216924], K.V.B.Vasanth Rayulu[12205149], hereby declare that the project work entitled ("**CYBER THREAT PREDICTIVE ANALYTICS FOR IMPROVING CYBER SUPPLY CHAIN SECURITY**") is an authentic record of our own work carried out as requirements of Capstone Project for the award of Master of Computer Application degree in School of Computer Application from Lovely Professional University, Phagwara, under the guidance of (**Asst. Prof. Anjana Sharma**), during June to July 2024. All the information furnished in this capstone project report is based on our own intensivework and is genuine.

Project Group Number:

Name of Student 1: Nadendla Sushmitha Chowdary

Registration Number: 12200798 (Signature of Student 1)

Name of Student 2: Mandala Pavani

Registration Number: 12200890 (Signature of Student 2)

Name of Student 3: Nadendla Tharun

Registration Number: 12204843 (Signature of Student 3)

Name of Student 4: Dwarapu Komal Harsha

Registration Number: 12216924 (Signature of Student 4)

Name of Student 5: K.V.B.Vasanth Rayulu

Registration Number: 12205149 (Signature of Student 5)

DECLARATION BY THE SUPERVISOR

This is to certify that Nadendla Sushmitha Chowdary [1220798], Mandala Pavani [12200890], Nadendla Tharun [12204843], Dwarapu Komal Harsha [12216924], K.V.B.Vasanth Rayulu [12205149], declaration statement made by this group of studentsis correct to the best of my knowledge and belief. They have completed this Capstone Project under my guidance and supervision. The present work is the result of their original investigation, effort, and study. No part of the work has ever been submitted for any other degree at any University. The Capstone Projectis fit for the submission and partial fulfilment of the conditions for the award of MCA degree in School of Computer Application (SCA) from Lovely Professional University, Phagwara.

Signature

Asst. Prof. Anjana Sharma

UID : 30771

School of Computer Application

Lovely Professional University

Phagwara, Punjab.

Journal of Emerging Technologies and Innovative Research

An International Open Access Journal Peer-reviewed, Refereed Journal

www.jetir.org | editor@jetir.org An International Scholarly Indexed Journal

Certificate of Publication

The Board of

Journal of Emerging Technologies and Innovative Research (ISSN : 2349-5162)

Is hereby awarding this certificate to

Nadendla Tharun

In recognition of the publication of the paper entitled

CYBER THREAT PREDICTIVE ANALYTICS FOR IMPROVING CYBER SUPPLY CHAIN SECURITY

Published In JETIR (www.jetir.org) ISSN UGC Approved (Journal No: 63975) & 7.95 Impact Factor

Published in Volume 11 Issue 7 , July-2024 | Date of Publication: 2024-07-30

EDITOR

EDITOR IN CHIEF

JETIR2407698 Research Paper Weblink http://www.jetir.org/view?paper=JETIR2407698 Registration ID : 545797

An International Scholarly Open Access Journal, Peer-Reviewed, Refereed Journal Impact Factor Calculate by Google Scholar and Semantic Scholar | AI-Powered Research Tool, Multidisciplinary, Monthly, Multilanguage Journal Indexing in All Major Database & Metadata, Citation Generator

Journal of Emerging Technologies and Innovative Research

An International Open Access Journal Peer-reviewed, Refereed Journal

www.jetir.org | editor@jetir.org An International Scholarly Indexed Journal

Certificate of Publication

The Board of

Journal of Emerging Technologies and Innovative Research (ISSN : 2349-5162)

Is hereby awarding this certificate to

Dwarapu Komal Harsha

In recognition of the publication of the paper entitled

CYBER THREAT PREDICTIVE ANALYTICS FOR IMPROVING CYBER SUPPLY CHAIN SECURITY

Published In JETIR (www.jetir.org) ISSN UGC Approved (Journal No: 63975) & 7.95 Impact Factor

Published in Volume 11 Issue 7 , July-2024 | Date of Publication: 2024-07-30

Parisa P

EDITOR

EDITOR IN CHIEF

ISSN 2349-5162

JETIR2407698 Research Paper Weblink http://www.jetir.org/view?paper=JETIR2407698 Registration ID : 545797

An International Scholarly Open Access Journal, Peer-Reviewed, Refereed Journal Impact Factor Calculate by Google Scholar and Semantic Scholar | AI-Powered Research Tool, Multidisciplinary, Monthly, Multilanguage Journal Indexing in All Major Database & Metadata, Citation Generator

Journal of Emerging Technologies and Innovative Research

An International Open Access Journal Peer-reviewed, Refereed Journal

www.jetir.org | editor@jetir.org An International Scholarly Indexed Journal

Certificate of Publication

The Board of

Journal of Emerging Technologies and Innovative Research (ISSN : 2349-5162)

Is hereby awarding this certificate to

Nadendla Sushmitha Chowdary

In recognition of the publication of the paper entitled

CYBER THREAT PREDICTIVE ANALYTICS FOR IMPROVING CYBER SUPPLY CHAIN SECURITY

Published In JETIR (www.jetir.org) ISSN UGC Approved (Journal No: 63975) & 7.95 Impact Factor

Published in Volume 11 Issue 7 , July-2024 | Date of Publication: 2024-07-30

EDITOR

EDITOR IN CHIEF

ISSN 2349-5162

JETIR2407698 Research Paper Weblink http://www.jetir.org/view?paper=JETIR2407698 Registration ID : 545797

An International Scholarly Open Access Journal, Peer-Reviewed, Refereed Journal Impact Factor Calculate by Google Scholar and Semantic Scholar | AI-Powered Research Tool, Multidisciplinary, Monthly, Multilanguage Journal Indexing in All Major Database & Metadata, Citation Generator

Journal of Emerging Technologies and Innovative Research

An International Open Access Journal Peer-reviewed, Refereed Journal

www.jetir.org | editor@jetir.org An International Scholarly Indexed Journal

Certificate of Publication

The Board of

Journal of Emerging Technologies and Innovative Research (ISSN : 2349-5162)

Is hereby awarding this certificate to

Mandala Pavani

In recognition of the publication of the paper entitled

CYBER THREAT PREDICTIVE ANALYTICS FOR IMPROVING CYBER SUPPLY CHAIN SECURITY

Published In JETIR (www.jetir.org) ISSN UGC Approved (Journal No: 63975) & 7.95 Impact Factor

Published in Volume 11 Issue 7 , July-2024 | Date of Publication: 2024-07-30

EDITOR

EDITOR IN CHIEF

ISSN 2349-5162

JETIR2407698 Research Paper Weblink http://www.jetir.org/view?paper=JETIR2407698 Registration ID : 545797

An International Scholarly Open Access Journal, Peer-Reviewed, Refereed Journal Impact Factor Calculate by Google Scholar and Semantic Scholar | AI-Powered Research Tool, Multidisciplinary, Monthly, Multilanguage Journal Indexing in All Major Database & Metadata, Citation Generator

Journal of Emerging Technologies and Innovative Research

An International Open Access Journal Peer-reviewed, Refereed Journal

www.jetir.org | editor@jetir.org An International Scholarly Indexed Journal

Certificate of Publication

The Board of

Journal of Emerging Technologies and Innovative Research (ISSN : 2349-5162)

Is hereby awarding this certificate to

Kosuri Veera Bhoga Vasanth Rayalu

In recognition of the publication of the paper entitled

CYBER THREAT PREDICTIVE ANALYTICS FOR IMPROVING CYBER SUPPLY CHAIN SECURITY

Published In JETIR (www.jetir.org) ISSN UGC Approved (Journal No: 63975) & 7.95 Impact Factor

Published in Volume 11 Issue 7 , July-2024 | Date of Publication: 2024-07-30

Parisa P

EDITOR

EDITOR IN CHIEF

ISSN 2349-5162

JETIR2407698 Research Paper Weblink http://www.jetir.org/view?paper=JETIR2407698 Registration ID : 545797

An International Scholarly Open Access Journal, Peer-Reviewed, Refereed Journal Impact Factor Calculate by Google Scholar and Semantic Scholar | AI-Powered Research Tool, Multidisciplinary, Monthly, Multilanguage Journal Indexing in All Major Database & Metadata, Citation Generator

ABSTRACT:

In the realm of cybersecurity, bolstering the rigidity of force chains against implicit risks is consummate. This study delves into the realm of predictive analytics, employing advanced machine learning ways including CatBoost, XGBoost, and mounding Classifier. using a comprehensive dataset encompassing a plethora of features analogous as IsBeta, RtpStateBitfield, and AVProductStatesIdentifier, among others, the disquisition aims to develop robust models suitable of soothsaying cyber risks. By checking pointers suchlike Has DetectionsandWdft_IsGamer, the models strive to identify patterns reflective of implicit security breaches. Through scrupulous analysis of attributes like Firewall statusandCensus_OSBuildNumber, the study seeks to give practicable perceptivity for enhancing cyber force chain security. By integrating different classifiers, this disquisition trials to offer a multifaceted approach to trouble discovery, thereby fortifying the cyber structure against arising risks and icing the integrity of force chain networks.

Keywords : Cat Boost, Stacking Classifier, Gradient Boost(GB) and AdaBoost.

1. INTRODUCTION

1.1 Motivation:

The provocation behind the title" CYBER THREAT PREDICTIVE ANALYTICS FOR IMPROVING CYBER SUPPLY CHAIN SECURITY" stems from the critical need to enhancethe rigidity of force chains against cyber risks. using advanced machine learning ways and a comprehensive dataset, the study aims to develop robust models suitable of soothsaying implicit security breaches. By integrating different classifiers, it seeks to give practicable perceptivity for fortifying cyber structure and icing the integrity of force chain Networks.

1.2 Problem Statement:

The research objective of the project entitled "CYBER THREAT PREDICTIVE ANALYTICS FOR IMPROVING CYBER SUPPLY CHAIN SECURITY" is to develop robust analytical models using advanced machine learning techniques such as CatBoost, XGBoost, Stacking Classifier and multiple data sets and analyzed. By identifying instances of security breaches, the study intends to strengthen the security of the cyber supply chain and increase the resilience of infrastructure to emerging threats.

1.3 Objective of the Project:

This project aims to build predictive analytics models using advanced machine learning techniques such as CatBoost, XGBoost, Stacking Classifier etc. to enhance cyber supply chain security By analyzing a comprehensive data set with various cybersecurity features, the project seeks to force potential threats and It seeks to enhance insights that can be used to enhance the resilience of supply chain networks to emerging threats.

1.4 Scope:

This study focuses on utilizing predictive analytics, including CatBoost, XGBoost, and Stacking Classifier, to enhance cyber supply chain security. It employs a comprehensive dataset encompassing various features to develop robust models for forecasting cyber threats. The research aims to identify patterns indicative of potential security breaches and provide actionable insights for enhancing cyber infrastructure resilience. It also explores amalgamating diverse classifiers to fortify supply chain networks against emerging risks.

1.5 Project Introduction:

In today's interconnected digital landscape, the security of cyber supply chains is of paramount importance. The proliferation of cyber threats poses significant challenges to organizations, necessitating proactive measures to safeguard their networks and data. Predictive analytics, powered by advanced machine learning algorithms, offers a promising approach to bolstering cyber supply chain security. This study explores the efficacy of predictive analytics in forecasting cyber threats, utilizing sophisticated techniques such as CatBoost, XGBoost, and Stacking Classifier. The research leverages a rich dataset comprising diverse features, including IsBeta, RtpStateBitfield, and AVProductStatesIdentifier, among others. These features serve as critical indicators of potential security vulnerabilities within supply chain networks. By analyzing attributes such as HasDetections and Wdft_IsGamer, the models aim to identify patterns indicative of cyber threats, enabling preemptive mitigation strategies. Furthermore, the study delves into the intricacies of Firewall status and Census_OSBuildNumber, dissecting their impact on cyber supply chain security. By examining these attributes in conjunction with other relevant variables, the research endeavors to provide actionable insights for enhancing resilience against emerging threats. A key aspect of this research lies in its holistic approach to threat detection. By amalgamating diverse classifiers, including Gradient Boost (GB) and AdaBoost, the study aims to develop robust models capable of detecting a wide range of cyber threats. This multifaceted approach not only enhances the accuracy of threat prediction but also strengthens the overall cyber infrastructure resilience. In essence, this project seeks to advance our understanding of predictive analytics in the context of cyber supply chain security. By harnessing the power of machine learning and comprehensive datasets, it aims to equip organizations with the tools and insights necessary to proactively mitigate cyber threats and safeguard their supply chain networks.

2. LITERATURE SURVEY

2.1 Related Work:

Ensuring the dependability of Mechanical Web of Things (IIoT) systems is basic for assembly partner desires and anticipating potential hurt. A reliable IIoT framework coordinating different security properties from IT frameworks, such as security, security, security, unwavering quality, and versatility. Conventional security instruments and strategies are inadequately for ensuring IIoT stages due to contrasts in conventions, constrained update alternatives, convention bungles, and the utilize of obsolete working frameworks in mechanical settings. This paper points to upgrade the dependability of IIoT systems, particularly supervisory control and information securing (SCADA) systems, through a tried and true and versatile cyberattack discovery demonstrate. The proposed arrangement leverages an gathering learning approach, combining a arbitrary subspace (RS) learning strategy with arbitrary tree (RT) strategies to distinguish cyberattacks utilizing arrange activity information from SCADA-based IIoT stages. The imaginative perspective of this demonstrate lies in its utilize of mechanical protocol-based arrange activity, utilizing RS to address unessential highlight affectability, and utilizing gathering RT to moderate overfitting. This comes about in a vigorous location motor custom fitted to mechanical conventions, accomplishing tall location rates. The proposed demonstrate has been assessed on 15 SCADA organize datasets, with exploratory comes about illustrating its prevalence over ordinary location strategies, in this manner upgrading the security and reliability of IIoT platforms[1] M. M. Hassan, A. Gumaei, S. Huda, and A. Almogren, ‘‘Increasing the trustworthiness in the industrial IoT networks through a reliable cyberattack detection model,’’ IEEE Trans. Ind. Informat., vol. 16, no. 9, pp. 6154–6162, Sep. 2020..

Summary: Defending SCADA-based mechanical systems from cyberattacks improves the reliability of IIoT systems. In any case, existing security strategies and program items stay wasteful and wrong for securing SCADA-based IIoT networks.

The expanding recurrence and modernity of cyberattacks require the utilization of machine learning for risk expectation. Conventional strategies such as spam channels, firewalls, and IDS/IPS setups are deficient as enemies utilize antagonistic machine learning methods to misuse vulnerabilities. This paper analyzes the achievability of utilizing machine learning to foresee malware assaults and make a classifier that naturally recognizes and names occasions as "Has Discovery" or "No Location." Utilizing a choice tree (DT) calculation, the ponder analyzes a dataset from the Microsoft Malware risk expectation site on Kaggle. The comes about illustrates that machine learning strategies can successfully distinguish cyberattacks on savvy lattices, foresee future patterns, and give profitable cyber risk intelligence[2] A. Yeboah-Ofori, ‘‘Classification of malware attacks using machine learning in decision tree,’’ IJS, vol. 11, no. 2, pp. 10–25, 2020..

Summary: This inquiry utilized machine learning and choice tree calculations to foresee and classify malware assaults, illustrating the adequacy of machine learning in distinguishing and labeling cyber threats.

Cybersecurity inside supply chains is basic for organizations to accomplish their trade goals safely. Whereas innovative headways have moved forward trade forms and decreased costs, they have moreover presented challenges due to expanded interdependencies among partners. These challenges incorporate need of third-party review components and cascading cyber dangers, driving to assaults such as control of plan determinations and modifications amid conveyance. This paper explores supply chain dangers, models and analyzes cyber supply chain (CSC) assaults, and proposes risk detailing instruments. Utilizing the STIX risk show, the paper assesses a savvy lattice case considered to illustrate the proposed model's adequacy in analyzing dangers and prescribing CSC controls to improve organizational security[3] A. Yeboah-Ofori and S. Islam, ‘‘Cyber security threat modelling for supply chain organizational environments,’’ MDPI. Future Internet, vol. 11, no. 3, p. 63, Mar. 2019 .

Summary: Cybersecurity in supply chains is challenging due to interconnected frameworks among partners. This considers models and analyzes supply chain dangers, and advertising suggestions to move forward security.

As cyber occurrences gotten to be unavoidable, proactive hazard forecast strategies are pivotal for minimizing harm. This paper presents Hazard Teller, a framework that analyzes parallel record appearance logs to anticipate which machines are at chance of contamination. By making comprehensive machine profiles and partnering them with chance levels through completely and semi-supervised learning strategies, Hazard Teller can foresee contaminations months in development. Assessments on a year-long dataset from 18 ventures appear that Hazard Teller accomplishes tall expectation exactness, permitting organizations to take proactive measures against cyber threats[4] L. Bilge, Y. Han, and M. D. Amoco, ‘‘Risk teller: Predicting the risk of cyber incidents,’’ in Proc. CCS, 2017, pp. 1299–1311,.

Summary: Hazard Teller is a proactive framework that predicts cyber dangers by analyzing machine profiles, empowering early mediation, and lessening potential harm from cyber incidents.

Power framework unsettling influences can stem from different characteristics and human-made sources, complicating administrator decision-making. In the setting of cyberattacks, human judgment is regularly questionable due to the tricky nature of assaults. This paper investigates the utilization of machine learning to recognize between distinctive sorts of control framework unsettling influences, particularly centering on identifying cyberattacks. By assessing different machine learning strategies, they think about points to improve existing control framework structures while moving forward unsettling influence segregation capabilities[5] R. C. B. Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, and S. Pan, ‘‘Machine learning for power system disturbance and cyber-attack discrimination,’’ in Proc. 7th Int. Symp. Resilient Control Syst. (ISRCS), Denver, CO, USA, Aug. 2014, .

Summary: The think about builds up benchmarks for applying machine learning to classify control framework unsettling influences, highlighting its potential to improve keen control network security.

The Savvy Lattice, a next-generation control framework, coordinates progressed computing and communication innovations to make strides in effectiveness and unwavering quality. In any case, the broad interconnection of electronic gadgets poses critical cybersecurity challenges. This paper gives a comprehensive overview of cybersecurity issues in the Keen Network, counting security prerequisites, arranging vulnerabilities, assault countermeasures, and secure communication conventions. The point is to extend understanding of Keen Network security vulnerabilities and arrangements, directing future inquiries about directions [6] W. Wang and Z. Lu, ‘‘Cyber security in smart grid: Survey and challenges,’’ Elsevier Comput. Netw., vol. 57, no. 5, pp. 1344–1371, Apr. 2013. .

Summary: This study highlights the developing consideration of Savvy Lattice cybersecurity, tending to vulnerabilities and arrangements to upgrade the security of interconnected control systems.

3. SYSTEM ANALYSIS

3.1 Existing System

Current methods for enhancing cyber supply chain security utilize predictive analytics through algorithms such as Decision Tree, Random Forest, and Support Vector Machine (SVM). While these approaches are effective, they come with challenges, including significant computational demands, the risk of overfitting, and difficulties in processing large and complex datasets. Additionally, these methods may be less effective in identifying sophisticated cyber threats that use advanced evasion strategies or exploit specific supply chain vulnerabilities.

3.2 Disadvantages

1. Significant computational demands: Using Decision Tree, Random Forest, and SVM algorithms for predictive analytics can be resource-intensive.

2. Risk of overfitting: These methods may not generalize well to unseen data, leading to overly complex and potentially inaccurate models.

3. Difficulty handling large datasets: Managing and analyzing large datasets with high dimensionality can pose challenges for these algorithms.

4. Limited detection of sophisticated threats: The existing system may not effectively identify advanced cyber threats employing evasion techniques or targeting specific vulnerabilities in supply chain networks.

5. Lack of adaptability: The chosen algorithms may not easily adapt to evolving cyber threats and may require frequent updates and adjustments to maintain effectiveness.

3.3 Proposed System

The proposed system provides cyber threat predictive analysis for supply chain security by incorporating machine learning techniques such as XGBoost, Stacking Classifier, Gradient Boosting (GB), AdaBoost algorithm etc. This system requires data types an extended set of features such as IsBeta, RtpStateBitfield, AV ProductStatesIdentifier to predict possible security breaches. By combining multiple classifications, the system provides a comprehensive approach to threat detection, strengthening the cyber infrastructure against evolving threats.

3.4 Advantages

1. Enhanced predictive accuracy: Integration of multiple algorithms improves the accuracy of cyber threat predictions.

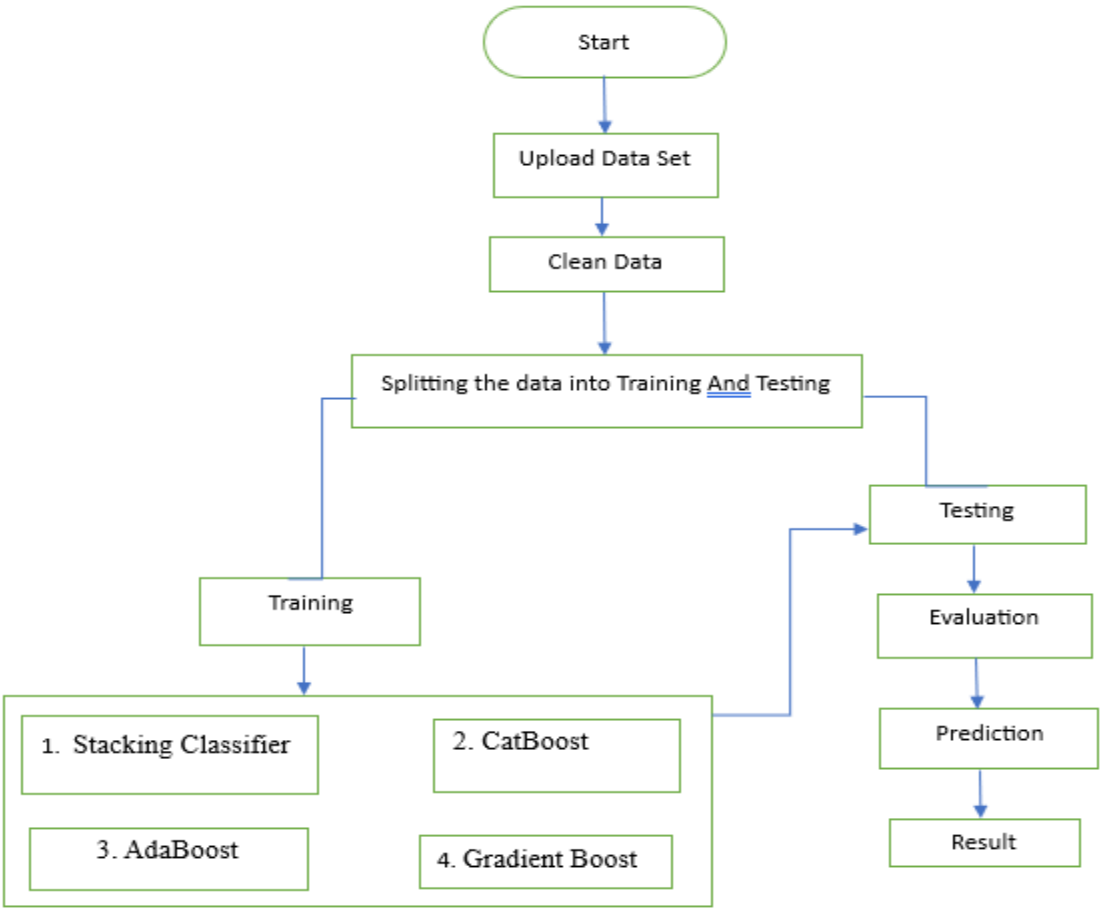
2. Comprehensive threat detection: The system analyzes diverse datasets to identify potential security breaches across various aspects of cyber supply chain.

3. Adaptive defense strategy: Leveraging machine learning techniques allows the system to adapt to evolving cyber threats and enhance defense mechanisms accordingly.

4. Multifaceted approach to security: Amalgamating different classifiers provides a holistic view of potential threats, enabling more effective security measures.

5. Proactive risk mitigation: By forecasting potential security breaches, the system empowers organizations to take proactive measures to mitigate risks and strengthen cyber supply chain security.

3.5 work Flow of Proposed system



4. REQUIREMENT ANALYSIS

4.1 Functional and non-functional requirements

Requirement evaluation is a critical procedure that performs a important function in determining the achievement of a machine or software program challenge. Requirements are normally divided into classes: practical and non-functional requirements.

Functional Requirements: These are the precise functions and functionalities that users anticipate the system to offer. These requirements should be included within the system as a part of the undertaking contract. They are commonly expressed in phrases of inputs furnished to the system, the operations to be completed, and the predicted outputs. Essentially, those are the requirements that customers can directly engage with and notice inside the very last product, not like non-useful requirements.

Examples of functional requirements:

1) 1. User authentication each time they log into the system.
2) 2. Automatic machine shutdown inside the event of a cyber-attack.
3) 3. Sending a verification e mail to customers upon their first registration on a software platform.

Non-functional requirements: These talk over with the satisfactory attributes that the device need to meet as per the venture agreement. The diploma to which these necessities are carried out can range relying on the specific mission. Non-useful necessities, regularly referred to as non-behavioral requirements, address aspects such as:

- Portability
- Security
- Maintainability
- Reliability
- Scalability
- Performance
- Reusability
- Flexibility

Examples of non-functional requirements:

1) Emails should be sent with a latency of no greater than 12 hours from such an activity.
2) The processing of each request should be done within 10 seconds
3) The site should load in 3 seconds whenever of simultaneous users are > 10000

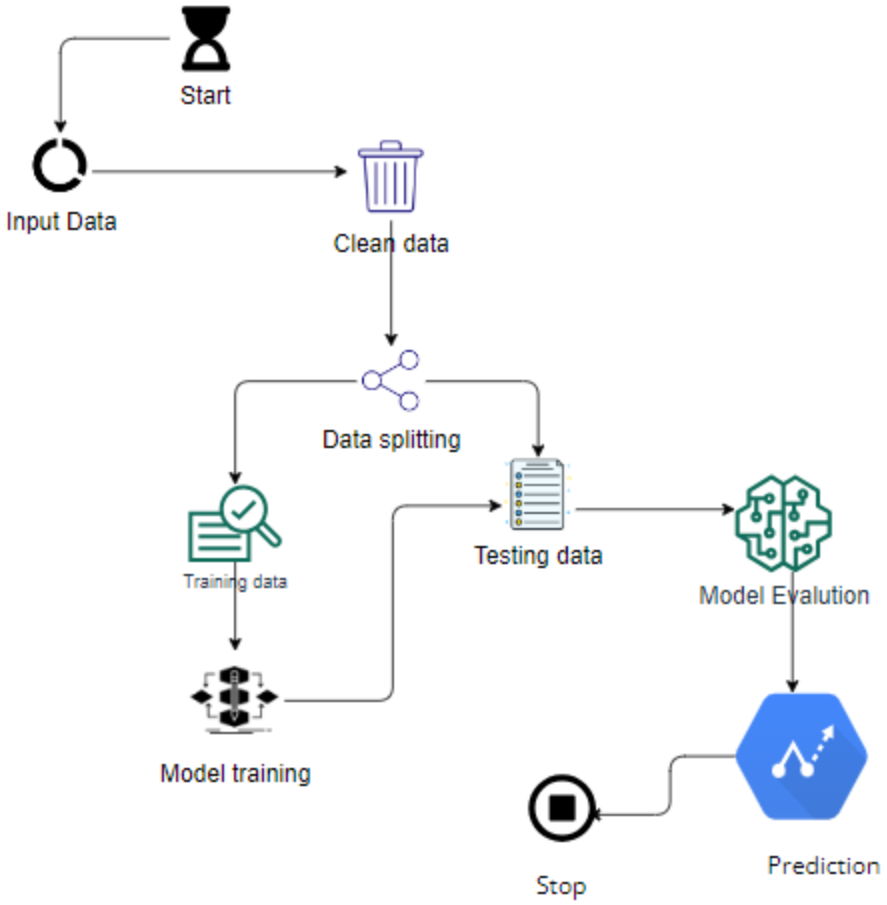
4.2 Hardware Requirements

5	Operating system	:	Windows 7 or 7+
6	RAM	:	8 GB
7	Hard disc or SSD	:	More than 500 GB
8	Processor	:	Intel 3rd generation or high or Ryzen with 8 GB Ram

4.3Software Requirements:

9	Software's	:	Python 3.6 or high version
10	IDE	:	PyCharm.
11	Framework	:	Flask

4.4 Architecture:



5. SYSTEM DESIGN

5.1 Introduction of Input Design:

In an data system, enter refers back to the raw statistics that is processed to generate the desired output. When designing the input stage, builders need to recall the various enter devices consisting of PCs, MICR, OMR, and others. The nice of the machine's enter at once impacts the

excellent of its output. Well-designed enter bureaucracy and screens ought to have the following characteristics:

- They have to efficaciously serve their meant cause, whether or not it is for storing, recording, or retrieving information.
- They must ensure accuracy and completeness at some stage in records access.
- They should be person-friendly, making them clean to fill out and easy to apply.It should focus on user's attention, consistency, and simplicity.
- All these objectives are obtained using the knowledge of basic design principles regarding
- The design should prioritize shooting the person's attention, maintaining consistency, and ensuring simplicity.
- Achieving these dreams entails making use of fundamental design standards, together with:
- Identifying the essential inputs for the gadget.
- Understanding how customers engage with exceptional elements of bureaucracy and screens.

Objectives for Input Design:

Input Design Objectives:

- The fundamental desires of input design are to:
- Develop statistics access and enter strategies.
- Minimize input volume.
- Create source files for facts capture or expand different records seize strategies.
- Design enter information records, facts access monitors, and person interface screens.
- Implement validation assessments and establish powerful enter controls.

Output Design

Output design is important for any machine. During this segment, builders determine the important forms of outputs, and recollect the required output controls and prototype file layouts.

Objectives of Output Design:

- The dreams of output design are to:
- Create output designs that satisfy their meant motive and save you the technology of useless outputs.
- Develop output designs that align with the necessities of the cease users.To deliver the appropriate quantity of output.
- Ensure the output is supplied in the right quantity.
- Format the output accurately and direct it to the right recipient.
- Make the output available promptly to assist powerful decision-making.

5.2 UML Diagrams:

UML, or Unified Modeling Language, is a standardized language used for preferred-purpose modeling in item-orientated software program engineering. Managed and created with the aid of the Object Management Group (OMG), UML objectives to be a familiar language for modeling item-oriented pc software.

UML presently includes two principal additives: a meta-model and a notation. In the future, it can consist of an related approach or technique.

Unified Modeling Language is a widespread language for specifying, visualizing, building, and documenting the artifacts of software structures, as well as for business modeling and different non-software program systems.

UML encapsulates a collection of nice engineering practices which have proven powerful in modeling huge and complex structures. It performs a crucial function in the development of item-orientated software and the software program improvement manner. UML ordinarily makes use of graphical notations to symbolize the layout of software program projects.

5.2.1 Use Case Diagram:

1. **Type of Diagram:** It is a sort of behavioral diagram inside the Unified Modeling Language (UML).

2. **Creation Basis:** Developed through use-case analysis.

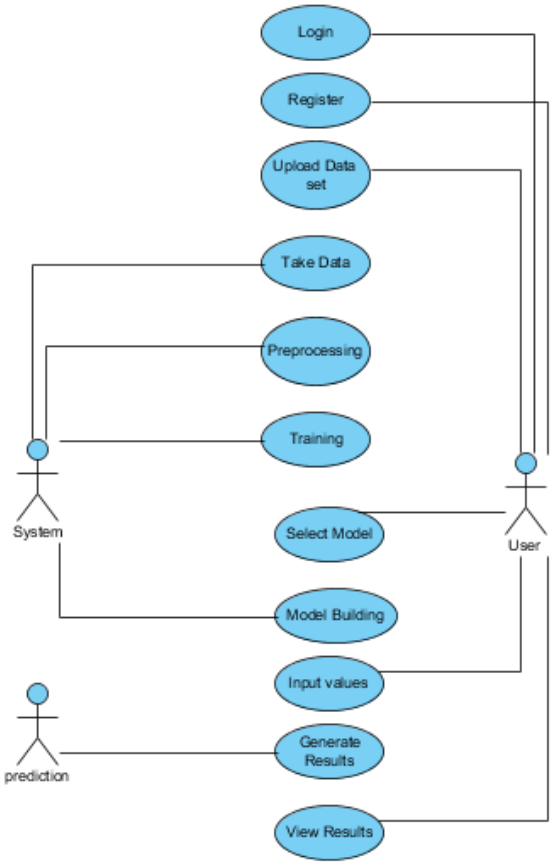
3. **Primary Purpose:** Provides a graphical evaluate of the gadget's functionality.

4. **Components:**

- Actors : Entities that engage with the system (e.G., customers or other systems).
- Use Cases: Goals or responsibilities that actors want to acquire the use of the system.
- Dependencies : Relationships and interactions between unique use cases.

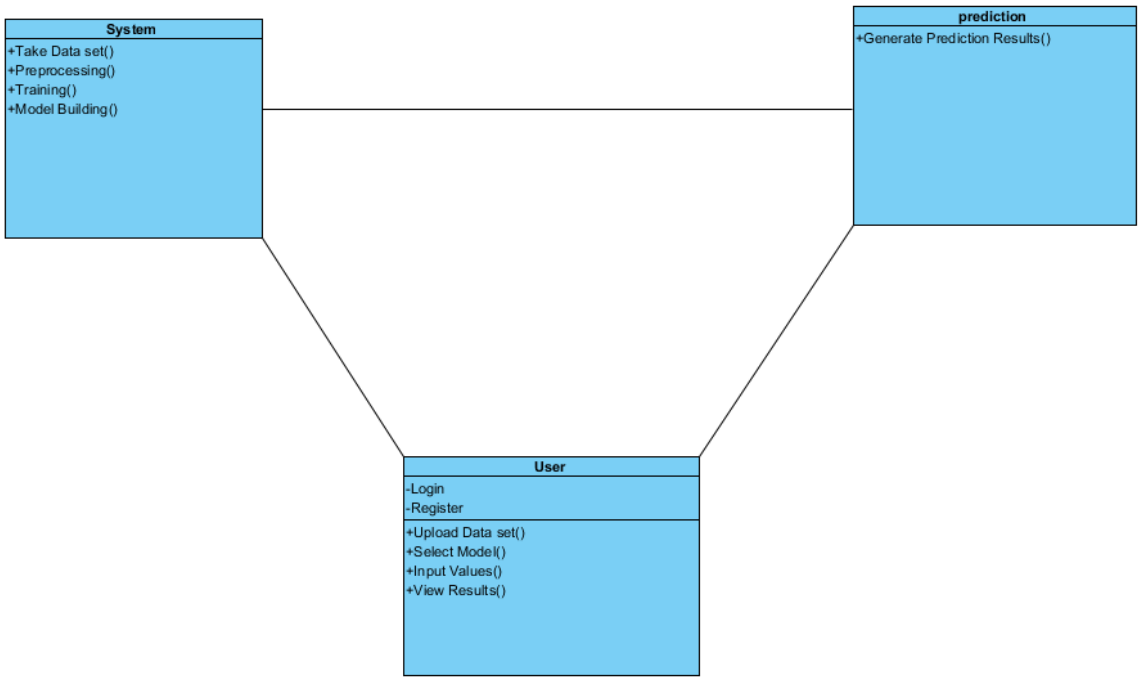
5. **Visualization :** Shows what device functions are completed for each actor.

6. **Role Depiction :** Highlights the jobs of different actors within the machine.

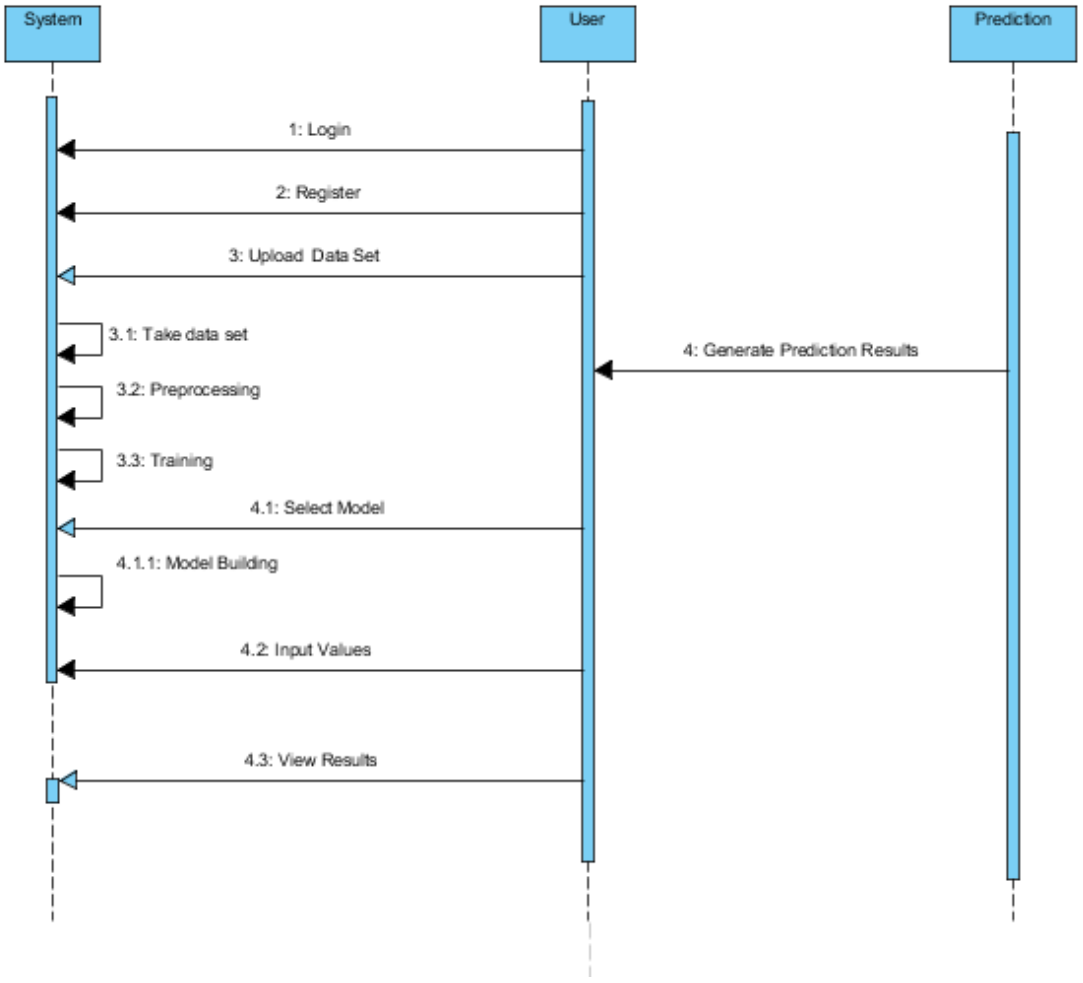


5.2.2 Class Diagram:

A Class Diagram in software engineering, specially inside the Unified Modeling Language (UML), is a form of static structure diagram. It illustrates the shape of a system through depicting the system's lessons, their attributes, operations (or techniques), and the relationships among the lessons. Essentially, it details how special instructions interact and which elegance holds unique pieces of statistics.

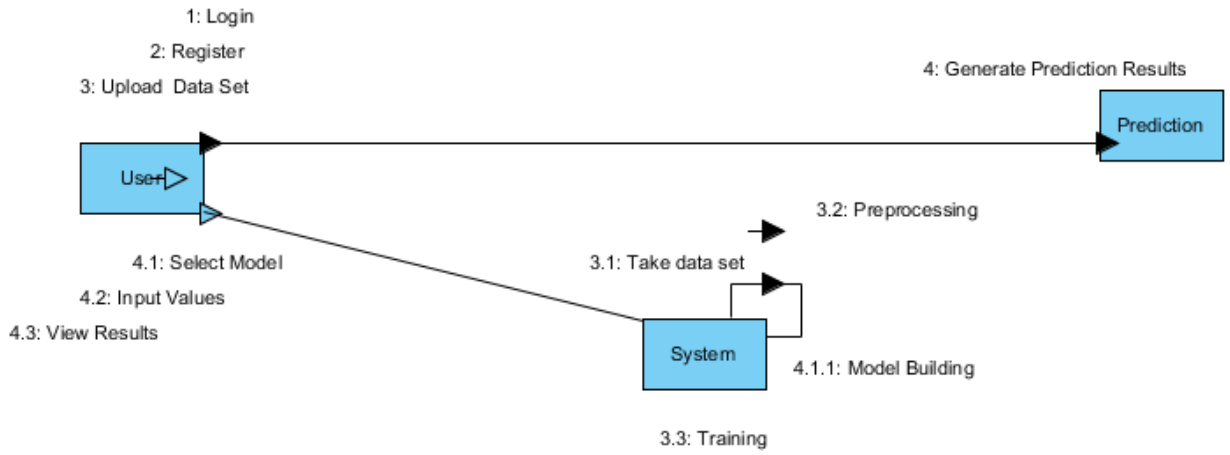


A Sequence Diagram in the Unified Modeling Language (UML) is a kind of interaction diagram that illustrates how procedures engage and the collection in which those interactions arise. It is derived from the Message Sequence Chart and is from time to time referred to as an occasion diagram, occasion state of affairs, or timing diagram.



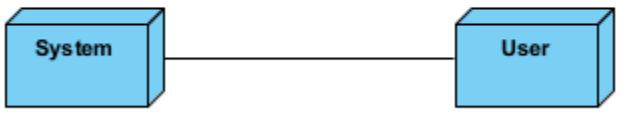
5.2.4 Collaboration Diagram:

A Collaboration Diagram, also known as a Communication Diagram, represents the collection of approach calls inside a device using a numbering approach. This numbering illustrates the order in which strategies are known as. Although it is similar to a Sequence Diagram in phrases of showing approach calls, a Collaboration Diagram differs in that it makes a speciality of the business enterprise of gadgets and their interactions, in preference to the collection of interactions.



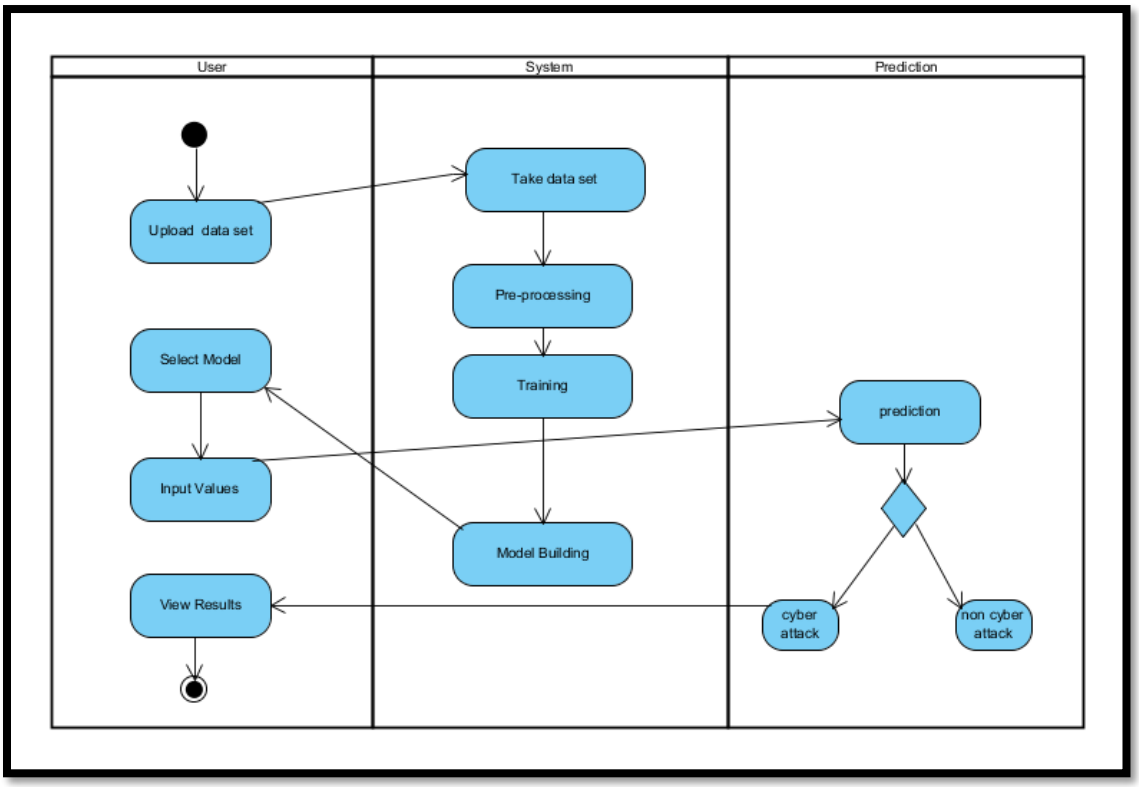
5.2.5 Deployment Diagram

A Deployment Diagram represents the deployment view of a system, focusing at the bodily components of the architecture. It is carefully associated with the Component Diagram, because it shows how components are deployed across various hardware nodes. In a Deployment Diagram, nodes represent physical hardware or gadgets used to set up and execute the software.



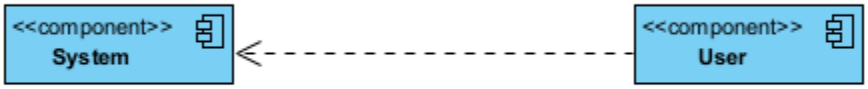
5.2.6 Activity Diagram:

An Activity Diagram is a graphical illustration of workflows that depict step-by way of-step activities and movements, consisting of alternatives for desire, new release, and concurrency. In the Unified Modeling Language (UML), Activity Diagrams are used to describe the distinctive workflows and operational techniques of additives inside a device, illustrating the general drift of manage..



5.2.7 Component Diagram:

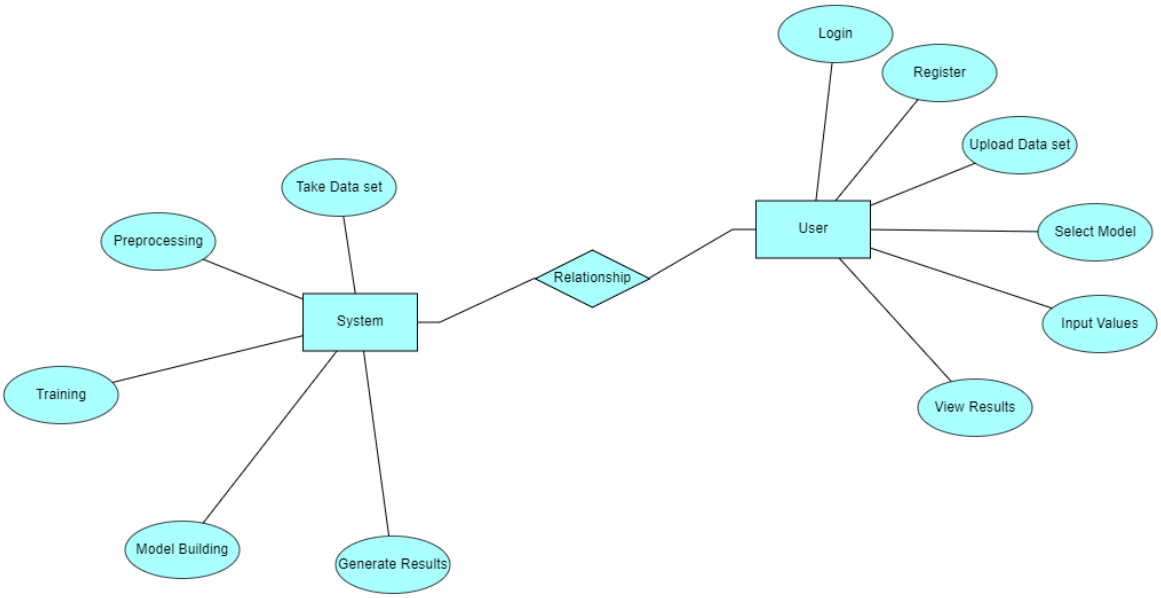
A Component Diagram, additionally known as a UML Component Diagram, illustrates the business enterprise and connections of bodily additives within a machine. It is used to model implementation details and ensure that each one components of the gadget's required functionality are addressed via the planned development.



5.2.8 ER Diagram:

An Entity-Relationship (ER) Diagram, or ER Model, represents the shape of a database via a visual diagram. This model serves as a layout blueprint for a database, that may later be implemented. The number one components of an ER version are entity sets and courting sets.

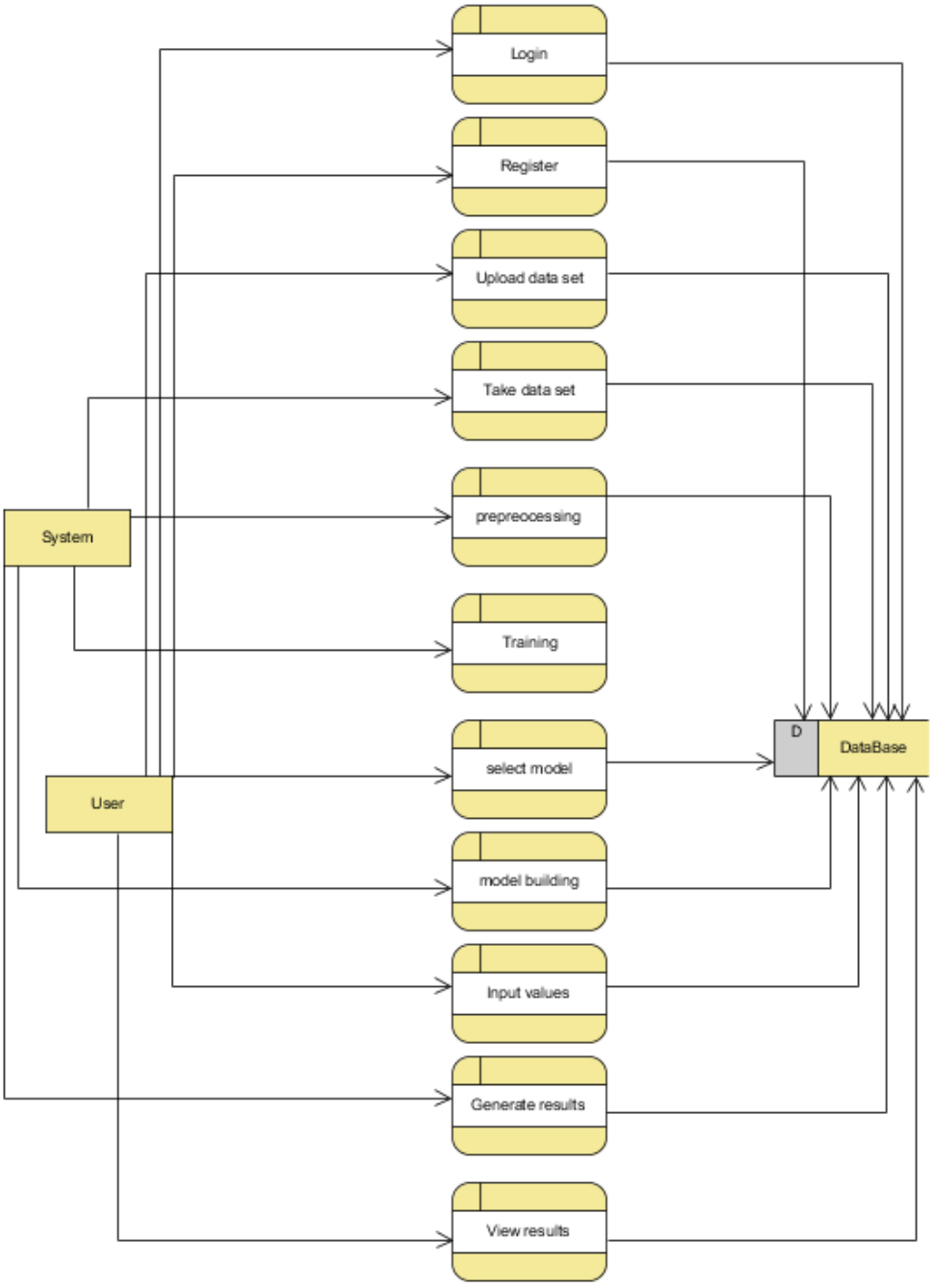
An ER Diagram illustrates the relationships amongst entity sets, where an entity set is a collection of similar entities that may have attributes. In the context of a Database Management System (DBMS), entities correspond to tables or attributes within those tables. By depicting the relationships between tables and their attributes, an ER Diagram outlines the complete logical shape of a database.



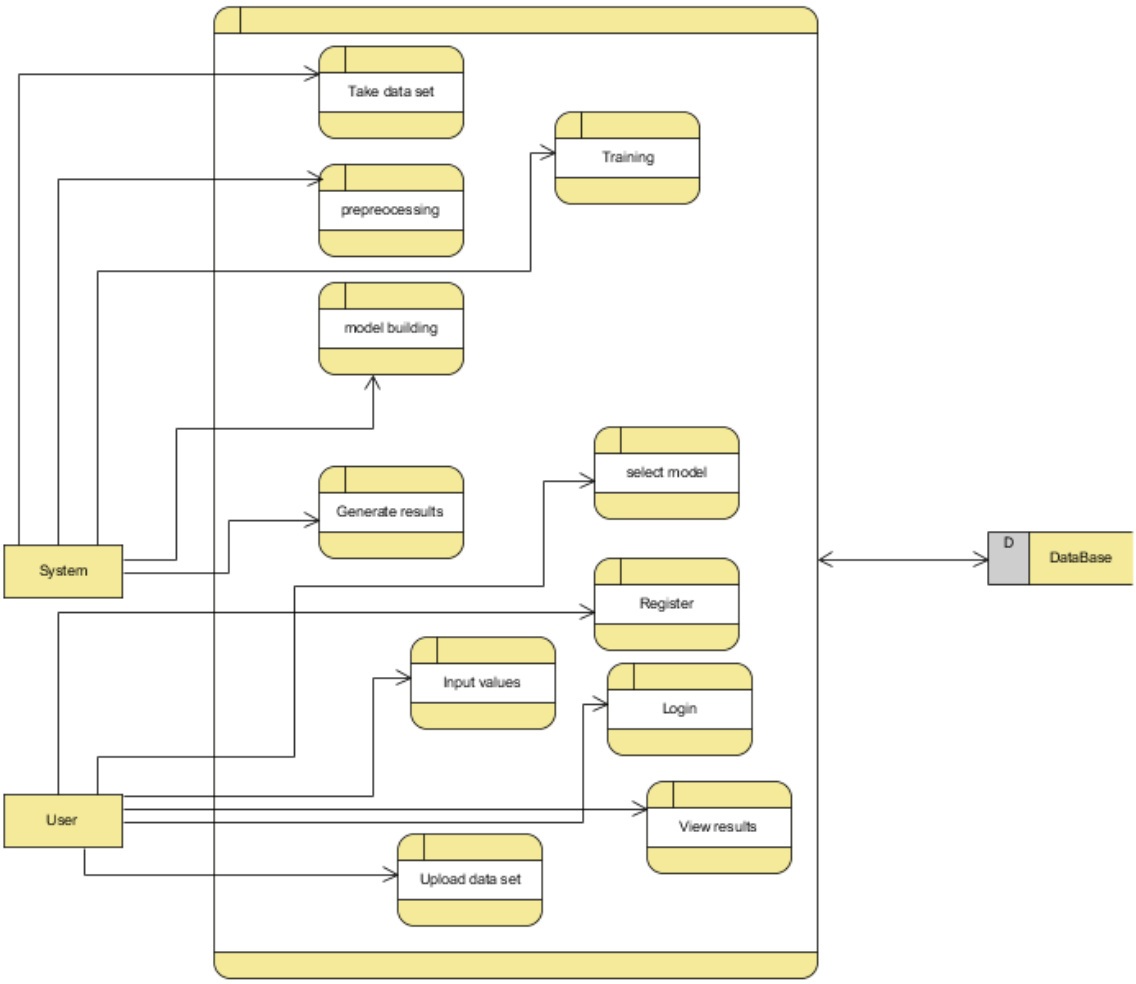
5.3 DFD Diagram:

A Data Flow Diagram (DFD) is a conventional method for illustrating the flow of information within a system. A well-constructed DFD can effectively represent a substantial portion of the system requirements visually. It can depict processes that are manual, automated, or a combination of both. The diagram illustrates how information enters and exits the system, the transformations it undergoes, and where it is stored. The main objective of a DFD is to define the scope and boundaries of an entire system. It serves as a communication tool between a systems analyst and stakeholders involved in the system, and can also be the initial step in redesigning a system.

1st Level Diagram:



2nd Level Diagram:



6. IMPLEMENTATION AND RESULTS

- **MODULES:**

1. User:

1.1 View Home Page:

Users can view the home page of the cyber threat application.

1.2 View about page:

Users can access the about page to learn more about the cyber threat platform.

1.3 View load page:

On the load data page, users can upload the dataset for modeling.

View Page:

Users can view the dataset.

1.4 Input Model:

Users must provide input values for certain fields to generate results.

1.5 View Results:

Users can see the results generated by the model.

1.6 View Score

Here user have ability to view the accuracy score in %

1.7Graph:

Comparison of accuracy foe every models

2. System

2.1 Working on dataset:

The system checks the availability of the dataset and loads the data from CSV files if available.

2.2 Pre-processing:

Data is pre-processed to improve the accuracy of the models and to provide better insights.

2.3 Training the data:

After pre-processing, the data is split into training and testing sets for model training.

2.4 Model Building

This module assists in creating a model that predicts the dataset with high accuracy.

2.5 Generated Score:

Users can view the accuracy score expressed as a percentage.

2.6 Generate Results:

The machine learning algorithm is trained on the data, and the system generates predictions based on this training.

6.2 Algorithms

CatBoost:

CatBoost is a machine learning library developed by Yandex, specifically designed to handle categorical features in gradient boosting algorithms. It is known for its efficiency and effectiveness in managing high-dimensional data that includes categorical variables. CatBoost employs a unique algorithm called Ordered Boosting, which leverages the natural order of categorical features to enhance training speed and accuracy. It also supports various loss functions, calculates feature importance, and includes tools for model interpretability. CatBoost is widely used across multiple domains such as finance, marketing, and healthcare, where dealing with categorical data and achieving accurate predictions are essential.

XGBoost:

XGBoost, short for eXtreme Gradient Boosting, is a popular open-source machine learning library developed by Tianqi Chen. Renowned for its efficiency and accuracy with structured data, XGBoost implements gradient boosting algorithms and offers extensive functionality for tasks such as classification, regression, and ranking. Its superior performance stems from its ability to handle missing data, support parallel computing, and utilize tree pruning techniques. XGBoost has gained significant popularity in fields like finance, healthcare, and technology, and has consistently performed well in competitions on platforms such as Kaggle. Its flexibility and scalability make it a preferred choice for many data scientists and machine learning professionals.

Stacking Classifier:

A Stacking Classifier is an ensemble learning technique that improves predictive performance by combining multiple classification models. It involves training a meta-classifier based on the predictions of base classifiers. These base classifiers generate individual predictions on the input data, and their outputs are used as features for training the meta-classifier. Stacking can mitigate the weaknesses of individual models by leveraging their combined strengths, often resulting in better accuracy and robustness compared to using a single classifier. This method is a powerful

tool in machine learning, particularly for tasks where combining predictions from multiple models can enhance overall performance.

Code:

```
from flask import Flask,render_template,request
import pandas as pd
import mysql.connector
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
from sklearn.ensemble import RandomForestRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error
from sklearn.ensemble import BaggingRegressor
import xgboost as xg
from sklearn.ensemble import GradientBoostingRegressor
import catboost as cb
from sklearn.svm import SVR
from sklearn.neighbors import KNeighborsRegressor
from sklearn.ensemble import ExtraTreesRegressor




mydb =
mysql.connector.connect(host='localhost',user='root',password='',port='3306',data
base='real_time')
cur = mydb.cursor()


app = Flask(__name__)

@app.route("/")
def index():
    return render_template('index.html')


@app.route('/about')
def about():
```

```
    return render_template('about.html')



@app.route('/login',methods=['GET','POST'])
def login():
    if request.method == "POST":
        email = request.form['email']
        psw = request.form['password']
        sql = "SELECT * FROM blood WHERE Email=%s and Password=%s"
        val = (email, psw)
        cur = mydb.cursor()
        cur.execute(sql, val)
        results = cur.fetchall()
        mydb.commit()
        if len(results) >= 1:
            return render_template('loginhome.html', msg='login succesful')
        else:
            return render_template('login.html', msg='Invalid Credentias')

    return render_template('login.html')

@app.route('/loginhome')
def loginhome():
    return render_template('loginhome.html')

@app.route('/registration',methods=['GET','POST'])
def registration():

    if request.method == "POST":
        print('a')
        name = request.form['name']
        print(name)
        email = request.form['email']
        pws = request.form['psw']
        print(pws)
        cpws = request.form['cpsw']
        if pws == cpws:
            sql = "select * from blood"
            print('abccccccccc')
            cur = mydb.cursor()
            cur.execute(sql)
            all_emails = cur.fetchall()
            mydb.commit()
            all_emails = [i[2] for i in all_emails]
```

```
            if email in all_emails:
                return render_template('registration.html', msg='a')
            else:
                sql = "INSERT INTO blood(name,email,password) values(%s,%s,%s)"
                values = (name, email, pws)
                cur.execute(sql, values)
                mydb.commit()
                cur.close()
                return render_template('registration.html', msg='success')
        else:
            return render_template('registration.html', msg='repeat')

    return render_template('registration.html')

@app.route('/upload',methods=['POST','GET'])
def upload():
    if request.method == "POST":
        file = request.files['file']
        print(file)
        global df
        df = pd.read_csv(file)
        print(df)
        return render_template('upload.html', columns=df.columns.values,
rows=df.values.tolist(),msg='success')
    return render_template('upload.html')

@app.route('/viewdata')
def viewdata():
    print(df.columns)
    df_sample = df.head(100)
    return render_template('viewdata.html', columns=df_sample.columns.values,
rows=df_sample.values.tolist())


@app.route('/preprocessing',methods=['POST','GET'])
def preprocessing():
    global X, y, X_train, X_test, y_train, y_test
    if request.method == "POST":
        size = int(request.form['split'])
        size = size / 10
        print(size)
        df.drop(['Time', 'Clock'], axis=1, inplace=True)
        df.Pulse.fillna(value=df.Pulse.mode()[0], inplace=True)
        df.SpO2.fillna(value=df.SpO2.mode()[0], inplace=True)
        df.Perf.fillna(value=df.Perf.mode()[0], inplace=True)
```

```
        df.awRR.fillna(value=df.awRR.mode()[0], inplace=True)
        df['NBP (Sys)'].fillna(value=df['NBP (Sys)'].mode()[0], inplace=True)
        df['NBP (Dia)'].fillna(value=df['NBP (Dia)'].mode()[0], inplace=True)
        df['NBP (Mean)'].fillna(value=df['NBP (Mean)'].mode()[0], inplace=True)
        df.etSEV.fillna(value=df.etSEV.mode()[0], inplace=True)
        df.inSEV.fillna(value=df.inSEV.mode()[0], inplace=True)
        df['Pleth'].fillna(value=df['Pleth'].mode()[0], inplace=True)
        X = df.iloc[:, :-1]
        y = df.iloc[:, -1]
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)
        print(X_train)
        print(X_train.columns)
        return render_template('preprocessing.html', msg='Data Preprocessed and
It Splits Succesfully')

    return render_template('preprocessing.html')



@app.route('/model',methods=['POST','GET'])
def model():
    if request.method=='POST':
        models = int(request.form['algo'])
        if models==1:
            print("==")
            model = DecisionTreeRegressor()
            model.fit(X_train, y_train)
            y_pred = model.predict(X_test)
            acc = r2_score(y_pred, y_test)
            dts = mean_squared_error(y_test, y_pred, squared=False)
            msg = 'Accuracy  for Decision Tree is ' + str(acc)
            a = 'mean_squared_error  for Decision Tree is ' + str(dts)
        elif models== 2:
            print("======")
            model = RandomForestRegressor()
            model.fit(X_train, y_train)
            y_pred = model.predict(X_test)
            acc = r2_score(y_pred, y_test)
            rfs = mean_squared_error(y_test, y_pred, squared=False)
            msg = 'Accuracy  for Random Forest is ' + str(acc)
            a = 'mean_squared_error  for Random Forest is '  + str(rfs)
            return render_template('model.html', msg=msg)
        return render_template('model.html')
```

```
if __name__=="__main__":
    app.run(debug=True)
```

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
```
[1] Python

```
# !pip install sklearn_relief
```
[2] Python

```
df = pd.read_csv(r'cyber121.csv')
```
[3] Python

```
df.head()
```
[4] Python

	IsBeta	RtpStateBitfield	IsSxsPassiveMode	DefaultBrowsersIdentifier	AVProductStatesIdentifier	AVProductsInstalled	HasTpm	CountryIdentifier	CityIdentifier	OrganizationIdentifier	...	Firewall
0	98503	173902	53447	102888	37194	8192	17134	17134	2543	165	...	768
1	1400	456924	53447	186069	64721	2048	16299	16299	2042	492	...	768
2	44488	29144	53447	317708	12536	2048	10586	10586	1848	1176	...	1080
3	129332	121027	53447	313374	38396	16384	17134	17134	3125	191	...	1080
4	18964	474552	43927	126133	39065	4096	17134	17134	3427	165	...	768

5 rows × 26 columns

```
df.shape
```
[5] Python

```
(702545, 26)
```

```
df.isnull().sum()
```
[6] Python

```
IsBeta                                          0
RtpStateBitfield                                0
IsSxsPassiveMode                                0
DefaultBrowsersIdentifier                       0
AVProductStatesIdentifier                       0
AVProductsInstalled                             0
HasTpm                                          0
CountryIdentifier                               0
CityIdentifier                                  0
OrganizationIdentifier                          0
GeoNameIdentifier                               0
LocaleEnglishNameIdentifier                     0
OsBuild                                         0
OsSuite                                         0
IsProtected                                     0
SMode                                           0
Firewall                                        0
Census_InternalPrimaryDiagonalDisplaySizeInInches    0
Census_InternalPrimaryDisplayResolutionHorizontal    0
Census_InternalPrimaryDisplayResolutionVertical      0
Census_OSBuildNumber                            0
Census_OSBuildRevision                          0
Census_OSInstallLanguageIdentifier              0
Wdft_IsGamer                                    0
```

```
[7]  df.RtpStateBitfield.fillna(value=df.RtpStateBitfield.mode()[0],inplace=True)                                   Python

[8]  df.DefaultBrowsersIdentifier.fillna(value=df.DefaultBrowsersIdentifier.mode()[0],inplace=True)                 Python

[9]  df.AVProductStatesIdentifier.fillna(value=df.AVProductStatesIdentifier.mode()[0],inplace=True)                 Python

[10] df.AVProductsInstalled.fillna(value=df.AVProductsInstalled.mode()[0],inplace=True)                             Python

[11] df.CityIdentifier.fillna(value=df.CityIdentifier.mode()[0],inplace=True)                                       Python

[12] df.OrganizationIdentifier.fillna(value=df.OrganizationIdentifier.mode()[0],inplace=True)                       Python

[13] df.GeoNameIdentifier.fillna(value=df.GeoNameIdentifier.mode()[0],inplace=True)                                 Python

     df.IsProtected.fillna(value=df.IsProtected.mode()[0],inplace=True)
```

```
[14] df.IsProtected.fillna(value=df.IsProtected.mode()[0],inplace=True)                                             Python

[15] df.SMode.fillna(value=df.SMode.mode()[0],inplace=True)                                                         Python

[16] df.Firewall.fillna(value=df.Firewall.mode()[0],inplace=True)                                                   Python

[17] df.Census_InternalPrimaryDiagonalDisplaySizeInInches.fillna(value=df.Census_InternalPrimaryDiagonalDisplaySizeInInches.mode()[0],inplace=True)   Python

[18] df.Census_InternalPrimaryDisplayResolutionHorizontal.fillna(value=df.Census_InternalPrimaryDisplayResolutionHorizontal.mode()[0],inplace=True)   Python

[19] df.Census_InternalPrimaryDisplayResolutionVertical .fillna(value=df.Census_InternalPrimaryDisplayResolutionVertical .mode()[0],inplace=True)   Python

[20] df.Wdft_IsGamer.fillna(value=df.Wdft_IsGamer.mode()[0],inplace=True)                                           Python
```

```
[22]  df.Census_OSInstallLanguageIdentifier .fillna(value=df.Census_OSInstallLanguageIdentifier .mode()[0],inplace=True)
                                                                                                                    Python

[23]  df.isnull().sum()
                                                                                                                    Python

...   IsBeta                                                0
      RtpStateBitfield                                      0
      IsSxsPassiveMode                                      0
      DefaultBrowsersIdentifier                             0
      AVProductStatesIdentifier                             0
      AVProductsInstalled                                   0
      HasTpm                                                0
      CountryIdentifier                                     0
      CityIdentifier                                        0
      OrganizationIdentifier                                0
      GeoNameIdentifier                                     0
      LocaleEnglishNameIdentifier                           0
      OsBuild                                               0
      OsSuite                                               0
      IsProtected                                           0
      SMode                                                 0
      Firewall                                              0
      Census_InternalPrimaryDiagonalDisplaySizeInInches     0
      Census_InternalPrimaryDisplayResolutionHorizontal     0
      Census_InternalPrimaryDisplayResolutionVertical       0
      Census_OSBuildNumber                                  0
      Census_OSBuildRevision                                0
      Census_OSInstallLanguageIdentifier                    0
      Wdft_IsGamer                                          0
      Wdft_RegionIdentifier                                 0
      HasDetections                                         0
      dtype: int64
```

```
[24]  df.info()
                                                                                                                    Python

...   <class 'pandas.core.frame.DataFrame'>
      RangeIndex: 702545 entries, 0 to 702544
      Data columns (total 26 columns):
       #   Column                                            Non-Null Count   Dtype
      ---  ------                                            --------------   -----
       0   IsBeta                                            702545 non-null  int64
       1   RtpStateBitfield                                  702545 non-null  int64
       2   IsSxsPassiveMode                                  702545 non-null  int64
       3   DefaultBrowsersIdentifier                         702545 non-null  int64
       4   AVProductStatesIdentifier                         702545 non-null  int64
       5   AVProductsInstalled                               702545 non-null  int64
       6   HasTpm                                            702545 non-null  int64
       7   CountryIdentifier                                 702545 non-null  int64
       8   CityIdentifier                                    702545 non-null  int64
       9   OrganizationIdentifier                            702545 non-null  int64
       10  GeoNameIdentifier                                 702545 non-null  int64
       11  LocaleEnglishNameIdentifier                       702545 non-null  int64
       12  OsBuild                                           702545 non-null  int64
       13  OsSuite                                           702545 non-null  int64
       14  IsProtected                                       702545 non-null  int64
       15  SMode                                             702545 non-null  int64
       16  Firewall                                          702545 non-null  int64
       17  Census_InternalPrimaryDiagonalDisplaySizeInInches 702545 non-null  int64
       18  Census_InternalPrimaryDisplayResolutionHorizontal 702545 non-null  int64
       19  Census_InternalPrimaryDisplayResolutionVertical   702545 non-null  int64
      ...
       24  Wdft_RegionIdentifier                             702545 non-null  float64
       25  HasDetections                                     702545 non-null  int64
      dtypes: float64(1), int64(25)
```

```
df.columns
```

```
Index(['IsBeta', 'RtpStateBitfield', 'IsSxsPassiveMode',
       'DefaultBrowsersIdentifier', 'AVProductStatesIdentifier',
       'AVProductsInstalled', 'HasTpm', 'CountryIdentifier', 'CityIdentifier',
       'OrganizationIdentifier', 'GeoNameIdentifier',
       'LocaleEnglishNameIdentifier', 'OsBuild', 'OsSuite', 'IsProtected',
       'SMode', 'Firewall',
       'Census_InternalPrimaryDiagonalDisplaySizeInInches',
       'Census_InternalPrimaryDisplayResolutionHorizontal',
       'Census_InternalPrimaryDisplayResolutionVertical',
       'Census_OSBuildNumber', 'Census_OSBuildRevision',
       'Census_OSInstallLanguageIdentifier', 'Wdft_IsGamer',
       'Wdft_RegionIdentifier', 'HasDetections '],
      dtype='object')
```

```
df.shape
```

```
(702545, 26)
```

```
x = df.iloc[:,:-1]
y = df.iloc[:,-1]
```

```
X_train,X_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
```

Adaboost

```
from sklearn.ensemble import AdaBoostClassifier
adb = AdaBoostClassifier()
adb.fit(X_train,y_train)
```

```
AdaBoostClassifier()
```

```
y_pred = adb.predict(X_train)
acc_adb1 = accuracy_score(y_train,y_pred)
acc_adb1
```

```
1.0
```

```
y_pred = adb.predict(X_test)
acc_adb2 = accuracy_score(y_test,y_pred)
acc_adb2
```

```
0.999966787496916
```

Stacking Classifier

```
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier, StackingClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.base import BaseEstimator, ClassifierMixin, clone

# Define base classifiers
base_classifiers = [
    RandomForestClassifier(n_estimators=100, random_state=42),
    GradientBoostingClassifier(n_estimators=100, random_state=42)
]

# Define meta classifier
meta_classifier = LogisticRegression()

# Create stacking classifier
stc = StackingClassifier(base_classifiers=base_classifiers, meta_classifier=meta_classifier)

# Train stacking classifier
stc.fit(X_train, y_train)
```

6.3 Results:

Home Page:

Fig2: Home Page

About:

Fig3: About

Registration:

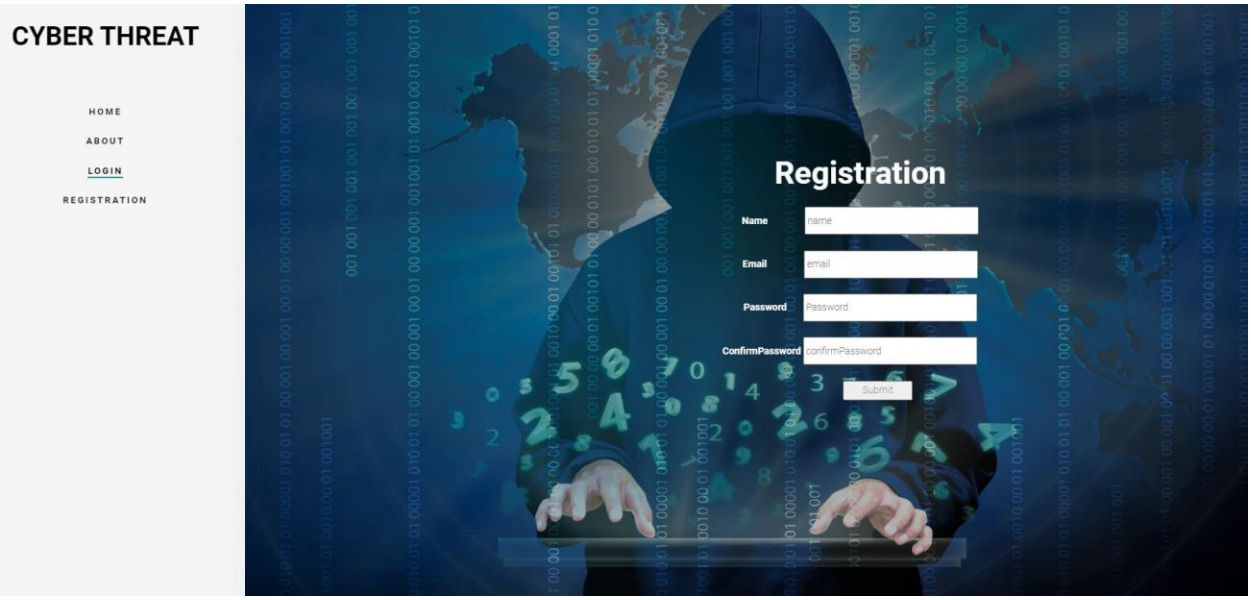


Fig4: Registration page.

Login:

Fig5: Login page.

Login Home Page:

Fig6: Login Home Page.

Upload Data Set:

Fig7: Upload page.

View Data:

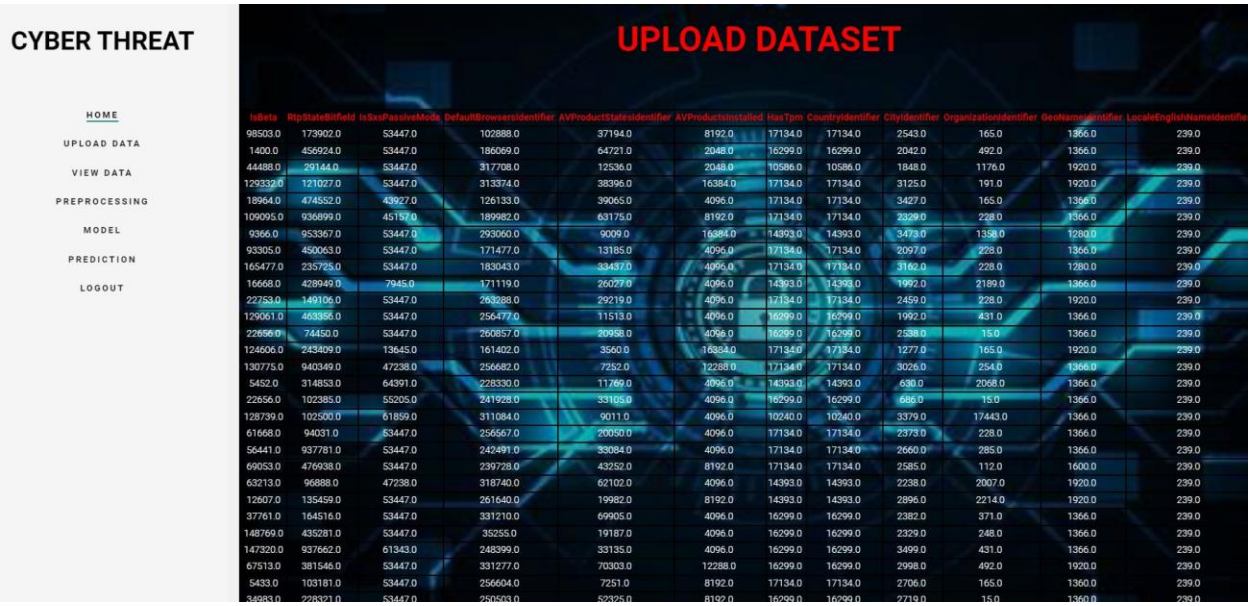


Fig8: View Data.

Preprocessing:

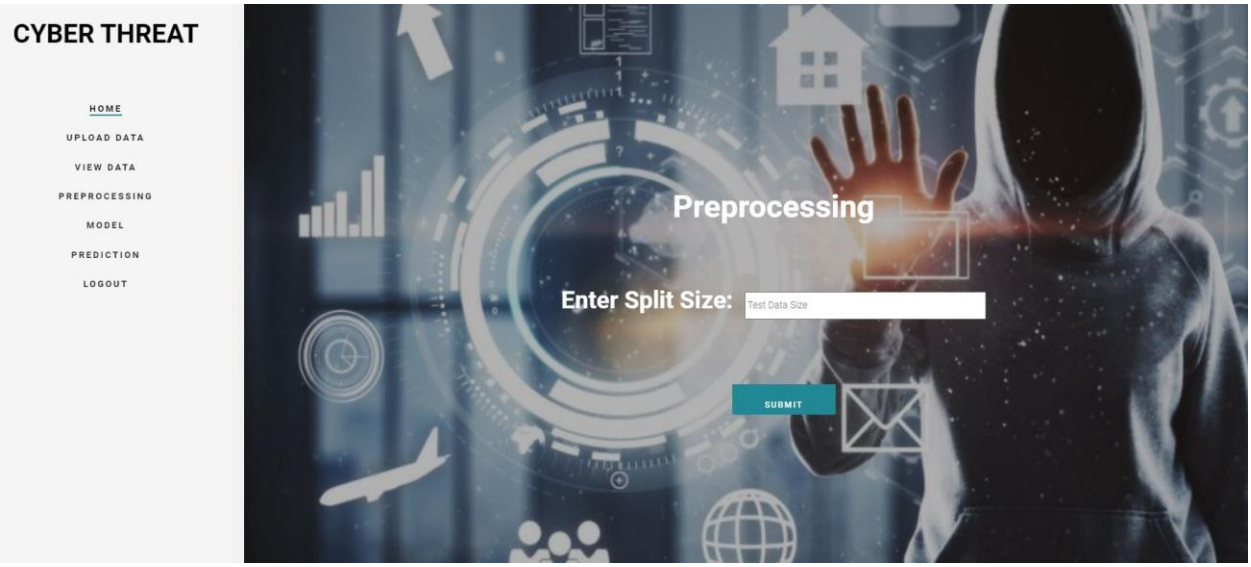


Fig9: Preprocessing page.

Training:

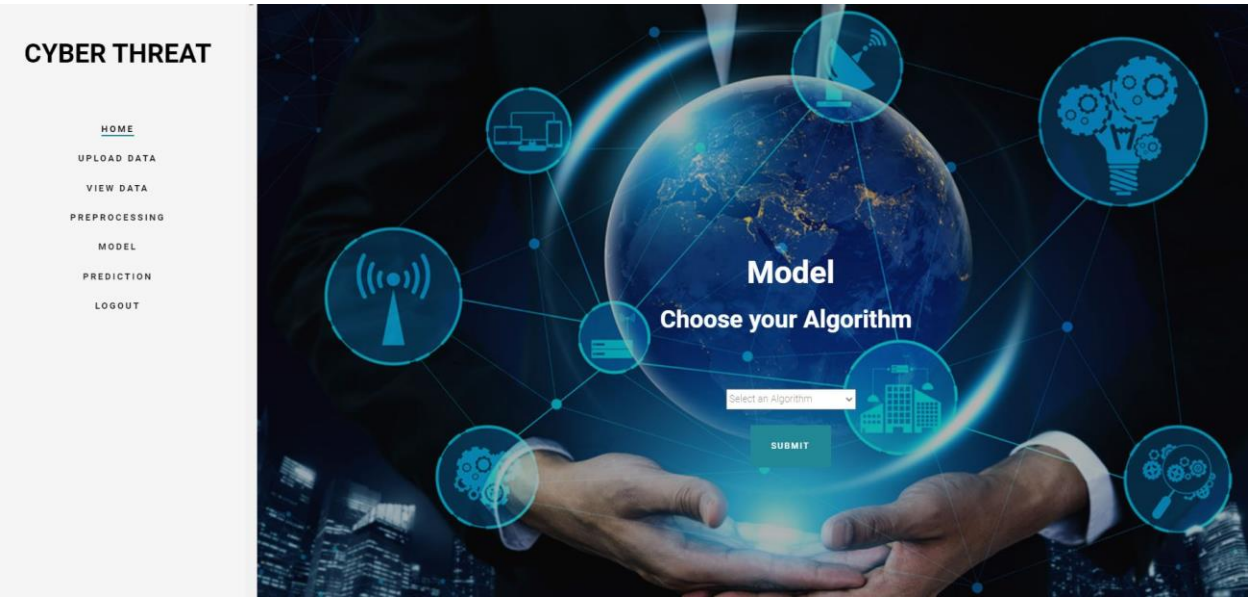


Fig4: Training Algorithm

Prediction:

Fig5: The cyber-attacks.

Fig6: The Not Cyber Attacks.

7. SYSTEM STUDY AND TESTING

7.1 Feasibility Study

In this phase, the feasibility of the project is assessed, and a business proposal is presented, including a preliminary project plan and cost estimates. The feasibility study during system analysis aims to ensure that the proposed system will not impose undue burdens on the company. To conduct a thorough feasibility analysis, a clear understanding of the system's major requirements is essential.

Three key aspects of feasibility analysis include:

Economic Feasibility

This analysis assesses the financial impact of implementing the system on the organization. Given the limited budget for research and development, it is crucial that expenditures are justified. The system must be developed within the allocated budget, which is feasible due to the use of many freely available technologies. Only customized products required specific purchases.Technical Feasibility

Technical Feasibility

This analysis evaluates the technical requirements of the system to ensure it does not place excessive demands on available resources. The system should have modest technical requirements, minimizing the need for significant changes to existing infrastructure. The goal is to ensure that the system can be implemented with minimal disruption to the client's current technical setup.

Social Feasibility

This aspect examines the user acceptance of the system. It involves training users to operate the system effectively and ensuring they perceive it as a valuable tool rather than a burden. Acceptance is influenced by the methods used to educate users and familiarize them with the system. Building user confidence is essential, allowing them to provide constructive feedback and integrate the system seamlessly into their workflow.

System Testing

The goal of system testing is to identify errors and ensure the quality of the software. It involves examining every possible fault or weakness in the product to ensure it functions as intended. Testing provides a method for evaluating the performance of individual components, sub-assemblies, assemblies, or the entire finished product. This process is designed to verify that the software meets its requirements and user expectations, and to prevent unacceptable failures. Various types of tests are conducted, each targeting specific aspects of the system's functionality and performance.

Types of Tests

72.1 Unit Testing

Unit testing involves creating test cases to validate the internal logic of individual software units, ensuring that inputs produce correct outputs and that all decision branches and code paths are functioning as expected. This type of testing is conducted on individual units of the application after their development but before integration with other units. It is a form of structural testing that requires detailed knowledge of the unit's construction and is inherently invasive. Unit tests are designed to verify specific business processes, application configurations, or system components, ensuring that each distinct path performs accurately according to documented specifications. The tests involve clearly defined inputs and expected results to confirm the correctness of the unit's functionality.

72.2 Integration Testing

Integration testing focuses on evaluating the interactions between integrated software components to ensure they work together as intended. This type of testing is event-driven and primarily concerned with verifying the overall outcome of integrated components, such as screens or fields. It aims to confirm that, while individual components may pass unit tests successfully, their combination functions correctly and consistently. Integration testing is specifically designed to identify issues that arise from the interaction between components. It involves incrementally integrating multiple components or software applications on a single platform, checking for errors caused by interface defects. The primary goal of integration testing is to ensure that combined components or applications interact seamlessly without errors.

Test Results: All the test cases were executed successfully, with no defects identified.

Acceptance Testing

User Acceptance Testing (UAT) is a crucial phase in any project that requires substantial involvement from end users. This phase ensures that the system meets all functional requirements as defined by the users.

Test Results: All test cases were successfully completed with no defects identified.

Functional Testing

Functional testing systematically verifies that the functions of the system perform as specified in the business and technical requirements, system documentation, and user manuals. This type of testing focuses on:

Valid Input: Ensuring that recognized categories of valid input are accepted.

Invalid Input: Confirming that recognized categories of invalid input are correctly rejected.

Functions: Testing all identified functions to ensure they operate as expected.

Output: Verifying that all expected application outputs are produced correctly.

Systems/Procedures: Ensuring that interfacing systems or procedures are properly invoked.

Functional testing is organized around requirements, key functions, and special test cases. It involves systematic coverage of business process flows, data fields, predefined processes, and successive processes. Additional tests are identified as needed, and the effectiveness of current tests is evaluated to ensure comprehensive coverage.

7.2.4 White Box Testing

White Box Testing involves examining the internal workings, structure, and code of the software. Testers have knowledge of the software's design and implementation, which allows them to test areas that are not accessible through black box testing methods. This approach ensures that the internal logic and functionality of the software are thoroughly evaluated.

7.2.5 Black Box Testing

Black Box Testing involves evaluating the software without any knowledge of its internal workings, structure, or code. In this approach, tests are designed based on external specifications or requirements documents, treating the software as a "black box" where only inputs and outputs are considered. The tester does not need to understand how the software processes the inputs to produce the outputs.

TEST CASES

Input	Output	Result
Dataset Input	Various models are evaluated using user-provided data on different algorithms.	Success
Gradient Boosting Classifier	Different user-provided inputs are tested on models created with various algorithms and datasets.	Success
Prediction	Predictions are generated using models built with the algorithm.	Success

- **Test cases for Model building:**

S.NO	Test cases	Input/Output	Expected Output	Actual Output	P/F
1	Dataset Loading	Dataset's path.	Path to the dataset	Dataset should be loaded successfully	Pass (If not, Fail due to incorrect format, e.g., not in .csv)
2	Data Preprocessing	Check for missing values/categories	Dataset Preprocessed Data preprocessing completed successfully	Data cleaned and processed	Pass (If not, Fail)

3	Model Training	Input data with specified algorithms for evaluation	Accuracy of each model represented as a percentage	Model accuracy metrics obtained for each algorithm	Pass (If not, Fail)
4	Prediction	User input values for prediction	Prediction results should be accurately produced using the chosen algorithm	Predictions generated successfully	Pass (If not, Fail)

8.CONCLUSION

In conclusion, this study demonstrates the effectiveness of predictive analytics, utilizing advanced machine learning techniques like CatBoost, XGBoost, and Stacking Classifier, to enhance cyber supply chain security. By analyzing a comprehensive dataset and leveraging diverse classifiers, the models developed in this research showcase the capability to forecast cyber threats with precision. Through the examination of various indicators and attributes, actionable insights are provided for fortifying supply chain networks against emerging risks. The amalgamation of classifiers such as Gradient Boost (GB) and AdaBoost offers a multifaceted approach to threat detection, ensuring the resilience and integrity of the cyber infrastructure. Overall, this study contributes to the ongoing efforts to improve cybersecurity in supply chain management.

9. FUTURE ENHANCEMENT

Future enhancements for improving cyber threat predictive analytics in the realm of supply chain security could involve incorporating more sophisticated anomaly detection techniques and integrating real-time data streams. By leveraging advanced anomaly detection algorithms such as Isolation Forest or Autoencoders, the models can better identify subtle deviations from normal behavior, thus enhancing their predictive capabilities. Additionally, integrating real-time data streams from various sources such as network traffic, endpoint logs, and threat intelligence feeds can provide timely insights into emerging threats, enabling proactive mitigation strategies. Furthermore, exploring ensemble methods that combine the strengths of different predictive models, including deep learning architectures, could further enhance the accuracy and robustness of the predictive analytics framework.

PUBLICATION DETAILS

Title: Cyber Threat Predictive Analytics for Improving Cyber Supply Chain Cecurity

Published Paper ID: JETIR2407698

Registration ID: 545797

Published In: Volume 11 | Issue 7 | Year July-2024

DOI (Digital Object Identifier):

Page No: g887-g901

Country: Jalandhar, Punjab, India .

Area: Science & Technology

ISSN Number: 2349-5162

Publisher: IJ Publication

Published Paper URL :: https://www.jetir.org/view?paper=JETIR2407698

Published Paper PDF: https://www.jetir.org/papers/JETIR2407698

7.95 impact factor calculated by Google scholar

10.REFERENCES

[1] National Cyber Security Centre. (2018). Example of Supply Chain Attacks. [Online] Available: https://www.ncsc.gov.uk/collection/supply chain-security/supply-chain-attack-examples

[2] A. Yeboah-Ofori and S. Islam, ‘‘Cyber security threat modelling for supply chain organizational environments,’’ MDPI. Future Internet, vol. 11, no. 3, p. 63, Mar. 2019. [Online]. Available: https://www.mdpi.com/1999-5903/11/3/63

[3] B. Woods and A. Buchman, ‘‘Supply chain in the software era,’’ in Scowcroft Centre for Strategic and Security. Washington, DC, USA: Atlantic Council, May 2018.

[4] Exploring the Opportunities and Limitations of Current Threat Intelligence Platforms, Version 1, ENISA, Dec. 2017. [Online]. Available: https://www.enisa.europa.eu/publications/exploring-the-opportunities and-limitations-of-current- threat-intelligence-platforms.

[5] C. Doerr, TU Delft CTI Labs. (2018). Cyber Threat Intelligences Standards—A High Level Overview. [Online]. Available: https://www.enisa.europa.eu/events/2018-cti-eu-event/cti-eu-2018-presentations/ cyber-threat-intelligence-standardization.pdf

[6] Research Prediction. (2019). Microsoft Malware Prediction. [Online]. Available: ttps://www.kaggle.com/c/microsoft-malware-prediction/data

[7] A. Yeboah-Ofori and F. Katsriku, ‘‘Cybercrime and risks for cyber physical systems,’’ Int. J. Cyber-Secur. Digit. Forensics, vol. 8, no. 1, pp. 43–57, 2019.

[8] CAPEC-437, Supply Chain. (Oct. 2018). Common Attack Pattern Enumeration and lassification: Domain of Attack. [Online]. Available: https://capec.mitre.org/data/definitions/437.html

[9] Open Web Application Security Project (OWASP). (2017). The Ten Most Critical Application Security Risks, Creative Commons Attribution-Share Alike 4.0 International incense . [Online] Available: https://owasp.org/ www-pdf-archive/OWASP_Top_10-2017_%28en%29.pdf.pdf

[10] US-Cert. (2020). Building Security in Software & Supply Chain Assurance. [Online]. Available: https://www.us-cert.gov/bsi/articles/ knowledge/attack-patterns

Project Plag report

by vishal choaudhary

Submission date: 31-Jul-2024 08:53PM (UTC+0530)
Submission ID: 2388216928
File name: FINAL_DOCUMENT_CAP769.pdf (3.31M)
Word count: 7281
Character count: 43843

ORIGINALITY REPORT

13%	10%	7%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	github.com Internet Source	3%
2	Submitted to Manipal University Student Paper	1%
3	Submitted to Universidad Carlos III de Madrid - EUR Student Paper	1%
4	www.ijraset.com Internet Source	1%
5	deepnote.com Internet Source	1%
6	Amir Shachar. "Introduction to Algogens", Open Science Framework, 2024 Publication	1%
7	Submitted to CSU Northridge Student Paper	1%
8	Submitted to Asia Pacific International College Student Paper	1%
9	Submitted to University of Bedfordshire	