THE VULNERABILITY OF CYBER SYSTEM UNDER STEALTHY ATTACKS

A Major Project Report Submitted in complete fulfillment of the

Requirements for the award of degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

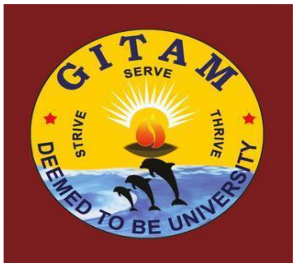
Submitted by

HEMANTH REDDY GALIGUTTA	**221810302021**
DUVVURI LAXMI PRANATHI	**221810302018**
CHINTAKINDI PRUTHVIRAJ	**221810302013**
SOMISETTY SAI CHARAN	**221810302056**

Under the esteemed Guidance of

MR. RAJ MOHAMMED

Assistant Professor, CSE Dept.



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM

(Deemed to be University)

HYDERABAD CAMPUS-502329

APRIL - 2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM SCHOOL OF TECHNOLOGY

GITAM

(Deemed to be University)

DECLARATION

We hereby declare that the major project report entitled "**THE VULNERABILITY OF CYBER SYSTEM UNDER STEALTHY ATTACKS**" done by us under the supervision of **MR. RAJ MOHAMMED, Assistant professor, Department of Computer Science Engineering** is submitted in complete fulfillment of the requirements for the award of Bachelor of Engineering in Computer Science Engineering.

Date:

Registration No(s).	Name(s)	Signature(s)
221810302021	**HEMANTH REDDY GALIGUTTA**	
221810302018	**DUVVURI LAXMI PRANATHI**	
221810302013	**CHINTAKINDI PRUTHVIRAJ**	
221810302056	**SOMISETTY SAI CHARAN**	

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM INSTITUTE OF TECHNOLOGY

GITAM

(Deemed to be University)

CERTIFICATE

This is to certify that this Major Project Report entitled "**THE VULNERABILITY OF CYBER SYSTEM UNDER STEALTHY ATTACKS**" is work of **HEMANTH REDDY GALIGUTTA (221810302021), DUVVURI LAXMI PRANATHI (221810302018), CHINTAKINDI PRUTHVIRAJ (221810302013)** and **SOMISETTY SAI CHARAN (221810302056)** in complete fulfillment of the requirement for the award of Bachelor of Technology in "COMPUTER SCIENCE ENGINEERING" at GITAM UNIVERSITY, Hyderabad.

Project Guide

Mr. RAJ MOHAMMED

Assistant Professor

Department of CSE

Head of the Department

Dr. S. PHANI KUMAR

Professor and HOD

Department of CSE

NAME AND SIGNATURE OF PROJECT PANEL MEMBERS:

1.

2.

3.

ACKNOWLEDGEMENT

Our project could no longer have been a achievement without the help of several people. We would love to thank the personalities who've been part of our challenge in numerous ways, individuals who gave us remarkable help from the start of the project.

We are exceedingly thankful to our respectable Pro-Vice-Chancellor, **Prof. N. Siva Prasad**, for presenting vital infrastructure and belongings for the accomplishment of our undertaking.

We are in particular obligated to **Prof. N. Seetharamaiah**, Principal, School of Technology, for his help in the course of the tenure undertaking.

We are very lots of grateful to our loved **Dr. S. Phani Kumar**, Head of the Department of Computer Science & Engineering, for providing the opportunity to undertake this assignment and encouragement in final touch of this assignment.

We hereby choice to specify our deep revel in of gratitude to **Mr. Raj Mohammed**, Assistant Professor, Department of Computer Science and Engineering, School of Technology, for the esteemed guidance, moral assistance, and beneficial advice provided with the resource of the use of the fulfilment of the undertaking.

We are also thankful to all of the personnel people of the Computer Science and Engineering department who've cooperated in making our task a fulfilment. Furthermore, we would love to thank all our dad and mom and buddies who extended their assistance, encouragement, and ethical assist both without delay or in a roundabout way in our project work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVATIONS

ABBREVATIONS		EXPANSION
AI	-	Artificial Intelligence
ML	-	Machine Learning
GUI	-	Graphical User Interface
PYPI	-	Python Package Index
API	-	Application Programming Interface
IPYNB	-	Interactive Python Notebook
RAM	-	Random Access Memory
CSV	-	Comma-Separated Values
UML	-	Unified Modeling Language
KW	-	Kilo Watt
RSS	-	Sum of squares of residuals
TSS	-	Total sum of squares
DFD	-	Data Flow Diagram

ABSTRACT

One of the number one annoying conditions in cyber safety is the delivery of an automated and effective cyber-threat detection technique. In the base paper, the linear gadget model of the cyber system is used on this have a look at to have a look at the impact of stealthy attacks at the cyber gadget. We superior an AI-SIEM gadget based mostly on a aggregate of event profiling for records preprocessing and one-of-a-type artificial neural network strategies, including CNN and LSTM. To compare the overall performance evaluation with current strategies, we performed experiments with the usage of the five traditional machine-mastering strategies like Support Vector Machine, K-nearest neighbors, Random Forest, Navies base, Decision tree. Vulnerability is described as a fixed of occasions that need to be met so as for it to exist. The experimental outcomes of this examination make certain that our proposed strategies are able to be hired as mastering-primarily based fashions for community intrusion-detection totally and display that even though it is hired withinside the actual world, the overall performance outperforms the traditional machine-mastering strategies.

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

Strict Stealthy and Stealthy from cyber systems are a number of the unique sorts of assaults that may be detected and labelled with the aid of using this mathematical linear system. It is sort of tough to locate a strict stealthy attack, even though it is feasible to locate a stealthy strike and hasn't any had an effect on harm or later sensor information after detection.

1.2 DEFINITION OF THE ISSUE

Because of this monitoring, humans will no longer be required to manually check the furnace temperature.

Another problem arises when a malicious attacker injects false readings into a sensor, causing the sensor to relay false data to the physical server, which in turn causes the physical server to take the wrong decision and result in losses.

The author of this research uses combat the aforementioned attacks, monitoring the state EIGEN values of each sensor's data to ensure that no attacks are identified even if NOISES are present. Strict Stealthy attacks can be difficult to distinguish from noise data if the attacker injects closed data into the typical range. When it comes to an assault that is so difficult to detect that even if it is removed, the system would still suffer if the attack is left in place.

1.3 AIM OF PROJECT

Strict Stealthy and Stealthy from Cyber System attacks can be detected and classified.

STRICT STEALTHY attacks might be difficult to detect if the attacker injects closed data within the typical range. It is nearly impossible to detect a strict covert attack like this, therefore removing it could result in noise or attack data, and leaving it will still have an impact on the system.

Eigen state values that fall within the normal range will be considered a STEALTHY attack and can be removed with no negative influence on the system after the attack has been eliminated.

Temperature can have a normal range of 15 to 45 degrees Fahrenheit defined as a default threshold, for example, and various sources can have a distinct normal range of readings. Attacks can be categorized based on deviations from the norm.

1.4LIMITATIONS OF PROJECT

The gathering and analysis of project requirements, the design of an application system and the actual implementation of that system
In-depth evaluation of my software by hand

1. System Application Deployment
2. Keep the project in good shape

1.5 STRUCTURE OF PROJECT

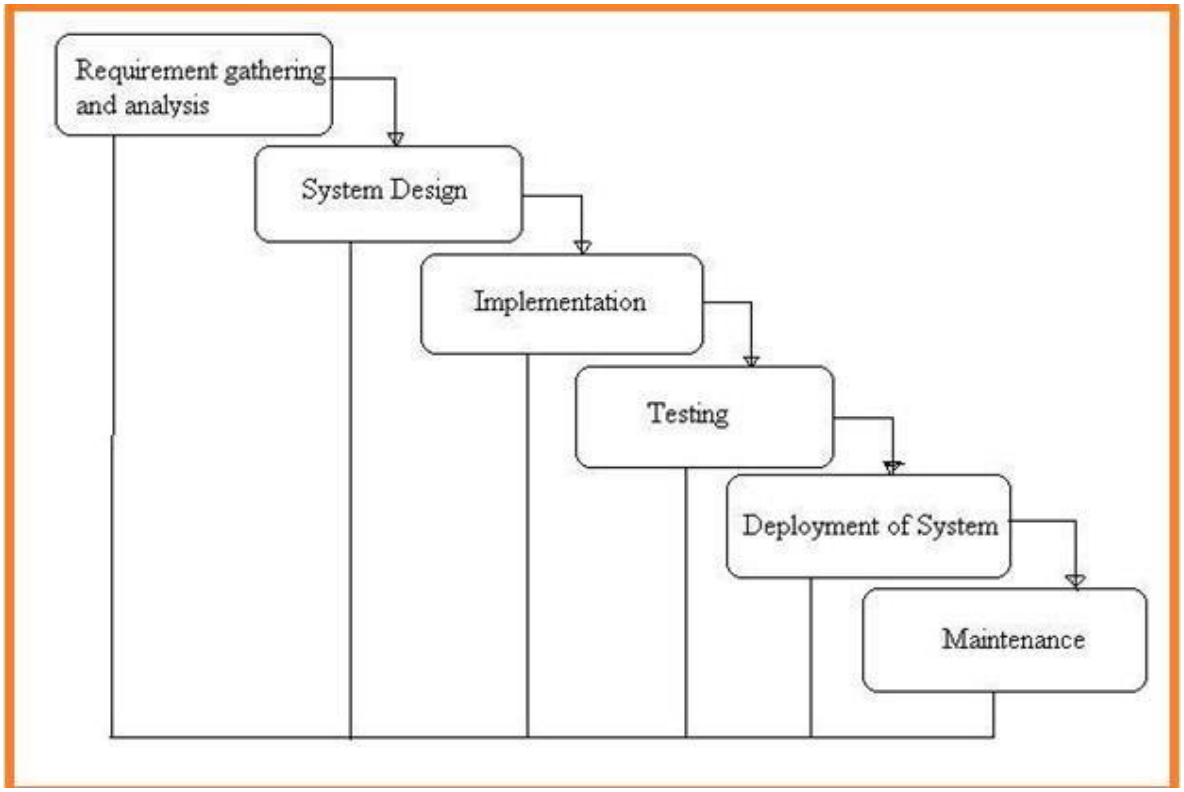


Fig: 1.1 SDLC

1.5.1 REQUISITES ACCUMULATING AND ANALYSIS

Our duty is a scholarly leave for basics; thus, we followed IEEE journals and amassed a plethora of IEEE relegated papers before separating a few at the final minute. Singular web revisitation by setting and substance importance input was assigned to the article, and for the research stage we used references from the paper and wrote reviews of particular papers, accumulating all of the requirements for the venture at this point.

1.5.2 SYSTEM DESIGN

There are three types of system design: GUI design, UML design, and a third type of system design that focuses on creating a project in an easy-to-understand manner with a wide variety of entertainers and its users. Class diagrams provide information on numerous classes in the project using strategies that must be employed in the project if our UML is to be used in this way. In the third step of the framework configuration, we attempt to plan the data base based on how many modules we have in our project.

1.5.3 IMPLEMENTATION

It is in this phase that much of the coding in business logic comes into play, making this phase the most critical and important component of the project's implementation.

1.6 TESTING

UNIT TESTING

Every stage of the project has been done by the developer, with the exception of this one, when we're going to fix all of the runtime errors.

MANUAL INSPECTIONS

In order to conduct manual testing because our project is currently on academic leave, we employ the endeavor and error way of testing

1.6.1 SYSTEM INSTALLATION AND MAINTENANCE

For our scholastic leave, we organized our school lab with all necessary software and a Windows OS only to send customer frameworks around the world.Our Project's Maintenance is a one-time activity.

1.7 FUNCTIONAL REQUIREMENTS

1. Data Gathering

2. Preparation of Data

3. Tests and training

4. Modeling

5. Predicting

1.8 NON-FUNCTIONAL REQUIREMENTS

A software system's quality trait is known as NON-FUNCTIONAL REQUIREMENT (NFR). These non-functional criteria are crucial to the success of the software system and are judged on their responsiveness, usability, security, and portability. "How quickly does the website load?" is an example of a nonfunctional demand. Non-functional requirements might lead to systems that don't meet the needs of the end user. Using non-functional requirements, you can restrict the system's design across several agile backlogs. Non-functional requirements. With a concurrent user count of more than 10,000, the site should be able to load in three seconds. Functional and non-functional criteria are equally important.

- Requirement for serviceability
- The need for manageability
- Recoverability is a pre-requisite
- The need for security
- Data Integrity is a need.
- Demand for capacity
- Requirement for accessibility
- The system must be scalable.
- Reliability is essential.
- Requirement for easy upkeep
- Requirement of the law
- Requirement of the environment

1.8.1 EXAMPLES

Examples of non-functional requirements include the following:

1. User datasets must be uploaded to the system.
2. An application's portability is essential. So, switching from one OS to another isn't a hassle.
3. Privacy, export restrictions, intellectual property rights, and other issues all need to be examined.

1.8.2 BENEFITS

Non-functional testing has the following advantages:

• Nonfunctional criteria ensure that the software system complies with legal and regulatory standards.

• They ensure the software system's stability, availability, and performance.

• They ensure that the program me is easy to use and provides a pleasant user experience.

• They assist in the development of the software's security policy.

1.8.3 DISADVANTAGES

Negative aspects of non-function requirements include:

- Software subsystems are not impacted by functional requirements.
- They necessitate additional work in the early stages of software architecture and high-level design.
- It is rare for their implementation to correspond to a specific software sub-system.
- After the architecture phase, it is difficult to change anything non-functional.

1.8.4 IMPORTANT LESSONS TO LEARN

The experimental results show that the proposed strategy can outperform most of the current methods.

CHAPTER 2

LITERATURE SURVEY

1. **TITLE: Invertibility of linear time-invariant dynamical systems**

AUTHORS: M. Sain and J. Mass

ABSTRACT: In many engineering applications, fault detection, isolation, and reconfiguration (FDIR) is a significant and difficult subject that is still being researched in the control community. An overview of the numerous model-based FDIR approaches developed during the past decade is shown here.. There are two parts to the FDIR problem: the detection and isolation (FDI) phase, and the controller reconfiguration (CR) phase. Different statistical strategies are discussed for FDI to test residuals for sudden changes in the form of a model-based residual that is robust to noise, unknown disturbances, and model uncertainties (or faults). In response to errors, we then discuss alternative methods for creating reconfigurable control strategy.

2. **TITLE: . Generic properties and control of linear structured syste**

AUTHORS: J. Dion, C. Commault, and J. Wou

ABSTRACT: A linear consensus network's trustworthy computation is addressed in this research. Motion coordination, clock synchronization, and cooperative estimation all benefit from a solution to this problem in multi-agent systems. Misbehaving agents can exist in a linear consensus network, and their actions will diverge from the nominal consensus progression. Our misbehavior detection and identification challenge is framed within an unknown-input system theoretic framework, in which we describe misbehaviors as network-wide effects with unknown and unmeasurable inputs. Our focus here is on agents that are either defective (non-colluding) or malignant (Byzantine). It is important to understand how misbehaving agents can influence the consensus network without being recognized or identified by other agents who are participating in the network. There are also limits on how many concurrently malfunctioning or hostile agents can be discovered and recognized. Precisely, for k malevolent (resp. faulty) agents to be traceable and recognized by every well-behaving agent, the consensus network must be 2 k +1 (resp. k +1) connected. As a last step, we calculate the impact of hidden inputs on the final consensus result. Finally, we develop three techniques for detecting and identifying bad actors. Using defect detection techniques, the first and second algorithms are able to provide complete detection and identification, albeit at a high computational cost. Local detection and identification of misbehaving agents whose behavior deviates above a threshold are provided by the third method, which takes advantage of the network's weakly coupled subparts.

3. **TITLE: The detection and identification of cyber-physical system attacks.**

AUTHORS: F. Pasqualetti, F. Dorfler, and F. Bullo are the authors of this book.

ABSTRACT: The use of cyber-physical systems is becoming increasingly common in a variety of sectors, including electric power generation and distribution, transportation, process automation, and other essential infrastructure. These systems have to be able to withstand both expected and unexpected failures, as well as hostile attacks from the outside. Mathematical frameworks for cyber-physical systems, attacks, and monitors are presented in this paper, as well as system- and graph-theoretic and graph-theoretic descriptions of fundamental monitoring limitations. Finally, centralised and distributed attack detection and identification monitors are proposed. Finally, we provide compelling instances to back up our assertions and findings alike.

4. **TITLE: Wireless control network: Detection of harmful activity**

AUTHORS: Pajic, Hadjicostis, R. Mangharam, and G. Pappas are the authors.

ABSTRACT: Stabilizing a plant with wi-fi nodes which have confined assets is a challenge. As a part of a associated study, we devised a protocol wherein every node time and again communicates a linear mixture of statistics from its instant surroundings. Our aim on this paper to layout and put into effect an IDS for this manage scheme, which video display units the communications of precise nodes and makes use of that statistics to each get better plant outputs for facts logging and diagnostic purposes, in addition to become aware of malicious behaviors through any of the community's wi-fi nodes. We show that the IDS best desires to look at a part of the community nodes to attain this aim if the community connectivity is excessive enough. An top restriction at the time required to acquire the applicable statistics is furnished through our approach.

5. **TITLE: Sensor and actuator attacks on control systems provide a significant security risk.**

AUTHORS: Paulo Tabuada and Suhas Diggavi are the authors of this book.

ABSTRACT: When a malevolent entity attacks part of the sensors or actuators in a linear system, the challenge of estimation and control becomes more difficult. The estimate problem was previously framed as a dynamic error correction problem with sparse attack vectors, which we studied in our earlier work. Our research into the role of inputs and control is furthered in this publication. State feedback can be used to strengthen the system's resilience against attacks by modifying the system dynamics while allowing (nearly) absolute freedom in the placement of the new poles. When it comes to stabilizing a plant via output-feedback despite attacks on sensors, we show that a notion of separation of estimation and control holds.

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM:

For example, in an industrial furnace, a sensor will measure the temperature and transmit the information to a physical server, which will then determine whether the temperature is excessive and send a cooling instruction to the sensor. Because of this sensor monitoring, humans will no longer be required to manually check the furnace temperature.

Another problem arises when a malicious attacker injects false readings into a sensor, causing the sensor to relay false data to the physical server, which in turn causes the physical server to take the wrong decision and result in losses.

Disadvantages:

If a malicious attacker manipulates a sensor, it could send false information to the server, causing the server to make an incorrect decision, which could result in financial loss.

3.2 PROPOSED SYSTEM:

Strict Stealthy and Stealthy from Cyber System are some of the different types of attacks that can be detected and classified by this mathematical stochastic linear system. It is nearly hard to detect a strict stealthy attack, although it is possible to detect a stealthy strike and have no effect on damage or later sensor data after detection.

Advantages:

A performance limit is provided for the difference between a healthy and an attacker system.

For the vulnerability to exist, we need to establish all the conditions that are both necessary and adequate.

3.3 SYSTEM STUDY

FEASIBILITY STUDY

During this stage, the project's viability is assessed, and a business proposal containing a high-level outline of the project and some rough cost estimates is presented. In order to determine whether or not the proposed system is viable, a feasibility study must be conducted. So that the suggested system isn't a burden on the business, this is a requirement. Understanding the system's primary requirements is critical while conducting a feasibility study.

COST COMFORTABILITY

To see how the system would affect the company's bottom line, this research is conducted. The corporation has a limited quantity of money to invest in the system's research and development. It's imperative that the costs be deemed necessary. In order to keep costs down, most of the technologies used in the designed system were free of charge. Customized goods were the only ones that had to be purchased.

EFFICIENCY IN TECHNOLOGY

The goal of this research is to look at the technological viability of the system, as well as the s ystem's technical requirements.The use of technological resources by a system must be maint ained to a minimum.This will result in a shortage due to the increased demand for technical r esources.As a result, the client will be held to unrealistic standards.Because relatively minor o r no alterations are required to execute this system, it must have a minimal demand.

SOCIAL COMFORT

The study's purpose is to determine how effectively the system is accepted by the intended au dience.This includes demonstrating how to get the most out of the technology to the user.The user should not be concerned about the system; rather, he or she should accept it as an unavoi dable evil.

3.4 CONSTRUCTION OF A PREDICTIVE MODEL

Machine gaining knowledge of wishes records collecting have lot of beyond records. Data collecting have enough historic records and uncooked records. Before records pre-processing, uncooked records can’t be used directly. It’s used to preprocess then, what form of set of rules with version. Training and trying out this version running and predicting effectively with minimal errors.

● Learning processes: This factor of ML programming makes a specialty of obtaining records and developing guidelines for a way to show the records into actionable information. The guidelines, which might be known as algorithms, offer computing gadgets with step-by-step commands for a way to finish a particular task.

● Reasoning processes: This factor of ML programming makes a specialty of selecting the proper set of rules to attain a favored outcome.

● Self-correction processes: This factor of ML programming is designed to always fine-track algorithms and make sure they offer the maximum correct outcomes possible. The assessment

turns into extra biased as talent at the validation dataset is included into the version configuration.

The validation set is used to evaluate a given model, but this is for not unusual place assessment. It as tool studying engineers uses this facts to fine-music the model hyper parameters. Data collection, facts analysis, and the way of addressing facts content, quality, and form can add as lots as a time-ingesting to-do list. During the way of facts identification, it permits to recognize your facts and its properties; this statistics will help you choose out which set of guidelines to use to assemble your model. A amount of diverse facts cleaning duties using Python's Pandas library and specifically, it interest on in all likelihood the biggest facts cleaning task, missing values and it able to greater short clean facts. It wants to spend lots much less time cleaning facts, and additional time exploring and modeling.

CHAPTER 4

SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

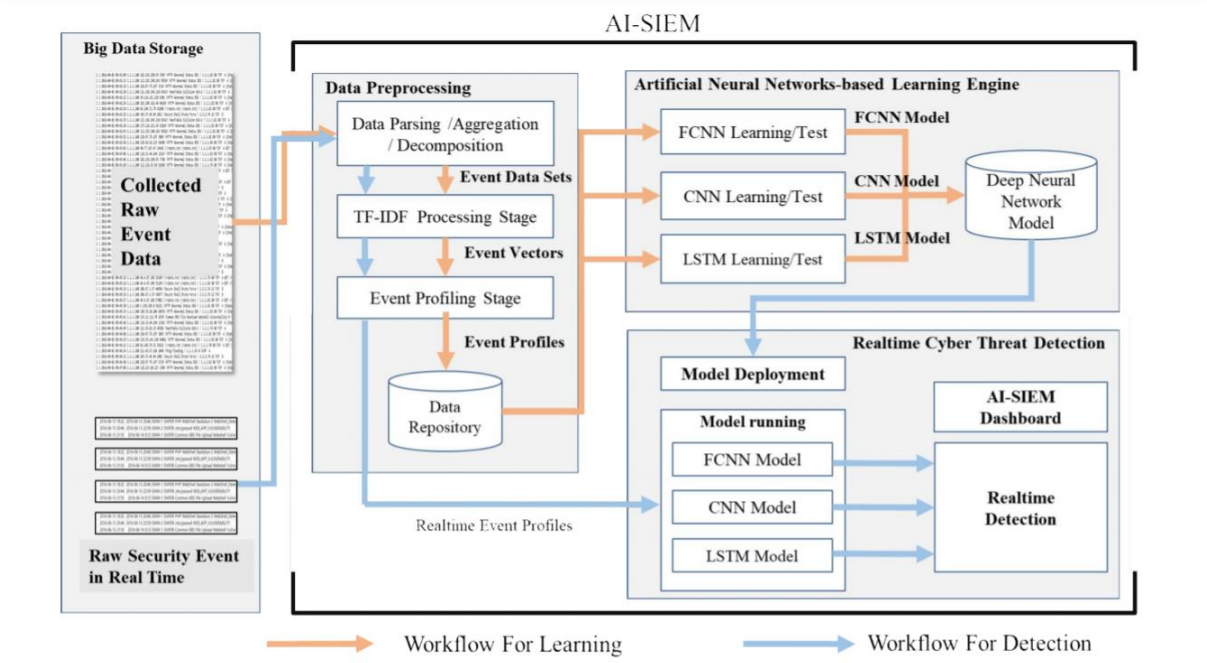


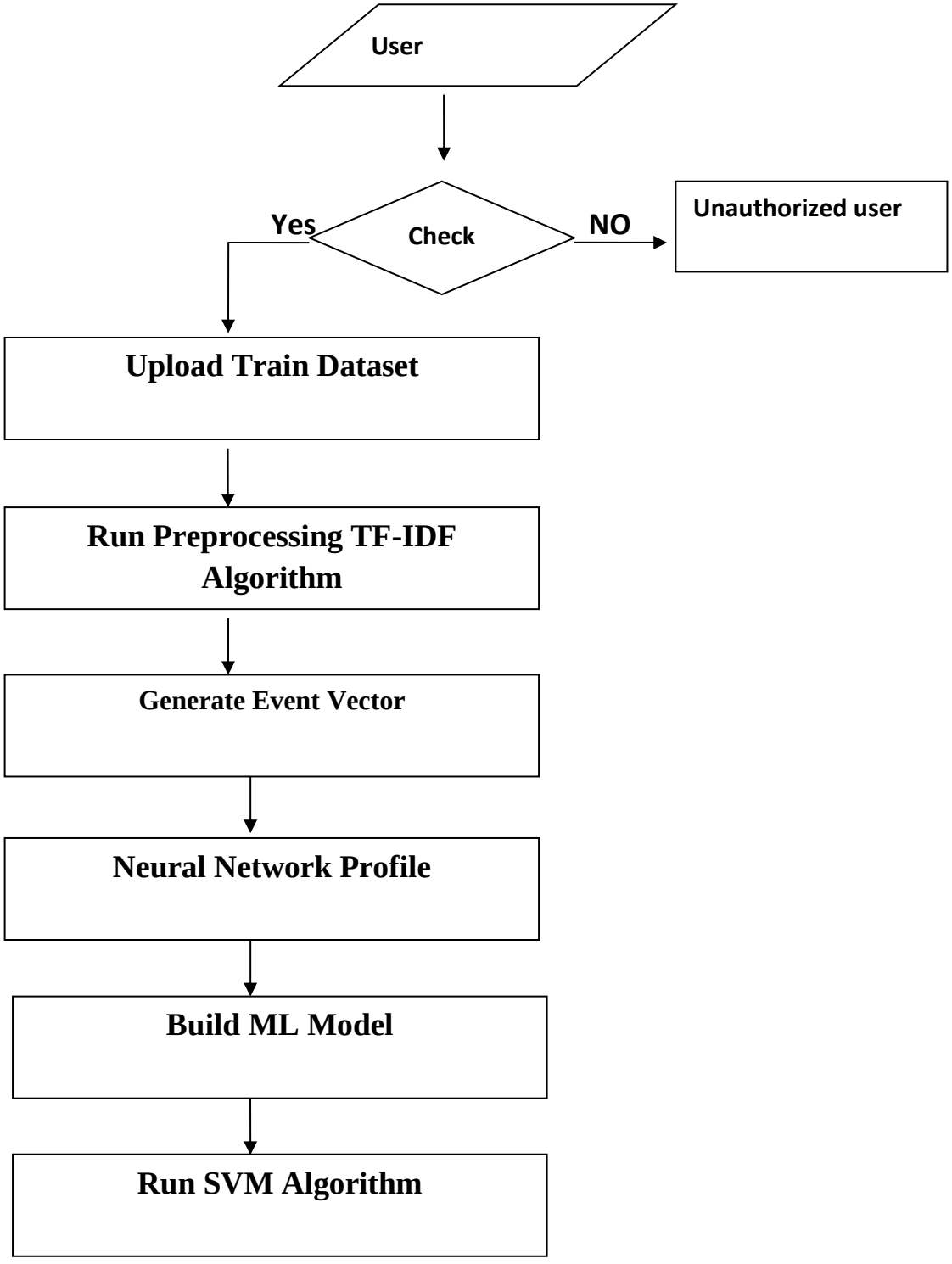
Figure 4.1: System architecture

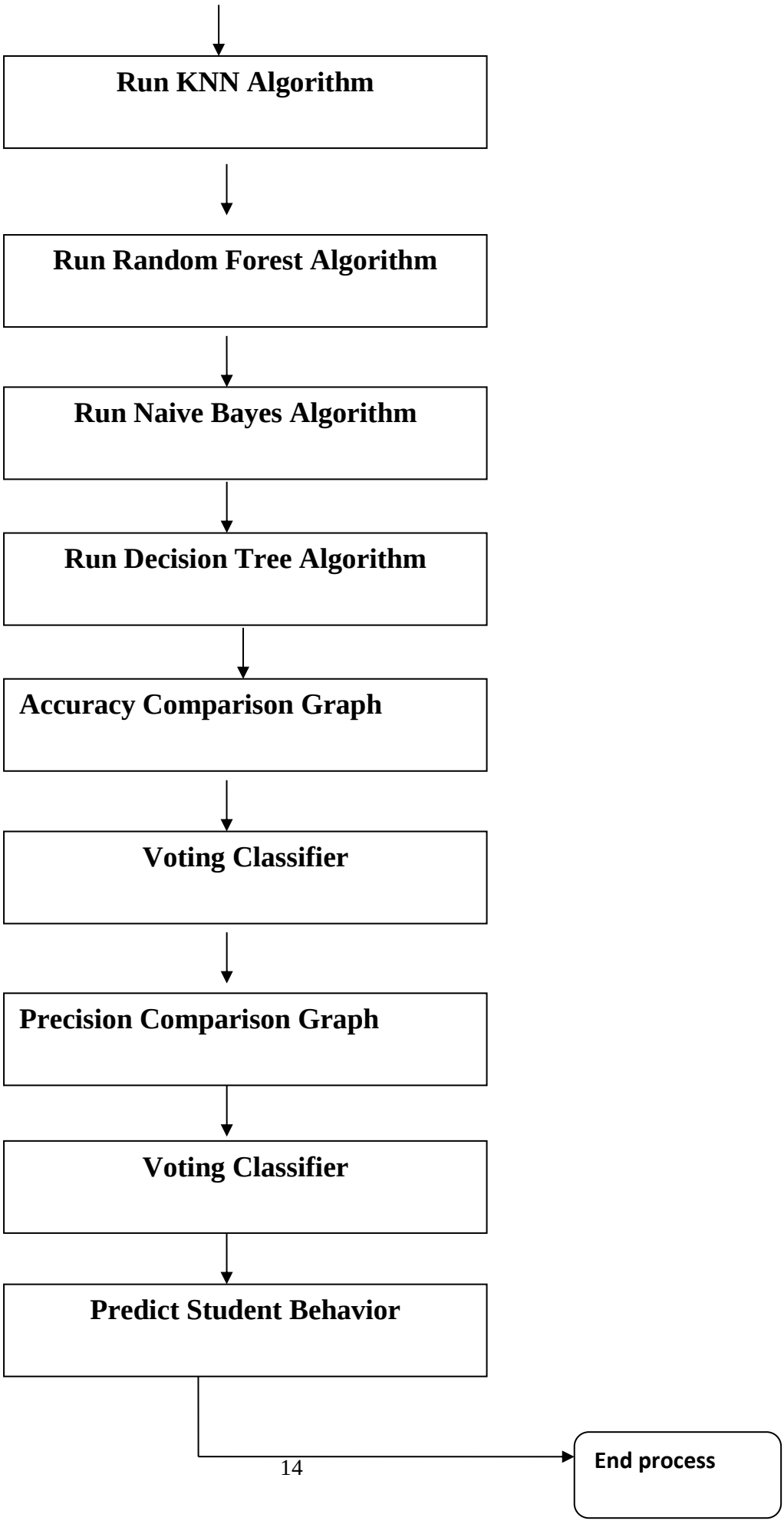
4.2 Model description

First, people suggest to gather the facts records. These are the enter facts records, after which shifting onto the exceeded records from each domain, we ought to bear the records from a selected department. Includes the records data of changing the enter facts. Finally, keep the exceeded records we should expect the use of the beyond records predicted in destiny enter fee those are the transmitted records. Then we calculate the pre-processing records for the primary modules, the primary module is pre-processing, and the second one module is records visualization. Finally, we evaluate the ml set of rules to expect each set of rules generating the

accuracy and which one has the very best accuracy. We should calculate; that is the device structure this is taking place to challenge workflow.

4.3 VARIABLE DATA FLOW DIAGRAM




Run KNN Algorithm
Run Random Forest Algorithm
Run Naive Bayes Algorithm
Run Decision Tree Algorithm
Accuracy Comparison Graph
Voting Classifier
Precision Comparison Graph
Voting Classifier
Predict Student Behavior
End process

1. The DFD is known as bubble chart. Input data, various processing operations on that data, and the output data generated by the system can all be represented graphically using simple formalism. In a word, yes.

2. When it comes to modelling, data flow diagram (DFD) is an essential tool. Using this application, you may see how things work in a system in a virtual environment. It is important to understand the system as a whole, including its processes, data, and external entities.

3. The data flow diagram (DFD) shows how information moves through the system and how it is transformed through a sequence of transformations. This technique depicts the flow of information and the adjustments that are performed as data moves from input to output.

4. The DFD can also be referred to as a bubble chart. A DFD can represent any level of abstraction. Based on the information flow and functional detail, a DFD can be divided into tears.

4.4 UML DIAGRAMS

The acronym UML stands for Unified Modeling Language. UML is a standard modelling language for object-oriented software engineering, and it is widely used in the industry. The Object Management Group manages and established the standard. Models of object-oriented computer software should be created using the UML language.

There are two key components to UML:

1. Meta-model and a notation in its current form. A method or process may also be added to, or related with, UML in the near future. If you're looking for a common way to describe and document the artefacts of any kind of software or non-software system, you can use the Unified Modeling Language (UML).
2. For large and complex systems, the UML is a compilation of the finest engineering approaches that have proven successful.

The UML is an essential aspect of object-oriented software development and the software development process. For the most part, software project design is expressed using graphical notations in the UML.

GOALS:

The following are the primary aims of the UML design:

1. Providing users with an expressive visual modelling language that is ready-to-use will allow them to create and exchange meaningful models.
2. Provide means for extending the fundamental concepts through extendibility and specialization.
3. The ability to work in any programming language or methodology is a must-have.
4. This step establishes a formal foundation for comprehending the modelling language.
5. Encourage the expansion of the market for OO tools.
6. Support higher-level ideas such as collaborations, patterns and components in the development process.
7. Best practices should be incorporated.

USE CASE DIAGRAM:

Unified Modeling Language (UML) use case diagrams are a specific sort of behavioral diagram that are produced as a result of doing a case study. Graphically depicting the system's actors, their goals (expressed as use cases), and any interdependencies across use cases is its primary goal. A use case diagram's primary goal is to explain how the system performs for each actor. The system's actors can be portrayed in detail.

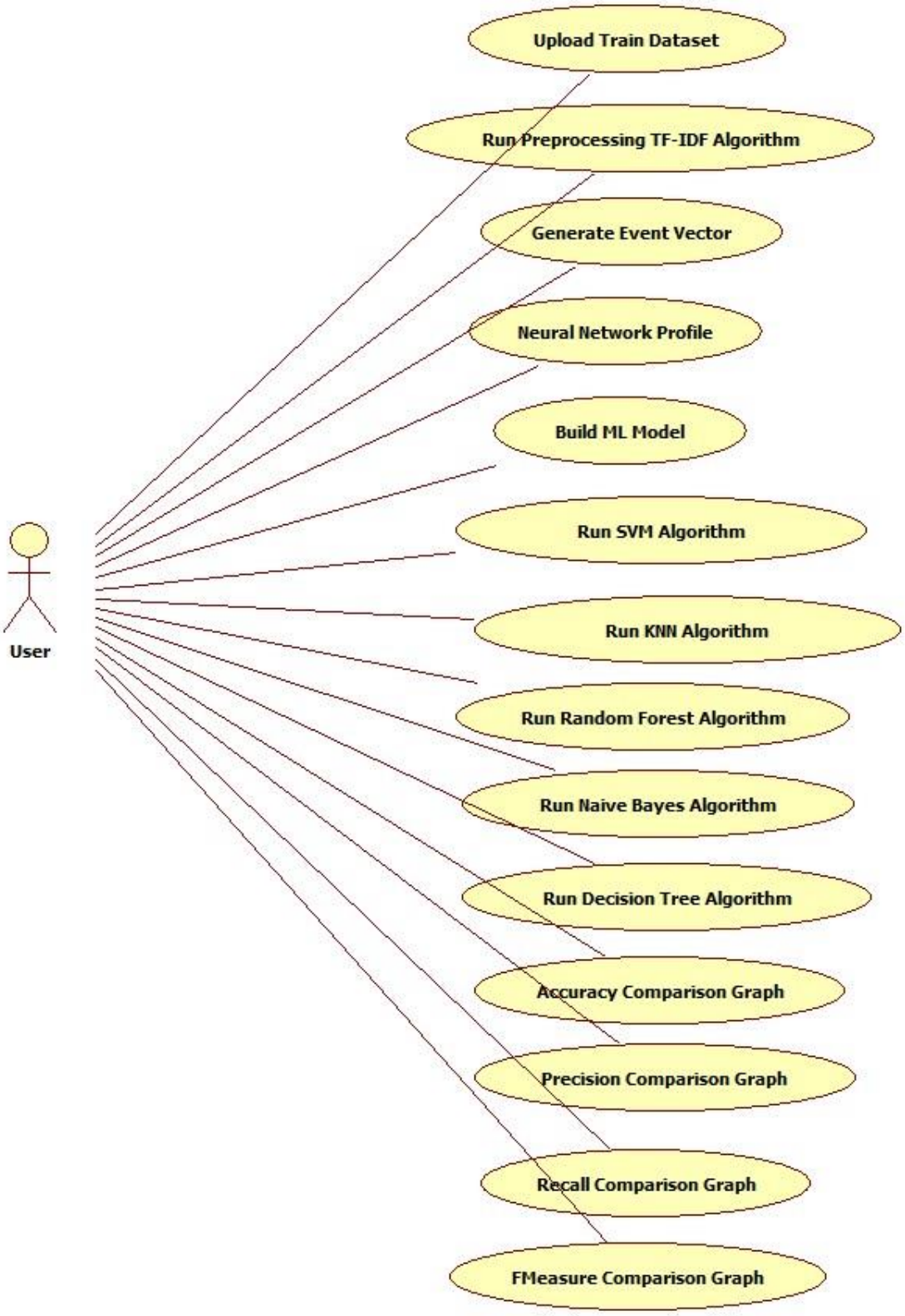


Figure 4.3: Use case diagram

Object diagram:

The object diagram is a subset of the class diagram, which is a particular case of the class diagram. A class's instance is an object. An object represents the current state of a class during the course of the system's operation. The object diagram depicts the current state of the system's many classes, as well as the connections or associations that exist between them at any one time.

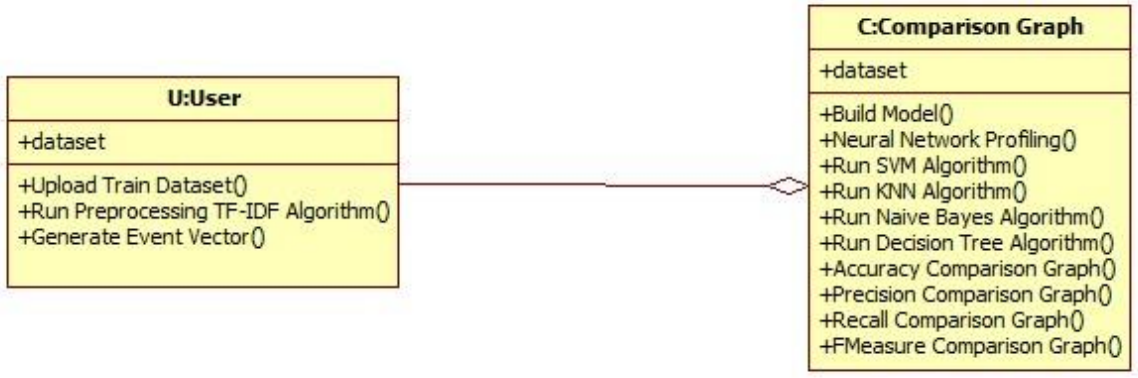


Figure 4.4: Object diagram

Activity diagram:

The activity diagram depicts how the system's many processes move back and forth. An activity diagram is similar to a state diagram in that it includes activities, actions, transitions, initial and final states, and guard conditions.

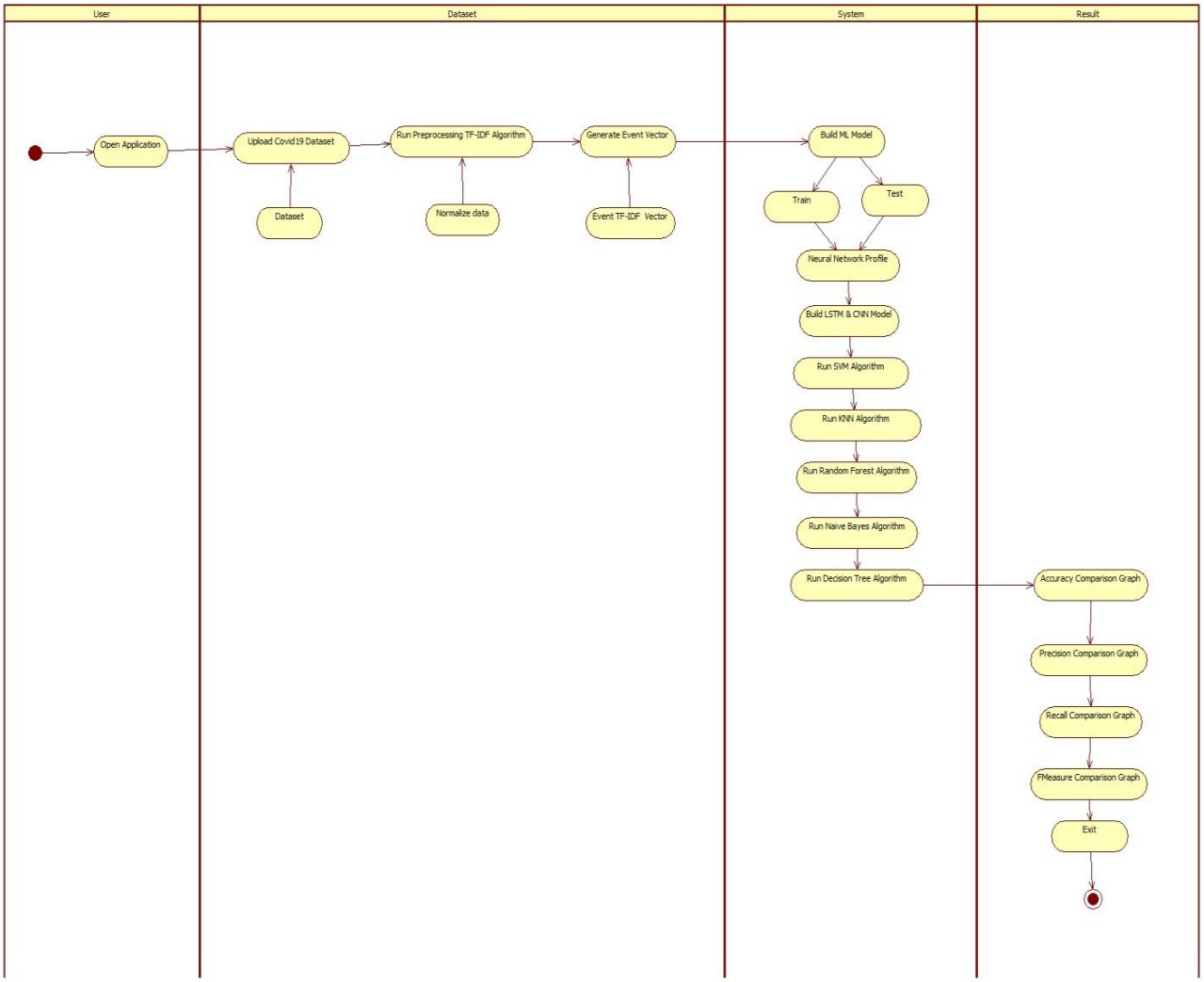


Figure 4.5: Activity diagram

Sequence diagram:

A sequence diagram is a visual representation of how various components in a system interact with one another. A sequence diagram's most critical feature is that it's chronologically arranged. A step-by-step representation of the interactions between the items is what this signifies. There are many "messages" that are exchanged between the various sequence diagram elements.

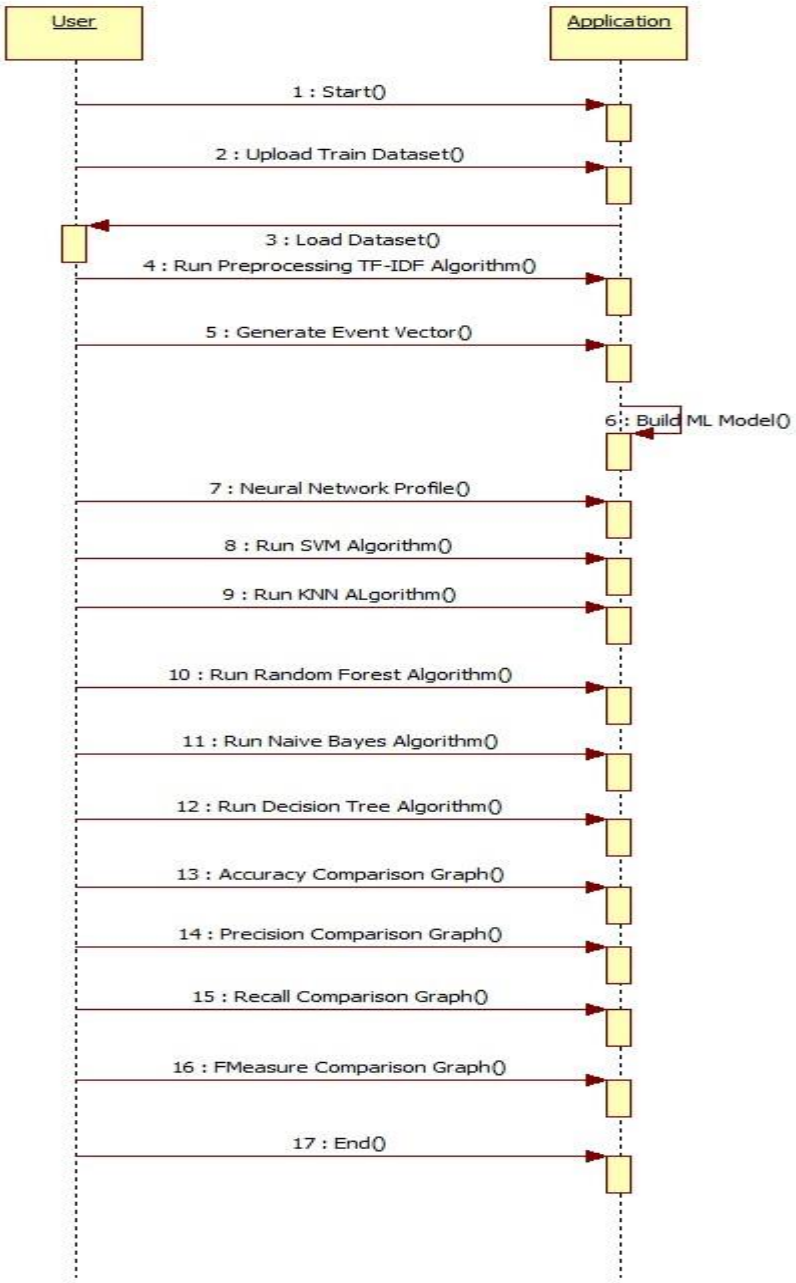


Figure 4.6: Sequence diagram

Collaboration diagram:

A cooperation diagram is a visual representation of how different items interact with one another. Numbered interactions make it easier to follow the chain of events that led up to the engagement. Using a collaboration diagram, you can see all the possible interactions that each object has with other things.

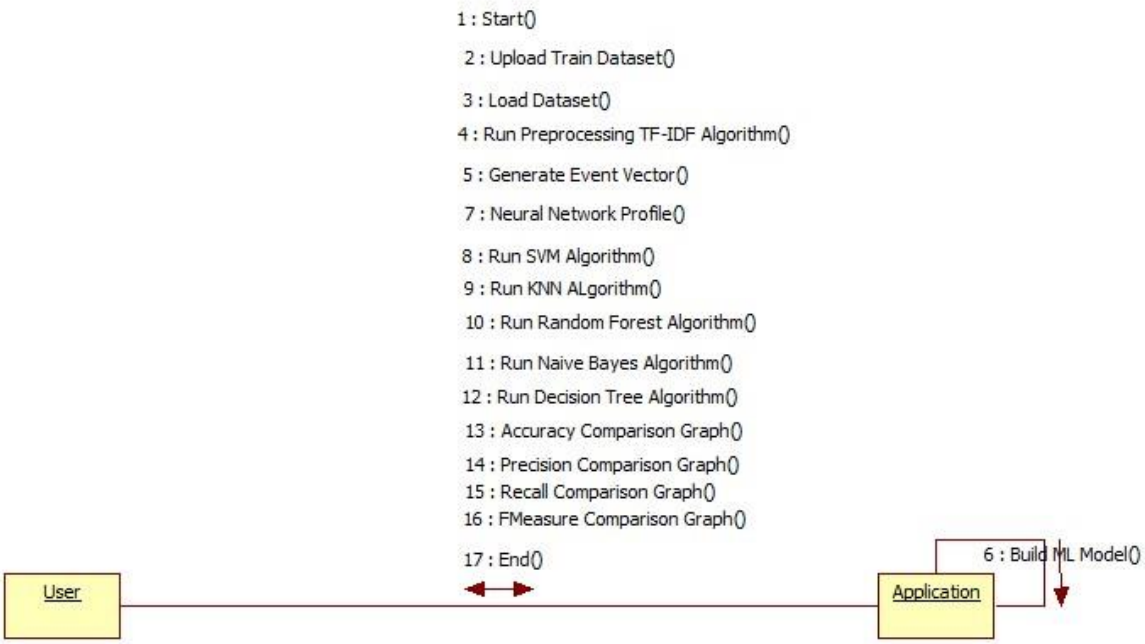


Figure 4.7: Collaboration diagram

CHAPTER 5

IMPLEMENTATION

Pre-processing manner statistics cleaning, which might also additionally require toggle statistics. It may be noisy statistics, nameless statistics, and null values may be there, so we must pre-method and easy the statistics.

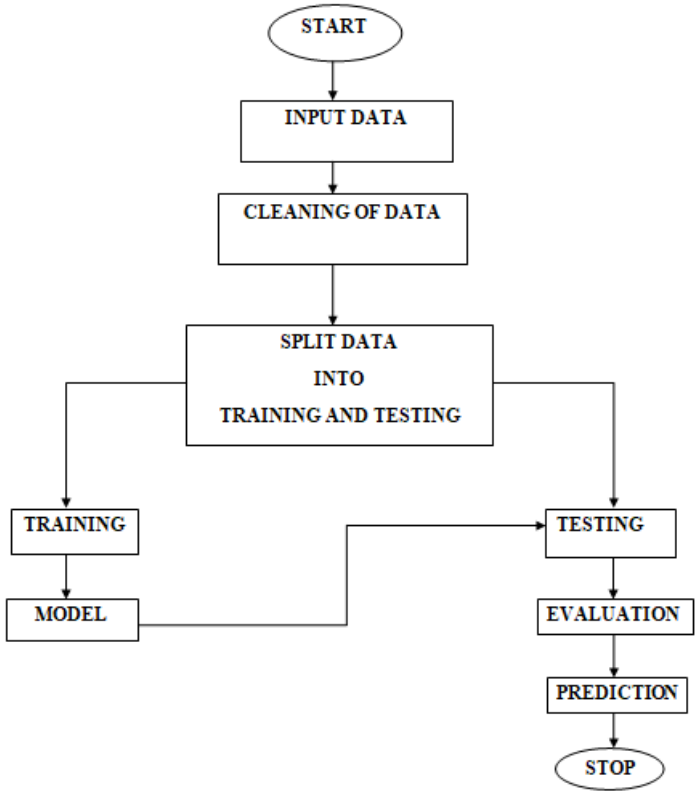


Figure 5.1: Block Diagram

5.1 PROBLEM STATEMENT

The prediction of the cyber-attack's detection. The maximum current research withinside the subject of intrusion detection have given multiplied consciousness to gadget studying and synthetic intelligence strategies for detecting attacks. Advancement in AI fields can facilitate the research of community intrusions with the aid of using protection analysts in a well timed and automatic manner.

5.2 DATA SET DESCRIPTION

In this study are performed by authors using two benchmark datasets (NSLKDD and CICIDS2017) and two datasets collected in the real world

5.3 OBJECTIVE OF THE CASE STUDY

The main objective of the proposed method is to cyber-attacks. we developed an AI-SIEM system based on a combination of event profiling for data pre-processing and different artificial neural network methods, including FCNN, CNN, and LSTM. The system focuses on discriminating between true positive and false positive alerts, thus helping security analysts to rapidly respond to cyber threats.

5.4 SYSTEM DESIGN

Pre-processing of The Data

Prepossessing of the data actually involves the following steps:

Getting the Dataset

We can get the data set from the database or we can get the data from client.

Importing the Libraries

we have to import the libraries as per the requirement of the algorithm.

5.5 ALGORITHMS IMPLEMENTATION

Naive Bayes Classifier: The Naive Bayes classifier is a one-of-a-kind algorithm based on the idea that all talents are separate and unrelated. It defines that a single function's reputation in a class has no influence on the reputation of other functions in the class. Because it is based on conditional danger, it is seen as a strong set of constraints employed for humanistic goals. It's particularly useful for records with balance issues or missing data. Naive Bayes, a tool mastering classifier, employs the Bayes Theorem.

Decision Tree Classifier: A Decision Tree is a supervised tool that develops a set of suggestions for dealing with various sorts of problems. The fundamental goal of using Decision Tree to those study artworks is to predict target attractiveness using a wish rule developed from past data. Nodes and internodes are used for prediction and type. Instances are classified by root nodes depending on their unique properties. Root nodes should have or bigger branches at the same time since leaf nodes represent type. The Decision Tree picks each node at each level by analyzing the rise in first-class statistics across all criteria. The Decision Tree approach's average overall performance was examined.

Support Vector Machine (SVM): The term "supervised learning model" refers to a collection of supervised learning models that are utilized in categorization. By disregarding hyperplane many of the two classes, the purpose of an assist vector tool is to acquire the best highest-margin out of a two-class training sample. Falsehoods regarding facts components do not belong in the opportunity class for improved generalization, according to Hyperplane. The hyperplane must be set up in such a way that the factual elements of each category are represented. The objects closest to the classifier's margin are the assist vectors. The correctness of the check is evaluated using the WEKA interface.

KNN nearest neighbor algorithm: K-Nearest Neighbors is one of the maximum simple but crucial category algorithms in Machine Learning. It belongs to the supervised gaining knowledge of area and reveals severe software in sample recognition, records mining and intrusion detection. It is broadly disposable in real-lifestyles situations due to the fact it's far non-parametric, meaning, it does now no longer make any underlying assumptions approximately the distribution of records.

Random forest: Atype of supervised learning algorithm that is commonly used in classificati on and regression issues.It constructs choice bushes from exceptional samples and uses the m ajority vote for category and common in regression.The Random Forest Algorithm's ability to take care of greater effects for category problems is one of its most important skills.

CHAPTER 6

TESTING AND TRAINING

Errors are discovered during testing. The goal of testing is to find any and all flaws in a product or service. Components and subassemblies, as well as finished products, can be tested to ensure their functionality. If you want to make sure that the software system doesn't break in an unacceptable way and meets all of its needs and expectations, it is the process of exercising software. There are a wide variety of exams to be taken. Each test type is designed to meet a certain need.

6.1 TYPES OF TESTS

6.1.1 UNIT TESTING:

When a program's logic is tested, it may be assured that the program's inputs and outputs are legitimate. It is essential to test each step of the decision-making process as well as the internal code flow. Piecemeal testing is done on the application. After each unit is completed, it is integrated into the whole. In order to do this invasive structural test, you must be familiar with the design of the system beforehand. In order to evaluate a specific business process, application, or system configuration at the component level, this type of test is utilized Unit tests are one method of verifying that a business process adheres to its documented criteria.

6.1.2 INTEGRATION TESTING

An integration test's job is to make sure that all of the software components that have been combined work together as a single unit. The importance of screen or field outcomes has waned in testing. However, even if each component has passed unit testing, it is still correct and consistent to put them together. Interfacing two or more pieces of software together is the primary goal of integration testing.

6.1.3 TEST OF FUNCTIONALITY

To guarantee that business and technical requirements, system documentation, and user guides are met, functional tests give systematic demonstrations that the evaluated functionalities are available in the required form.

Functional tests are organized and prepared in accordance with the requirements, important functionalities, or specific test cases that they are designed to evaluate. The system must also be tested for its capacity to cover all of the required data fields, pre-programmed processes, and subsequent operations. Before functional testing is finished, new tests are discovered and the value of the ones that already exist is assessed.

6.1.4 SYSTEM TEST

It is only via system testing that you can be certain that your integrated software system is up to par with the required specifications. A collection of tests is run on a particular setup to ensure that the results are in line with what was predicted. A system testing method is the configuration-oriented system integration test. Pre-driven process links and integration points are the primary focus of system testing.

6.2 SPLITTING TRAIN/TEST:

Preprocessing, split test and dataset, cut up reaction variable. We have a few values withinside the statistics set, then we must separate the schooling set, and the trying out set. Here schooling set is extra than the trying out set, on the way to provide extra accuracy. The intake is the goal column method the output column. After the label encoder method, integer values modified into comprehensible pc values. As intake is our goal column, we must educate the entirety besides intake, and we must lock those into the trying out process.

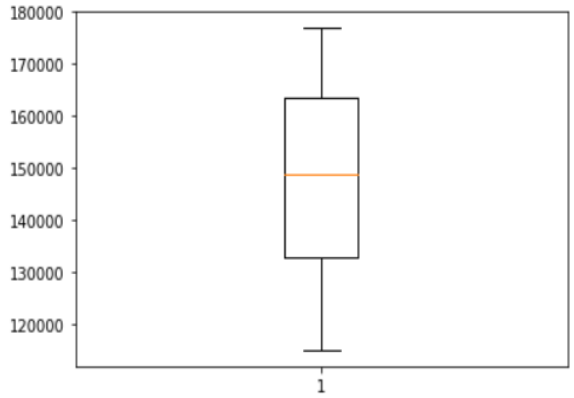


Figure 6.1: Splitting train/test

Boxplot manner if our dataset carries any out layer, it'll get extra errors, so we've to test out layers for our given data. There aren't any out layers. If it's far present, it will likely be represented as a dotted layer. If we've out layers, we should eliminate them to keep away from complex errors. Here we haven't any out layers. Everything proven in diagrammatical visualization.

6.3 OVERVIEW TEST AND TRAIN DATA

The method starts from variable identity like structured and unbiased variable wherein we discover the goal column. Then the pre-processing strategies are carried out like handling the lacking values the pre-processed records then used to construct a version via way of means of dividing the dataset is used for education reason this is version learns the sample and the closing trying out records is used to check the overall performance of records. The type version additionally may be used to are expecting the power intake of the electrical vehicle.

Here records may be made easy via way of means of becoming it to a class function. The class used can be linear (having one impartial variable) or more than one (having more than one impartial variable). The outliers can be undetected, or they may fall outdoor the clusters. This technique organizations comparable records in a cluster. In this process, we are attempting to test the bull records gift withinside the dataset and eliminate the ones general values from the dataset. In our dataset, we've a complete of 14 columns and 1285 records gift approximately the electrical vehicle. In our dataset, we haven't any null values gift, so that is a smooth dataset so that you can be beneficial for the task want now no longer alternate the features. Data

visualization is the graphical illustration of facts and records. Using records visualization gear presents an on-hand manner to look and apprehend trends, outliers, and styles in records.

6.4 EXECUTION PLAN

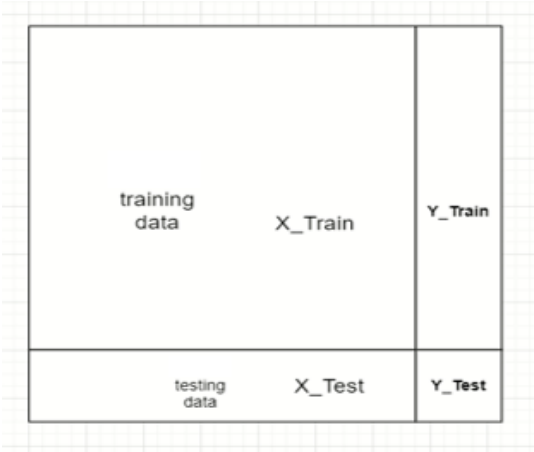


Figure 6.2: Training and testing data execution

In the given picturized of our education dataset, is our education statistics, and checking out statistics. Assume that it's miles our dataset, and we need to anticipate the X education set and y education set, and we've got X check and y check. A y check is our final results column as y is likewise our final results column, however it's miles an education set and is wanted check expected price is there. Now, we need to break up X and Y, X is each enter price, and Y approach output goal column. Here we're simply splitting the statistics only.

CHAPTER 7

RESULT ANALYSIS

7.1 ACCURACY RESULTS

After running python file CyberThreatDetection.py in command prompt we get console screen as below:

Figure 7.1: Console Screen

In above displayed console screen check on 'Upload Train Dataset' and add KDD-Cup dataset.

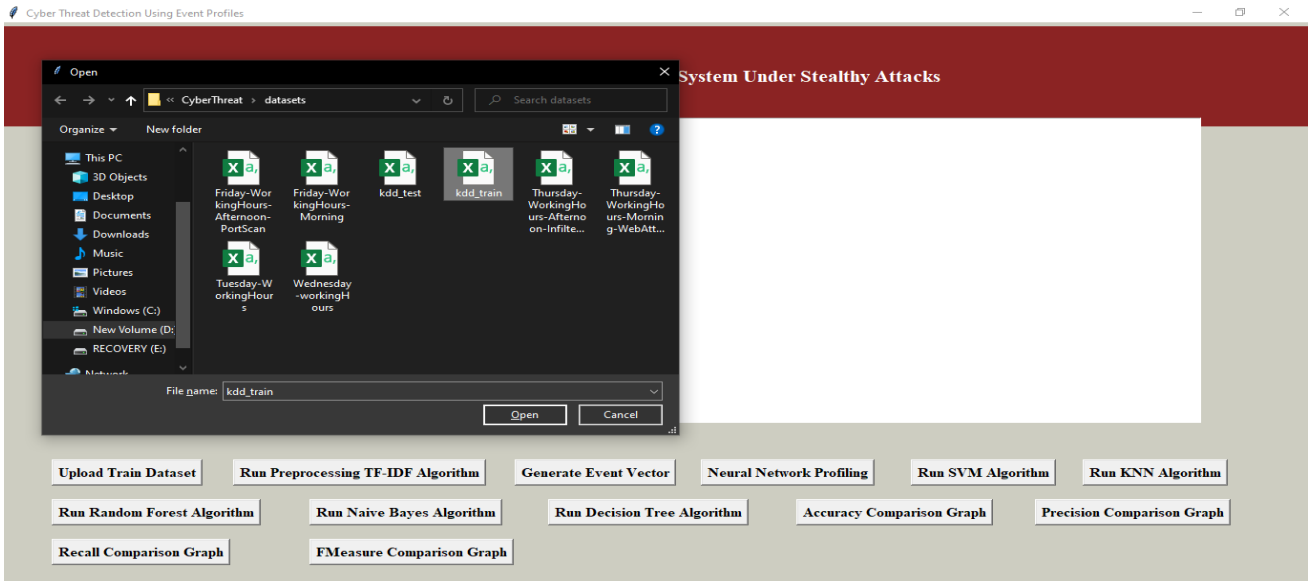


Figure 7.2: Uploading Data-set

In displayed console screen upload ‘kdd_train.csv’ (It’s a KDD cup 99 dataset of network intrusion dataset) dataset and after add gets under display screen

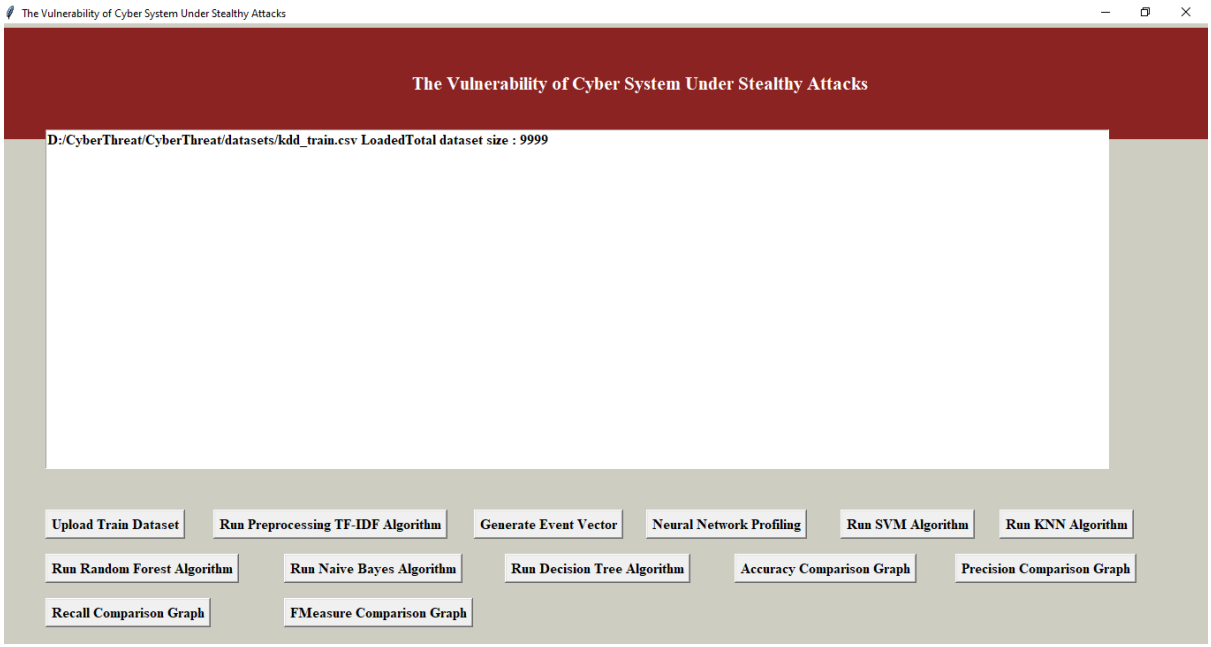


Figure 7.3: Loaded Data

In above displayed we are able to see dataset consists of 9999 statistics and check on ‘Run Preprocessing TF-IDF Algorithm’ transform uncooked dataset into TF-IDF values.

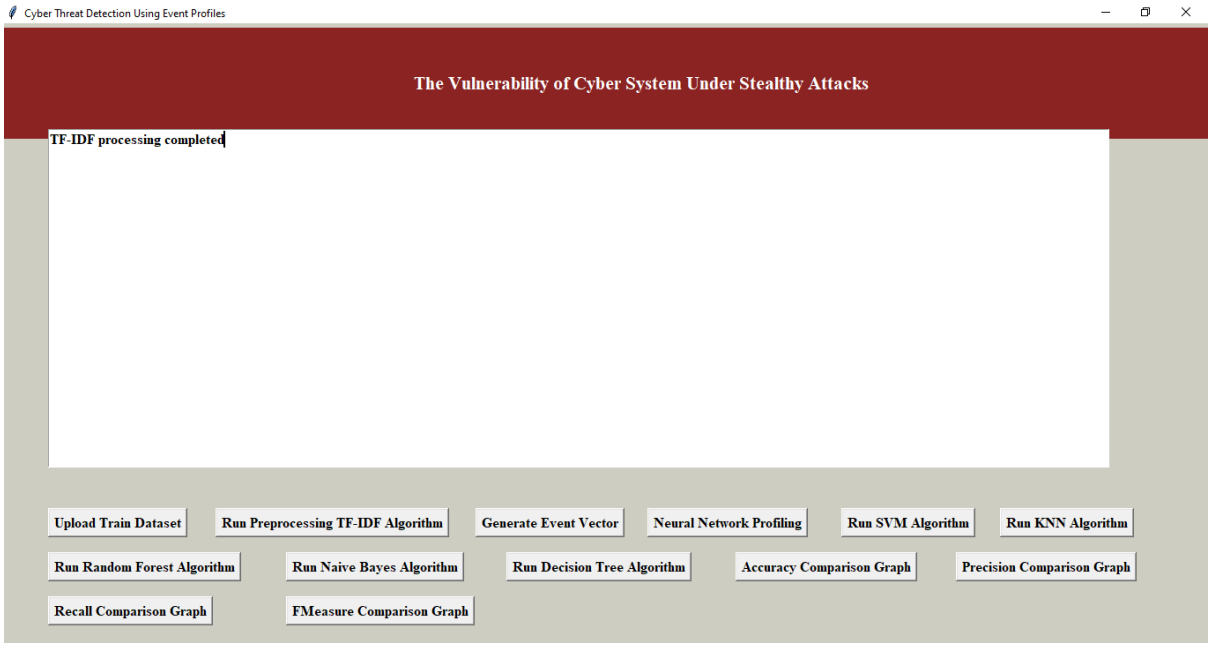


Figure 7.4: TF-IDF Process

In display Console screen TF-IDF processing finished and check on ‘Generate Event Vector’ to create a vector with specific events

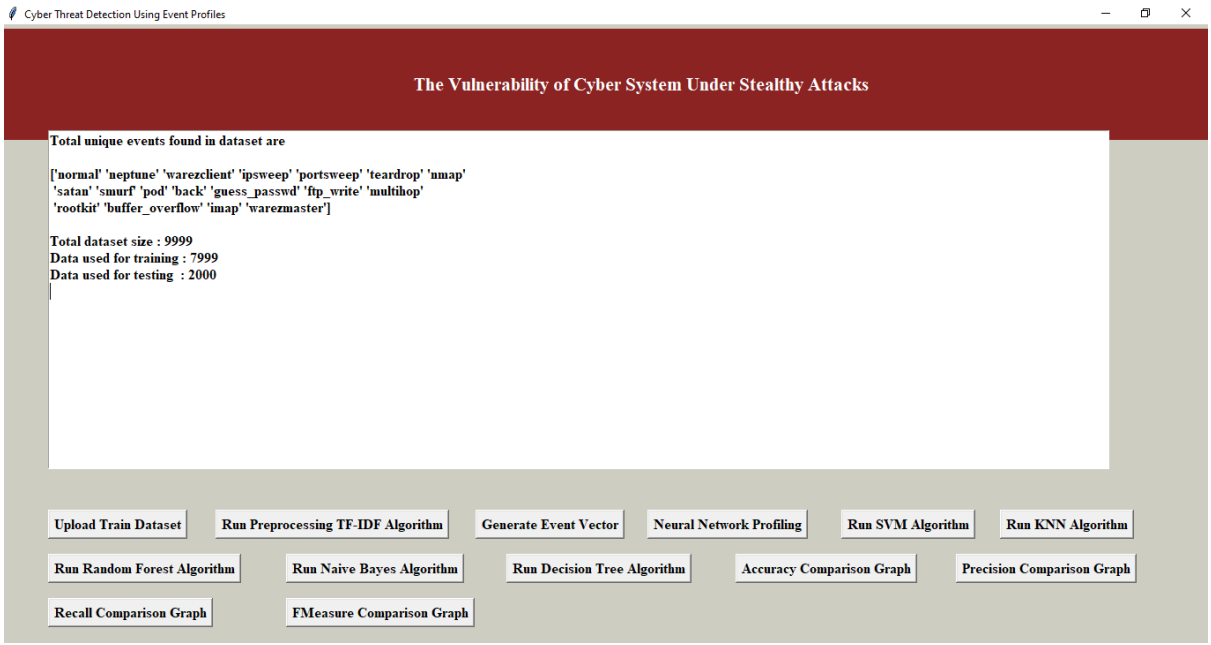


Figure 7.5: Generate Event Vector

In above display screen we are able to see definitely one of a kind specific occasions names and in underneath we are able to see dataset overall length and alertness the usage of 80 percentage dataset (with 7999 records) for schooling and the usage of 20 percentage dataset (with 2000 Columns) for testing. Now KDD-cup dataset teach and check occasions version geared up and checking on ‘Neural Network Profiling’ button it will be creating Long short-term memory and Convolutional Neural Network version

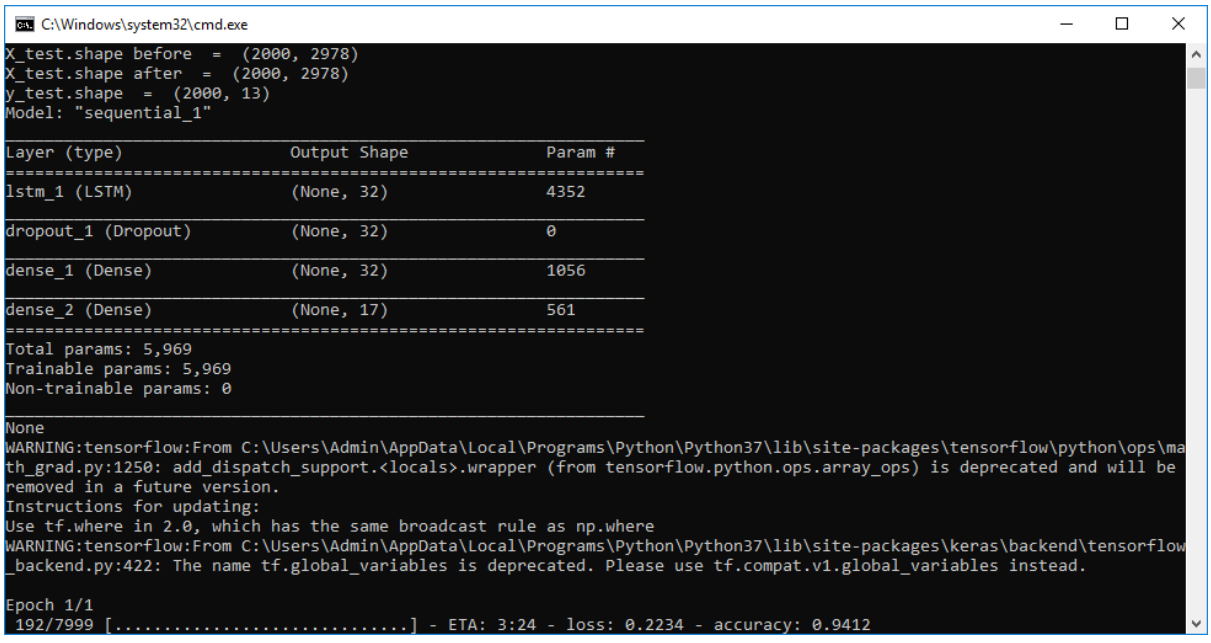


Figure 7.6: CNN Data Processing

In display Long short-term memory version is generated by ten epoch strolling additionally commenced and with accuracy is 0.94. Executing for whole dataset might also additionally take time for LSTM and Convolutional Neural Network schooling procedure to get executed. The KDD-cup dataset consists of 7999 data information and Long short-term memory will iterate all information to clear out and construct version.

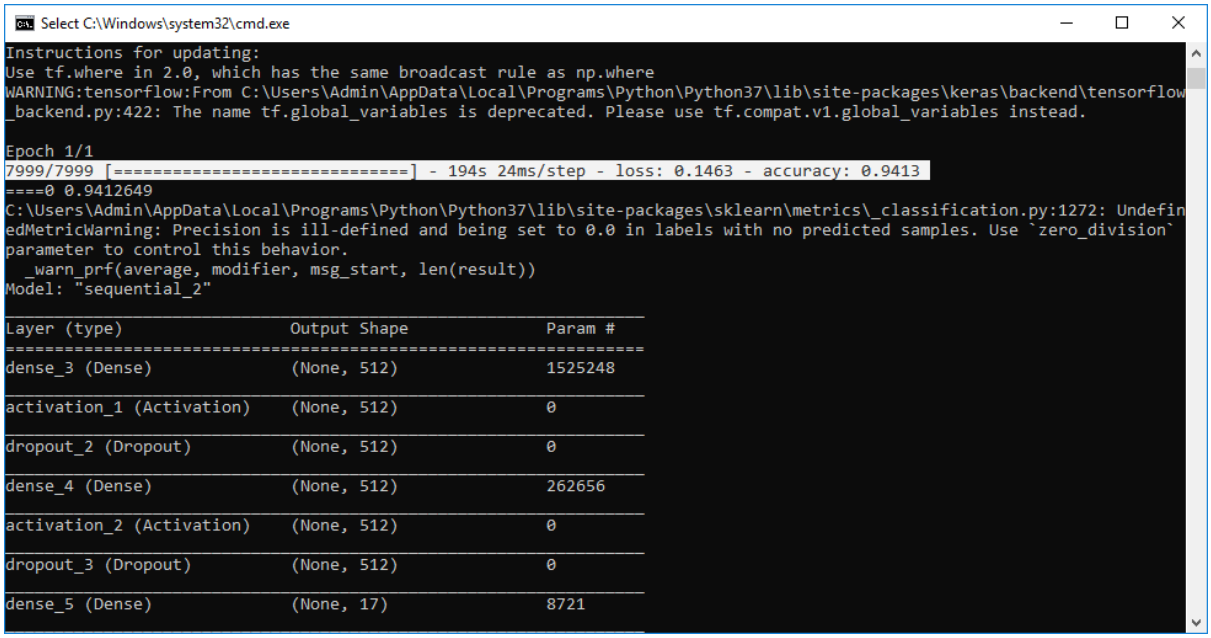


Figure 7.7: CNN Accuracy

In above decided on textual content we will see Long short-term memory whole all iterations and in beneath strains we will see Convolutional Neural Network version additionally begins offevolved execution

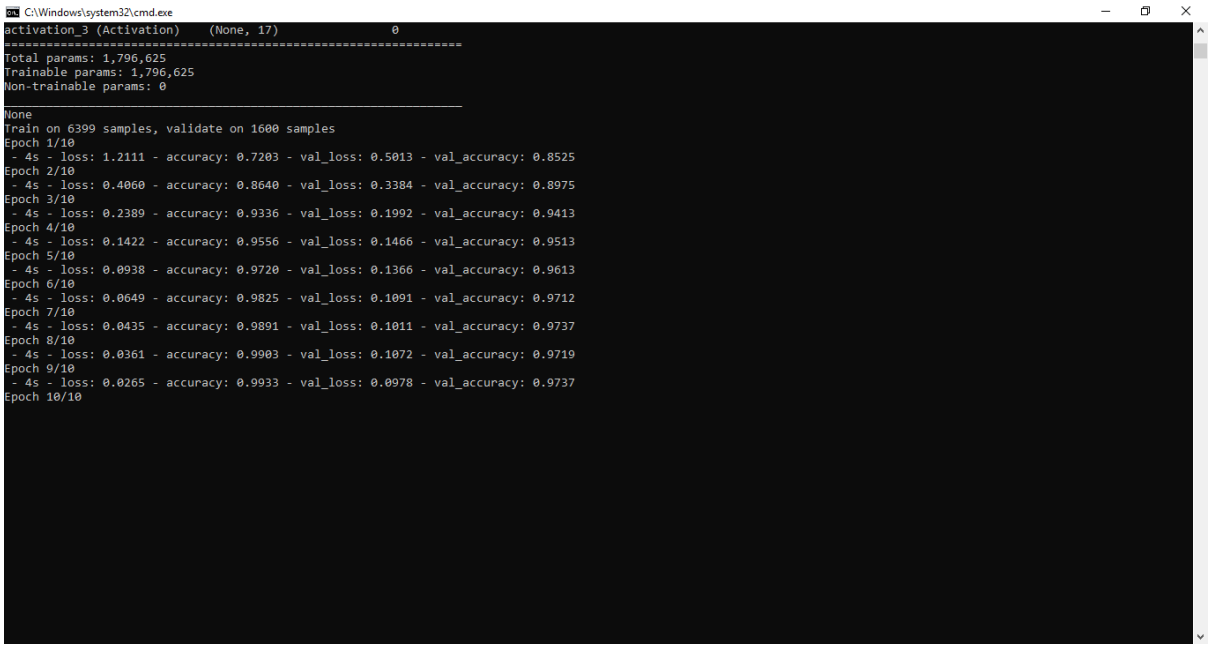


Figure 7.8: LSTN Data Processing

In above display screen Convolutional Neural Network additionally begins offevolved first new release with accuracy as seventy-two and after finishing all loops that we were given progressed accuracy as ninety-nine and multiply with the aid of using a hundred will provide us ninety-nine accuracies. So, CNN is giving higher accuracy examine to Long short-term memory and now see beneath Graphical user interface display screen with all details.

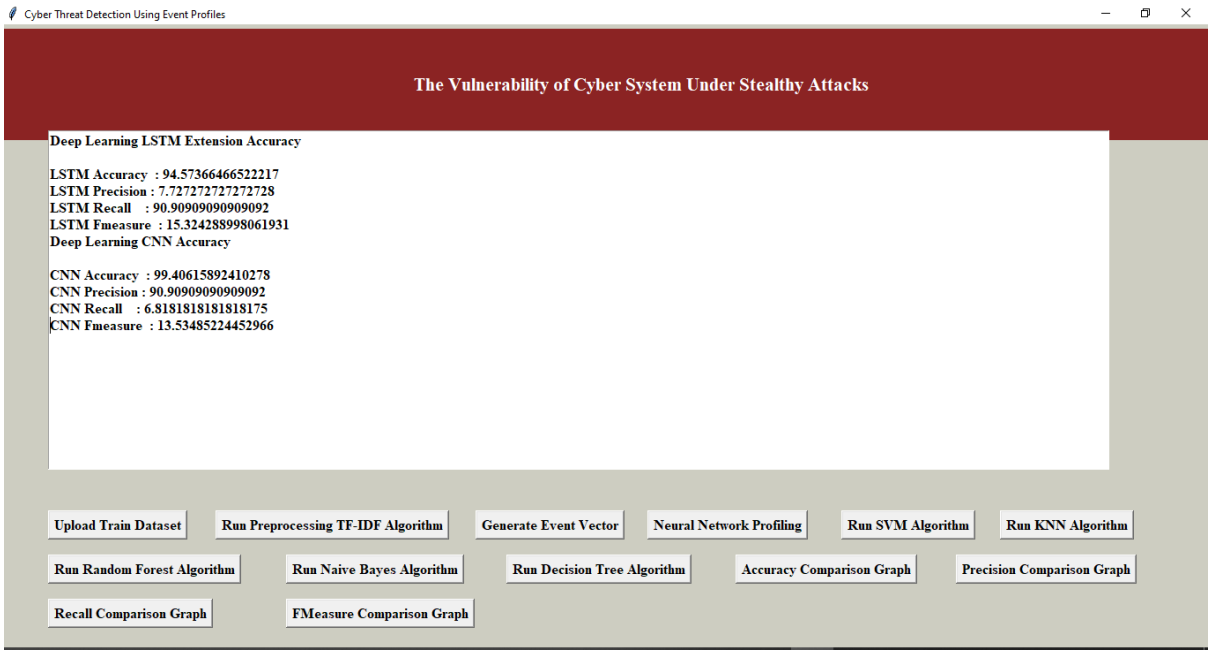


Figure 7.9: CNN and LSTN Accuracy

In display screen we see algorithms Accuracy, Precision, Recall and F-Measure values. Checking on 'Run SVM Algorithm' button to run current Support Vector Machine algorithm

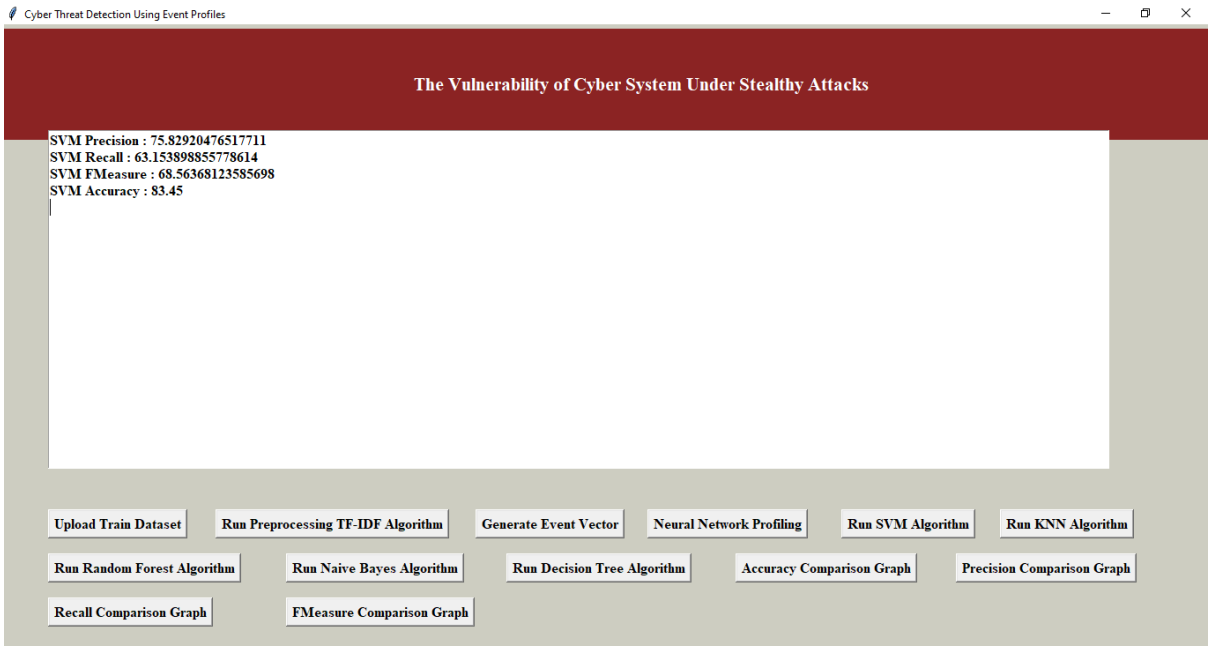


Figure 7.10: SVM Accuracy

In above display screen we are able to see Support Vector Machine set of rules output values and checking on ‘Run KNN Algorithm’ to run k-nearest neighbors set of rules.

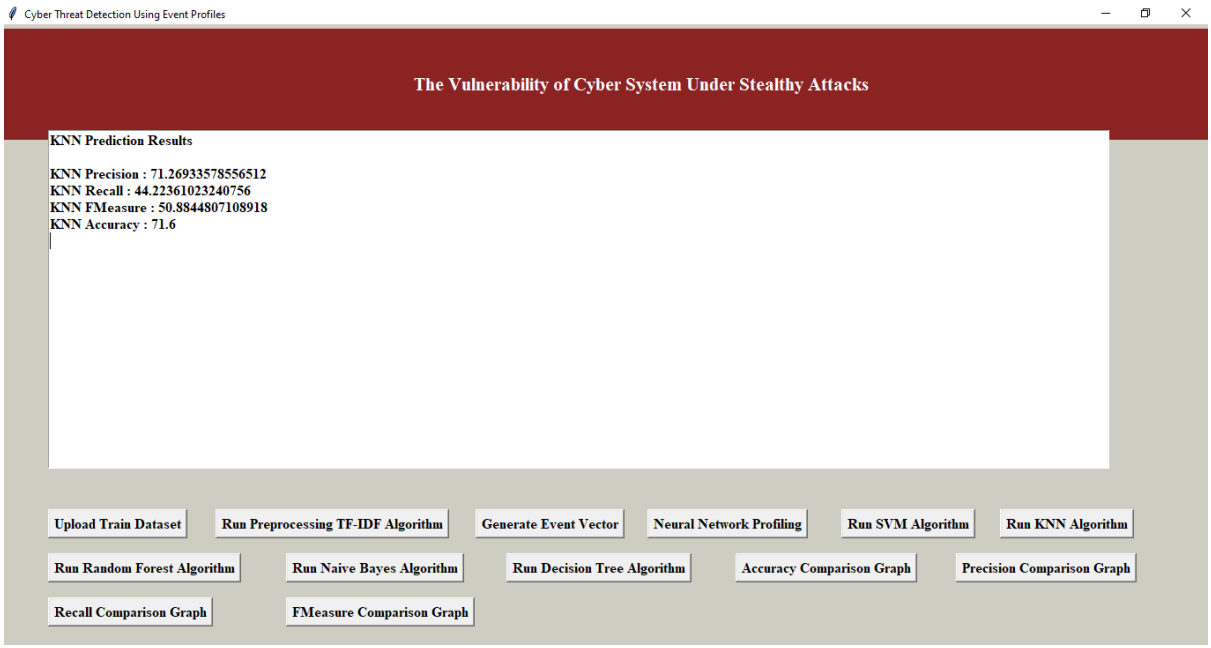


Figure 7.11: KNN Accuracy

In above display we will see k-nearest neighbors set of rules output values and checking on ‘Run Random Forest Algorithm’ to set of rules of algorithm.

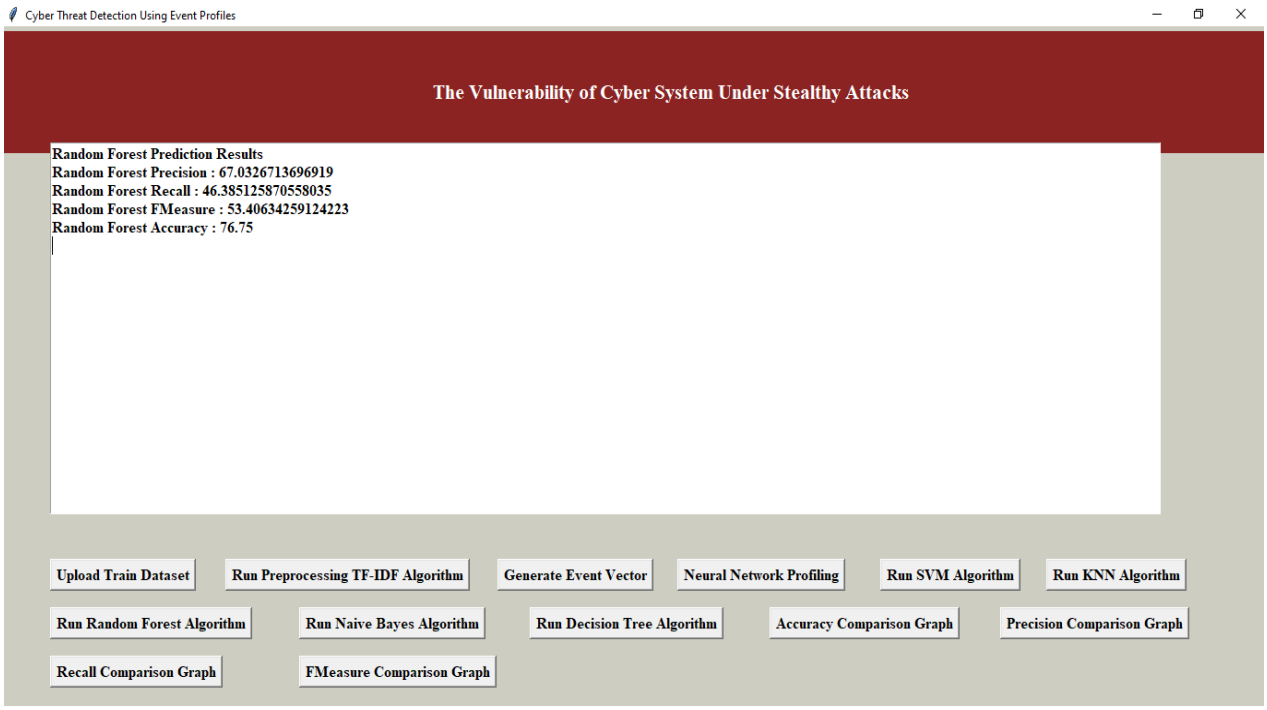


Figure 7.12: Random Forest Accuracy

In display screen Random Forest set of rules output values and checking on ‘Run Naïve Bayes Algorithm’ to run set of rules of algorithm

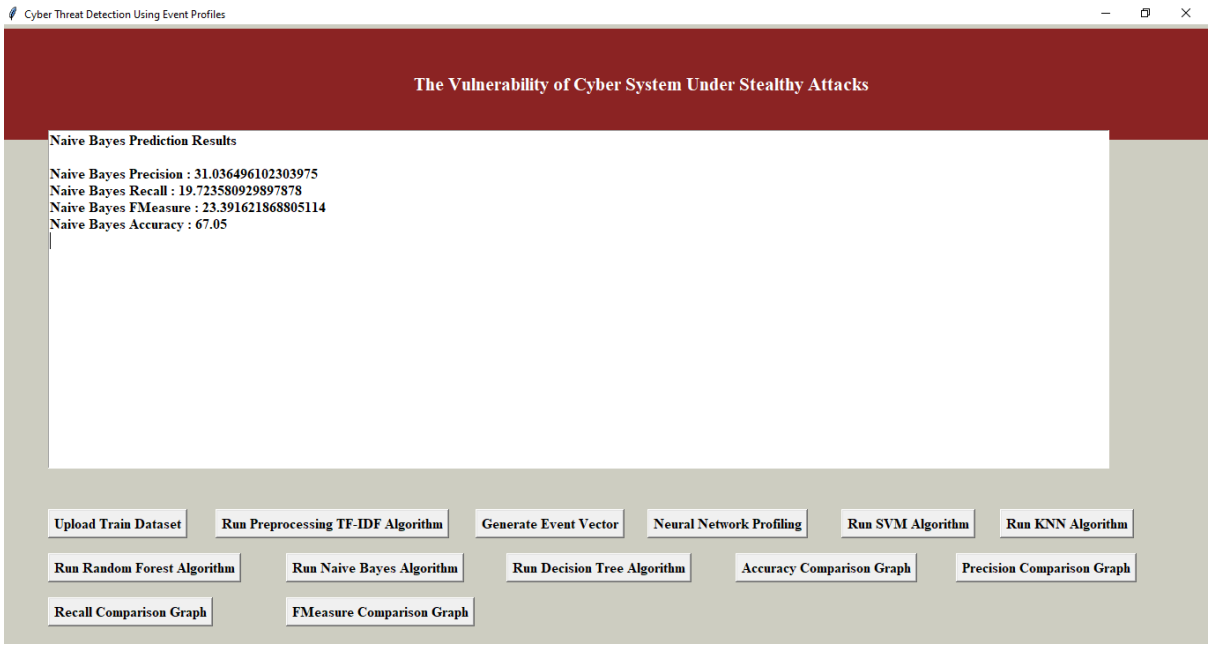


Figure 7.13: Naïve Bayes Accuracy

In display we will see Naïve Bayes set of rules output values and checking on ‘Run Decision Tree Algorithm’ to run Decision Tree Algorithm

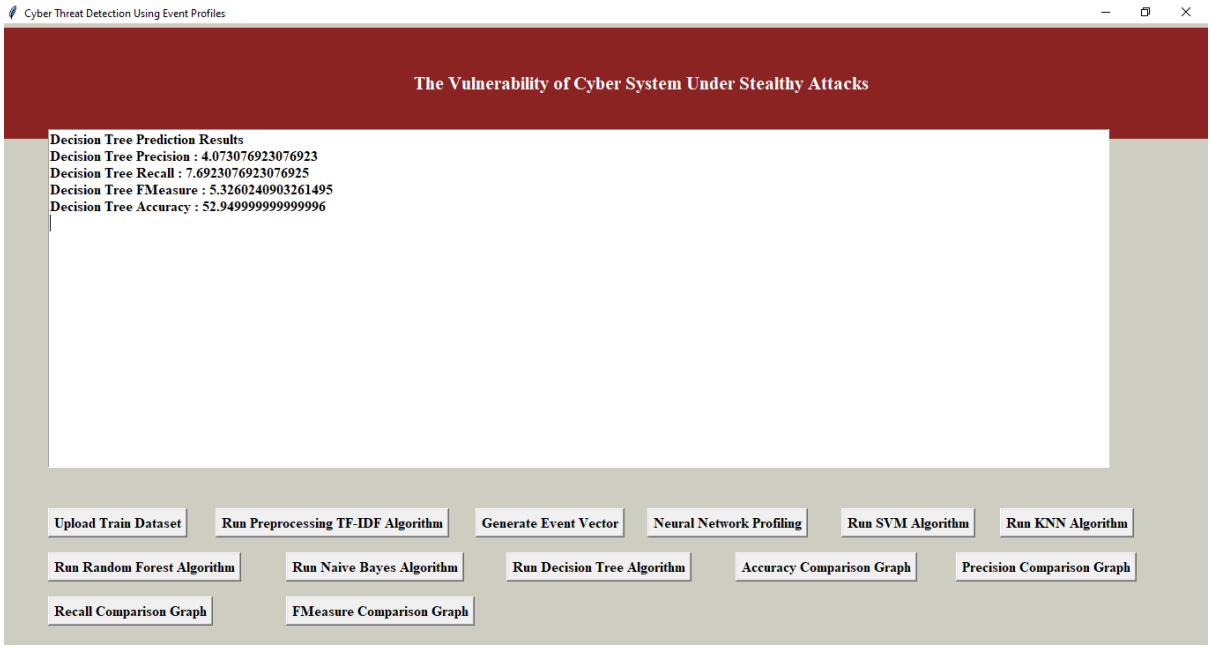


Figure 7.14: Decision Accuracy

7.2 ANALYSIS

By checking on ‘Accuracy Comparison Graph’ to get an accuracy graph of all algorithms i.e., Long short-term memory, Convolutional Neural Network, Support Vector Machine, k-nearest neighbors, Random Forest, Decision Tree, Naïve Bayes.

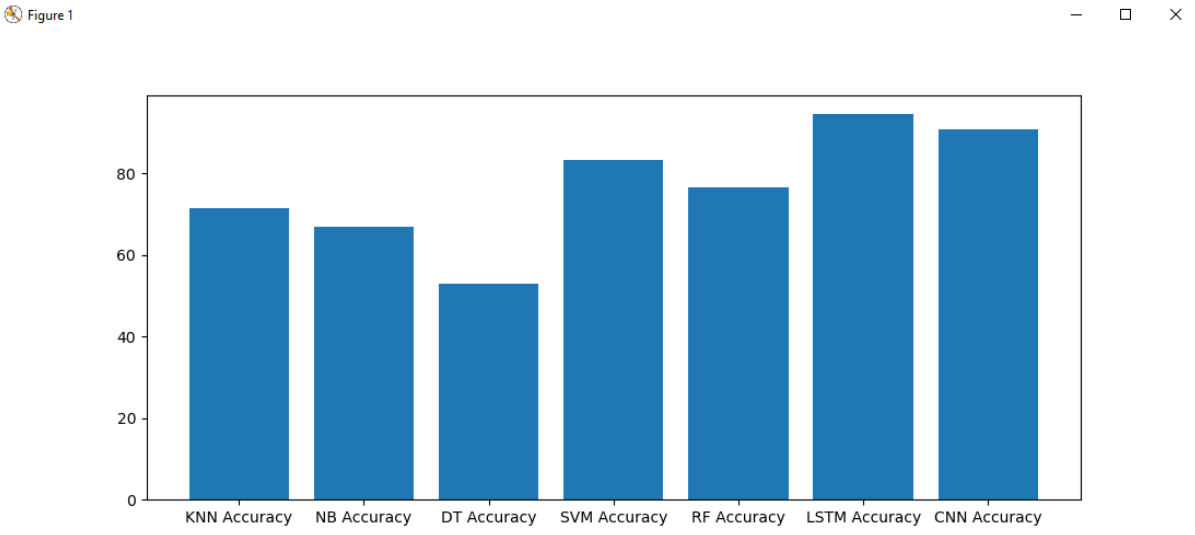


Figure 7.15: Accuracy

As shown in above graph x-axis represents set of rules call and y-axis represents accuracy of algorithms and as in graph able to finish that Convolutional Neural Network and Convolutional Neural Network carry out well. By checking on 'Precision Comparison Graph’ to get beneath graph.

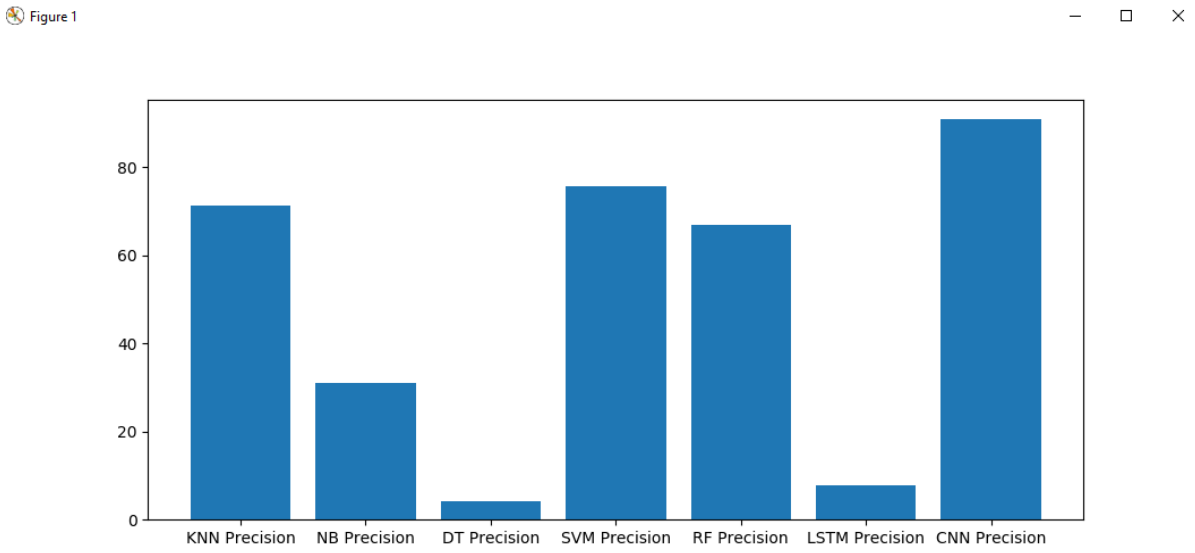


Figure 7.16: Precision

As shown in graph Convolutional Neural Network is acting nicely and checking on ‘Recall Comparison Graph’

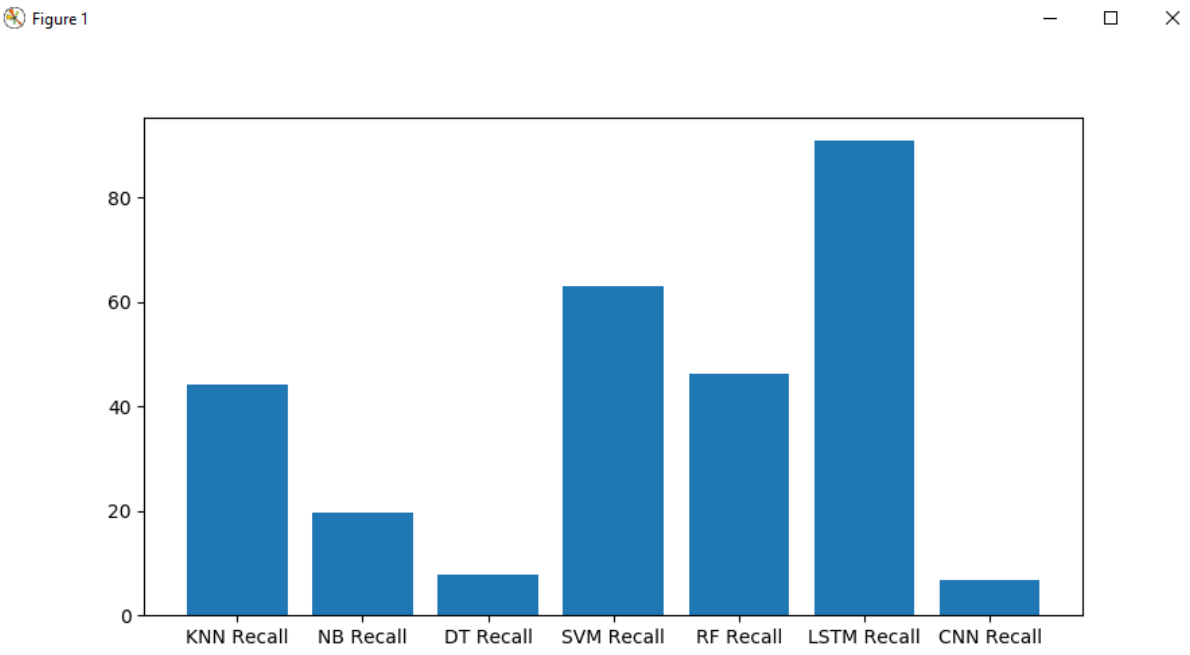


Figure 7.17: Recall

As shown in graph Convolutional Neural Network is appearing properly and checking on ‘F-Measure Comparison Graph’ to get underneath graph

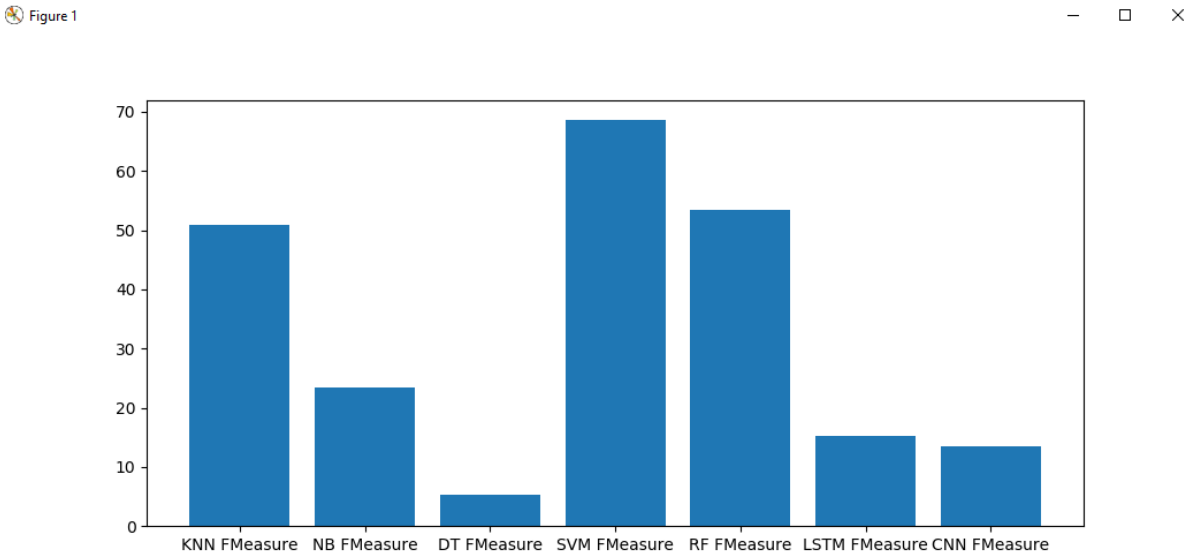


Figure 7.18: F-Measure

From all contrast graph we are able to see Long short-term memory and Convolutional

Neural Network appearing properly with accuracy graph, recall graph and precision graph.

CHAPTER 8

CONCLUSION & FUTURE SCOPE

We have proposed an AI-SIEM system the usage of occasion profiles and artificial neural networks. The novelty of our paintings lies in condensing very big-scale records into occasion profiles and the usage of the deep learning-primarily based totally detection techniques for more advantageous cyber-threat detection ability. The AI-SIEM system permits the safety analysts to cope with considerable safety indicators directly and correctly via way of means of evaluating long-time period safety records. By lowering fake fantastic indicators, it could additionally assist the safety analysts to unexpectedly reply to cyber threats dispersed throughout a big variety of safety events. For the assessment of overall performance, we finished a overall performance assessment the usage of benchmark datasets and datasets amassed withinside the actual world.

First, primarily based totally at the assessment test with different techniques, the usage of widely recognized benchmark datasets, we confirmed that our mechanisms may be implemented as one of the studying-primarily based totally fashions for community intrusion detection. Second, thru the evaluation using real datasets, we presented promising consequences that our era moreover outperformed conventional system reading strategies in terms of accurate classifications.

9. REFERENCES

[1] J. Farwell and R. Rohozinski. Stuxnet and the future of cyber war. Survival, 53(1):23–40, 2011.

[2] J. Slay and M. Miller. Lessons learned from the maroochy water breach. in Proceeding of Critical Infrastructure Protection, 253:73–82, 2007.

[3] J. Conti. The day the samba stopped. Engineering & Technology, 5(6):46–47, 2010.

[4] S. Kuvshinkova. Sql slammer worm lessons learned for consideration by the electricity sector. North American Electric Reliability Council, 1(2):5, 2003.

[5] G. Richards. Hackers vs slackers. Engineering & Technology, 3(19):40– 43, 2008.

[6] A. Cardenas, S. Amin, and S. Sastry. Research challenges for the security of control systems. In HotSec, 2008.

[7] P. Huber. Robust statistics. Springer Berlin Heidelberg, 2011.

[8] K. Zhou, J. Doyle, and K.Glover. Robust and optimal control. New Jersey: Prentice hall, 1996.

[9] A. Willsky. A survey of design methods for failure detection in dynamic systems. Automatica, 12(6):601–611, 1976.

[10] M. Massoumnia, G. Verghese, and A. Willsky. Failure detection and identification. IEEE Transactions on Automatic Control, 34(3):316–321, 1989.

[11] I. Hwang, S. Kim, Y. Kim, and C. Seah. A survey of fault detection, isolation, and reconfiguration methods. IEEE Transactions on Control System Technology, 18(3):636–653, 2010.

[12] Y. Liu, P. Ning, and M. Reiter. False data injection attacks against state estimation in electric power grids. ACM Transactions on Information and System Security, 14(1):1–16, 2011.

[13] F. Pasqualetti, A. Bicchi, and F. Bullo. Consensus computation in unreliable networks: A system theoretic approach. IEEE Transactions on Automatic Control, 57(1):90–104, 2010.

[14] F. Pasqualetti, F. Dorfler, and F. Bullo. Attack detection and identification in cyber-physical systems. IEEE Transactions on Automatic Control, 58(11):2715–2729, 2013.

[15] S. Sundaram, M. Pajic, C. Hadjicostis, R. Mangharam, and G. Pappas. The wireless control network: Monitoring for malicious behavior. in Proceedings of IEEE Conference on Decision and Control, pages 5979– 5984, 2010