A

Major Project

On

## CYBER THREAT DETECTION BASED ON ARTFICIAL NEURAL NETWORKS USING EVENT PROFILES

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

BANDARI SAIKUMAR (197R1A05C9) GUDIMETLA SRIKANTH (197R1A05E0) JOGANNAGARI SAINATH REDDY (197R1A05E2)

Under the Guidance of

### A.UDAY KIRAN

(Assistant Professor)

### DEPARTMENT OF COMPUTER SCIENCEAND ENGINEERING

**CMR TECHNICAL CAMPUS UGC AUTONOMOUS**

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi) Recognized Under Section 2(f) & 12(B) of the UGCAct.1956, Kandlakoya (V), Medchal Road, Hyderabad-501401.

**2019-2023**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



## CERTIFICATE

This is to certify that the project entitled **“CYBER THREAT DETECTION BASED ON ARTIFICIAL NEURAL NETWORKS USING EVENT PROFILES ”** being submitted by**B.SAIKUMAR(197R1A05C9), G.SRIKANTH(197R1A05E0)&J.SAINATHREDDY(19**

**7R1A05E2)** in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by them under our guidance and supervision during the year 2022-23.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

|  |  |
| --- | --- |
| **A.Uday Kiran** | **Dr. A. Raji Reddy** |
| (Assistant Professor) INTERNAL GUIDE | DIRECTOR |
| **Dr. K. Srujan Raju**  HOD | **EXTERNAL EXAMINER** |
| **Submitted for viva voice Examination held on** | |

### ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

We take this opportunity to express my profound gratitude and deep regard to my guide **A.Uday Kiran,** Assistant Professor for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to the Project Review Committee (PRC) **Dr. Punyaban Patel, Ms. Shilpa, Dr.M . Subha Mastan Rao & J. Narasimharao** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to **Dr. K. Srujan Raju,** Head, Department of Computer Science and Engineering for providing encouragement and support for completing this project successfully.

We are obliged to **Dr. A. Raji Reddy,** Director for being cooperative throughout the course of this project. We also express our sincere gratitude to Sri. **Ch. Gopal Reddy,** Chairman for providing excellent infrastructure and a nice atmosphere throughout the course of this project.

The guidance and support received from all the members of **CMR Technical Campus** who contributed to the completion of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement, without which this assignment would not be completed. We sincerely acknowledge and thank all those who gave support directly and indirectly in the completion of this project.

#### B. SAI KUMAR (197R1A05C9)

**G.SRIKANTH (197R1A05E0)**

**J.SAINATH REDDY (197R1A05E2)**

## ABSTRACT

One of the major challenges in cybersecurity is the provision of an automated and effective cyber-threats detection technique. In this project, we present an AI technique for cyber-threats detection, based on artificial neural networks. The proposed technique converts multitude of collected security events to individual event profiles and use a deep learning-based detection method for enhanced cyber-threat detection. For this work, we developed an AI-SIEM system based on a combination of event profiling for data preprocessing 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. All experiments in this study are performed by authors using two benchmark datasets (NSLKDD and CICIDS2017) and two datasets collected in the real world. To evaluate the performance comparison with existing methods, we conducted experiments using the five conventional machine-learning methods (SVM, k-NN, RF, NB, and DT). Consequently, the experimental results of this study ensure that our proposed methods are capable of being employed as learning-based models for network intrusion-detection and show that although it is employed in the real world, the performance outperforms the conventional machine-learning methods.

## LIST OF FIGURES/TABLES

|  |  |  |
| --- | --- | --- |
| FIGURE NO | FIGURE NAME | PAGE NO |
| Figure 3.1 | Project Architecture for cyber threat detection based on artificial neural networks using event profiles | 10 |
| Figure 3.2 | Use Case Diagram for cyber threat detection based on artificial neural networks using event profiles | 12 |
| Figure 3.3 | Class Diagram for cyber threat detection based on artificial neural networks using event profiles | 13 |
| Figure 3.4 | Sequence Diagram for cyber threat detection based on artificial neural networks using event profiles | 14 |
| Figure 3.5 | Activity Diagram for cyber threat detection based on artificial neural networks using event profiles | 15 |

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| ABSTRACT | i |
| LIST OF FIGURES | ii |
| 1. INTRODUCTION | 1 |
| 1.1 PROJECT SCOPE | 1 |
| 1.2 PROJECT PURPOSE | 1 |
| 1.3 PROJECT FEATURES | 1 |
| 2.LITERATURE SURVEY | 2 |
| 2.1 ENHANCED NETWORK ANOMLY DETECTION | 2 |
| 2.2 NETWORK INTRUSION DETECTION | 3 |
| 2.3 HAST-IDS | 3 |
| 2.4 DATASECURITY ANALYSIS FOR DD0S | 4 |
| 3. SYSTEM ANALYSIS | 5 |
| 3.1 PROBLEM DEFINITION | 5 |
| 3.2 EXISTING SYSTEM | 5 |
| 3.2.1 DISADVANTAGES OF THE EXISTING SYSTEM | 6 |
| 3.3 PROPOSED SYSTEM | 6 |
| 3.3.1 ADVANTAGES OF PROPOSED SYSTEM | 6 |
| 3.4 FEASIBILITY STUDY | 7 |
| 3.4.1 ECONOMIC FEASIBILITY | 7 |
| 3.4.2 TECHNICAL FEASIBILITY | 7 |
| 3.4.3 BEHAVIORAL FEASIBILITY | 8 |
| 3.5 HARDWARE & SOFTWARE REQUIREMENTS | 9 |
| 3.5.1 HARDWARE REQUIREMENTS | 9 |
| 3.5.2 SOFTWARE REQUIREMENTS | 9 |
| 4. ARCHITECTURE | 10 |
| 4.1 PROJECT ARCHITECTURE | 10 |
| 4.2 DESCRIPTION | 11 |
| 4.3 USE CASE DIAGRAM | 12 |
| 4.4 CLASS DIAGRAM | 13 |
| 4.5 SEQUENCE DIAGRAM | 14 |
| 4.6 ACTIVITY DIAGRAM | 15 |
| 5. CONCLUSION | 16 |
| 6.REFERENCES | 17 |
|  |  |

# 1.INTRODUCTION

## 1 INTRODUCTION

#### PROJECT SCOPE

With the emergence of artificial intelligence (AI) techniques, learning-based approaches for detecting cyber attacks, have become further improved, and they have achieved significant results in many studies. However, owing to constantly evolving cyber attacks, it is still highly challenging to protect IT systems against threats and malicious behaviors in networks. Because of various network intrusions and malicious activities, effective defenses and security considerations were given high priority for finding reliable solutions.

#### PROJECT PURPOSE

Fourth, some hackers can deliberately cover their malicious activities by slowly changing their behavior patterns. Even when appropriate learning-based models are possible, attackers constantly change their behaviors, making the detection models unsuitable. Moreover, almost all security systems have been focused on analyzing short-term network security events. To defend consistently evolving attacks, we assume that over long-term periods, analyzing the security event history associated with the generation of events can be one way of detecting the malicious behavior of cyber attacks.

#### PROJECT FEATURES

Our proposed system can help security analysts rapidly to respond cyber threats, dispersed across a large amount of security events. For this, the proposed the AI-SIEM system particularly includes an event pattern extraction method by aggregating together events with a concurrency feature and correlating between event sets in collected data. Our event profiles have the potential to provide concise input data for various deep neural networks. Moreover, it enables the analyst to handle all the data promptly and efficiently by comparison with longterm history data.

# 2.LITERATURE SURVEY

## 2 LITERATURE SURVEY

### Enhanced Network Anomaly Detection Based on Deep Neural Networks

**Abstract :**Due to the monumental growth of Internet applications in the last decade, the need for security of information network has increased manifolds. As a primary defense of network infrastructure, an intrusion detection system is expected to adapt to dynamically changing threat landscape. Many supervised and unsupervised techniques have been devised by researchers from the discipline of machine learning and data mining to achieve reliable detection of anomalies. Deep learning is an area of machine learning which applies neuron-like structure for learning tasks. Deep learning has profoundly changed the way we approach learning tasks by delivering monumental progress in different disciplines like speech processing, computer vision, and natural language processing to name a few. It is only relevant that this new technology must be investigated for information security applications. The aim of this paper is to investigate the suitability of deep learning approaches for anomaly-based intrusion detection system. For this research, we developed anomaly detection models based on different deep neural network structures, including convolutional neural networks, autoencoders, and recurrent neural networks. These deep models were trained on NSLKDD training data set and evaluated on both test data sets provided by NSLKDD, namely NSLKDDTest+ and NSLKDDTest21. All experiments in this paper are performed by authors on a GPU-based test bed. Conventional machine learning-based intrusion detection models were implemented using well-known classification techniques, including extreme learning machine, nearest neighbor, decision-tree, random-forest, support vector machine, naive-bays, and quadratic discriminant analysis. Both deep and conventional machine learning models were evaluated using well-known classification metrics, including receiver operating characteristics, area under curve, precision-recall curve, mean average precision and accuracy of classification. Experimental results of deep IDS models showed promising results for real-world application in anomaly detection systems.

### Network Intrusion Detection Based on Directed Acyclic Graph and Belief Rule Base

**Abstract:** Intrusion detection is very important for network situation awareness. While a few methods have been proposed to detect network intrusion, they cannot directly and effectively utilize semi‐quantitative information consisting of expert knowledge and quantitative data. Hence, this paper proposes a new detection model based on a directed acyclic graph (DAG) and a belief rule base (BRB). In the proposed model, called DAG‐BRB, the DAG is employed to construct a multi‐layered BRB model that can avoid explosion of combinations of rule number because of a large number of types of intrusion. To obtain the optimal parameters of the DAG‐ BRB model, an improved constraint covariance matrix adaption evolution strategy (CMA‐ES) is developed that can effectively solve the constraint problem in the BRB. A case study was used to test the efficiency of the proposed DAG‐BRB. The results showed that compared with other detection models, the DAG‐BRB model has a higher detection rate and can be used in real networks.

### HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection

**Abstract:**The development of an anomaly-based intrusion detection system (IDS) is a primary research direction in the field of intrusion detection. An IDS learns normal and anomalous behavior by analyzing network traffic and can detect unknown and new attacks. However, the performance of an IDS is highly dependent on feature design, and designing a feature set that can accurately characterize network traffic is still an ongoing research issue. Anomaly-based IDSs also have the problem of a high false alarm rate (FAR), which seriously restricts their practical applications. In this paper, we propose a novel IDS called the hierarchical spatial-temporal features-based intrusion detection system (HAST-IDS), which first learns the low-level spatial features of network traffic using deep convolutional neural networks (CNNs) and then learns high-level temporal features using long short-term memory networks. The entire process of feature learning is completed by the deep neural networks automatically; no feature engineering techniques are required. The experimental results show that the HAST-IDS outperforms other published approaches in terms of accuracy, detection rate, and FAR, which successfully demonstrates its effectiveness in both feature learning and FAR reduction.

### Data security analysis for DDoS defense of cloud based networks

**Abstract:**Distributed computing has become an effective approach to enhance capabilities of an institution or organization and minimize requirements for additional resource. In this regard, the distributed computing helps in broadening institutes IT capabilities. One needs to note that distributed computing is now integral part of most expanding IT business sector. It is considered novel and efficient means for expanding business. As more organizations and individuals start to use the cloud to store their data and applications, significant concerns have developed to protect sensitive data from external and internal attacks over internet. Due to security concern many clients hesitate in relocating their sensitive data on the clouds, despite significant interest in cloud-based computing. Security is a significant issue, since data much of an organizations data provides a tempting target for hackers and those concerns will continue to diminish the development of distributed computing if not addressed. Therefore, this study presents a new test and insight into a honeypot. It is a device that can be classified into two types: handling and research honeypots. Handling honeypots are used to mitigate real life dangers. A research honeypot is utilized as an exploration instrument to study and distinguish the dangers on the internet. Therefore, the primary aim of this research project is to do an intensive network security analysis through a virtualized honeypot for cloud servers to tempt an attacker and provide a new means of monitoring their behavior

# SYSTEM ANALYSIS

**SYSTEM ANALYSIS**

1. **SYSTEM ANALYSIS**

System Analysis is the important phase in the system development process. The System is studied to the minute details and analyzed. The system analyst plays an importantrole of an interrogator and dwells deep into the working of the present system. In analysis,a detailed study of these operations performed by the system and their relationships withinand outside the system is done. A key question considered here is, “what must be done to solve the problem?” The system is viewed as a whole and the inputs to the system are identified. Once analysis is completed the analyst has a firm understanding of what is to bedone.

### PROBLEM DEFINITION

It is still difficult to recognize and detect intrusions against intelligent network attacks owing to their high false alerts and the huge amount of security data. Hence, the most recent studies in the field of intrusion detection have given increased focus to machine learning and artificial intelligence techniques for detecting attacks. Advancement in AI fields can facilitate the investigation of network intrusions by security analysts in a timely and automated manner. These learning-based approaches require to learn the attack model from historical threat data and use the trained models to detect intrusions for unknown cyber threats.

### EXISTING SYSTEM

Cyber security has recently received enormous attention in today’s security concerns, due to the popularity of the Internet-of-Things (IoT), the tremendous growth of computer networks, and the huge number of relevant applications. Thus, detecting various cyber-attacks or anomalies in a network and building an effective intrusion detection system that performs an essential role in today’s security is becoming more important. However, many previous studies used benchmark dataset, which, though accurate, are not generalizable to the real world because of the insufficient features. To overcome these limitations, an employed learning model requires to evaluate with datasets that are collected in the real world. Third, using an anomaly-based

method to detect network intrusion can help detect unknown cyber threats; whereas it can also cause a high false alert rate.

#### DISADVANTAGES OF EXISTING SYSTEM

* + - * Data leakage is take place using Cyber attacks.
      * It is less Secured from cyber attacks

### PROPOSED SYSTEM

we present an AI technique for cyber-threats detection, based on artificial neural networks. The proposed technique converts multitude of collected security events to individual event profiles and use a deep learning-based detection method for enhanced cyber-threat detection. For this work, we developed an AI-SIEM system based on a combination of event profiling for data preprocessing 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 threat. we conducted experiments using the five conventional machine-learning methods (SVM, k-NN, RF, NB, and DT). Consequently, the experimental results of this study ensure that our proposed methods are capable of being employed as learning-based models for network intrusion-detection.

#### ADVANTAGES OF PROPOSED SYSTEM

* + - * it is employed in the real world, the performance outperforms the conventional machine-learning methods.
      * In this project data is secured from attacker.

### FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and a business proposal is putforth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. Three key considerations involvedin the feasibility analysis:

* + - EconomicFeasibility
    - TechnicalFeasibility
    - SocialFeasibility

**3.4.1 ECONOMIC FEASIBILITY**

The developing system must be justified by cost and benefit. Criteria to ensure thateffort is concentrated on a project, which will give best, return at the earliest. One of the factors, which affect the development of a new system, is the cost it would require.

The following are some of the important financial questions asked during preliminary investigation:

* + - 1. The costs conduct a full system investigation.
      2. The cost of the hardware and software.
      3. The benefits in the form of reduced costs or fewer costly errors.

Since the system is developed as part of project work, there is no manual cost to spend for the proposed system. Also all the resources are already available, it give an indication that the system is economically possible for development.

**3.4.2 TECHNICAL FEASIBILITY**

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

#### BEHAVIORAL FEASIBILITY

This includes the following questions:

* + - 1. Is there sufficient support for the users?
      2. Will the proposed system cause harm?

The project would be beneficial because it satisfies the objectives when developedand installed. All behavioral aspects are considered carefully and conclude that the projectis behaviorally feasible

### HARDWARE & SOFTWARE REQUIREMENTS

#### HARDWARE REQUIREMENTS:

Hardware interfaces specify the logical characteristics of each interface between the software product and the hardware components of the system. The following are some hardware requirements.

* + - 1. Processor : Pentium IV or higher processor
      2. Hard disk : minimum 512MB space in Hard Disk.
      3. RAM : 256 MB RAM
      4. Input devices : Keyboard, mouse.

#### SOFTWARE REQUIREMENTS:

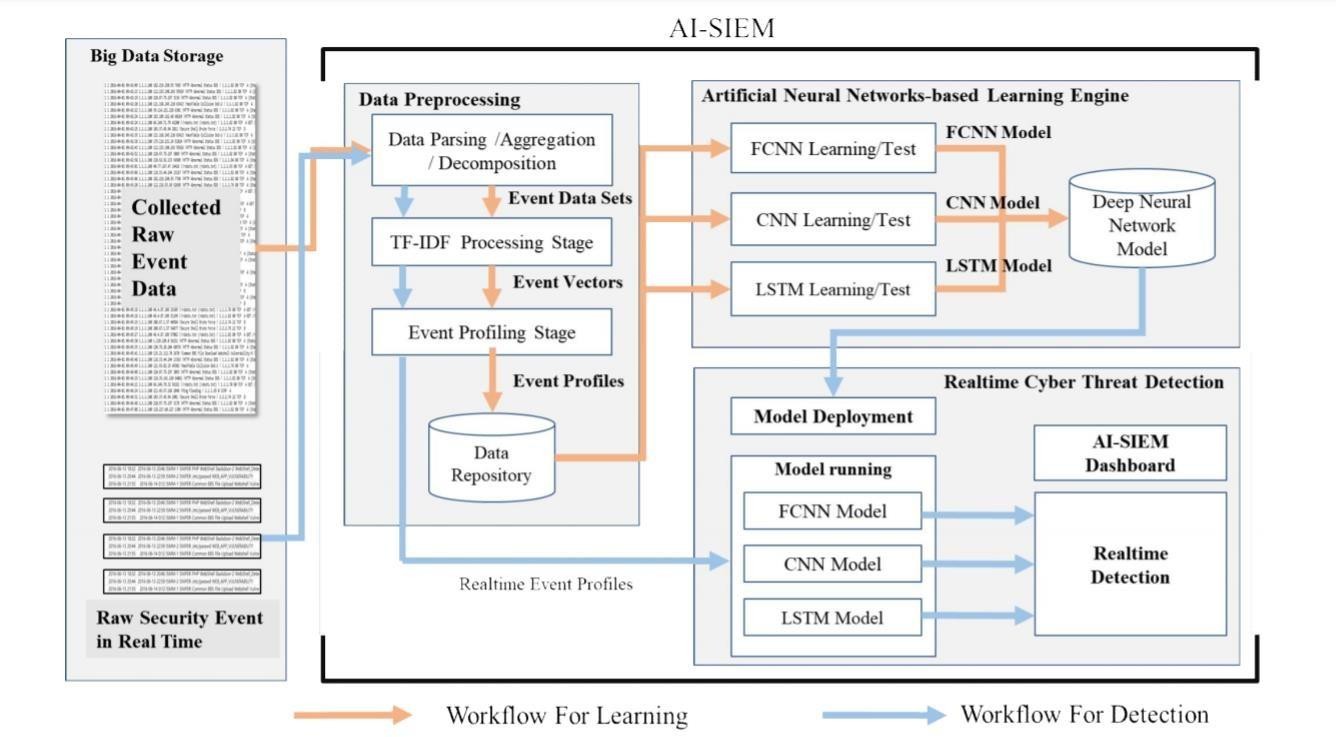
Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirements,

* + - * Languages : Python,tkinter
      * Operating system:Windows 8 and above
      * Tools : Python IDEL33.7 version,vscode

# ARCHITECTURE

## PROJECT ARCHITECTURE

#### PROJECT ARCHITECTURE

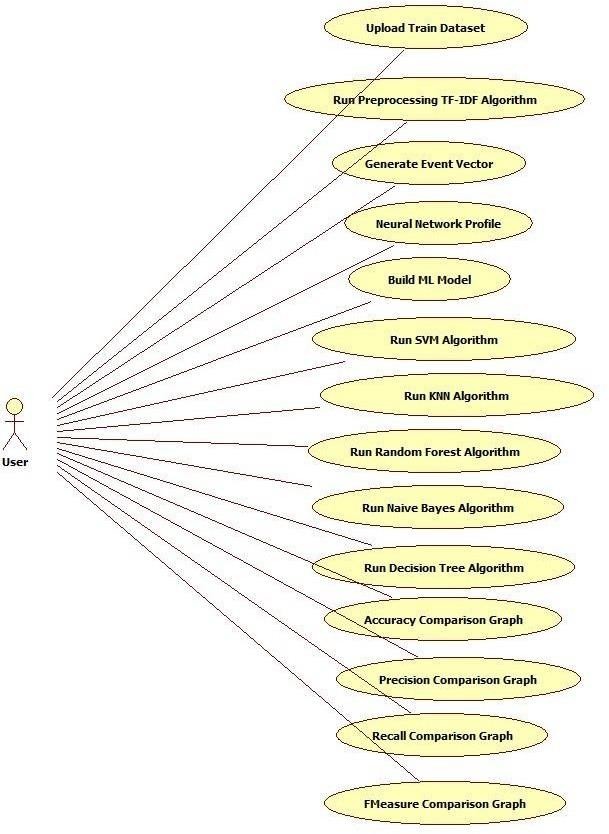


* 1. **DESCRIPTION**

The workflow and architecture for the developed artificial intelligent (AI)-based SIEM system. The AI-SIEM system comprises three main phases: The data preprocessing, artificial neural networks-based learning engine, and real-time threat detection phase. The first preprocessing phase in the system, termed event profiling, aims at providing concise inputs for various deep neural networks by transforming raw data. In the data preprocessing phase, data aggregation with parsing, data normalization stage using TF-IDF mechanism, and event profiling stage are consecutively performed in the AI-SIEM system. Each stage generates event data sets, event vectors, and event profiles, respectively, and the output is utilized in next each stage, as shown in Figure. This phase not only precedes the data learning stage but also precedes the conversion of raw security events to the deep-learning engine’s input data when the system operates on detecting network intrusions in real time. The second AI-based learning engine employs three artificial neural networks for modeling. For the data learning stage, the preprocessed data are fed into the three artificial neural networks, and each ANN performs learning to find the most accurate model. Finally, in real-time threat detection, each ANN model mechanically classifies each security raw event using the trained model, and the dashboard shows the only recognized true alerts to security analysts for reducing false ones.

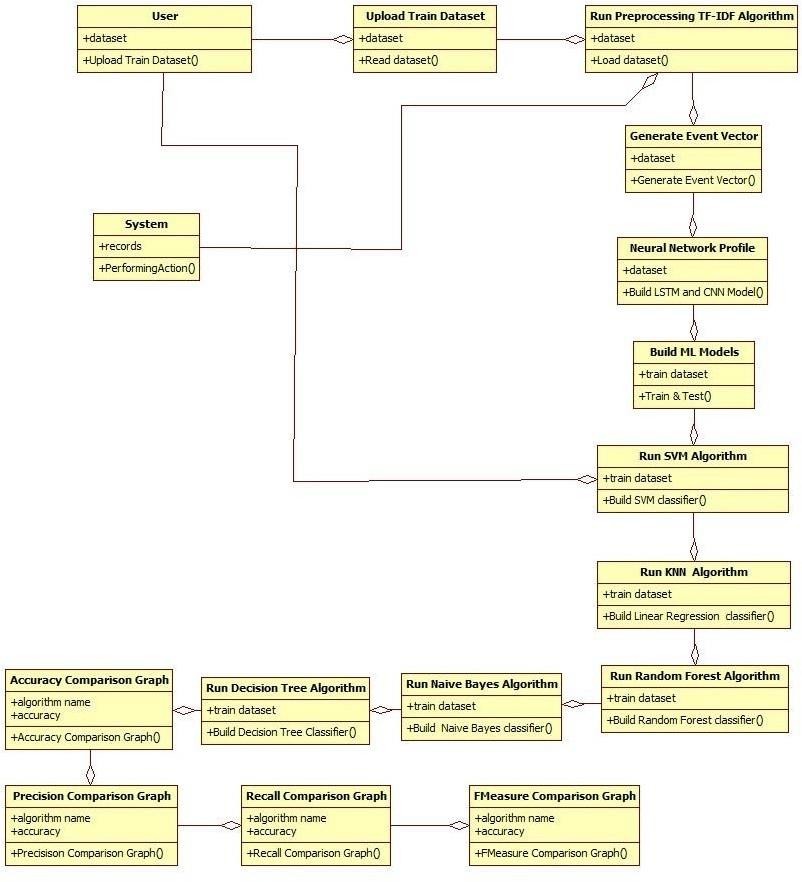
#### USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



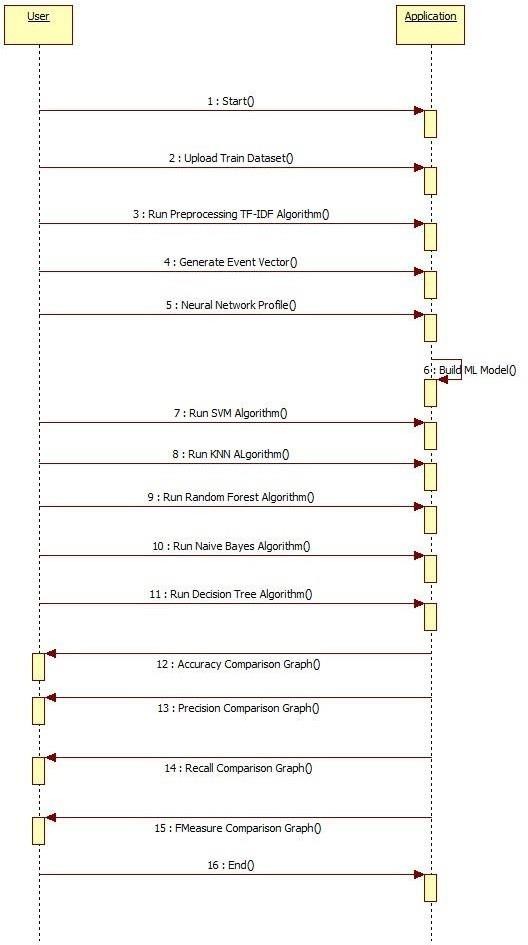
#### CLASS DIAGRAM

The class diagram is used to refine the use case diagram and define a detailed design of the system. The class diagram classifies the actors defined in the use case diagram into a set of interrelated classes. The relationship or association between the classes can be either an "is-a" or "has-a" relationship. Each class in the class diagram may be capable of providing certain functionalities. These functionalities provided by the class are termed "methods" of the class. Apart from this, each class may have certain "attributes" that uniquely identify the class.

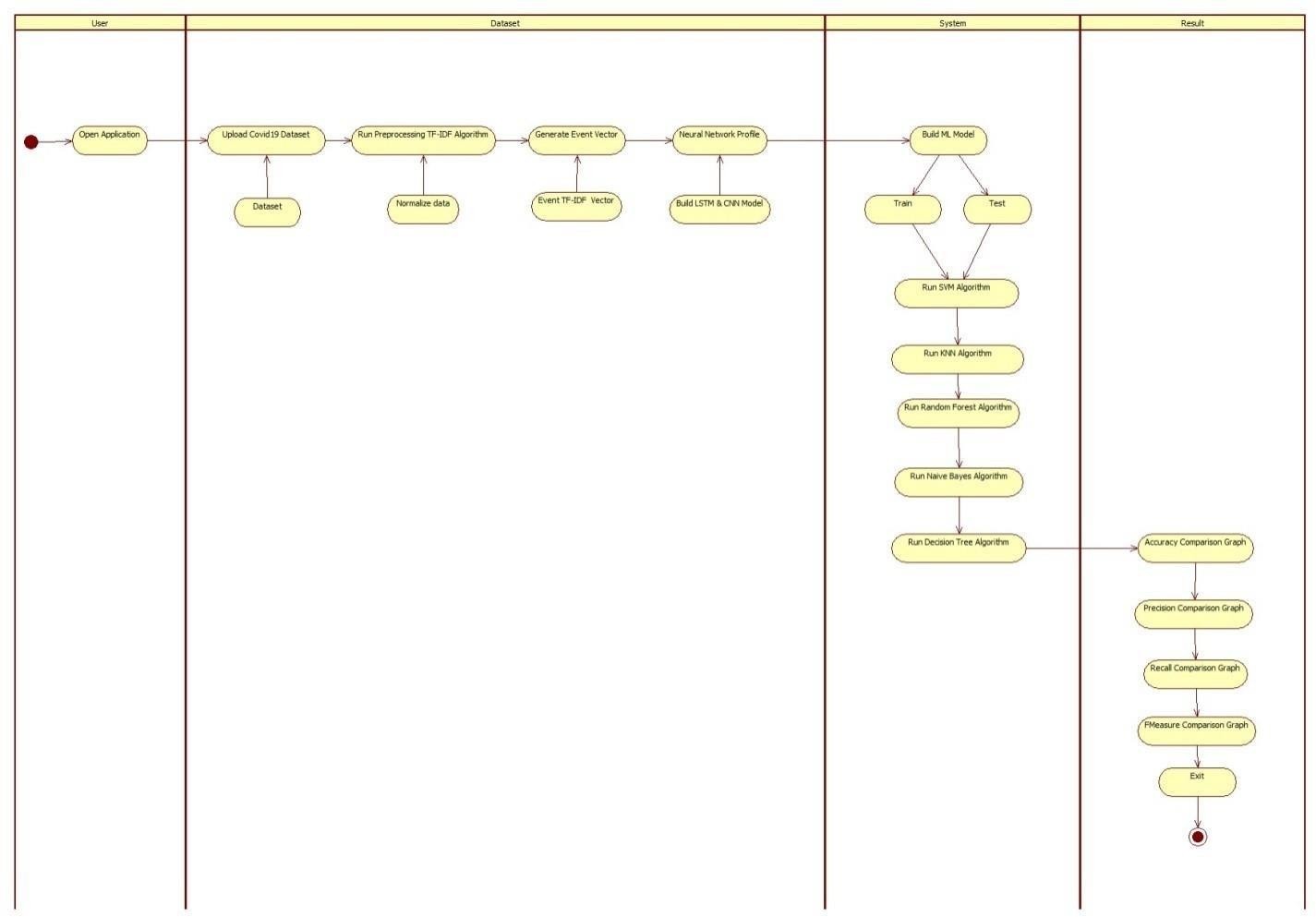


#### SEQUENCE DIAGRAM

A sequence diagram represents the interaction between different objects in the system. The important aspect of a sequence diagram is that it is time-ordered. This means that the exact sequence of the interactions between the objects is represented step by step. Different objects in the sequence diagram interact with each other by passing "messages".



#### 4.6 ACTIVITY DIAGRAM

The process flows in the system are captured in the activity diagram. Similar to a state diagram, an activity diagram also consists of activities, actions, transitions, initial and final states, and guard conditions.

**5.CONCLUSION**

In this project, we have proposed the AI-SIEM system using event profiles and artificial neural networks. The novelty of our work lies in condensing very large-scale data into event profiles and using the deep learning-based detection methods for enhanced cyber-threat detection ability. The AI-SIEM system enables the security analysts to deal with significant security alerts promptly and efficiently by comparing longterm security data. By reducing false positive alerts, it can also help the security analysts to rapidly respond to cyber threats dispersed across a large number of security events.

For the evaluation of performance, we performed a performance comparison using two benchmark datasets (NSLKDD, CICIDS2017) and two datasets collected in the real world. First, based on the comparison experiment with other methods, using widely known benchmark datasets, we showed that our mechanisms can be applied as one of thelearning-based models for network intrusion detection. Second, through the evaluation using two real datasets, we presented promising results that our technology also outperformed conventional machine learning methods in terms of accurate classifications.

**6.REFERENCES**

[1] S. Naseer, Y. Saleem, S. Khalid, M. K. Bashir, J. Han, M. M. Iqbal,K. Han, "Enhanced Network Anomaly Detection Based on DeepNeural Networks," *IEEE Access*, vol. 6, pp. 48231-48246, 2018.

[2] B. Zhang, G. Hu, Z. Zhou, Y. Zhang, P. Qiao, L. Chang, "NetworkIntrusion Detection Based on Directed Acyclic Graph and Belief RuleBase", *ETRI Journal*, vol. 39, no. 4, pp. 592-604, Aug. 2017

[3] W. Wang, Y. Sheng and J. Wang, "HAST-IDS: Learning hierarchicalspatial-temporal features using deep neural networks to improveintrusion detection," *IEEE Access*, vol. 6, no. 99, pp. 1792-1806,2018.

[4] M. K. Hussein, N. Bin Zainal and A. N. Jaber, "Data security analysisfor DDoS defense of cloud based networks," *2015 IEEE StudentConference on Research and Development (SCOReD)*, KualaLumpur, 2015, pp. 305-310.

[5] S. Sandeep Sekharan, K. Kandasamy, "Profiling SIEM tools andcorrelation engines for security analytics," *In Proc. Int. Conf.Wireless Com., Signal Proce. and Net.(WiSPNET)*, 2017, pp. 717-721.

[6] N.Hubballiand V.Suryanarayanan,‘‘False alarm minimizationtechniques in signature-based intrusion detection systems: Asurvey,’’ *Comput. Commun.*, vol. 49, pp. 1-17, Aug. 2014.

[7] A. Naser, M. A. Majid, M. F. Zolkipli and S. Anwar, "Trusting cloudcomputing for personal files," *2014 International Conference onInformation and Communication Technology Convergence (ICTC)*,Busan, 2014, pp. 488-489.