IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DESIGN OF INTRUSION DETECTION SYSTEM FOR IOT ENVIRONMENT WITH MACHINE LEARNING APPROACH

[1]Mr.B.Srinivas Raja, [2]Dasari Naga Venkata Lakshmi Kanaka Akshaya, [3]Challa Ayyappa Reddy, [4]Vepa Rohit Sai Naga Kumar, [5]Neelam Nikhil

[1] Professor, [2 3 4 5] Student

[1]Electronics and Communication Engineering,

[1]Godavari Institute of Engineering and Technology, Rajahmundry, India

Abstract: This project presents novel ideas on boosting security and thus reducing the risk in Internet of Things (IoT) based on implementing Intrusion Detection Systems leveraging machinelearning. The project details the designing of two IoT devices based on DHT11 sensors that are used to measure the temperatureand humidity levels. These devices use datagram communication,for which data is transmitted through a Mosquitto broker using the MQTT protocol. In the server side, a machine learning modelmade out of Random Forest and Support Vector Machine (SVM)algorithms analyzes the information that passes it through for patterns hinting at intrusion or unusual behavior. Whenever alarm is identified from a possible security threat, notifications are there forwarded on the Telegram platform within the relevantauthorities. This project integrates IoT, machine learning, and secured communications technologies together in order to enhance the robustness of connected environments that can very quickly detect threats and respond accordingly in a timely manner.

***Index Terms* – Intrusion Detection, Feature Selection, Machine Learning , Random Forest Ensemble Approach ,Support Vector Machine.**

I. INTRODUCTION

The way we engage with and use technology has undergonea radical paradigm shift with the introduction of the Inter-net of Things (IoT). It consists of an extensive network of linked sensors, devices, and systems that exchange data and communicate with each other in a seamless manner to auto- mate tasks, boost productivity, and facilitate better decision- making. This interconnection has completely changed the waywe live, work, and interact with our surroundings. It has revolutionized a number of industries, including healthcare, transportation, agriculture, and smart homes.The origins of theInternet of Things can be found in the early 2000s, when earlysensor networks and RFID (Radio Frequency Identification) technology first came into being. But the Internet of Things didn’t really take off until the development of low-cost, low-power microcontrollers, together with improvements in wireless communication protocols and cloud computing.IoT growth was further propelled by the widespread use of smart- phones and high-speed internet connectivity, which provided the infrastructure and connectivity required to support a wide range of networked devices. These days, Internet of Things (IoT) gadgets are present in almost every area of our life, ranging from smart thermostats and wearable fitness trackers to industrial machinery and self-driving cars.The fundamental promise of the Internet of Things is its capacity to gather, process, and act upon massive volumes of data in real-time, providing previously unheard-of insights and optimization opportunities. IoT systems can identify patterns, anticipatetrends, and automate decision-making by utilizing sensor data and machine learning algorithms. This results in increasedproductivity, lower costs, and better user experiences.

IoT devices in the healthcare industry remotely monitorpatients’ vital signs, allowing for the early identification of health problems and individualized interventions. Sensors in agriculture monitor environmental factors and soil moisture content to maximize crop yields and irrigation schedules. IoT-enabled automobiles in the transportation sector interact with other cars and infrastructure to increase road safety and efficiency.

II. Related Work

[1]Sajad M. Khan, Faheem Syeed Masoodi ”Intrusion Detection System for IoT Environment using Ensemble Approaches”, proposed a system in feb 2013 To ensure that important assets are available and secure within a protected network architecture, Intrusion Detection Systems (IDS) are commonly used. However, current IDS algorithms often struggle to perform effectively. To address this, machine learning has been employed to enhance IDS efficiency. The main challenge with IDS classification is the large amount of irrelevant and redundant data in high-dimensional datasets, making it impossible for a single classifier to identify all types of attacks effectively. Thus, a novel ensemble IDS approach was proposed in this study. The approach involved using Random Forest (RF) for dimensionality reduction to select the optimal subset of the initial dataset. An ensemble learning method was then used for intrusion detection and identification. The proposed RF method outperformed other state-of- the-art approaches in several parameters with an accuracy of 99% as demonstrated by experimental results on the IoTID20 dataset. The approach was evaluated using several performance criteria, including Accuracy, Precision, Recall, and F1-score The accuracy of the ensemble classifiers was higher than that of the individual models. This improvement can be attributed to ensemble learning strategies that leverage diverse learning mechanisms with varying capabilities. By combining these strategies, we were able to enhance the reliability of our predictions while reducing the occurrence of classification errors. The experimental results show that the framework can improve the efficiency of the Intrusion Detection System, achieving an accuracy rate of 0.9863.

[2]Cristiano Antonio de Souza,Carlos Becker Westphall ”Two-step ensemble approach for intrusion detection and identification in IoT and fog computing environments”, proposed a system in march 2022 Due to Internet of Things devices resource limitations, security often does not receive enough attention. Intrusion detection approaches are important for identifying attacks and taking appropriate countermeasures for each specific threat. This work presents a two-step approach for intrusion detection and identification. The first step performs a traffic analysis with an Extra Tree binary classifier. Events detected as intrusive are analyzed in the second stage by an ensemble approach consisting of Extra Tree, Random Forest, and Deep Neural Network. An extensive evaluation was performed with the Bot-IoT, IoTID20, NSL- KDD, and CICIDS2018 intrusion datasets. The experiments demonstrated that the proposed approach could achieve similar or superior performance to other machine learning techniques and state-of-the-art approaches in all databases, demonstrating the robustness of the proposed approach The Internet of Things (IoT) is expanding quickly and is becoming increasingly important in our daily lives. Internet-connected IoT nodes can connect to the internet by using an IP address. As a result, users of various social networking platforms will be able to connect to and share devices . There is a concern about security and privacy with this broad range of IoT applications. Without a secure and up of a wide variety of devices that generally have a small size, and internet connectivity as their main characteristics [1]. Due to their small size, these devices generally have limited resources, low processing capacity, and limited memory. Thus, to carry out the storage, processing, and analysis of the generated data, it is necessary to send them to a device with greater computational power. The high traffic generated by these devices and the latency hampers the sending to cloud computing.

[3]Umaira Ahad, Yashwant Singh, Pooja Anand ,”Intrusion Detection System Model for IoT Networks Using Ensemble Learning”,proposed a sytem in 2022 The capacity to identify breaches and malicious activity inside the Internet of Things (IoT) networks is important for network infrastructure resilience as the dependence on IoT devices and services grows. Intrusion detection systems (IDS) are basic components of network security. IDSs monitor and analyze the activity of a system in a network to identify intrusions. Existing intrusion detection systems (IDS) gather and utilize large amounts of data with irrelevant, unneces- sary, and unsuitable characteristics, resulting in long detection times and low accuracy. In this paper, we present an IDS model based on a Random Forest (RF) classifier. NSL-KDD dataset is used to test the performance of the model and the satisfying performance is obtained in terms of accuracy, detection rate, and false alarm rate. The proposed model has attained an average accuracy of 99.3% and 98% for binary classification and multiclass classification, respectively. To demonstrate the efficacy of the suggested model, its accuracy was compared with some existing approaches that utilize other models such as AIDS, ELM and PCA, MapReduce-based hybrid architecture, and DRNN.ud. In this way, it can provide faster processing and response for IoT devices.Currently, many electronic devices can be connected to the Internet and provide data and services to users. The Internet of Things (IoT) environments are evolving and becoming popular. The number of devices connected to the Internet continues to grow.

III.Challenges and Concerns

IoT has enormous potential, but it also comes with a lot of difficulties and worries, especially when it comes to security and privacy. Because of their interconnectedness, Internet of Things devices have a larger attack surface, which leaves them open to online dangers including malware outbreaks, hacking, and data breaches. Furthermore, privacy and data protection are put at risk by the massive volumes of sensitive data created by IoT devices, which makes strong security protocols and legal frameworks necessary to secure user data.

Standards compliance and interoperability are two more issues that the IoT ecosystem must deal with. Making sure that a multitude of devices from various manufacturers work together seamlessly and with a variety of communication protocols can be a challenging undertaking. Furthermore, as IoT installations continue to increase in size and complexity, scalability and sustainability become more difficult to achieve. To reduce their negative effects on the environment, effective resource management and infrastructure optimization are needed.

Support Vector Machine Algorithm - Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

Random Forest Algorithm - In machine learning, random forest is a potent ensemble learning technique that is applied to both classification and regression problems. It is a member of the decision tree based algorithm family and is used extensively because of its efficiency, simplicity, and resilience.

IV. EXISTING SYSTEM

The present system for our environmental monitoring equipment is based on traditional communication protocols with no sophisticated security features. Currently, two IoT devices equipped with DHT11 sensors for temperature and humidity monitoring are connected in a simple manner. These devices interact using a normal datagram protocol, passing data through a Mosquitto broker that use the MQTT protocol. The device collect data from their surroundings and exchange it with cloud.

V. PROPOSED SYSTEM

The proposed system aims to greatly improve the security of our IoT environmental monitoring devices by including sophisticated features, notably an Intrusion Detection System (IDS) based on machine learning methods. Two IoT devices equipped with DHT11 sensors for temperature and humidity measurement remain at the heart of environmental monitoring. However, the connectivity between these devices under goes significant improvement. Datagram connection is preserved, how ever data is now transmitted over a Mosquitto broker using the MQTT protocol, giving a more secure communication channel than the previous method.

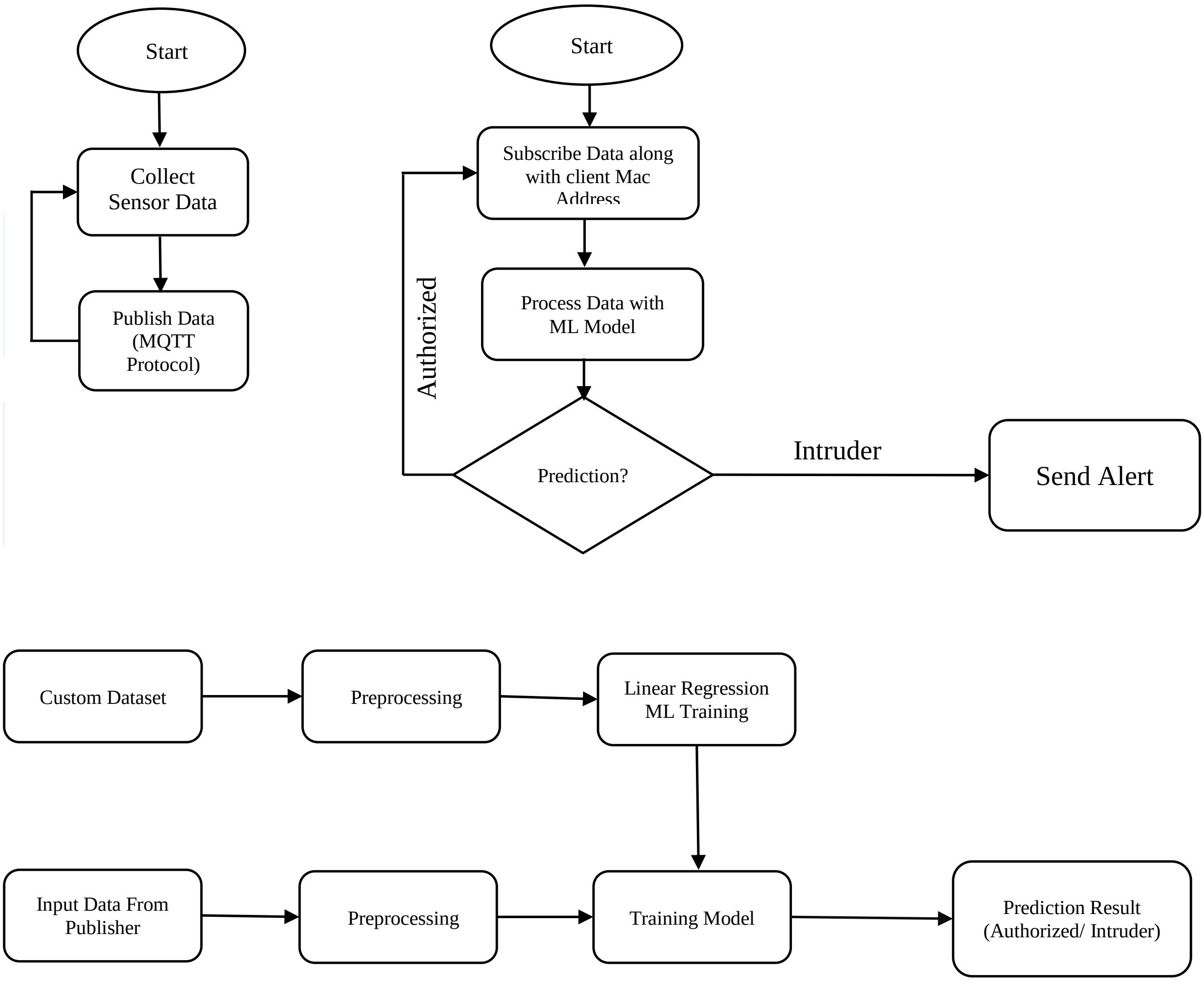

Start
Collect Sensor Data
Publish Data (MQTT Protocol)
Start
Subscribe Data along with client Mac Address
Process Data with ML Model
Authorized
Prediction?
Intruder
Send Alert
Custom Dataset
Preprocessing
Linear Regression ML Training
Input Data From Publisher
Preprocessing
Training Model
Prediction Result (Authorized/ Intruder)

Fig 1: Flow diagram of Model

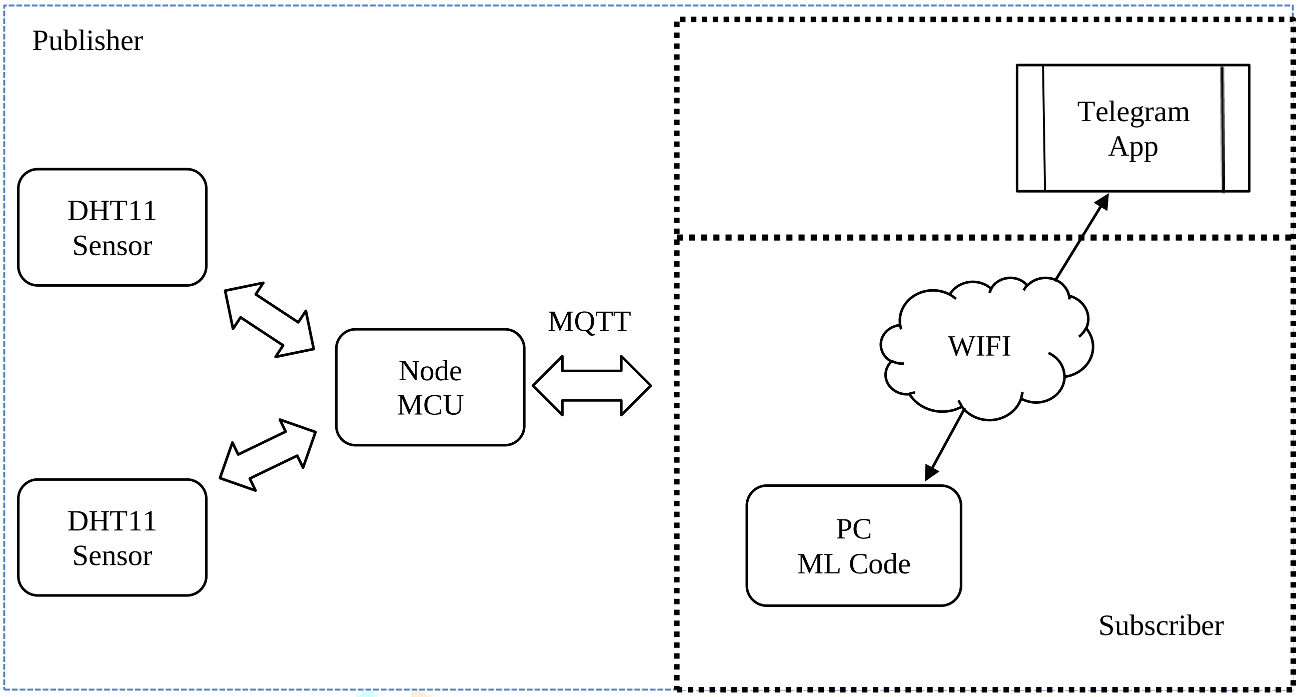


Fig 2: Block diagram of Proposed System

VI. Hardware Requirements

4.1. Node MCU :

The NodeMCU is a popular option for professionals and hobbyists alike because to its many features and capabilities. The ESP8266 microcontroller, which powers the NodeMCU, has a 32-bit Tensilica Xtensa LX106 CPU that can operate at up to 80 MHz. The NodeMCU's potent processor and inte- grated Wi-Fi enable it to easily connect to wireless networks and communicate with other devices via the internet.

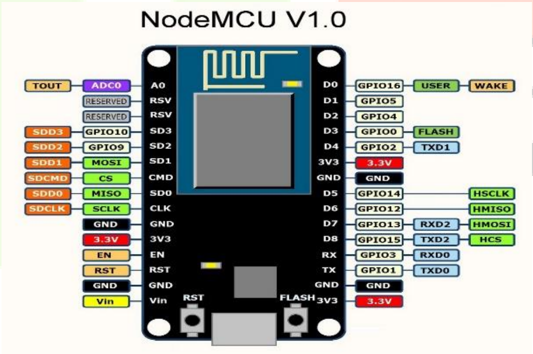


FIG 3: Node MCU

4.2. DHT11 sensors:

The DHT11 sensor is a digital temperature and humid- ity sensor module that is widely utilized in many different applications, such as home automation systems and weather stations.The analog-to-digital converter (ADC), temperature sensor, and humidity sensor are the three main components of the DHT11 sensor, which are all combined into one single package. The sensor's interface with microcontrollers and other electronic equipment is made simpler by its single-wire digital operation.Through a single-wire digital interface, the DHT11 sensor can communicate with external devices like microcontrollers or development boards. Through the use of this interface and a particular communication protocol, the host device receives data packets from the sensor that contain temperature and humidity readings. After that, the host device decodes these data packets and handles the data appropriately.

`

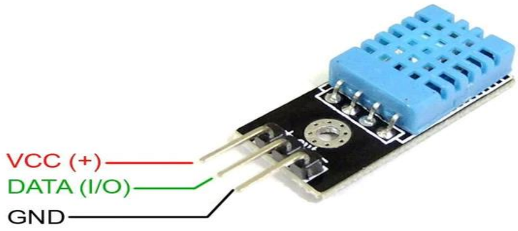


FIG 4: DHT11 sensors

VII. SOFTWARE REQUIREMENTS

The detailed list of Software requirements for the proposed system are as follows:

- Operating System: Windows 10 or Windows 11 (64-bit preferred).
- Programming Language: Python 3.x ,C.
- Machine Learning algorithms: SVM, Random forest

VIII. RESULTS AND DISCUSSION

In the experiment, the assessment of accuracy, a widely used multi-class performance metric, is conducted to analyze the performance of the executed models. Precision, Recall, and F1-score scores are also computed. The results indicate that among the ensemble models RF and SVM, SVM attains the highest test accuracy, scoring 98%Additionally, the MQTT protocol and datagram connection provide secure and effective data transfer between IoT devices and the server. The system's responsiveness and efficacy in identifying and addressing security events are supported by the framework's dependable communication architecture. The implementation of sophisticated machine learning techniques on the server side, such as Random Forest and Support Vector Machine (SVM), is the central component of the project. These algorithms enable the system to proactively detect and mitigate potential security threats by analyzing incoming data streams for patterns suggestive of intrusion or anomalous behavior.

	Accuracy	**F1 Score**	**Precision**			
Mintial	1.0	1.0	1.0			
Iteration -1	0.97	0.99	0.99			
Iteration -2	0.96	0.95	0.96			
After Intrusion	0.80	0.98	0.97			

TABLE NO 1: PERFORMANCE METRICS

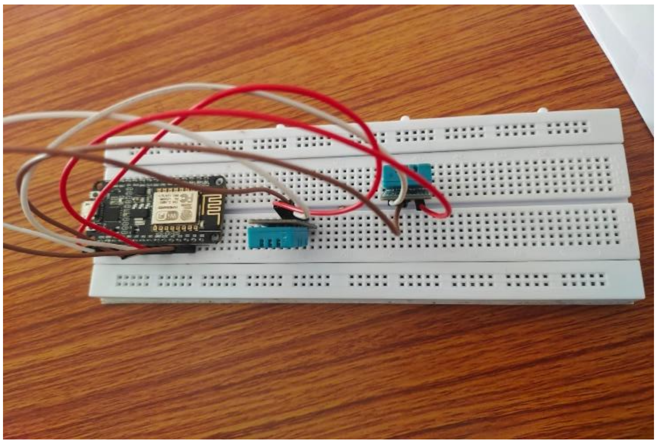
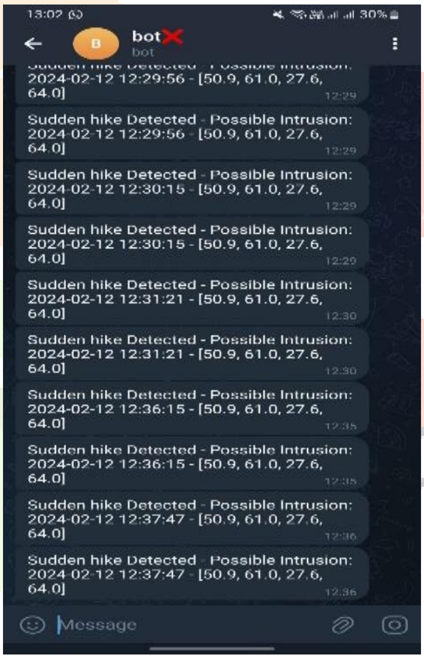


FIG 5: HARDWARE MODEL

When the system detected any intrusion it alert us via telegram as shown in the below diagram.



Sudden hike Detected - Possible Intrusion:
2024-02-12 12:29:56 - [50.9, 61.0, 27.6, 64.0] 12:29

Sudden hike Detected - Possible Intrusion:
2024-02-12 12:29:56 - [50.9, 61.0, 27.6, 64.0] 12:29

Sudden hike Detected - Possible Intrusion:
2024-02-12 12:30:15 - [50.9, 61.0, 27.6, 64.0] 12:29

Sudden hike Detected - Possible Intrusion:
2024-02-12 12:30:15 - [50.9, 61.0, 27.6, 64.0] 12:29

Sudden hike Detected - Possible Intrusion:
2024-02-12 12:31:21 - [50.9, 61.0, 27.6, 64.0] 12:30

Sudden hike Detected - Possible Intrusion:
2024-02-12 12:31:21 - [50.9, 61.0, 27.6, 64.0] 12:30

Sudden hike Detected - Possible Intrusion:
2024-02-12 12:36:15 - [50.9, 61.0, 27.6, 64.0] 12:35

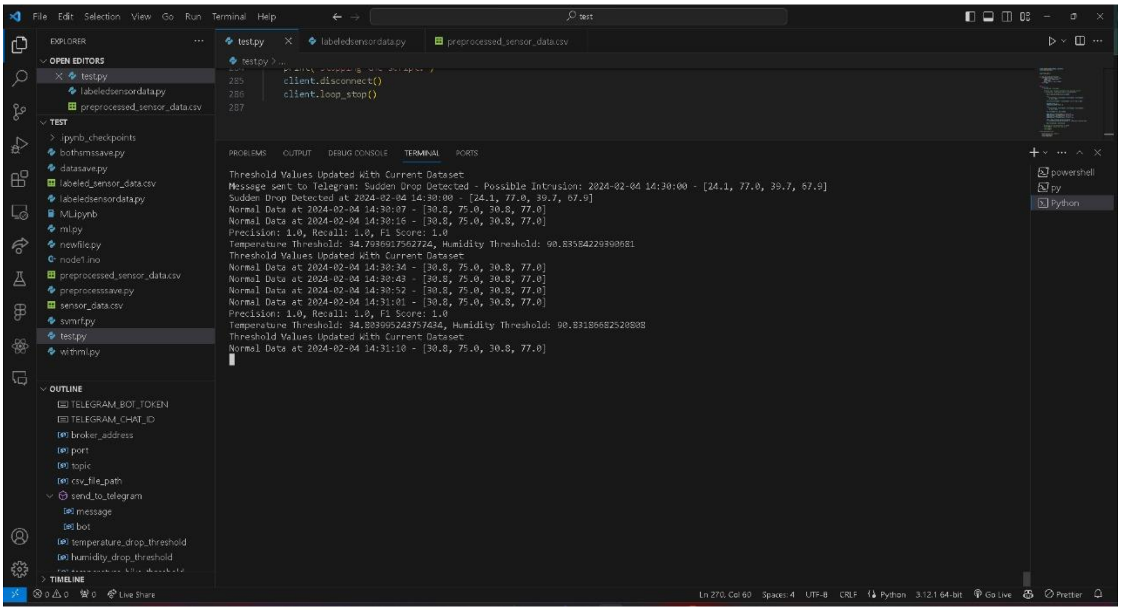
Sudden hike Detected - Possible Intrusion:
2024-02-12 12:36:15 - [50.9, 61.0, 27.6, 64.0] 12:35

Sudden hike Detected - Possible Intrusion:
2024-02-12 12:37:47 - [50.9, 61.0, 27.6, 64.0] 12:36

Sudden hike Detected - Possible Intrusion:
2024-02-12 12:37:47 - [50.9, 61.0, 27.6, 64.0] 12:36

FIG 6: TELEGRAM NOTIFICATION

System detects the hike increment in temperature are shown below figure.


File Edit Selection View Go Run Terminal Help
test.py
labeledsensordata.py
preprocessed_sensor_data.csv
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
    client.disconnect()
    client.loop_stop()
```

```
Threshold Values Updated With Current Dataset
Message sent to Telegram: Sudden Drop Detected - Possible Intrusion: 2024-02-04 14:30:00 - [24.1, 77.0, 39.7, 67.9]
Sudden Drop Detected at 2024-02-04 14:30:00 - [24.1, 77.0, 39.7, 67.9]
Normal Data at 2024-02-04 14:30:07 - [30.8, 75.0, 30.8, 77.0]
Normal Data at 2024-02-04 14:30:16 - [30.8, 75.0, 30.8, 77.0]
Precision: 1.0, Recall: 1.0, F1 Score: 1.0
Temperature Threshold: 34.7936917562724, Humidity Threshold: 90.83584229390681
Threshold Values Updated With Current Dataset
Normal Data at 2024-02-04 14:30:34 - [30.8, 75.0, 30.8, 77.0]
Normal Data at 2024-02-04 14:30:43 - [30.8, 75.0, 30.8, 77.0]
Normal Data at 2024-02-04 14:30:52 - [30.8, 75.0, 30.8, 77.0]
Normal Data at 2024-02-04 14:31:01 - [30.8, 75.0, 30.8, 77.0]
Precision: 1.0, Recall: 1.0, F1 Score: 1.0
Temperature Threshold: 34.803995243757434, Humidity Threshold: 90.83186682520808
Threshold Values Updated With Current Dataset
Normal Data at 2024-02-04 14:31:10 - [30.8, 75.0, 30.8, 77.0]
```

FIG 7: HIKE DETECTION

The temperature over time. The x-axis is labeled "Time," and the y-axis is labeled "Temperature (°C)" are shown in the figure.

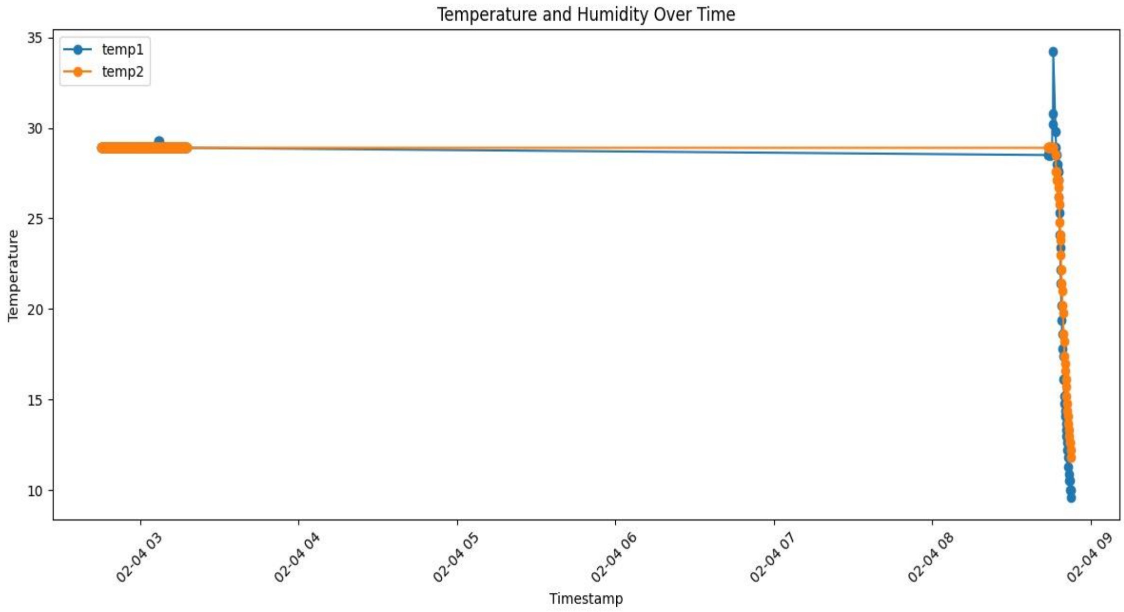


FIG 8: TEMPERATURE OVER TIME

The humidity over time. The x-axis is labeled "Time," and the y-axis is labeled "Humidity (%)" are shown in the figure

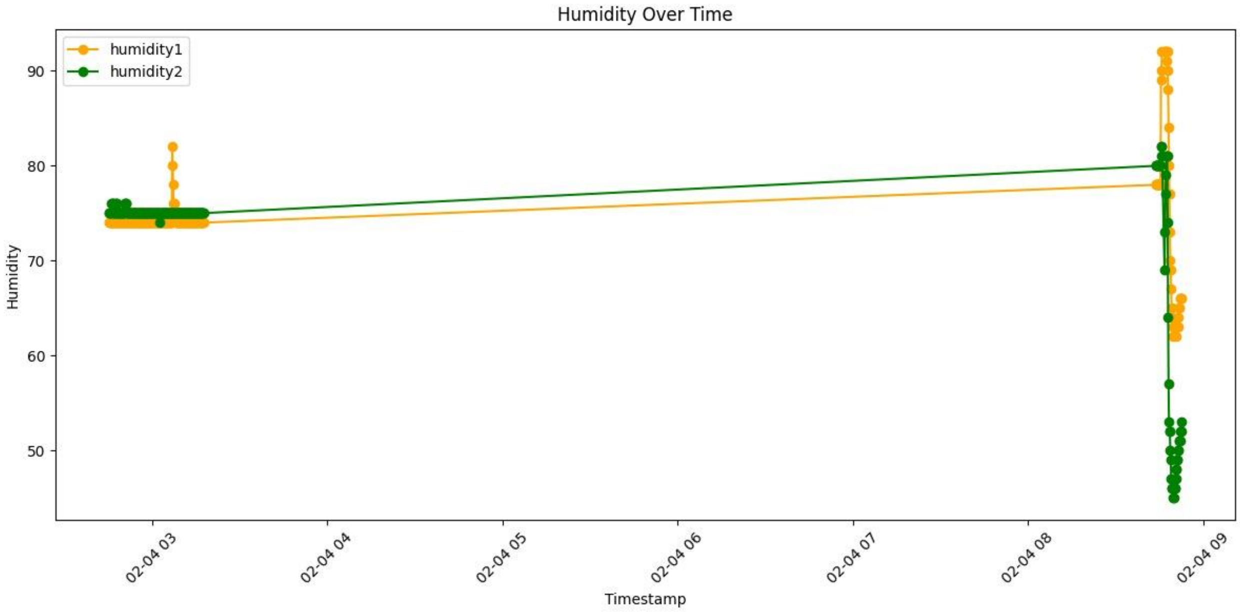


FIG 9: HUMIDITY OVER TIME

IX. CONCLUSION

The project is an excellent example of a comprehensive and progressive strategy for enhancing security in the context of the Internet of Things (IoT). Through the integration of DHT11-equipped IoT devices, machine learning-powered Intrusion Detection Systems (IDS), and secure communication protocols, the system creates a mutually beneficial synergy that improves the resilience and robustness of networked settings.Additionally, the MQTT protocol and datagram con- nection provide secure and effective data transfer between IoT devices and the server. The system's responsiveness and efficency in identifying and addressing security events are supported by the framework's dependable communication architecture.These algorithms enable the system to proactively detect and mitigate potential security threats by analyzing incoming data streams for patterns suggestive of intrusion or anomalous behavior.Through the utilization of machine learning and the combined intelligence of IoT devices, the system improves situational awareness and gives stakeholders the ability to take proactive measures to protect against new risks.

REFERENCES

[1] Sajad M. Khan, Faheem Syeed Masoodi Proposed ”In- trusion Detection System for IoT Environment using Ensemble Approaches” was presented on 15-17 March 2023. It was added to IEEE Xplore on 04 May 2023. The electronic ISBN for the paper is 978-93-80544-47-2, and the Print on Demand (PoD) ISBN is 978-1-6654-7703-1.”

[2] Cristiano Antonio de Souza, Carlos Becker Westphall, Re- nato Bobsin Machado Proposed ”Two-step ensemble approach for intrusion detection and identification in IoT and fog com- puting environments”,Maintained by Elsevier, Published in the journal Computers & Electrical Engineering. The content is classified as an article and is protected by copyright © 2022 Elsevier Ltd. All rights reserved.”

[3] Umaira Ahad, Yashwant Singh, Pooja Anand, Zakir Ahmad Sheikh, and Pradeep Kumar Singh Proposed ”Intrusion Detection System Model for IoT Networks Using Ensemble Learning”, is published in the Journal of Interconnection Networks, Volume 22, Issue 03