iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51583

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Real-Time IoT Data Analysis Using Machine Learning

Sathwika M[1], Vinni B[2], Sai Sindhu P[3], Ravi Kumar S[4]
[1, 2, 3]*Student,* [4]*Assistant Professor, Electronics and Communication Department, Maturi Venkata Subba Rao Engineering College,Osmania University, Hyderabad, India*

Abstract: ***The world is changing rapidly and many online businesses depend on collecting data, which is used for future predictions. IoT systems can access many devices and large amounts of data can be stored in IoT clouds like Thingspeak. This project aims to collect real-time data using DHT11, Gas level sensors, and analyze it using machine learning algorithms(Random forest, Decision tree classifier, Linear discriminant analysis). We are comparing the performance of algorithms using metrics like accuracy, confusion matrix, precision, recall, and score to find the best algorithm which detects the presence of attacks or data anomalies more accurately.***
Keywords: Thingspeak, DHT11, Gas sensor, Machine learning algorithms, Metrics.

I. INTRODUCTION

The Internet of Things describes physical objects that are implanted with sensors, computing ability, software, and other technologies that connect and share data with other devices and systems over the Internet without any human interaction. With the fast growth and development of Internet of Things(IoT) devices, there is an increase in cyber-attacks bugging these devices. Unauthorized users continue to find new ways and loopholes for exploiting the existing IoT- HUB for illegal purposes. Machine Learning algorithms are proposed for the detection of malicious attacks or data anomalies. The framework uses three classification-based ML algorithms, namely Random Forest(RF), Decision tree classifier, and Linear discriminant analysis for detecting the presence of attacks.

II. METHODOLOGY

1) *Hardware Configuration:* The system combines the DHT11, Gas sensor with NODEMCU(ESP8266) Wi-fi module.
2) *Arduino IDE software:* A basic Embedded C program code is uploaded.
3) *Thingspeak IoT:* The data generated using sensors is uploaded to the cloud and data sets are downloaded in Excel sheets for further use in algorithms.
4) *Machine Learning:* Jupyter Notebook software is used for the ML part of the project. Required libraries are imported and the data set is split into training and testing sets. These sets are fitted into different ML algorithms. Performance metrics are compared to find out the best algorithm that detects the attacks more accurately.

III. COMPONENTS REQUIRED

The following components are necessary for the proposed system:

1) *NODEMCU:* NodeMCU is an open-source that is especially used for IoT-based Applications. This comprises components that run on the ESP8266 Wi-Fi SoC from Espressif Systems, and hardware that is based on the ESP-12 module. ESP8266EX has 17 GPIO pins which are dedicated to different functions by programming the appropriate registers. This component is perfect for IoT based projects where Arduino is wireless. It has an inbuilt wifi module

Fig1.Nodemcu

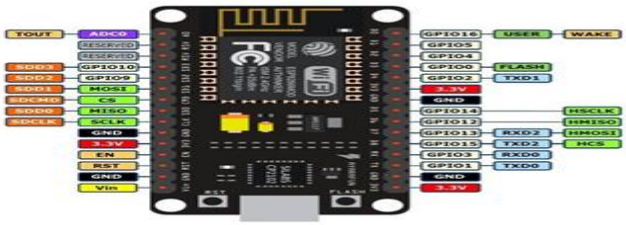


Fig2.Pin diagram of Nodemcu

2) *DHT11 Sensor:* DHT11 is a low-value virtual sensor for sensing temperature and humidity. It has a sensing element and a thermistor to measure the temperature. Humidity values are recorded with a change in capacitance. This sensor can be easily interfaced with any microcontroller such as Arduino, Raspberry Pi, etc… to measure humidity and temperature at particular periods. DHT11 sensor has 4 pins- VCC, GND, Data Pin, and a now no longer linked pin. It also has a pull-up resistor of 5-10k ohms for microcontroller communication.

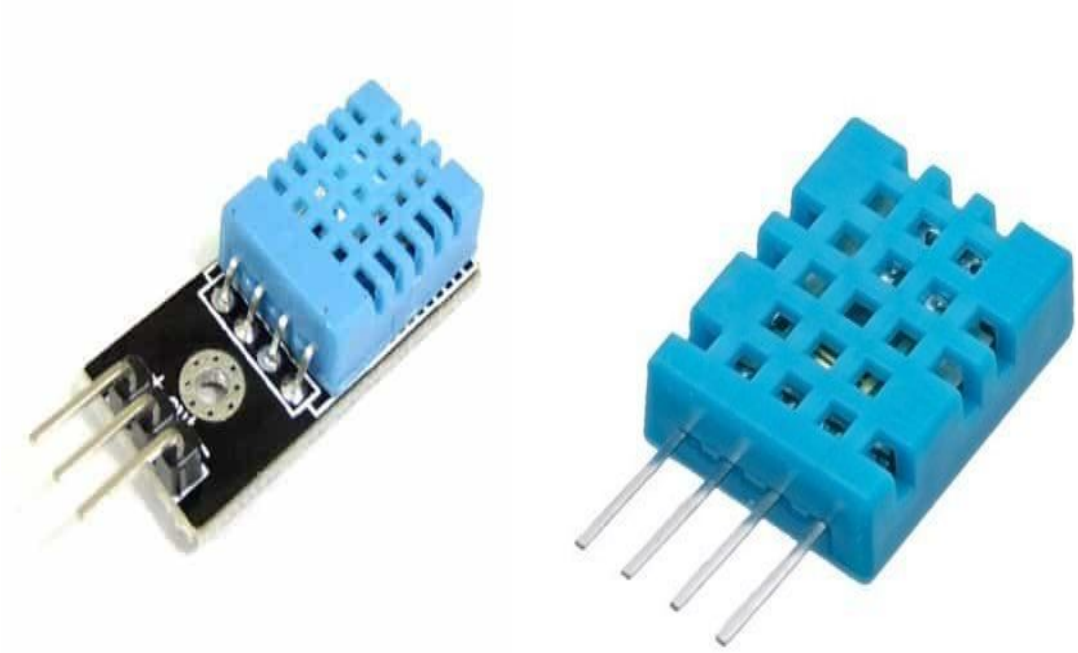


Fig 3.Dht11 sensor

3) ***Gas Sensor:*** A sensor that is used to detect the presence or concentration of gases in the air is called a gas sensor. Gas sensors are used to detect dangerous leaks like petroleum gas or CNG to avoid major accidents. All gas sensors include a sensing element that contains some components. The components are the Gas sensing layer, Heater coil, Electrode, etc. Depending on the level of concentration of the gas the sensor gives the respective change in potential energy by varying the resistance of the material inside the sensor. That change in potential can be termed output voltage. Depending upon the voltage value the type and concentration level of the gas can be evaluated. The sensing material present inside the sensor can detect the type of gas.

Fig 4. Gas sensor

IV. WORKING

The system collects real-time data using dht11 and a gas level sensor. These values can be displayed on LCD and NODEMCU has a in-built wi-fi module. The system is interfaced with the Arduino IDE and the cable from NODEMCU is connected to the laptop The data set is uploaded to Thingspeak and it can be downloaded in the form of an Excel sheet.

Block Diagram:

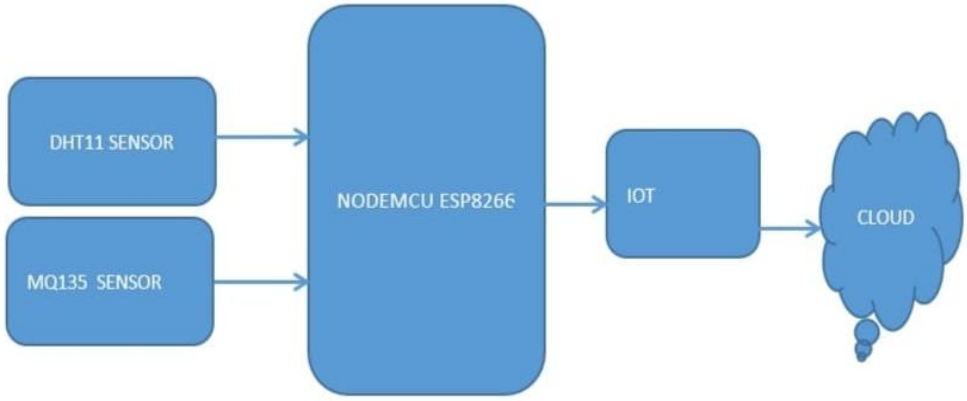


Fig 5.Block diagram

After the data set is downloaded, it is converted into a CSV file and then it is used for the machine learning algorithm.
It involves below steps:

1) Data sets are uploaded.
2) Required libraries must be imported.
3) Datasets are imported.
4) Any Missing Data should be taken care of.
5) Categorical Data is encoded.
6) Dataset is split into training and test set.
7) Fitting the model.

After the model is fitted, the performance of all algorithms is evaluated based on some metrics.

V. RESULTS

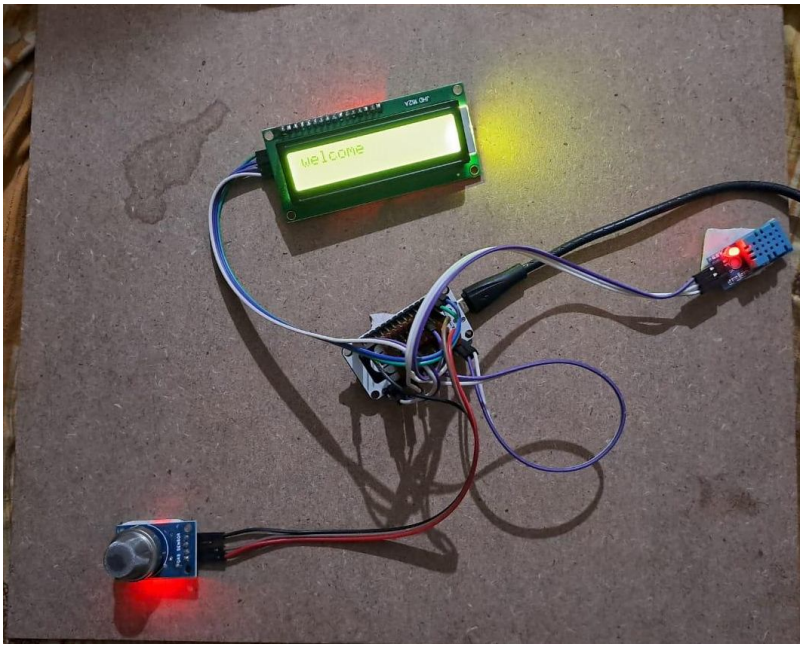


Fig 6. Hardware Configuration

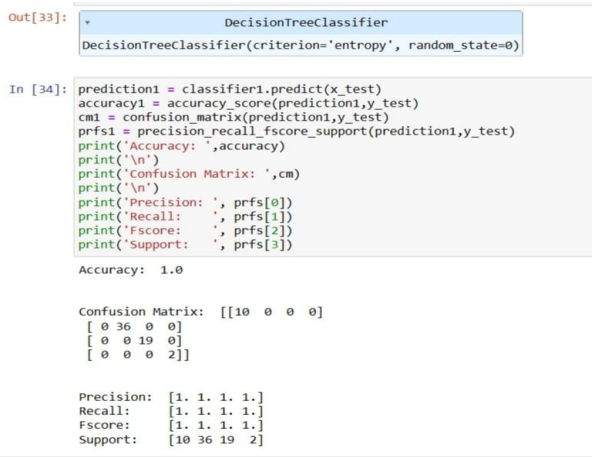


Fig 7. Results of a Decision tree classifier

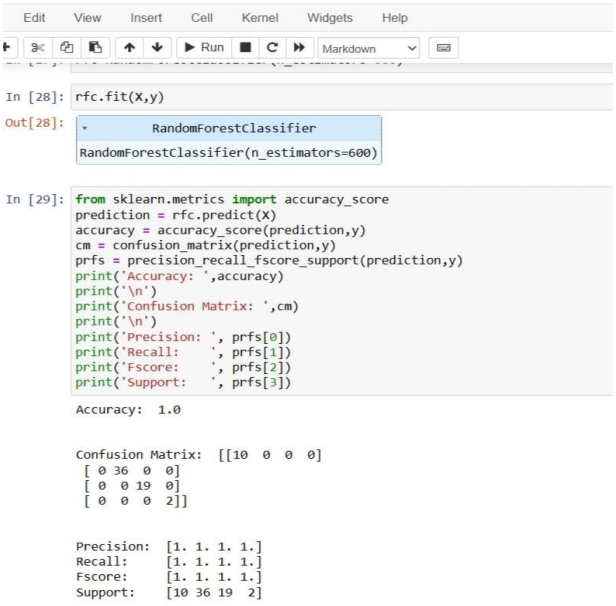


Fig 8. Results of Random forest algorithm

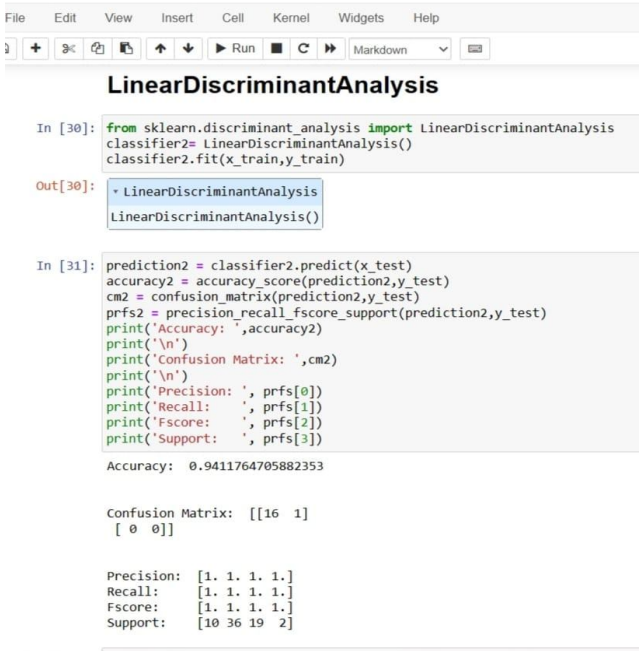


Fig 9 Results of Linear discriminant analysis

VI. CONCLUSION

The main aim of our proposed system is to find the best algorithm that provides good performance for attack detection. After comparing the results, we concluded that Random forest is the best one with 100 percent accuracy. This value may change depending on the data set. It has been concluded that large data sets are handled by random forest over the other algorithms. The problem is when the results combined from various decision trees are eliminated by Random Forest. We understood that Random Forest is fast and highly accurate.

VII. FUTURE SCOPE

We will collect more nonlinear data and include additional features from the packet capture for each type of device to provide more accurate classifications. For this, we use a random forest algorithm which performs better with nonlinear data. Random Forest is intrinsically suited for multiclass problems. For multiclass problems, you will need to reduce them into multiple binary classification problems. We can tune and optimize the parameters and compare random forest and Decision tree classifiers and linear discrimination analysis algorithms with the confusion matrix and other performance metrics like accuracy, precision, and recall.

REFERENCES

[1] Abeshu, A., Chilamkurti, N., 2018. Deep Learning: The Frontier for Distributed Attack Detection in Fog-toThings Computing. IEEE Communications Magazine 56, 169–175. https://doi.org/10.1109/MCOM.2018. 1700332

[2] Agatonovic-Kustrin, S., Beresford, R., 2020. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of Pharmaceutical and Biomedical Analysis 22, 717– 727. https://doi.org/10.1016/S0731-7085(99)00272-1

[3] Ali, T., Nauman, M., Jan, S., 2018. Trust in IoT: dynamic remote attestation through efficient behavior capture. Cluster Computing 21, 409–421.https://doi.org/ 10.1007/s10586-017-0877-5

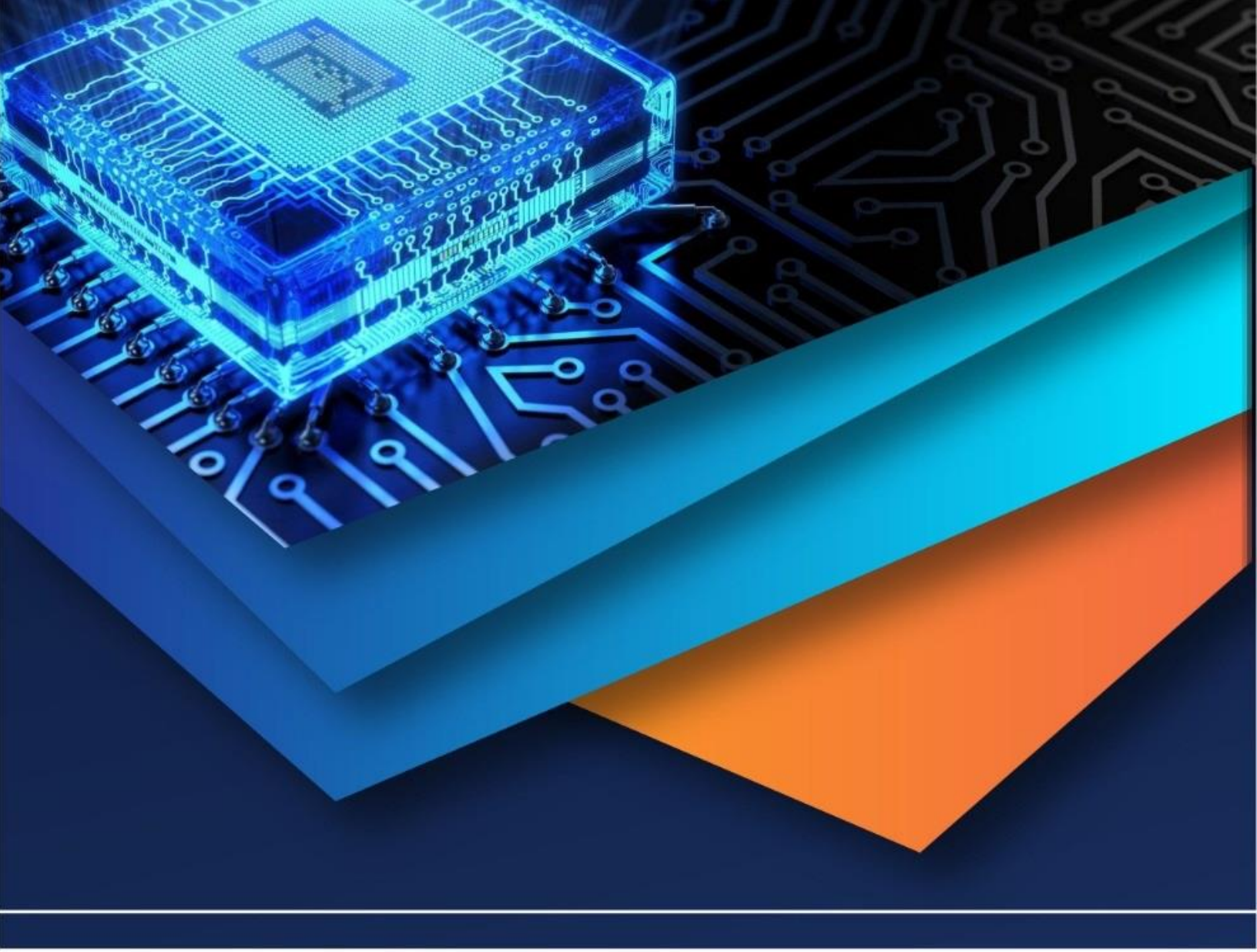
[4] Aminanto, M. E., Choi, R., Tanuwidjaja, H. C., Yoo, P. D., Kim, K., 2018. Deep Abstraction and Weighted Feature Selection for Wi-Fi Impersonation Detection. IEEE Transactions on Information Forensics and Security 13, 621–636. https://doi.org/10.1109/TIFS. 2017.2762828

[5] Amor, N. B., Benferhat, S., Elouedi, Z., 2021. Naive Bayes vs decision trees in intrusion detection systems. ACM Press, p. 420. https://doi.org/10.1145/967900. 967989

[6] Bhunia, S.S., Gurusamy, M., 2019. Dynamic attack detection and mitigation in IoT using SDN, in 2017 27th International Telecommunication Networks and Applications Conference (ITNAC). Presented at the 2017 27th International Telecommunication Networks and Applications Conference (ITNAC), IEEE, Melbourne, VIC, Australia, pp. 1–6. https://doi.org/ 10.1109/ATNAC.2017.8215418.

[7] Wu, M., Song, Z., Moon, Y.B., 2017. Detecting cyber-physical attacks in the cyber Manufacturing systems with machine learning methods. Journal of Intelligent Manufacturing. https://doi.org/10.1007/s10845-017- 1315-5. 61

[8] Wang, L. (Ed.), 2020. Support vector machines: theory and applications, Studies in fuzziness and soft computing. Springer, Berlin.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY