

IoT BASED AIR AND SOUND POLLUTION MONITORING SYSTEM USING MACHINE LEARNING ALGORITHMS

M. Ramana Reddy,

Department of Electronics and communication engineering,
Chaitanya Bharathi Institute of Technology,
Osmania University,
Hyderabad, Telangana, India.
Email: mramanareddy_ece@cbit.ac.in

Abstract: Air pollution is the largest environmental and public health challenge in the world today. Air pollution leads to adverse effects on human health, climate and ecosystem. Air is getting polluted because of release of Toxic gases by industries, vehicular emissions and increased concentration of harmful gases and particulate matter in the atmosphere. In order to overcome these issues an IoT based air and sound pollution monitoring system is designed. To design this monitoring system, machine learning algorithms K-NN and Naive Bayes are used. K-Nearest Neighbour and Naive Bayes are machine learning algorithms used to predict the status of pollution present in the environment. In this system, analog to digital converter, global service mobile communication, temperature sensor, humidity sensor, carbon monoxide and sound sensors are interfaced with raspberry pi using serial cable. The sensor data is uploaded in thinkspeak (IoT) and webpage. This data is compared with the trained data to check accuracy. To calculate the accuracy of both algorithms, Python code is developed using python software tool.

Keywords: IoT, Temperature, Humidity, Carbon Monoxide, Sound, Raspberry Pi.

1. Introduction

Air and sound pollution monitoring system is IoT (Internet of Things) based application, used to detect the pollution using machine learning algorithms[1-2]. The temperature sensor, humidity sensor, carbon monoxide sensor and sound sensor interfaced with raspberry pi to detect pollution in the environment. GSM SIM800L is a miniature cellular module which allows for GPRS transmission, sending and receiving messages. The system proposed a comparative study between various techniques for prediction. This work focuses on developing an optimized system model which predicts future weather. There are some models that predict weather during real time, or monthly period. This system that carries out weather prediction using previous or historical weather data having attributes (Date, Temperature, Humidity, Carbon monoxide and Sound). In machine learning, K-Nearest Neighbour and Naive Bayes algorithms are used for the purpose of weather forecasting. Comparison based on evaluation parameters to identify which model has more accurately performed the predictions. The main objective of the project is to design and implement a real-time monitoring system for pollution applications. The system

is based on a single-chip microcontroller with a set of sensors and a wireless communication unit. Sensor data is loaded directly into ThinkPick and the web server[3-4].

2. Components Required

2.1 Temperature and Humidity Sensor (DTH11)

DHT11 is a humidity and temperature sensor. It can be used to monitor the temperature and humidity levels in a region. It can be interfaced with a Raspberry Pi module and can give immediate results. In this project, we are using this sensor to monitor the varying humidity and temperature levels.

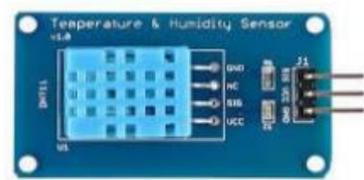


Figure 1: Temperature and humidity sensor (DTH11)

1.2 Carbon Monoxide (Mq-7) Gas Sensor

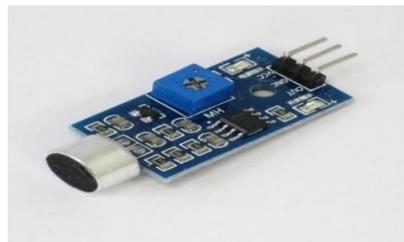
This semiconductor gas sensor detects the presence of carbon monoxide in concentrations ranging from 10 to 10,000 ppm. The simple analog voltage interface of the sensor requires a single analog input pin of the microcontroller.



Figure 2: Carbon monoxide sensor (MQ7)

This carbon monoxide (CO) gas sensor detects CO concentrations in the air and produces its reading as an analog voltage. The sensor can measure concentrations of 10 to 10,000 ppm. Sensor - Operates at temperatures from 10 to 50 ° C and consumes less than 150 mA at 5 V.

Microphone used to capture the audio signal. Sound is a sensitive material that converts sound into electrical energy. Amplifier used to amplify the PM358 to amplify the electrical signal generated.



Figur.3: Sound sensor (LM393)

1.3 Analog-To-Digital Converter(Mcp3208) IC

Microchip technology Inc. The MCP3208 devices are analog-to-digital (A / D) converters with sequential sampling in a 12-bit approximation with built-in sampling and maintenance circuits. The MCP3208 is programmed to provide four pairs of pseudo-differential inputs or eight unbalanced inputs. Differential nonlinearity (DNL) is specified in ± 1 LSB, and integral nonlinearity (INL) is available in versions ± 1 LSB (MCP3204 / 3208-B) and ± 2 LSB (MCP3204 / 3208-C).



Figure .4: Analog-to-Digital Converter IC (MCP3208)

1.4 GSM

Packet-based standard for mobile data transmission in 2G and 3G cellular communication network's Global Mobile Communication System (GSM). GPRS was created by the European Telecommunications Standards Institute (ETSI) in response to previous packet-switched i-mode cellular CDPD technologies.



Figure.5: Global System for mobile communication (GSM)

The SIM800L is a small cellular module that allows you to send GPRS, send and receive SMS, as well as make and receive voice calls. Low cost, small footprint and support for four frequency bands make this module an ideal solution for any project that requires a long distance connection. After connecting the power supply module, it loads, searches for the cellular network and automatically enters the system.

3. Algorithms

3.1 Machine Learning

Machine learning (ML) is a machine that can be used to improve the performance of a machine that implies the Artificial intelligence methods by which human seems to identify, recognize and the expertise. Machine Learning (ML) is a sub-field of Artificial Intelligence (AI) which concerns with developing computational theories of learning and building learning machines. Machine learning algorithms can divide into three different groups based on their learning style[5-6]:

- i. Supervised learning
- ii. Unsupervised learning
- iii. Reinforcement learning

3.2 Supervised Learning

Supervised learning occurs when an algorithm learns from input data also known as training data which has known target responses or labels that can be a numeric value or string. And prepare a model through training or learning process that assesses the correct answer when given a new example. The monitored approach is divided into two: classification and regression. In classification, the algorithm predicts the class to which the given tests data fall into whereas regression predicts a numeric value for target variable. The classification is part of the supervised learning model in which the algorithm envisages the class in which the new data are found, where the class should not be a numerical value, it will deal with three important classification algorithms:

- i. K-Nearest Neighbour
- ii. Naive Bayes

3.3 K-Nearest Neighbour Algorithm

K-NN is slow supervised learning algorithm, more time to get trained classification like other algorithm is divided into two steps training from data and testing it on new instance. The K Nearest Neighbour working principle is based on assignment of weight to each data point which is called as neighbour. In K, the distance of the nearest neighbour is calculated for the training data for each of the data points coming from the nearest K, now the ranking is based on the majority votes. The KNN Euclidean Distance, Manhattan, Minkowski Measured three types of distances, in which most of the following formulas are used to calculate the distance.

$$\text{Euclidian distance} = D(x, y) = \sqrt{\sum_{k=1}^K (x_k - y_k)^2}$$

K=number of clusters and (x, y) are coordinate sample spaces.

3.4 Naive Bayes Algorithm

It is a classification technique based on [Bayes' Theorem](#) with an assumption of independence among predictors. : Simply put a naive Bayes classification uses that the existence of a particular attribute in a class is not related to the existence of other attributes. For example, a fruit can be considered an apple if it has a red, rounded and 3 inch diameter. Whether or not these traits depend on one another, all of these traits contribute independently to the likelihood that the fruit is an apple and is therefore called "innocent". The Naive Bayes model is easy to build and is particularly useful for very

large data sets.. The Bayes theory provides a way to calculate the posterior probabilities $P(C | X)$ from $P(C)$, $P(X)$ and $P(X | C)$. So, the equation is:

$$p(c|x) = \frac{p(x|c)P(c)}{P(x)}$$

Above,

- i. The posterior probability of class is $P(c/x)$ and (c, target) given predictor (x, attributes).
- ii. $P(c)$ is the prior probability of class.
- iii. The probability of predictor given class is $P(x/c)$ is the likelihood.
- iv. The $P(x)$ is the prior probability of predictor.

4. Methodology

An IoT [7-8]based air and sound pollution monitoring system is developed by interfacing temperature and humidity sensor, sound sensor, gas sensor, GPRS with the Raspberry pi Microcontroller. The microcontroller forms the heart of the project because it controls the devices connected to each other and communicates with the devices according to the written program. System uses air sensors to sense presence of harmful gas in the air and constantly transmit this data to raspberry pi.

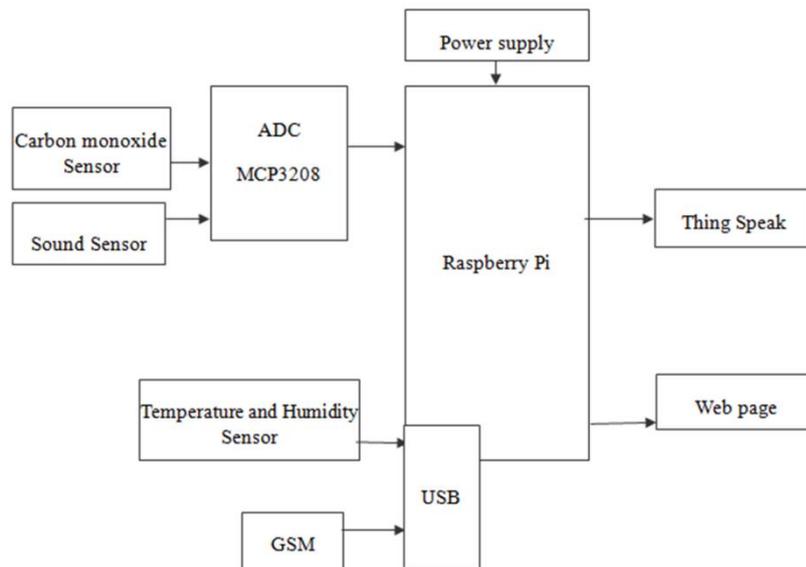


Figure 6: Block Diagram of Air and Sound Pollution Monitoring System

Also, system keeps measuring temperature, humidity, gas and sound level and reports it to the online server over IoT. The sensors interact with microcontroller which processes this data and transmits it over cloud. In webpage, data mining techniques are used for this purpose of prediction such as K-Nearest Neighbour and naive Bayes. Comparison based on evaluation parameters identifies which model made the predictions more accurate.

5. Results and Discussion

In this section, we have presented the results of the experimental trials performed to validate the behaviour and evaluate the performance of the developed hardware and the software. The experimental setup consists of a PCB with a sensor network [9-11] that takes samples monitoring on the serial monitor of Raspberry pi. For the real-time monitoring, we are using a Wi-Fi module and we will be updating the Thing Speak server forever 20s with different parameters. After uploading sensors [12] data into the thinkspeak results is shown by open the website www.thingspeak.com. Clicking on signup fill your details. After this verifies your E-mail id and click on continue.

5.1 Hardware Setup

In this system we are used GSM, DTH11 sensor, LM393 sensor, MQ7 sensor, ADC which are interfaced to microcontroller. The hardware model of system shows in fig:

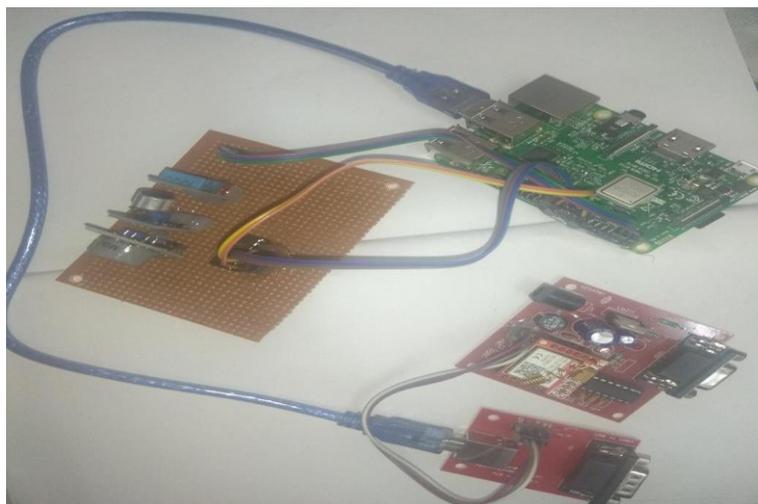


Figure 7: Complete hardware setup

5.2 Python Software Results

The reading sensors values are displayed and uploaded to thinkspeak and webpage directly. If any abnormal changes occur in the environment alert message will be sent to the registered mobile number.

```
pi@raspberrypi: ~
login as: pi
pi@192.168.43.245's password:
Linux raspberrypi 4.14.79-v7+ #1159 SMP Sun Nov 4 17:50:20 GMT 2018 armv7l

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Mon Jun 24 14:49:46 2019 from 192.168.43.35

SSH is enabled and the default password for the 'pi' user has not been changed.
This is a security risk - please login as the 'pi' user and type 'passwd' to set
a new password.

pi@raspberrypi:~ $ python weather.py -m 8639399876
Reading Sensor Values

CO: 131
Sound: 120
Temperature = 29.0*C
Humidity = 70.0%

Posting Values to Thingspeak Server

Posting Values to Cloud

Updated values in server
Alert: CO is High
Sending SMS

Alert: Sound is High
Sending SMS
```

Figure 8: Output readings of sensors values in Python

5.3 Thinkspeak Results

Click on the channel Air and sound pollution monitoring system be displayed by clicking on that temperature, humidity, carbon monoxide and sound sensors graphs will be displayed on webpage.

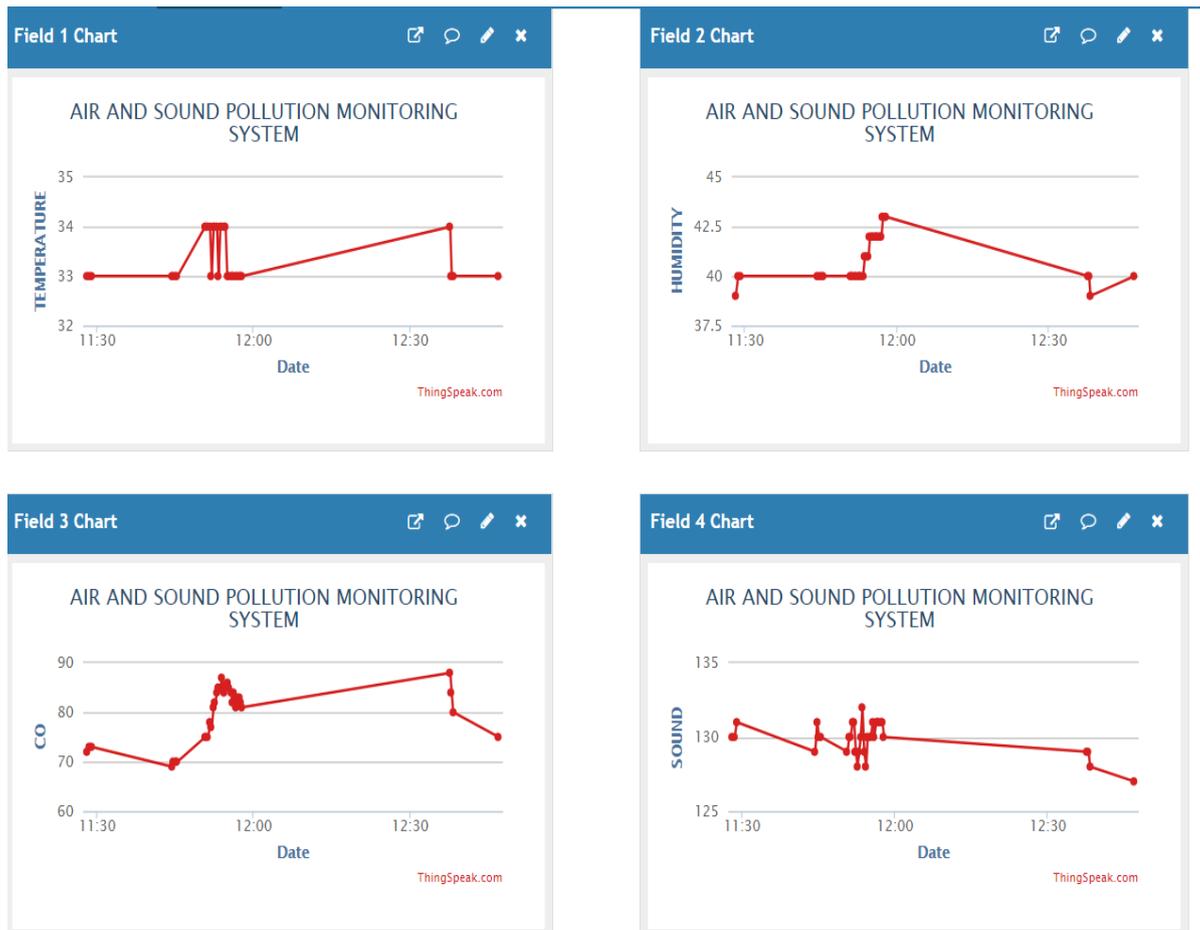


Figure 9: output of thinkspeak

5.4 Mobile App Results

To install IoT app in mobile from play store. Give API key and sign in. Channel will be opened which is created in thinkspeak. Open the channel the results are obtained. This is another way to check the results using mobile app.

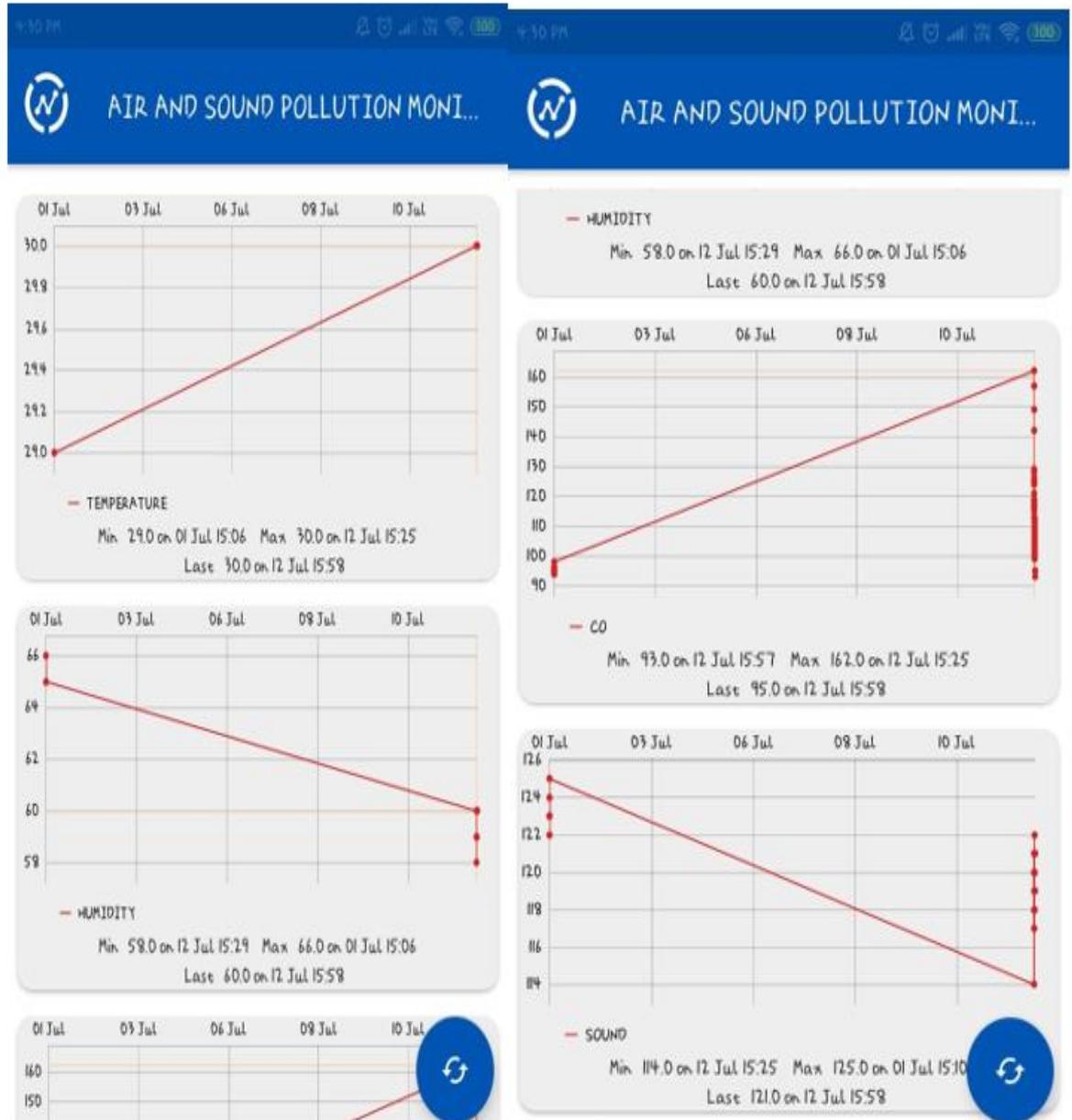


Figure 10: Output of mobile app

5.5 Machine Learning Algorithm Results

After performing the K-NN and Naive Bayes algorithm, air and sound pollution monitoring system predicted results by using train data set and test data set are shown below:

Temperature	30
Humidity	60
CO	95
Sound	121
<input type="button" value="SUBMIT"/>	

Figure 11: Predicted output

The train data set prepared in (excel sheet) air sound. xl is uploaded in the webpage. On applying algorithms with train data and test data used to predict the status pollution present in the environment.

SNO	TEMPERATURE	HUMIDITY	CO	SOUND	KNN	NAIVE_BAYES
13	36	19	48	131	CONTINUOUS NOISE	INTERMITTENT NOISE
14	36	19	48	131	CONTINUOUS NOISE	INTERMITTENT NOISE
15	36	20	54	131	CONTINUOUS NOISE	INTERMITTENT NOISE
16					NO SOUND	ERROR=LIST INDEX OUT OF RANGE
17	30	59	125	118	CONTINUOUS NOISE	CONTINUOUS NOISE
18	30	59	124	118	CONTINUOUS NOISE	CONTINUOUS NOISE
19	30	60	95	121	CONTINUOUS NOISE	CONTINUOUS NOISE
20	30	60	95	121	CONTINUOUS NOISE	CONTINUOUS NOISE
21	30	60	95	121	CONTINUOUS NOISE	CONTINUOUS NOISE

Figure 12: Output from K-NN and Naive Bayes Algorithm

By using K-NN, Naive Bayes algorithms accuracy is calculated for various data sizes as shown below:

Table 1.0 Comparison of K-NN and Naive Bayes Algorithm

Data size	K-NN	Naive Bayes
14	90	70
61	98	75
100	100	75

7. Conclusion

IoT based air and sound pollution monitoring system using machine learning algorithms are designed with analog to digital converter, global service mobile communication, various sensors are interfaced with raspberry pi. The sensors data obtained from the design are uploaded in thinkspeak (IoT) and webpage. This data is compared with trained data by applying algorithms and on comparison, the accuracy is improved by 100% using K-NN algorithm and by 75% using Naive Bayes algorithm.

References

- [1] Sumithra, J. Jane Ida, K. Karthika, S. Gavaskar. (2016). A Smart Environmental Monitoring System Using Internet of Things. *International Journal of Scientific Engineering and Applied Science (IJSEAS)*, Volume-2, Issue-3.
- [2] Giovanni B. Fioccola, Raffaele Sommese, ImmaTufano, Roberto Canonico, Giorgio Ventre, Polluino (2016). An efficient cloud-based management of IoT devices for air quality monitoring. *IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*.

- [3] SRM.ArthiShri, NB. Keerthana, S. Sandhiyaa, P. Deepa, D. Mythili (2017). Noise and Air Pollution Monitoring System Using IOT. *SSRG International Journal of Electrical and Electronics Engineering– (ICETM-2017)*, Special Issue.
- [4] Ren, J., et al. (2009). Naive Bayes classification of uncertain data. *Ninth IEEE International Conference on in Data Mining*.
- [5] Jha, Mukesh, Prashanth Reddy Marpu, Chi-Kin Chau, and Peter Armstrong (2015). Design of sensor network for urban micro-climate monitoring. *First IEEE International Conference in Smart Cities (ISC2)*, pp.1-4.
- [6] Nastic, Stefan, SanjinSehic, Duc-Hung Le, Hong-Linh Truong,and Schahram Dustdar, (2014). Provisioning software-defined IoT cloud systems. *International Conference in Future Internet of Things and Cloud (FiCloud)*, pp.288-295.
- [7] Duraipandian, M., and Mr R. Vinothkanna (2019). Cloud Based Internet of Things for Smart Connected Objects. *Journal of ISMAC*, 1(02), 111-119.
- [8] Valanarasu, M. R. (2019). Smart and Secure IoT and AI Integration Framework for Hospital Environment. *Journal of ISMAC*, 1(03), 172-179.
- [9] Pandian, A. P. (2019). Enhanced Edge Model for Big Data in The Internet Of Things Based Applications. *Journal of trends in Computer Science and Smart technology (TCSST)*, 1(01), 63-73.
- [10] <https://securedstatic.greenpeace.org/india/Global/india/Airpocalypse--Not-just-Delhi--Air-in-mostIndian-cities-hazardous--Greenpeace-report.pdf>
- [11] <content/uploads/2008/04/5v-regulator-using7805.JPG>
- [12] <https://store.arduino.cc/arduino-uno-rev3>
- [13] Shakya, S. (2019). An Efficient Security Framework for Data Migration in a Cloud Computing Environment. *Journal of Artificial Intelligence*, 1(01), 45-53.

Authors Profile



Author: Dr. M. Ramana Reddy, a graduate in ECE from IETE, New Delhi and M.Tech (I&C) from JNTUK, and PhD from JNTUK, Kakinada. He is presently working as an Assistant Professor in the department of ECE, CBIT, and Hyderabad since 2004. His key area of research is VLSI Design in analog and digital systems, and has 12 publications in peer-reviewed international journals and conference proceedings. As a facilitator for continuous learning, he has brought together Faculty, Researchers, students and Industry