International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Novel Computational and Forecasting Strategy For Environment Quality Monitoring using Deep Learning

Shaikh Shakeela, K. Uday Kiran, K. Rajesh Kumar, N. Shravan Kumar, M. Sree Ram Reddy

***Abstract:** Air pollution is the serious issue that one must think about and is caused by harmful gases present in the atmosphere such as Carbon Dioxide, Carbon Monoxide, Sulphur Dioxide etc. since level of pollution varies from one place to another. According to WHO (World Health Organization) air pollution is the fifth major cause for deaths after heart diseases, high blood pressure, poor nutrition and tobacco smoking. Monitoring and detection of the amount of harmful gases over particular area can reduce the chances of endanger to human beings and warn to take precautionary measures and do necessary remedies to regulate the emission of poisonous atmospheric gases. The present paper deals with the monitoring of the disastrous gases using gas sensor which is embedded with NodeMCU. The observed levels sent through internet to cloud platforms using MQTT protocols. The data is stored in the ThingSpeak cloud which can be further analyzed from anywhere in world. Data is processed using Machine learning (ML) algorithm called Long Short-Term Memory Network (LSTM) which is the state-of-the-art technique in the field of data analytics and majorly used for data forecasting.*

***Index Terms-** Gas sensors, NodeMCU, MQTT, LSTM, air pollution monitoring,*

I. INTRODUCTION

Air is essential element for humans and other living organisms like plants, animals for their survival. According to medical standards, pure air is a combination of 78% nitrogen, 21% oxygen, less then 1% of argon, carbon dioxide and other gases. Nowadays air is getting polluted due to several factors and it is one of the major issues in today's environment and majority of the metropolitan cities and urban areas, as they are affected by poor quality of air. Pollution control board indicates that the major causes for pollution are vehicular emissions from cars, trucks etc. and industrial pollutants which cause increase in the level of harmful gases like Sulphur dioxide, methane etc.

Manuscript published on 30 September 2019
* Correspondence Author
Shaikh Shakeela*, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India – 522502
K. Uday Kiran, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India – 522502
K. Rajesh Kumar, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India – 522502
N. Shravan Kumar, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India – 522502
M. Sree Ram Reddy, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India – 522502

However, it is very difficult to monitor the level of pollution over a particular area manually as it consumes lot of time and troubleshooting these issues is much more difficult. In order to solve this issue, a monitoring system using a Machine to Machine communication (M2M) is implemented with the essence of advancements in Internet of Things (IoT) [4]. The parameters involved in the IoT set up are handled and can be accessed through laptops, mobiles and gadgets with internet access in Real Time.

Similar technology is applied for satisfying needs in various domains such as smart homes, transportation, environmental monitoring, control of electrical power, agriculture and others. Sensor technology and its emerging importance has major applications in various fields and helps in environmental monitoring by detecting the amount of harmful gases through low cost and highly efficient sensors [3]. These does not harm the environment with its working nature but safeguards environment from pollution, they gather real time data which is then processed to provide sensible information regarding the polluting gases over the particular area. NodeMCU [8] a low-cost open source platform for IoT and ESP-12E is a microcontroller which works like Arduino with ESP8266 Wi-Fi module. Programming in NodeMCU is simple, flexible and interactive. Most of the IoT applications can be implemented with NodeMCU alone. This chip can be used to interface with sensors and send the data to cloud to monitor the situation for analysing server data. And it has dedicated protocol architecture for IoT data transfer i.e. MQTT (Message Queue Telemetry Transport) [5] which is different from conventional data transfer protocols like TCP/IP. In this project the sensor data is uploaded to ThingSpeak [1] server for analysis. And then comes the most crucial part of data analysis for predicting the future values of the levels of air pollution that can be performed about the complete system architecture and the strategies for air quality monitoring and predicting the future levels of gases in air. Section II deals with the structure of the system and section III is general methodology, focuses on how various technologies are working together to build the complete system which includes ESP8266 module communication with various gas sensors, MQTT protocol and ThingSpeak cloud and finally incorporating LSTM model for forecasting [10]. Section IV gives the experimental analysis and results for the proposed method. And finally, section V will have conclusions and future scope of the work which could be carried out.

II. ARCHITECTURE

NodeMCU is an open source IoT application platform that runs on ESP8266 Wi-Fi Systems, and equipment that depends on the ESP-12 module. It aims to provide with low-cost and easy manner for beginners and specialists to create gadgets that interact with their surroundings using sensors and actuators. These are commonly used in the areas of simple robots, thermostats and movement detectors. The availability of different types of gas sensors helps to measure the corresponding gas levels in the environment.

In the present work MQ series sensors, MQ-135[2] is integrated with NodeMCU [7] for monitoring the pollution level as it possess a wide detecting scope,highly sensitive nature and suites in the long run for detection of toxic gases like carbon dioxide, smoke, ammonia, benzene etc. These are electro chemical sensors that generally possess a small heater inside the sensors which is sensitive in detecting gases of different ranges according to their concentration in the outdoor environment and the recorded values of sensors are noticed in NodeMCU. This firmware utilizes MQTT protocol [6]. MQTT is crucial protocol which can be ideal for interconnecting physical world to real world. The use of MQTT as a communication protocol can save power and bandwidth as It's a lightweight protocol that possess a small message size. MQTT architecture using Publish / Subscribe is more suitable for use in IoT than other protocols that use Request / Response because the client on MQTT does not require a request update, that results in saving bandwidth as well as increases battery life of the device.

In this proposed model the sensor data will be preserved in ThingSpeak cloud and further analysis can be performed by machine learning algorithms made available in various cloud platforms as SaaS (Software as a Service) or PaaS (Platform as a Service) so in cloud itself the analytics can be made automatic for real time monitoring so human intervention all the time doesn't required. Platforms like Microsoft Azure, IBM Watson IoT, Amazon Web Services providing services and packages to access, and build IoT project with ease.

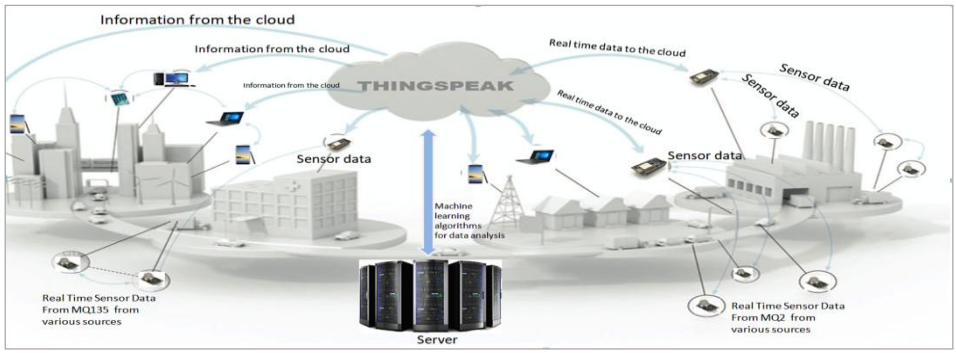


Fig. 1 A complete architecture of Air Quality Monitoring System

III. METHODOLOGY

A. NODEMCU AND MQTT PROTOCOL

The uniqueness of employing IOT in Machine to machine communication(M2M) is that there are various gas sensors present and a single storing device is present. MQTT[6] protocol gathers information from NodeMCU publishing unit and examines in such a way that the message is sent successfully, and similar process is carried out in between every sender and receiver devices until the agreement is cancelled by which it states that the subscriber does not need to participate in the transmission process. This mode is highly benevolent for low power systems as it less expensive and dependable. MQTT processing involves a sequential number of steps out of which MQTT client and MQTT broker[9] plays a deciding role of the protocol.

MQTT client: a device when subjected as client in MQTT receives data that varies from micro controller to server. In order to receive data, the client must be subscribed to specific topic and a library has to be setup in any network. The library functions may include codes written in C, C++, IOS etc.

MQTT Broker: Broker is the central entity as it handles the communication between MQTT clients and distributing messages. It possess a capability of handling multiple number of clients at once and has various tasks to implement and authorization of clients is one such example.

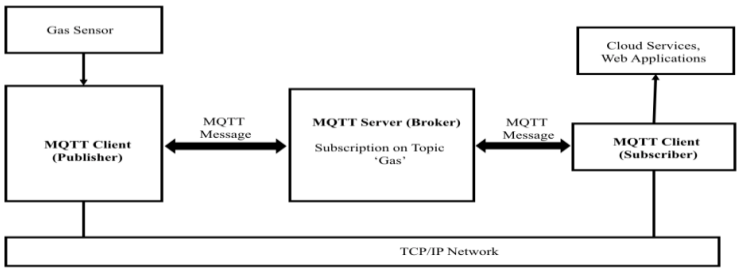
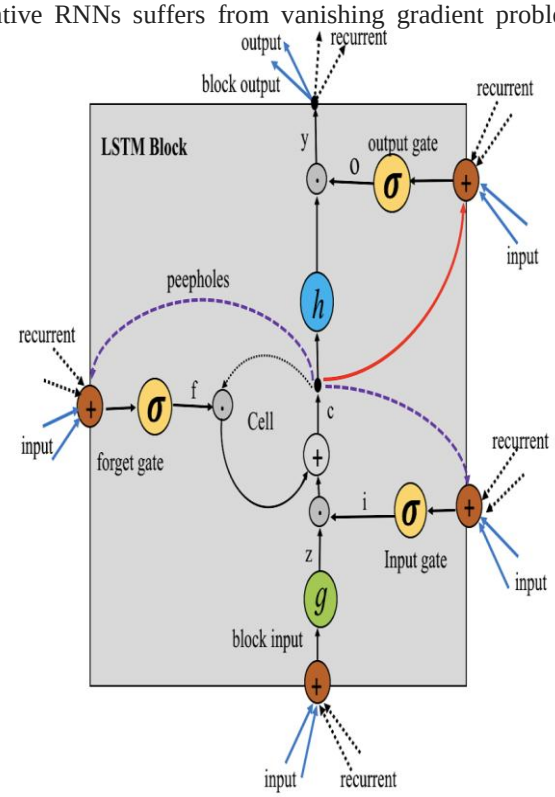


Fig. 2 MQTT protocol model

B. LONG SHORT-TERM MEMORY NETWORK (LSTM)

IoT platform will facilitates the data to pass through network and it has to land on the proper destination for further analysis. The proposed work includes forecasting, so data channelling should be towards the platform where various computations mechanisms can be employed without being weigh down the complete system. The conventional computational strategies are complex, cumbersome and requires high end processors, the alternate way to do the same task with more ease and flexibility is to have the leverage of recent technologies like Machine Learning (ML) algorithms. ML have wide range of models, for time series analysis especially LSTMs performs aptly among all of the models. Unlike Feed Forward Networks (FFN), LSTMs have memory and states are recurrent. The native RNNs suffers from vanishing gradient problem so LSTM is the modified version of it [11]. A single node of LSTM consists of various other units. Basic LSTM block consisted of input, output gates and multiple cells and depends on the necessity of the experiment other units like input gates, input activation function or bias units were included or discarded. Backpropagation through time and recurrent learning was used for training of the network. The LSTMs modified over time and currents equipped with some more extra units those Forget gate and Peephole connections. Forget gates resets the states for LSTMs and allow recurrent learning. Peepholes are responsible for precise timings to be learned properly by LSTMs [12]. Learning in LSTM involves forward pass and backpropagation over time. In proposed work data for the processing generated by the gas sensor MQ135 and passed to the network using NodeMCU, all the variations in gas (CO_2) level fed to the LSTM network for forecasting.



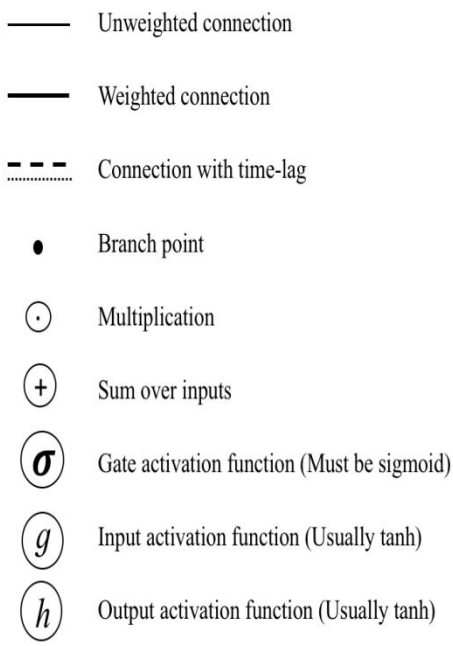


Fig. 3 LSTM Block

This network incorporates following two propagations in order to train the network. Let x^t be the input data vector at time t, $\mathcal{M}$ is the number of inputs and $\mathcal{N}$ is the number of LSTM blocks, the input weights $\mathbf{W}_z, \mathbf{W}_o, \mathbf{W}_f, \mathbf{W}_o \in \mathbb{R}^{N\times M}$. And recurrent weights are $\mathcal{R}_z, \mathcal{R}_s, \mathcal{R}_f, \mathcal{R}_o \in \mathbb{R}^{N\times N}$, Peephole weights $\mathcal{P}_z, \mathcal{P}_s, \mathcal{P}_f, \mathcal{P}_o \in \mathbb{R}^{N}$, bias weights $\mathbf{b}_z, \mathbf{b}_s, \mathbf{b}_f, \mathbf{b}_o \in \mathbb{R}^{N}$. The forward propagation expressions for the LSTM layer as follows:

$$\hat{z}^t = \mathbf{W}_z x^t + \mathcal{R}_z y^t + \mathbf{b}_z$$

block input, $z^t = g(\hat{z}^t)$

$$\hat{i}^t = \mathbf{W}_i x^t + \mathcal{R}_i y^{t-1} + \mathcal{P}_i \quad c^{t-1} + \mathbf{b}_i$$

input gate, $i^t = \sigma(\hat{i}^t)$

Gate uses sigmoid whereas tanh generally used by input and output activation functions.

$$\hat{f}^t = \mathbf{W}_f x^t + \mathcal{R}_f y^{t-1} + \mathcal{P}_f \quad c^{t-1} + \mathbf{b}_f$$

forget gate, $f^t = \sigma(\hat{f}^t)$

cell, $c^t = z^t \quad i^t + c^{t-1} \quad f^t$

$$\hat{o}^t = \mathbf{W}_o x^t + \mathcal{R}_o y^{t-1} + \mathcal{P}_o \quad c^{t-1} + \mathbf{b}_o$$

output gate, $o^t = \sigma(\hat{o}^t)$

block output, $y^t = h(c^t) \quad o^t$

Where σ, g and h are non-linear activation functions which are sigmoid and hyperbolic tangent.

Backpropagation over time done by finding the gradients in other words calculating the deltas of each back pass over all LSTM blocks.

$$\delta y^t = \Delta^t + \mathcal{R}_z^T \delta z^{t+1} + \mathcal{R}_i^T \delta i^{t+1} + \mathcal{R}_f^T \delta f^{t+1} + \mathcal{R}_o^T \delta o^{t+1}$$

$$\delta o^t = \delta y^t \quad h(c^t) \quad \sigma'(o^{-t})$$

$$\delta c^t = \delta y^t \quad o^t \quad h'(c^t) + \mathcal{P}_o \quad o^t + \mathcal{P}_i \quad \delta i^{t+1} + \mathcal{P}_f \quad \delta f^{t+1} + \delta c^{t+1} \quad f^{t+1}$$

$$\delta f^t = \delta c^t \quad c^{t-1} \quad \sigma'(\hat{f}^t)$$

$$\delta i^t = \delta c^t \quad z^t \quad \sigma'(\hat{i}^t)$$

$$\delta z^t = \delta c^t \quad i^t \quad g'(\hat{z}^t)$$

Here, Δ^t is the delta vector which is passed downward from higher layers. Suppose E is loss function and it is usually $(\partial E / \partial y^t)$, and there are no recurrent dependencies. So, deltas of inputs are computed as:

$$\delta x^t = W_z^T \delta z^t + W_i^T \delta i^t + W_f^T \delta f^t + W_o^T \delta o^t$$

And gradients of weights for the back pass can be calculated for $\{z, i, f, o\}$ and $\langle *_1, *_2 \rangle$ are outer product of two vectors and given as:

$$\delta w_* = \sum_{t=0}^{T} \delta *^t, x^t \qquad \delta p_i = \sum_{t=0}^{T-1} c^t \quad \delta i^{t+1}$$

$$\delta R_* = \sum_{t=0}^{T-1} \delta *^{t+1}, y^t \qquad \delta p_f = \sum_{t=0}^{T-1} c^t \quad \delta f^{t+1}$$

$$\delta b_* = \sum_{t=0}^{T} \delta *^t \qquad \delta p_o = \sum_{t=0}^{T} c^t \quad \delta o^t$$

IV. EXPERIMENTATION AND RESULT ANALYSIS

The proposed work concentrates on obtaining the air quality levels from atmosphere by MQ135 sensor and sending the levels to distant peers with NodeMCU and MQTT protocols over the network. Further analysis of the levels comprises Deep learning model LSMT[13], for predicting the air quality levels. Beginning of the algorithm divides the data into two sets, training and test. The algorithm works on 1191 levels or in other words continuous time steps, shown in fig. 4 (a) and further divided as 900 levels for training and remaining 291 levels for testing the model. Before feeding the data to LSTM it has to be prepossessed to avoid anomalies and outlier values. Then the time series data will be made ready for supervised learning by making the single variable data for next time step value label. So univariate analysis performed on LSTM supervised learning and data after labelling are shown in fig. 4 (b) and fig. 4 (c), values are then scaled between (-1, 1) fig. 4 (d) and fig. 4(e). Now the data trained in LSTM network with the batch size of 1, number of epochs are 100 and number of neurons 500. Model fitting on the data should be performed to predict the future values of the air quality level after invert scaling the data. The predicted values against test data levels are shown in fig. 4(f) and 4(g). And results shows that models gives good performance on test data and gives RMSE (Root Mean Square Error) as 15.716. The resulted error from experiment can be reduced by training the model with huge data set and parameters should be modified.

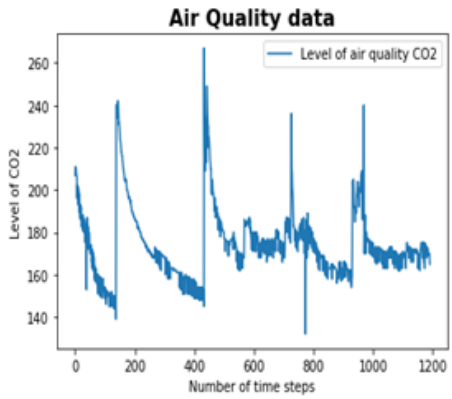


Fig. 4(a)

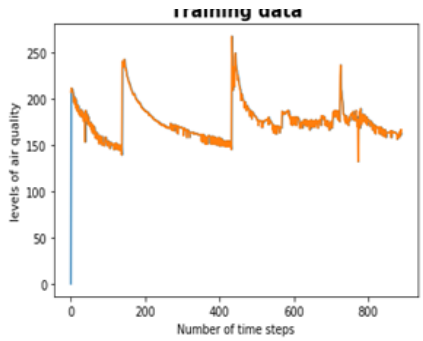


Fig. 4(d)

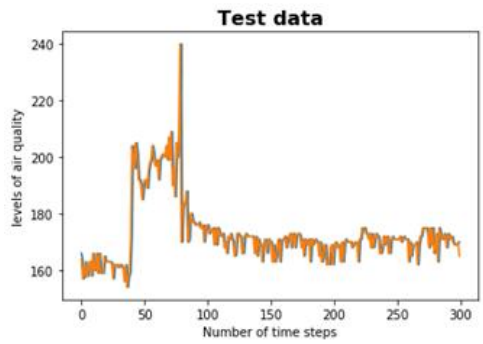


Fig. 4(c)

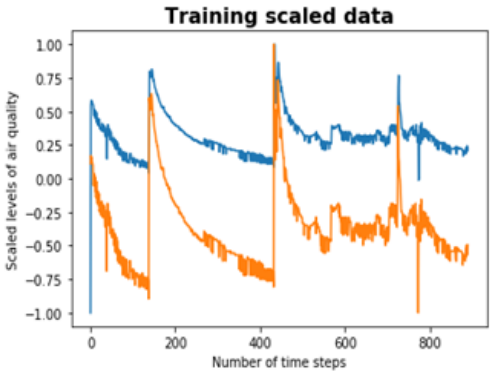


Fig. 4(d)

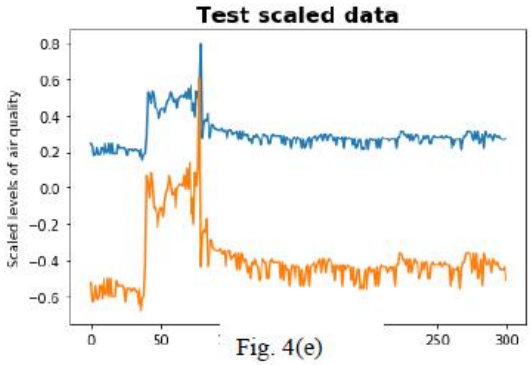


Fig. 4(e)

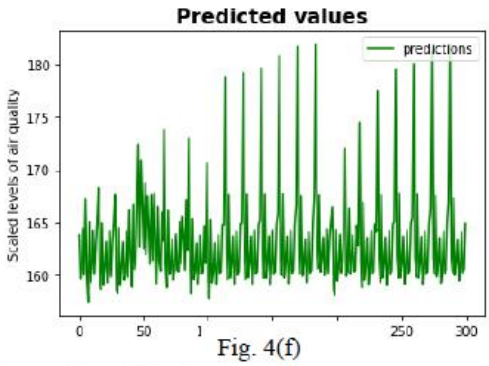


Fig. 4(f)

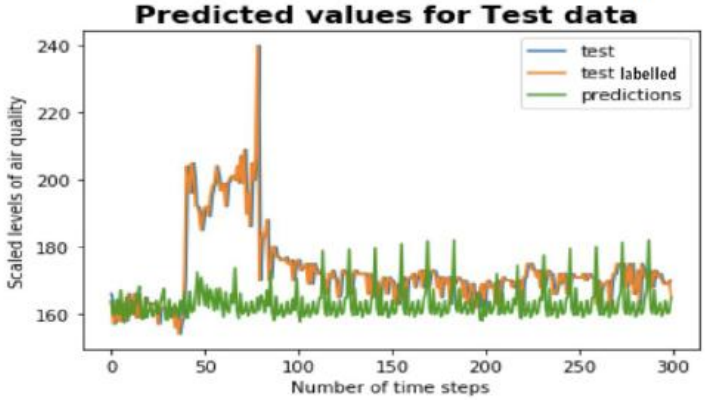


Fig. 4(g)

Fig. 4 Results (a) Air quality levels (b) training data (c) test data (d) Scaled training data (e) Scaled test data (f) Predicted air quality levels (g) Test vs Predicted air quality levels

IV. CONCLUSION AND FUTURE SCOPE

Paper focused on developing an air pollution monitoring system of an area and monitoring the values of toxic gases present and integrated model data in cloud platform through NodeMCU which avails the data accessibility anytime and can be analysed from anywhere whenever necessary. This analysis can help in taking preventive measures to minimize polluting gases in the environment. Since smart phones have revolutionized the human lives, which can be integrated to the sensors such that the data analysis can be implemented in mobiles and everyone would be aware of the amount of pollution caused and would help in Controlling the pollution.

REFERENCES

1. Meena S., Rajan M. A., Shivraj V. L. and Balamuralidhar P "Secure MQTT for Internet of Things (IoT)", International Conference on Communication Systems and Network Technologies, IEEE Computer Society, 2015.
2. Lamling, P. Rajalakshmi, Ayo Afonja, G. McPhillips, Russell B., Lawrence Cheng, Stephen H., "On the Development of a Sensor Module for real-time Pollution Monitoring", IEEE, 2011.

3. Nagaraj S., Rajashree V. Biradar, “Applications of Wireless Sensor Networks in the Real-time ambient air pollution monitoring and air quality metropolitan cities” IEEE, 2017.
4. Tejaswini roy c., Sri lakshmi D., Anirudh kumar G., Vishwas H. N. “Smart environment using IOT”, IEEE, 2017.
5. M. H. Asghar, N. Md. Zadeh, “Design and simulation of energy efficiency in node based on MQTT protocol in IOT”, IEEE, 2015.
6. P. Alqinsi, N. Ismail, I. Joseph, M. Edward, Wahyudin D., “IOT-based ups monitoring system using MQTT protocol”, IEEE,2018.
7. S. Chanthakit, C. Rattanapoka, “MQTT based air quality monitoring system using NodeMCU and Node-RED”, IEEE,2018.
8. Madhu G. M, C. Vyjayanthi, “Implementation of cost-effective smart home controller with arduino application using MCU and Internet of things (IOT)”, IEEE, 2019.
9. Kresimir G, Ivan S, Ivan H, "A Web Based IOT Solution for Monitoring Data Using MQTT Protocol”, IEEE, 2016.
10. Weicong Kong, Zhao Yang Dong, Youwei Jia "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network" IEEE, 2019.
11. Weishan Zhang, Wuwu Guo, Xin Liu, Yan Liu, Jiehan Zhou, Bo Li, Qinghua Lu, And Su Yang, "LSTM-Based Analysis of Industrial IoT Equipment", IEEE, 2019.
12. Shengzhe Dai, Li Li, Zhiheng Li, "Modeling Vehicle Interactions via Modified LSTM Models for Trajectory Prediction", IEEE, 2019.

AUTHORS PROFILE

Shaikh Shakeela She has completed her Master’s degree from VIT, Vellore, India in communication Engineering and presently working as assistant professor in the department of ECE at KLEF, Guntur. She is working in the area of Signal Processing, Machine Learning and Time Series Analysis.

Uday Kiran Kasi He completed his Master’s degree from NIT, Kurukshetra and presently working as assistant professor in the department of ECE at KLEF, Guntur. He is working in research area of Seismic Signal Processing and Advance Statistical Signal Processing GPSTEC for Seismic Precursors.

K. Rajesh Kumar He is currently pursuing his Bachelor degree in the department of ECE at KLEF, Guntur. He is working in the area of Internet of Things (IOT), Communication systems.

N. Shravan Kumar He is currently pursuing his Bachelor degree in the department of ECE at KLEF, Guntur. He is working in the area of communication and protocols.

M. Sreeram Reddy He is currently pursuing his Bachelor degree in the department of ECE at KLEF, Guntur. He is working in the area of communication and protocols for this project.