IoT based Machine Learning Techniques for Climate Predictive Analysis

M.K. Nallakaruppan, U. Senthil Kumaran

Abstract: ***The continuous research in the fields of Internet of Things and Machine Learning has offered ascend to various weather forecast models. However, the issue of precisely foreseeing or anticipating the weather still perseveres. This paper is an application of Internet of Things and Machine Learning algorithms like Decision Tree and Time Series Analysis. The Internet of Things actually signifies 'things' (e.g. sensors and other shrewd gadgets) which are associated with the web. Despite the fact that this may appear to be irrelevant, 'things' represent a new and progressively, critical foundation requiring their own particular devoted technological system. The obtained results from the Machine Learning demonstrated that the time series method forecasts the weather more accurately for a larger duration of time***

Keywords: Weather Prediction, Machine Learning, Internet of Things, Decision Tree, Support Vector Machines, Time Series.

I. INTRODUCTION

Weather forecasting implies foreseeing the weather conditions (states of atmosphere) of a specific given territory or area. All the more critically, exact climate forecast is essential to seek after every day exercises. Living and non-living things are reliant on weather forecasts.

Indeed, even following quite a while of weather forecasting, the climate industry in India is still in its underlying stage, confronting numerous snags. One of the real deterrents that weather forecasting faces is the subjective and ill-suited desires from the nature.

An automated weather station is a gadget that is utilized to gauge and record the known parameters of atmosphere without association of people [1][2]. The weather conditions are required to be checked to keep up the sound development in crops and to guarantee the protected workplace in enterprises, and so forth. Because of technological development, the way towards reading the natural parameters wound up less challenging contrasted with the past days. The sensors are the scaled down electronic gadgets used to gauge the physical and ecological parameters. By utilizing the sensors for checking the weather conditions, the outcomes will be exact and the whole framework will be faster and less power devouring.

The framework proposed in this paper portrays the executed stream of the weather checking station. It incorporates the wireless communication technology IEEE 802.11 b/g (Wi-Fi) for communication. The framework screens the weather circumstances and updates the data to the remote database. The purpose for sending the information to a remote database is that the weather conditions of a specific place can be known from anyplace on the planet. The framework comprises of rain measure, temperature, humidity and pressure sensors. Every one of these sensors measures the corresponding weather parameters. The framework is planned to use in extensive residential buildings and manufacturing firms.

Machine learning is the capacity of computer to learn without being expressly customized. It enables machines to find concealed patterns and insights. In supervised learning, we assemble a model in view of training information. The model is then utilized for mapping new illustrations. In view of the watched weather patterns from the past, a model can be constructed and used to predict the weather.

Time Series Analysis [16][17] and Decision Tree[12][13] models are widely used prediction models. Mostly, these models take the present weather conditions and process it to build a model for predicting the weather.

This research work centers around comprehending the weather prediction inconsistencies and in-proficiency in light of linear regression algorithms and time series model. The significant commitment of this examination work is to formulate a productive weather prediction model based on Decision Tree and Time Series Analysis.

II. RELATED WORK (LITERATURE SURVEY)

In this section, an analysis is carried with the current weather prediction strategies accessible in the literature. Linear Regression is the most fundamental and regularly utilized prescient model for investigation. Regression estimates are for the most part used to depict the information and illustrate connection between at least one independent and dependent factor. Linear regression finds the best-fit through the points, graphically. The best-fit line through the focuses is known as the regression line. Here, the line can be straight or curved relying upon the data. The best-fit line can likewise be a quadratic or polynomial which gives us better solution to our inquiries. Two of the algorithms used as a part of this research are Decision Tree and Time Series Analysis

Weather prediction has been a major challenge from early days; new methodologies cluster everyday replacing the old ones. Literature studies have shown that machine learning techniques achieved better performance than traditional statistical methods. The next wave in the era of computing will be outside the realm of the traditional desktop. In the Internet of Things (IoT) paradigm, many of the objects that surround us will be on the network in one form or another. Machine Learning is closely related to internet of things.

Manuscript received January 25, 2019.

M.K. Nallakaruppan, Assistant Professor, Vellore Institute of Technology, Vellore, India.

U. Senthil Kumaran, Associate Professor, Vellore Institute of Technology, Vellore, India.

A perfect combination of them can promote fast development of agricultural modernization, realize smart agriculture and effectively solve the issues concerning agriculture, countryside and farmers.

III. ARCHITECTURE

Utilizing the accessible open source platforms in terms of specification, cost and development devices it was watched the best accessible gadget is the Raspberry PI. In this project Raspberry PI 3 is utilized as the principle base platform for the project. The atmosphere parameters must be transmitted to the gadget and after that to the MySQL server for storing the values which are later used as input dataset for applying Machine Learning algorithms.

Wi-Fi is utilized as the medium through which the information is sent from the Raspberrypi to the server. Where the user can see the information in the MySQL server and download the information and apply Machine Learning Techniques.

The weather forecasting system consists of the following components:

- Raspberry Pi 3 Model B
- DHT-11 Temperature & Humidity Sensor
- Rainfall Sensor
- BMP-180 Pressure Sensor
- Wireless Access Adapter

For the remote transmission of the information through WiFi use of remote access connector and 802.11n WiFi connector is utilized. The server which is given by db4free.net is utilized for the storage of sensor information from which the user can get the information globally for Machine Learning purposes.

The following is the block diagram of the proposed weather forecasting system in Figure-1

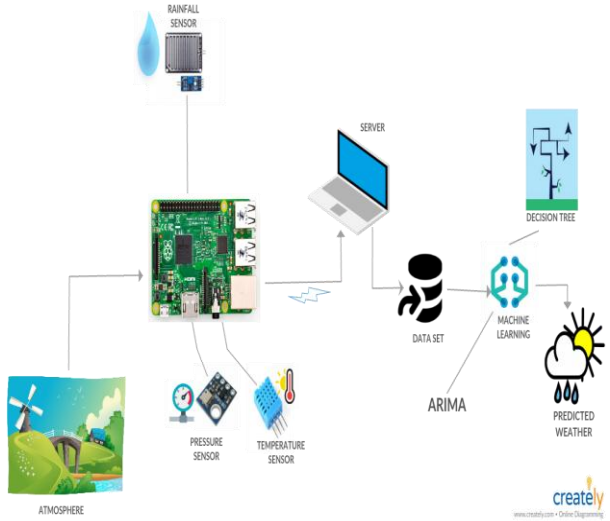


Figure 1: Architecture

A. Raspberry Pi

Raspberry PI is a card-sized ARM fueled Linux PC development board[1][4][6]. There is altogether of 5 sorts of different board with various specifications, for the proposed Weather forecasting framework Raspberry PI 3 Model B is utilized as the fundamental development board which is appeared in Figure-2.

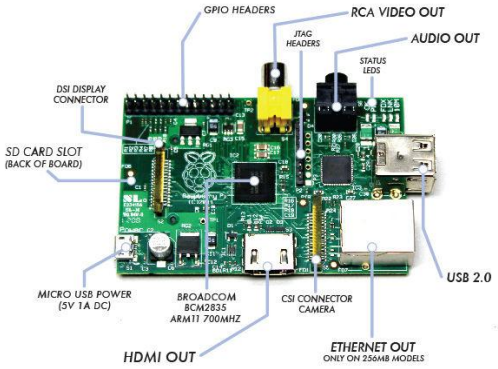


Figure 2: Raspberry PI model 3 board

The raspberry pi comprises of four USB Ports and one 10/100 Base T Ethernet Socket. Forty pins GPIO Header are available in the raspberry pi board which is utilized for associating with Analog to Digital converter chip (MCP3008) to which the sensors are associated. A 5V smaller scale USB control port is available to which the power supply is given for the gadget. A HDMI port is available through which interfacing of the screen and the Raspberry pi should be possible and the USB ports for the keyboard and mouse interfacing. At the last a Micro SD Card Slot is given where the Micro SD card is to be embedded with the Raspbian OS.

B. MCP3008

The MCP3008 is the Analog to Digital converter which is a 16 pinned and 8 channeled chip which changes over the analog voltages to 10 bit paired code.[1][5][6] The chip is associate with the external clock which is given by the Raspberry pi. The ADC chip is given a working voltage (VDD) of 3.3V and voltage reference (VREF) the same as (VDD). The chip has two pins for the input/yield for advanced information transmission. The chip gives 8 channels (CH0-CH7) to which 8 analog sensors can be associated at the same time.

C. Sensors

Sensors are utilized for the discovery of the different parameters in the atmosphere for the forecasting of the weather conditions over and over. The following are the depiction of the sensors in detail.

- **DHT-11**-It comprises of a humidity detecting segment, a NTC temperature sensor (or thermistor) and an IC on the rear of the sensor.[2][4][5] The humidity detecting segment has two anodes with dampness holding substrate between them. So as the humidity changes, the conductivity of the substrate changes or the protection between these electrode changes which are estimated and prepared by the IC and humidity value is computed. As the temperature builds the NTC thermistor resistance diminishes resulting in the expansion of output voltage which at that point prepared by the IC what's more, the temperature value is ascertained.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

- **Rain Sensor**-The rain sensor module is a simple instrument for rain identification. It can be utilized as a switch when raindrop falls through the sprinkling board and furthermore for estimating precipitation intensity. The module includes, a rain board and the control board appended for additional comfort, power LED. The analog output is utilized as a part of discovery of drops in the measure of rainfall. Associated with 3.3V/5V power supply and the sensor works in light of the level of the water interfacing the rain board, the output voltage of the gadget varies on the length of the rain board being wet which is changed over to digital through ADC chip.
- **BMP180-** It is a barometric pressure sensor with an I2C ("Wire") interface. Barometric pressure sensors measure the absolute pressure of the air around them. This pressure varies with both the weather and altitude. Depending on how you interpret the data, you can monitor changes in the weather, measure altitude, or any other tasks that require an accurate pressure reading.

A modem is connected to raspberry pi 3 board which connects to the wireless local areanetwork (WLAN) which is provided by the Wi-Fi routerwhich provides the internet to the Raspberry PI.

IV. SOFTWARE DEVELOPMENT

A. Coding the Raspberry pi

Python Language is used to code the raspberry pi. The code once executed gets data from all the sensors connected to the raspberry pi and stores the data in MySQL server. Python is used to interact with the sensors and manipulate the data received from them.

B. Interfacing with Server

For the system a data base is created on an external server called "db4free.net" where the data sent by the Raspberry pi is stored periodically in the server's data base.

V. EXPERIMENTAL SETUP

The prototype setup of the system is shown in Figure-3 connected all components on the Bread Board.

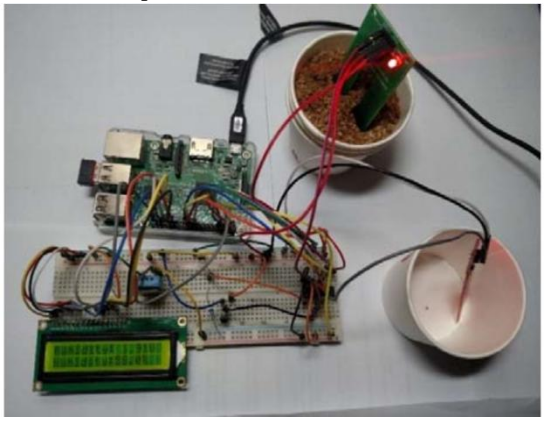


Figure 3: Prototype of the weather forecasting system.

The expected result of the model is all sensors appropriately working and identifying the parameters precisely and exchanging the information to raspberry pi which is sent out to the server through the Wi-Fi. The Figure-8 demonstrates the identification of the sensors and the values showing on raspberry pi which is sent out to the server through the Wi-Fi.

temperature	humidity	rainfall	pressure	date_time
27	80	0	1013	2018-03-13 23:43:23
27	80	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
25	83	0	1013	2018-03-13 23:43:23
28	83	0	1013	2018-03-13 23:43:23
28	83	0	1013	2018-03-13 23:43:23
28	83	0	1013	2018-03-13 23:43:23
28	83	0	1013	2018-03-13 23:43:23
28	83	0	1013	2018-03-13 23:43:23
28	83	0	1013	2018-03-13 23:43:23
28	83	0	1013	2018-03-13 23:43:23
28	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
25	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23
25	83	0	1013	2018-03-13 23:43:23
27	83	0	1013	2018-03-13 23:43:23

Fig. 4: Data recorded on the server

VI. MACHINE LEARNING METHODOLOGY

This project uses an application (IDE) called 'R-studio' for programming. The programming language used is 'R-language'. Two algorithms have been applied on the dataset created using Raspberry pi namely Decision Tree and Time Series Analysis.

Three sorts of weather parameters are predicted:

Minimum Temperature, Maximum Temperature and Mean Temperature. Temperature is the measure of hotness or coldness, for the most part estimated utilizing DHT-11. Units of temperature most of the time utilized is Celsius.

Table I demonstrates data from the training set, an enormous dataset is fed to the model. Parameters are trained and they respond to the model. At last, one might say that the machine has been trained.

Table I: Sample Training Data

DATE	MAX TEMP	MIN TEMP	MEAN TEM	HUMIDITY	RAINFALL	PRESSURE
1-Feb	31	17	26	65	0	1013
2-Feb	33	18	26	69	0	1013
3-Feb	31	18	26	65	0	1015
4-Feb	31	20	25	59	0	1013
5-Feb	30	20	25	65	0	1014
6-Feb	32	21	26	77	0	1013
7-Feb	32	23	27	79	0	1014
8-Feb	33	24	27	77	1	1014
9-Feb	32	22	26	78	0	1015
10-Feb	32	23	27	75	0	1014
11-Feb	32	22	26	75	0	1014
12-Feb	32	23	27	73	0	1013

VII. DECISION TREE

A decision tree regularly utilized in light of the fact that it is easy to comprehend and translate. Classification and Regression Trees - CART - is a procedure shaped by an accumulation of rules in view of estimations of specific variables in the modeling data. This project displays a small application of CART for whether prediction.[12][15]

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Decision trees models are normally utilized as a part of information mining to look at the information and to induce the tree and its rules that will be utilized to make forecasts. Various diverse calculations might be utilized for building decision trees including CHAID (Chi-squared Automatic Interaction Recognition), CART (Classification And Regression Trees), Quest, and C5.0.[12][13]

In this project, we use CART to build a regression tree to predict the temperature values around VIT-Vellore.

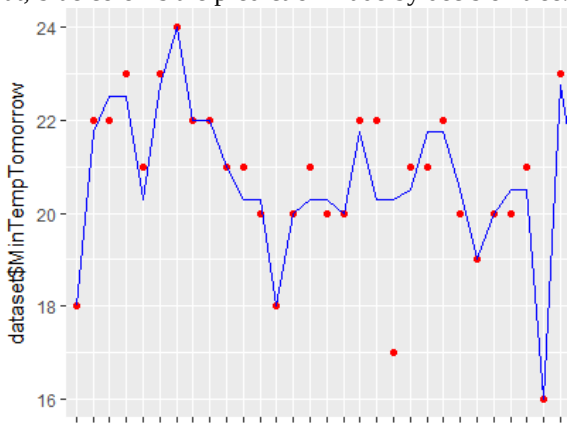
Hypothesis for our proposed model can be shown by the Equation 1:

Tomorrow's Temperature ~ (Today's Temperature + Today's Humidity + Today's Rainfall + Today's Pressure).(1)

VIII. RESULTS

The evaluation and simulation is carried out in R-studio. The results of the decision tree algorithm show very little deviation from the actual output. It can be deduced from the graphs that the algorithm is very accurate.

In the following figures, red color shows the actual output, blue color is the prediction made by decision tree.



x-axis : DATE

Fig. 4: Actual Output V Predicted Output

IX. TIME SERIES ANALYSIS - ARIMA

In real world applications, numerous processes can be represented utilizing the time series as follows:

X(t-p),…x(t-2),x(t-1),x(t)

For making a prediction using time series, a great variety of approaches are available. Prediction of scalar time series refers to the task of finding estimate of next future sample x(n+1) based on the knowledge of the history of time series, i.e. samples x(n), x(n-1)[17][16]

Traditionally, a time series forecasting problem is handled utilizing linear techniques such as Auto Regressive Moving Average (ARMA) and Auto Regressive Integrated Moving Average (ARIMA) models popularized by Box and Jenkins.

In practice many time series are non stationary and so we cannot apply stationary AR, MA or ARMA processes directly. One possible way of handling non-stationary series is to apply differencing so as to make them stationary. The first differences, namely (Xt − Xt−1) = (1 − B)Xt, may themselves be differenced to give second differences, and so on. The dth differences may be written as (1−B) dXt. If the original data series is differenced d times before fitting an ARMA (p, q) process, then the model for the original indifferences series said to be an ARIMA(p, d, q) process where the letter 'I' in the acronym stands for integrated and d denotes the number of differentiations applied.[16]

The following figure shows the dataset after defferentiating the logarithm of the dataset. This is done so that the entire data set has constant mean difference and standard deviation.

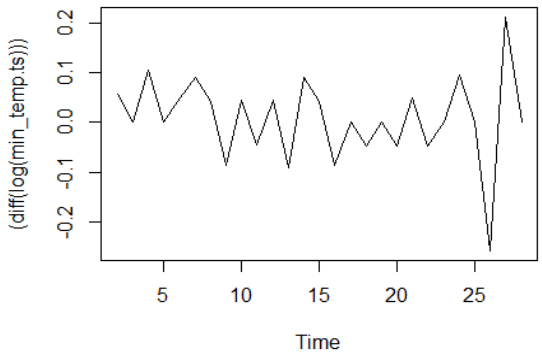


Fig. 5: Time V Diff(Log(Series))

ARIMA model can only be applied if the mean difference and standard deviation is constant at each interval of time.The number of times differentiated gives us the p value , here p = 1 since the dataset is differentiated only once.

The function Acf computes (and by default plots) an estimate of the autocorrelation function of a (possibly multivariate) time series. Function Pacf computes (and by default plots) an estimate of the partial autocorrelation function of a (possibly multivariate) time series.

The following graph gives the autocorrelation values of the time series.

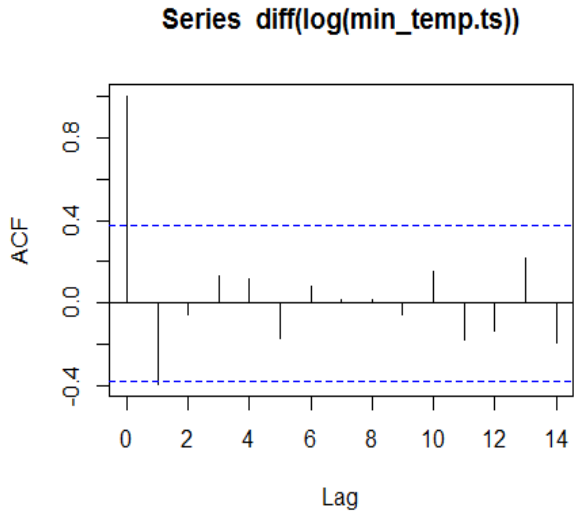


Fig. 5: Lag V ACF

The value of q is the index of the first line before the first inverted line which is 0 in this case so q=0.The following graph gives the partial autocorrelation values of the time series.

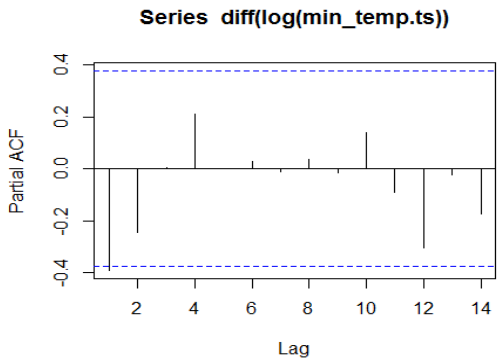


Fig. 7: Lag V PCF

The value of d is the index of the first line before the first inverted line which is 0 in this case so d=0.Hypothesis for our proposed model can be shown by the Equation 2:

Model = arima(log(min_temp.ts) , c(0,1,0) Where p = 0, q=0, d=1…….(2)

X. RESULTS

The evaluation and simulation is carried out in R-studio. The results of the ARIMA algorithm show very little deviation from the actual output. It can be deduced from the graphs that the algorithm is very accurate.

ARIMA model predicts the mean temperature of next 5 days using the formula in (2)

In the following figures, red color shows the actual output, blue color is the prediction made by ARIMA model.

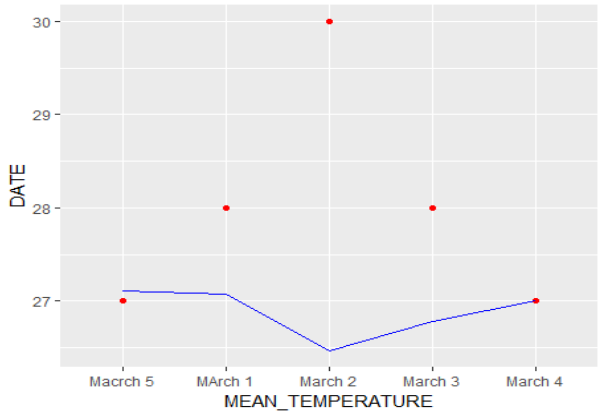


Fig. 8: Actual Output V Predicted Output

XI. CONCLUSION

The main motto was to use the inexpensive components and attain to maximum best accurate system which couldmonitor the weather in real time application in agriculturallands and use this data to make more accurate future weather predictions using machine learning algorithms. Compared to the Decision Tree, ARIMA can more efficiently capture dynamic behavior of the weather temperature, resulting in a more compact and natural internal representation of the temporal information contained in the weather profile.

Weather prediction has been a major challenge from early days, new methodologies cluster everyday replacing the old ones. The next wave in the era of computing will be outside the realm of the traditional desktop. In the Internet of Things (IoT) paradigm, many of the objects that surround us will be on the network in one form or another. Machine Learning is closely related to internet of things. A perfect combination of them can promote fast development of agricultural modernization, realize smart agriculture and effectively solve the issues concerning agriculture, countryside and farmers.

REFERENCES

1. Deshmukh A. D. &Shinde U. B. 2016, August. A low cost environment monitoring system using raspberry Pi and arduino with Zigbee. In: Inventive Computation Technologies (ICICT), International Conference on. 3: 1-6. IEEE.
2. Jindarat S. &Wuttidittachotti P. 2015, April. Smart farm monitoring using Raspberry Pi and Arduino.In: Computer. Communications, and Control Technology (I4CT), 2015 International Conference on . IEEE. pp. 284-288.
3. Savić T. &Radonjić M. 2015, November. One approach to weather station design based on Raspberry Pi platform. In: Telecommunications Forum Telfor (TELFOR), 23rd . IEEE. pp. 623-626.
4. Wang Y. & Chi Z. 2016, July. System of Wireless Temperature and Humidity Monitoring Based on Arduino Uno Platform. In: Instrumentation & Measurement, Computer, Communication and Control (IMCCC), 2016 Sixth International Conference on. IEEE. pp. 770-773.
5. Saini H., Thakur A., Ahuja S., Sabharwal N. & Kumar N. 2016, February. Arduino based automatic wireless weather station with remote graphical application and alerts. In: Signal Processing and Integrated Networks (SPIN), 2016 3rd International Conference on. IEEE. pp. 605-609.
6. Kumar N. P. &Jatoth R. K. 2015 May. Development of cloud based light intensity monitoring system using raspberry Pi. In: Industrial Instrumentation and Control (ICIC), 2015 International Conference on.IEEE. pp. 1356-1361.
7. Srinivasan V. S., Kumar T. &Yasarapu D. K. 2016, May. Raspberry Pi and iBeacons as environmental data monitors and the potential applications in a growing BigData ecosystem. In: Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE InternationalConference on. IEEE. pp. 961-965.
8. Ibrahim M., Elgamri A., Babiker S. & Mohamed A. 2015, October. Internet of things based smart environmental monitoring using the raspberry-pi computer. In: Digital Information Processing and Communications (ICDIPC), 2015 Fifth International Conference on. IEEE. pp. 159-164.
9. Folea S. C. &Mois G. 2015. A low-power wireless sensor for online ambient monitoring. IEEE Sensors Journal. 15(2): 742-749
10. Sandeep V., Gopal K. L., Naveen S., Amudhan A. & Kumar L. S. 2015, August. Globally accessible machine automation using Raspberry pi based on Internet of Things. In: Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on. IEEE. pp. 1144-1147.
11. Princy S. E. & Nigel K. G. J. 2015, November. Implementation of cloud server for real time data storage using Raspberry Pi. In: Green Engineering and Technologies (IC-GET), 2015 Online International Conference on . IEEE. pp. 1-4.
12. Chapman, L. & Thornes, J.E. (2011) What resolution do we need for a route-based road weather decision support system? Theoretical & Applied Climatology 104:551-559.
13. Chapman, L. (2012) Probabilistic road weather forecasting. Proceedings of the 16th SIRWEC Conference, Helsinki, Finland, May 2012.
14. Chapman, L. & Thornes, J.E. (2006) A geomatics based road surface temperature prediction model. Science of the Total Environment 360:68-80
15. Mahoney, W.P. & O'Sullivan, J.M. (2013) Realizing the Potential of Vehicle-Based Observations. Bulletin of the American Meteorological Society 94:1007–1018.
16. Box,G.E.P. and G.M.Jenkins, 1976.Time Series Analysis : Forecasting and Control. Holden Day Inc. San Francisco,CA.
17. Cook, D.F. and Wolfe, M.L., 1991, "A backpropagation neural network to predict average air temperatures", AI Applications, 5, p.p.40-46.