2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT)
Manipal University Jaipur, Sep 28-29, 2019

IoT Based Automated Weather Report Generation and Prediction Using Machine Learning

Anubha Parashar
Computer Science and Engineering
Manipal University Jaipur
Jaipur, India
anubha.parashar@jaipur.manipal.edu

***Abstract*— The weather can have great impact on lives. Weather changes can influence wide range of human activities and affect agriculture and transportation. The main aim of this paper is to monitor and report weather conditions so that one is informed beforehand and necessary actions can be taken to reduce the damage by any calamity by forecasting it. Here we are using various sensors in order to collect the data and previous data is used in order to train the system and with current data collection we do the prediction. We will be analyzing temperature, pollutants, humidity and pressure and will predict the weather. Existing models are expensive in contrast to ours and hence it will make monitoring local area feasible as it will be cheaper.**

Keywords—Internet of things, Regression, machine learning, data collection through sensors

I. INTRODUCTION

Day to Day atmospheric variations can be demonstrated by the term weather. All the day to day activities like temperature change, humidity, cloud cover and precipitation can be illustrated by weather monitoring. When the weather data is collected for a period of time, that weather statistics can be used to categories the climate of that locality. The assessment can be done on weekly, monthly or even on daily basis. So we can say that climate is a prolonged average of weather. Weather change plays a significant role in a lot of sectors like agriculture, food security etc.

Weather prediction is becoming more challenging in the modern world. Due to climate shift natural disasters like cyclones, storms, earth quakes, tsunami etc. occur almost on a daily basis and cause destruction on large scale. In these adverse times, correct rainfall prediction and weather monitoring is very important and useful. The countries which are dependent on agriculture, like India, correct rainfall prediction can be helpful in producing a bounty harvest [1].

In literary terms, there are many prediction models for rainfall and weather forecasting. For yearly forecasting or for large scale prediction these models do not provide accurate results due to vigorous climate change [2]. Forecasting is a very important analysis topic as it takes both science and technology to go hand in hand for correct results. Even though there are some very accurate prediction systems but there is no estimation of future damage that could be caused by the calamities.

The estimate of damage can still be accounted and predicted (in numerical terms) by collecting the dataset from Meteorological Departments (data center) for a specific location and conducting regression analysis on the dataset [3].

Hence, a better and more accurate weather monitoring and prediction system is needed which could give the estimates of the damage Caused by natural calamities and could also give a heads up before any natural disaster is about to occur[4].

II. PROPOSED METHODOLOGY

The model will sense different weather parameters like pressure, humidity, rainfall, temperature, dust particles and light etc. and will display current day's temperature on a web portal. The model will also be able to predict the maximum and minimum temperatures of the next day using past three days' data as shown in figure 1.

It has a huge scope in today's world as our current models are expensive and we are providing a cheaper alternative. Also, all the predictions eye a large scale but with our model local area monitoring will be made possible.

Though there are few weather reporting models which already exist but in today's world except for weather, concentration of various components of air and its quality are also gaining importance [5].

Till date there is no such model which can forecast weather, determine various components in air and can check or predict air quality. Existing models of weather forecasting has many mathematical calculations but they are not designed appropriately in order to get higher classification rate and it's not feasible to get the exact local area weather because covering local areas with central area and then the analysis of weather makes the system inaccurate up to 18%.

Previous weather forecasting models used the complicated blend of mathematical instruments which was insufficient to get higher classification rate and also covering local areas was not feasible. Using different components to collect and assess data will be expensive [6].

A. Architecture of Proposed Model

Previous models were used to gather information about the components used and researched about better and cheaper alternatives. All the information about the components needed to develop the desired model (in figure 1.).

978-1-7281-1711-9/19/$31.00 ©2019 IEEE

Post that, we analyzed the techniques previously used in an attempt to develop a similar model and came up with the best approach. A platform to store all the generated data was required. A portal having a decent UI was needed to retrieve the data from the platform and display it. Alongside, a huge database was required to train the model [7].

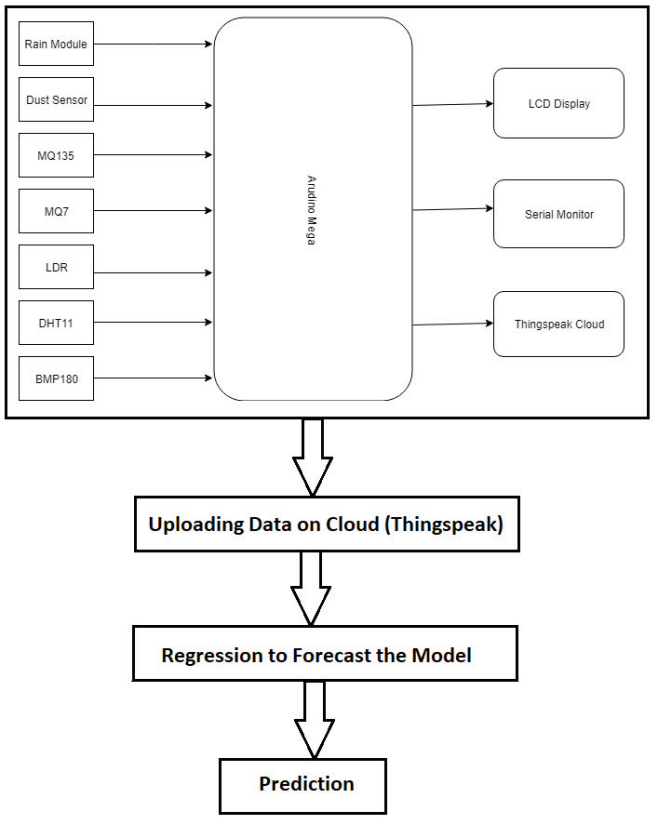

Fig. 1. Architecture of proposed system

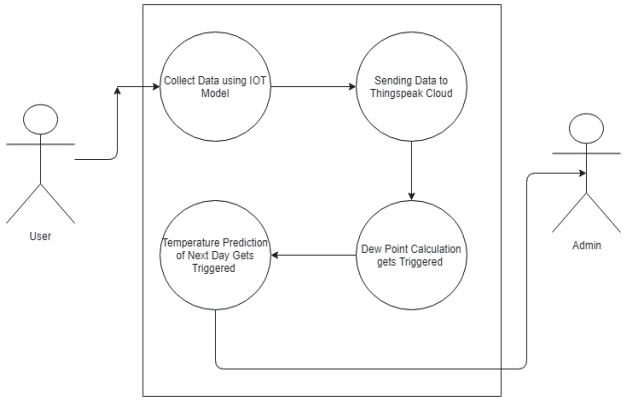

Fig. 2. Proposed system

B. *Proposed System*

Once the need was identified, a model with 8 sensors, display was devised. Machine learning (Linear Regression) using MATLAB analysis was thought of as a desirable approach. Wixsite was considered for developing the portal as it contains thousands of attractive templates. Dataset of past two years was collected and was cleaned to obtain only relevant data (in figure 2).

C. *System Requirements*

The system/model should be able to sense the current temperature, light intensity, pressure, amount of gasses and display it on the LCD display. Alongside, it should also send the data to the cloud from where it can be retrieved and displayed on a web portal. No only this the model should also be able to predict the next day's temperature [8].

Sensors are required to perform the sensing function and sensed data can be showed on the display attached to Arduino board. Internet connection will be required to send the data to the cloud (Thingspeak) from it can be retrieved using APIs [9].

D. *Software Requirements*

- Arduino IDE is used in order to code for Arduino mega in figure 3.

Fig. 3. Code in arduino mega

- Wixsite is made in order to display the data (in figure 17, 18).
- Thingspeak cloud platform is used in order to upload the data from sensors and which can be further used for analysis and prediction [10].
- MATLAB is used to make the model for prediction. Regression technique is used in this paper.
- Jupyter Lab is used to run all the machine learning algorithms.

E. *Hardware Requirements*

- **Wi-Fi module (ESP 8266):** ESP8266 is a chip that provides the Wi-Fi and dual mode Bluetooth and it is used in order to upload data on the cloud (in figure 4).

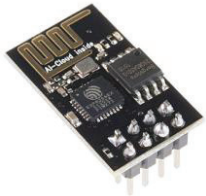

Fig. 4. Wi-Fi Module

- **Arduino Uno Mega**: Arduino Mega 2560 is a microcontroller board that is based on the ATmega2560 (in figure 5). It is used to connect the sensors and obtain the analog input from different sensors and upload on the cloud using Wi-Fi module [11-14].

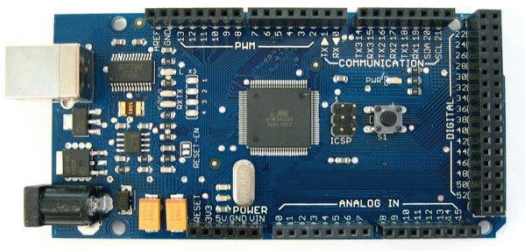

Fig. 5. Arduino Mega

- **LCD Display:** LCD Display displays all the values sensed by integrated sensors (in figure 6).

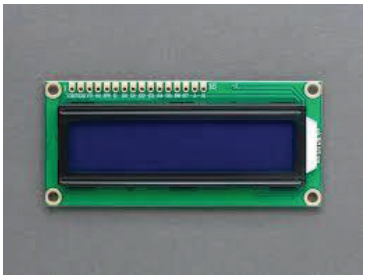

Fig. 6. LCD Display

- **MQ-135:** MQ-135 is a gas sensor. It's layer is made up of Tin Dioxide and it has lower conductivity to clean air while higher conductivity to polluted air in presence of which its conductivity increases (in figure 7).

Fig. 7. MQ135 Sensor

BMP-180: It works on the base of air weight. Our surrounding air has certain weight which has specific pressure which is sensed by BMP180 (in figure 8). The value is controlled and sent towards controller with the help of 12C interface [15].

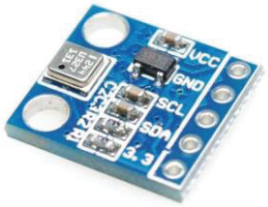

Fig. 8. BMP180 Sensor

- **Dust Sensor**: Dust sensor operates on principle of light scattering. A photo-detector and LED emitter oppose each other at an angle within the rectangular package of the sensor which has a dust through hole hole on either side, Air containing dust particles flows into the sensor chamber and causes the light from the LED emitter to be scattered towards the photo-detector. More the particles are, more will be the scattering of light and more will be the voltage (in figure 9.

Fig. 9. Dust Sensor

- **Rain Sensor**: It detects water if the drops is poured on the sensor plate. The wetter the board is the higher the conductivity (in figure10).

Fig. 10. Rain Sensor

- **LDR Sensor**: It detects the light and resistance varies and value is given by LDR (in figure 11).

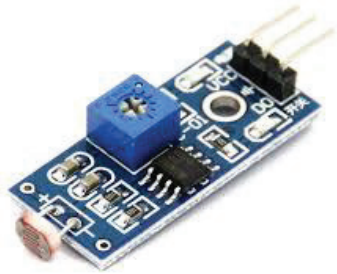

Fig. 11. LDR Sensor

- **MQ-7:** MQ-7 has two components (in figure 12).

1. Heating circuit having the control function.

2. Signal output circuit, it can accurately respond changes in surface resistance of the sensor. When it is in clean air, there is lower conductivity. Conductivity increases along rising gas concentration and low temperature. At high temperature, it cleans the other gases that is, nullify their concentration [16].

Fig. 12. MQ7 Sensor

- **DHT11:** This sensor calculates yemperature and humidity. It is less costly and reliable (in figure13) .

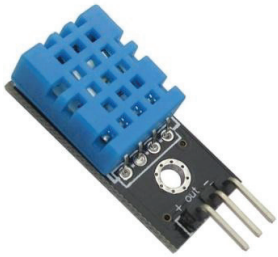

Fig. 13. DHT11 Sensor

III. SYSTEM DESIGN

The SDLC model used here i.e. **Iterative SDLC model** (in figure.14)

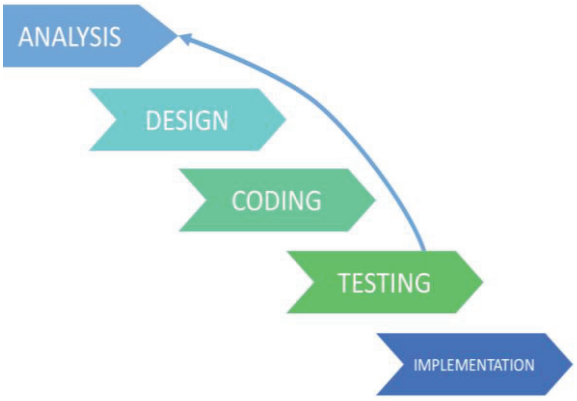

Fig. 14. Iterative SDLC

A. User Interface Design

Interface has been designed on wixsite with little help of CSS and JavaScript. It contains a soothing background with temperatures displayed and website information posted on it. It also contains information about the various sensors used in the model [17].

B. Preliminary Product Description

This model collects data from various sensors such as MQ135, MQ7, DHT11, BMP180 which gives us reading of different parameters in the surrounding like Temperature, Humidity, Pressure and Air quality. This data is then sent to the Thingspeak cloud using ESP 8266 Wi-Fi module. The data gets stored in designated channels which is then used to calculate the dew point of that time. After the dew point calculation machine learning will be used to predict the temperature of the following day.

IV. METHODOLOGY AND FRAMEWORK

A. Dataset Collection

Dataset was collected from Kaggle which was further extracted from weather underground which is currently not providing free service. The dataset provides the data from 1st May 2016 to 11th March 2018. It gives a list of parameters:

- Meantemp
- Meandewpoint
- Meanpressure
- Maxhumidity
- Minhumidity
- Maxtemp
- Mintemp
- Maxdewpoint
- Mindewpoint
- Maxpressure
- Minpressure

a) Selection: Highlight all author and affiliation lines.

b) Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

c) Deletion: Delete the author and affiliation lines for the extra authors.

B. Feature Selection

Number of parameters were increased to take into account weather conditions of past three days. Analysis of data was done and relevant parameters were identified. After conducting the following analysis, the following parameters were identified:

For max temperature prediction:

- Maxtempm_1
- Maxpressure_1
- Mintempm_3
- Maxpressure_3
- Meanpressurem_3
- For min temperature prediction:
- Mintempm_1
- Meantempm_3
- Maxtempm_1
- Maxtempm_2
- Maxdewptm_1
- Maxdewptm_3
- Meandewptm_1
- Meandewptm_2
- Meanpressure_1
- For mean temperature prediction:
- Meantemp_3
- Meanpressure_3

- Maxtemp_1
- Maxtemp_2
- Mintemp_1

C. Description of Algorithm Designed

We are using multiple linear regression (MLR), it is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable [19].

In essence, multiple regression is the extension of ordinary least-squares (OLS) regression that involves more than one explanatory variable [20].

The formula for multiple linear regression is given by equation below:

$$Y = a + b_1*X_1 + b_2*X_2 + \ldots + b_p*X_p \quad (1)$$

Where, Y= Dependent variable, a= Y-intercept, x= Explanatory variables, b= slope coefficients for each explanatory variables.

D. Training Set Description

The model was trained using the data collected, it held past 2 years' data and was sufficient to train the model with high accuracy. Main dataset was split into training set and testing set with ratio 0.8 to 0.2.

E. Technology Used

a) Internet of Things: To collect data.

b) Machine learning (Linear Regression): To train data.

c) Data Analytics: To analyze the predicted model.

d) Cloud Computing: To upload data on thingspeak.

V. DISCUSSION AND RESULTS

A. Testing

The model was tested multiple times. A few times it was left on the hostel terrace to sense the values and the same was performed at multiple locations as well. At first, system showed temperature value with an error of 1 degree Celsius but later was rectified. The same operation was performed at hostel rooms to understand any discrepancy and the result was expected.

B. Maintenance

The model has been fitted in a box to accommodate all the components together and keep them integrated rather than drifting apart (in figure15, 16). Loose connections can be rectified easily. Only continuous power supply and internet is required to run the model for a long time.

Fig. 15. Final Model

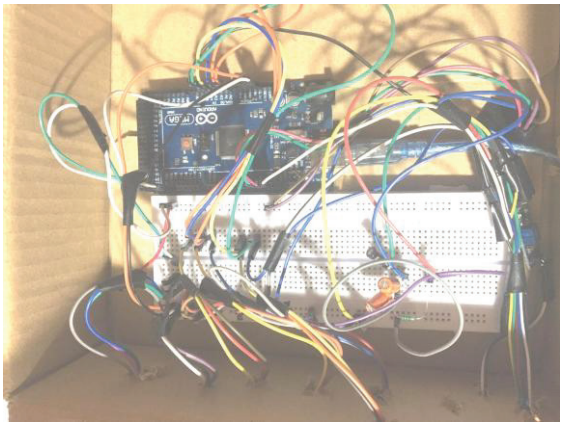

Fig. 16. Connections

C. Result Analysis

The parameter values collected from DHT11 sensor are stored on internet using cloud, which is used for further analysis using MATLAB. The data collected through sensors can be viewed on a web portal (in figure 17, 18).

TABLE I. ACCURACY ACHIEVED

Parameters	Current Temperature Accuracy is 99.05%		
	Next day Min Temperature Accuracy	Next day Max Temperature Accuracy	Next day Mean Temperature Accuracy
Explained variance	0.94	0.93	0.95
Mean absolute error	1.35 degree Celsius	1.29 degree Celsius	1.10 degree Celsius
Median absolute error	1.09 degree Celsius	0.97 degree Celsius	0.90 degree Celsius

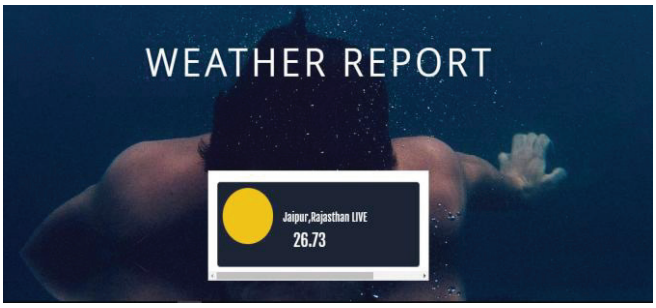

Fig. 17. Website Screenshot displaying current weather

Fig. 18. Website Screenshot of weather prediction

Current day's temperature and values from different sensors are displayed on the web portal. For prediction of following day's temperature, the parameters, through correlation functions on the parameters and using P values were sorted that is, from the existing dataset only relevant fields for the project were picked up. Eventually, these values were used to train the model and then it was able to predict the future values of temperature. The correlation analysis is carried out on three parameters, they are; **Temperature, Humidity and Pressure** (In Table I)**.**

The correlation between temperature and humidity is around **0.6-1 and -0.6 to -1**. Hence the correlation between those two parameters is good which shows that the existence of pressure parameter doesn't affect the value of temperature. The correlation parameter chosen for is in the range 0.6-1 and -0.6 to -1 and p value to segregate the parameters used is 0.05. The humidity and pressure are also correlated around 0.2. Hence further only temperature and humidity parameters are considered.

The standard error indicates that the predicted value nearly plus or minus of error value.

Current Temperature Accuracy: Temperature shown contains an error of 0.5 degree Celsius.

D. Conclusion

Main aim is the integration of various sensors to collect data using which temperature can be predicted. Here, more sensors are used to assess more parameters for high accuracy. Machine learning algorithms is used which will help in the prediction of temperature of the coming day based on the historical data hence aiming at higher accuracy at lower cost. Also, it will the first setup which will monitor air quality along with forecasting weather and present it in real time.

REFERENCES

[1] M. Zhang, "Application of Data Mining Technology in Digital Library", Journal of Computers, vol. 6, no. 4, (2011) April, pp. 761-768.

[2] Z. Danping and D. Jin, "The Data Mining of the Human Resources Data Warehouse in University Based on Association Rule", Journal of Computers, vol. 6, no. 1, (2011) January, pp. 139-146.

[3] L. M. Saini and M. K. Soni, "Artificial neural network-based peak load forecasting using conjugate gradient methods," IEEE Transactions on Power Systems, vol. 12, no. 3, pp. 907–912, . 2002.

[4] S. Fan, C. X. Mao, and L. N. Chen, "Peak load forecasting using the self-organizing map," in Advances in Neural Network-ISNN 2005. New York: Springer-Verlag, 2005, pt. III, pp. 640–649.

[5] Kourentzes, N., "Intermittent demand forecasts with neural networks", International Journal of Production Economics, Volume 143, Number 1, pages 198-206, 2013.

[6] Elia G. P., 2009, "A Decision Tree for Weather Prediction", Universitatea Petrol-Gaze din Ploiesti, Bd. Bucuresti 39, Ploiesti, Catedra de Informatică, Vol. LXI, No. 1.

[7] A. R. Finamore; V. Calderaro; V. Galdi; A. Piccolo; G. Conio; S. Grasso, "A day-ahead wind speed forecasting using data-mining model– a feed-forward NN algorithm", IEEE International Conference on Renewable Energy Research and Applications, 2015, pp. 1230-1235.

[8] E. Erdem, J. Shi, "ARMA based approaches for forecasting the tuple of wind speed and direction", Applied Energy 88, ELSEVIER, 2011, pp. 1405–1414.

[9] IOP Conf. Series: Journal of Physics: Conf. Series 910 (2017) 012020 doi :10.1088/1742-6596/910/1/012020.

[10] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, "Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction," IEEE Transaction on Power Systems, vol. 29, no. 2, 2014.

[11] Martin T. H., Howard B. D, Mark B., 2002, Neural Network Design, Shanghai: Thomson Asia PTE LTD and China Machine Press.

[12] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, "Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction," IEEE Transaction on Power Systems, vol. 29, no. 2, 2014.

[13] R. R. B. de Aquino, H. T. V. Gouveia, M. M. S. Lira, A. A. Ferreira, O.N. Neto, M. A. Carvalho Jr., "Wind Forecasting and Wind Power Generation: Looking for the Best Model Based on Artificial Intelligence", IEEE World Congress on Computational Intelligence, 2012.

[14] Roger A. Pielke Sr., "Mesoscale Meteorological Modelling", International Geophysics Series, Volume 98, oct, 2013.

[15] David J. Stensrud ,"Parameterization schemes – Keys To Understanding Numerical Weather Prediction Models", Cambridge University Press, 2007.

[16] A Parashar, A Parashar, S Goyal, B Sahjalan, "Push recovery for humanoid robot in dynamic environment and classifying the data using K-mean", Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies.

[17] A Parashar, A Parashar, S Goyal, "Classifying gait data using different machine learning techniques and finding the optimum technique of classification" Information and Communication Technology for Sustainable Development, 305-313.

[18] A Parashar, D Goyal, "Clustering Gait Data Using Different Machine Learning Techniques and Finding the Best Technique" International Conference on Smart Trends for Information Technology and Computer Communications.

[19] A Parashar, A Parashar, S Goyal, "Tracing Gesture and Extracting Gait Feature to Recognize Parkinson's Disease Using Multilayered Back Propagation Algorithm" International Conference on Innovative Technologies, IN-TECH 2017, 9

[20] A Parashar, A Parashar, S Goyal, "Identification of gait data using machine learning technique to categories human locomotion, Proceedings of the 10th International Conference on Security of Information and Networks.