International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

ARIMA Based Weather Prediction Model using IoT and Open Source Data

Gunjan Rawat[1], Dhruv Patel[2], Prajwal Khapare[3], Prof Sandhya Deshpande[4]

[1]Department of Electronics and Telecommunication Engineering, KJ Somaiya Institute of Engineering and Information Technology(Autonomous), (Affiliated to University of Mumbai),Mumbai,India)
2Professor, Department of Electronics and Telecommunication Engineering, KJ Somaiya Institute of Engineering and Information Technology(Autonomous), (Affiliated to University of Mumbai),Mumbai,India)

Abstract -The main objective of our research is to integrate weather parameters data collected from IMD(Indian Meteorological Department) and AccuWeather.com and combine it with our own data collected from sensors to create a database. This database is then used to predict and monitor the weather of our area. The prediction done in this manner will lead to a more accurate result of the area where the sensors are located and therefore results in better accuracy and is more suitable. As it is with respect to an area, it will be more precise than other Web applications which predict weather of a whole region compared to a specific area. This could find it's application in a variety of fields such as it can be used by farmers for better prediction, schools, universities and industries/offices where exist extreme weather conditions. IMD data is extracted using Python and a python library created specifically for this purpose called imdlib. This data constitutes the whole of India so QGIS can be used for mapping it to a certain area and the resultant can be stored in the form of a CSV file. The second part of the database is gathered from sensors which are connected to a microcontroller which in turn transfers this data to a wifi module and stores it in the form of CSV file.The database can then be used to work with the ARIMA model to predict the data. ARIMA model is one of the most accurate and easy to implement linear regression model.

KeyWords: Machine- Learning, IoT, Python ,Arduino, ARIMA, Linear Regression, Time Series Forcasting

1. INTRODUCTION

Climate change and environmental monitoring have received much attention recently. With our project we are proposing a Weather Prediction Model which uses data from our own sensors(IoT)[1][4] and data available from IMD(Indian Meteorological Department) as it's dataset. IoT-Internet of things is basically a network of physical objects(sensors) embedded with software and other technologies. QGIS is a free and open source geographic information system which allows to analyze and edit spatial information. Machine learning allows applications to become more accurate at predicting outcomes. With our project we are hoping to create a system which will be able to monitor the weather conditions over controlled areas like house, industry, agricultural area. Capture weather parameters like temperature, humidity and rain using sensors and finally predict the rainfall, temperature and humidity by implementing Machine learning Algorithms on the data received from sensors and from IMD.

1.1 Block Diagram

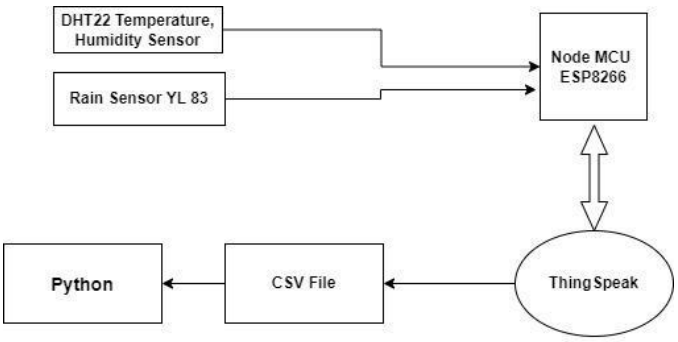

Figure 1- Sensor Flowchart

Figure 1 is the block diagram which explains the extraction of data using sensors. The DHT22 sensor senses the Temperature and humidity while the YL83 is a rain sensor. This information is then fed to Node MCU .This information is forwarded further by interfacing the NodeMCU with its inbuilt ESP8266 Wifi module which then transfers this data to Thing Speak. The data is then downloaded in CSV format.[3][4]This is then fed as one of

the inputs for our dataset. The second input will come in the form of data which is extracted from IMD.

For the second dataset we will be using an inbuilt python library called imdlib which is specifically used to extract data from IMD. Using that we will receive the data in CSV format. Thus the second input for our dataset will be ready

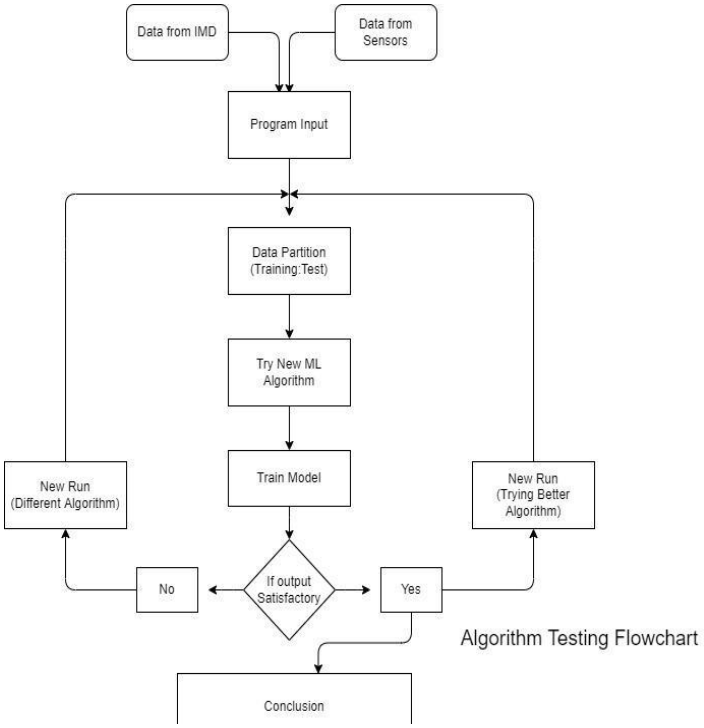

Figure 2- Algorithm Testing Flowchart

After the dataset is ready, the second part comprises testing the machine learning algorithm. A supervised machine learning model will be used for this prediction and after testing a number of different algorithms, choosing the one with the best accuracy. This will all be done using Python and its libraries.

2. Methodology

Data will be collected from two streams[2] : sensors and IMD free available data.

DATA COLLECTION-SENSORS[1][3][4]:

For temperature and humidity we will be using DHT22 sensor, Rain sensor YL 83. These sensors will in turn be connected to a microcontroller- Arduino UNO. We will use Arduino IDE for programming this. The arduino will be connected to a ESP8266 WiFi Module which will transfer the data to a google sheet/excel as a CSV file using ThingSpeak. This data will be one of our datasets.

DATA COLLECTION-IMD[2]:

For collecting the IMD data we will be extracting the data from the .GRD format. An inbuilt python library will be used to extract the data called 'imdlib'. This will help us to get the entire data of a single year . This dataset will then be mapped to our area using QGIS. This will be our second dataset.

DATA CLEANING:

All the data has to then be formatted and integrated into one database.

PREDICTION MODEL:

The final step will be integrating both these datasets and training our model. We will be using ARIMA time series forecasting for this which is a type of linear regression model. This will all be done using Python and its libraries.

3. ARIMA Model

Auto Regressive Integrated Moving Average, abbreviated as ARIMA, is an Algorithm for forecasting that is centered on the concept that the data in the previous values of the time series can alone be utilized in order to predict the future values.

The primary step is to make the time series stationary in order to build an ARIMA model. This is because the term 'Auto Regressive' in ARIMA implies a Linear Regression Model using its lags as predictors. And as we already know, Linear Regression Models work well for independent and non-correlated predictors.

In order to make a series stationary, we will utilize the most common approach that is to subtract the past value from the present value. Sometimes, depending on the series complexity, multiple subtractions may be required.

4. Practical Implementation

As the methodology suggests , the flow is followed and the database is created using the sensor and IMD data. The connections of the system are as shown in figure 3. The dht22 is on the breadboard and yl83 independent , both connected to ESP8266. This is then connected to the Arduino IDE using a micro USB cable.

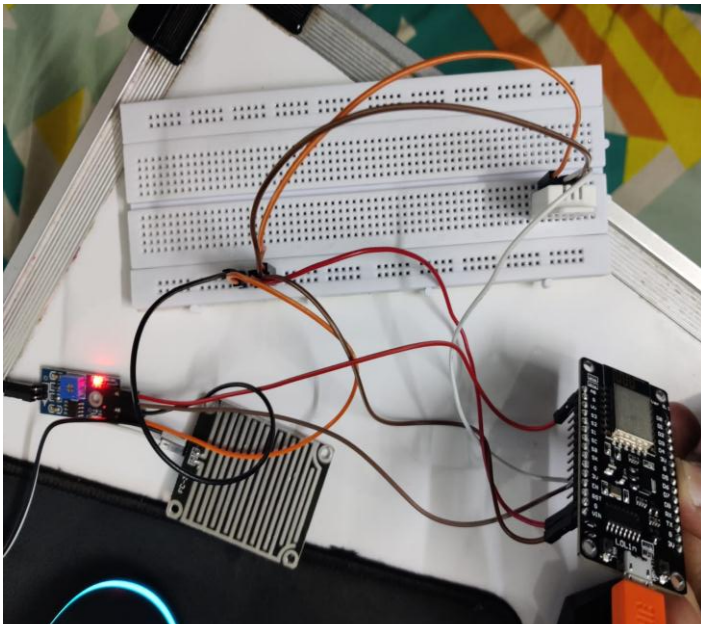

Figure 3- Sensor Connections

This sensor data is displayed on ThingSpeak

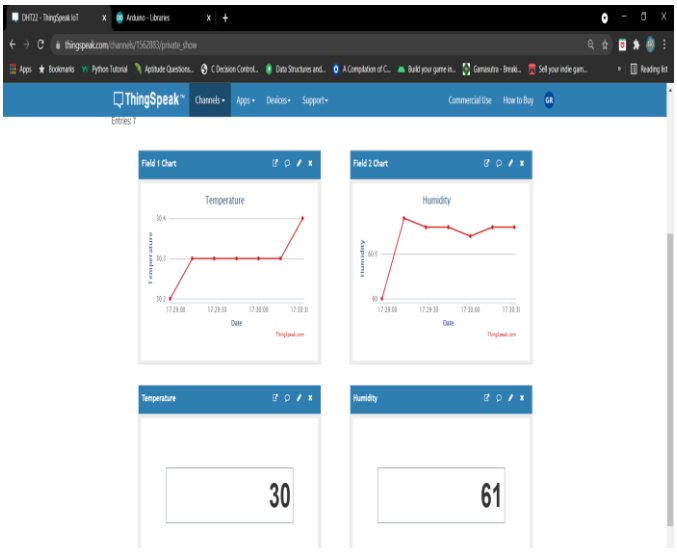

Figure 4 - ThingSpeak Visualization

After following the methodology and completing the entire process, the prediction for the maximum temperature for the month of april is displayed in figure 4. In a similar way the prediction for the other parameters can be carried out which are minimum temperature, rainfall and relative humidity. Jupyter Notebook along with libraries such as pandas ,numpy, statsmodels, matplotlib were used.

```
In [99]: future_df['forecast'].tail(30)

Out[99]: 2022-04-01    33.590730
         2022-04-02    33.403989
         2022-04-03    33.237173
         2022-04-04    33.088158
         2022-04-05    32.955042
         2022-04-06    32.836131
         2022-04-07    32.729907
         2022-04-08    32.635018
         2022-04-09    32.550254
         2022-04-10    32.474534
         2022-04-11    32.406894
         2022-04-12    32.346471
         2022-04-13    32.292495
         2022-04-14    32.244279
         2022-04-15    32.201207
         2022-04-16    32.162732
         2022-04-17    32.128362
         2022-04-18    32.097659
         2022-04-19    32.070232
         2022-04-20    32.045732
         2022-04-21    32.023846
         2022-04-22    32.004295
         2022-04-23    31.986830
         2022-04-24    31.971229
         2022-04-25    31.957293
         2022-04-26    31.944843
         2022-04-27    31.933722
         2022-04-28    31.923788
         2022-04-29    31.914914
         2022-04-30    31.906986
         Name: forecast, dtype: float64
```

Figure 4 - Maximum temperature prediction

5. Conclusions

Weather of the coming days is correctly predicted(85-90% accuracy expected) with the help of the entire setup. Data is successfully collected from the sensors and successfully extracted from IMD. The resultant datasets are then formatted, all null/not required values are removed. Both the datasets are then integrated to form a database which will serve as input to the algorithm. All this will be processed in python and using its vast amount of libraries for machine learning to train the ARIMA model.

6. References

[1] Development of IoT Based Weather Reporting System"-A F Pauzi1 and M Z Hasan , International Conference on Technology, Engineering and Sciences (ICTES) 2020,18 April 2020.

[2] "Smart Weather Forecasting Using Machine Learning:A Case Study in Tennessee"- A H M Jakaria,Md Mosharaf Hossain,Mohammad Ashiqur Rahman,arXiv-Cornell University,25 August 2020.

[3] "Internet of Things (IOT) based Weather Monitoring System"-Girija C,Harshalatha H,International Journal of Engineering Research & Technology (IJERT),2018.

[4] "IOT Based Weather Monitoring and Reporting System Project"=Anita M. Bhagat, Ashwini G. Thakare, Kajal A. Molke, Neha S. Muneshwar, Prof. V. Choudhary,International Journal of Trend in Scientific Research and Development (IJTSRD),March 2019.