REAL TIME WEATHER PREDICTION

IOT + AI = SMART WORLD

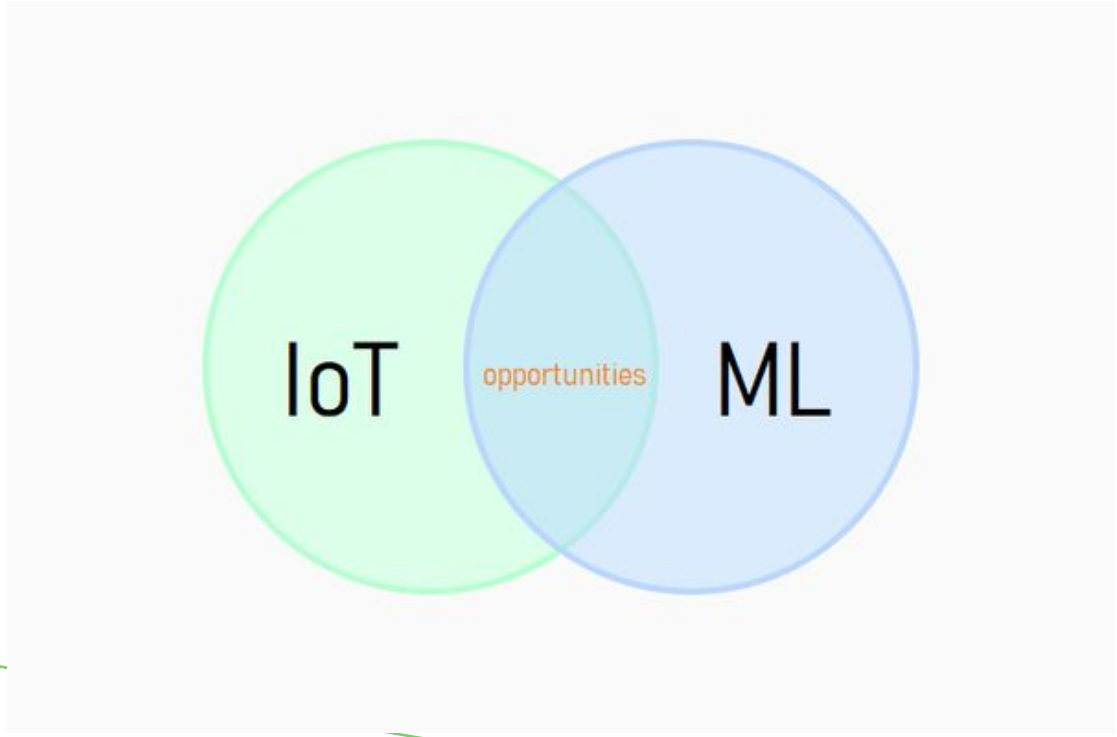
Aryan Jain

ABSTRACT

This project implements the real time weather prediction system. It can be used in number of applications like homes, industries, agriculture, stadiums etc. for predicting the weather information. The system utilizes a temperature and humidity sensor i.e. DHT11. A model is used for setting up the machine learning environment. This model is trained using the pre- recorded values of sensor data.

MOTIVATION

Being extremely interested in everything having a relation with the Machine Learning, the independent project was a great occasion to give me the time to learn and confirm my interest for this field.



INTRODUCTION

- This project proposes a method for forecasting weather conditions and predicting weather by means of machine learning.
- The aim of my project is to monitor several aspects of weather using IoT based smart system and to predict the future values.
- The main advantage of using this concept is it helps in monitoring weather of a local area and thus it helps in developing a microclimate system.
- Based on the monitored database it predicts the future weather of a particular zone thereby making the result more accurate and relevant for the zone.
- This model is train with the recorded data values of temperature and humidity.

PROBLEM STATEMENT

- Predicting weather using different machine learning algorithms in IOT.
- Which algorithm is optimal to use for best accuracy?
- How accurate are long term forecasts compared to short term forecasts?

OBJECTIVE

The main objective of this project is to predict weather using IOT sensors and ML

- To design a facility that can help the user to plan their day to day activities.
- The purpose of a weather forecast is to provide as accurate as possible prediction of what the weather will be like in the near future.

LITERATURE REVIEW

Sl No.	Year of Publication	Title	Authors	Technique / Methodology used	Reference
1.	2020	weather prediction done by using linear regression	M.K. Nallakaruppan and U. Senthil Kumaran	• It uses Raspberry Pi, consisting 4 USB ports and an ethernet port. • It consists of many sensors, and a power supply also given to this, a HDMI port is available which gives connection interfaceable screen. • A Micro SD is attached to Raspberry OS. Here coding used is python, data stored in MYSQL server. Here ARIMA used instead of Decision tree to dynamic behavior capturing.	International Journal of Recent Technology and Engineering (IJRTE)

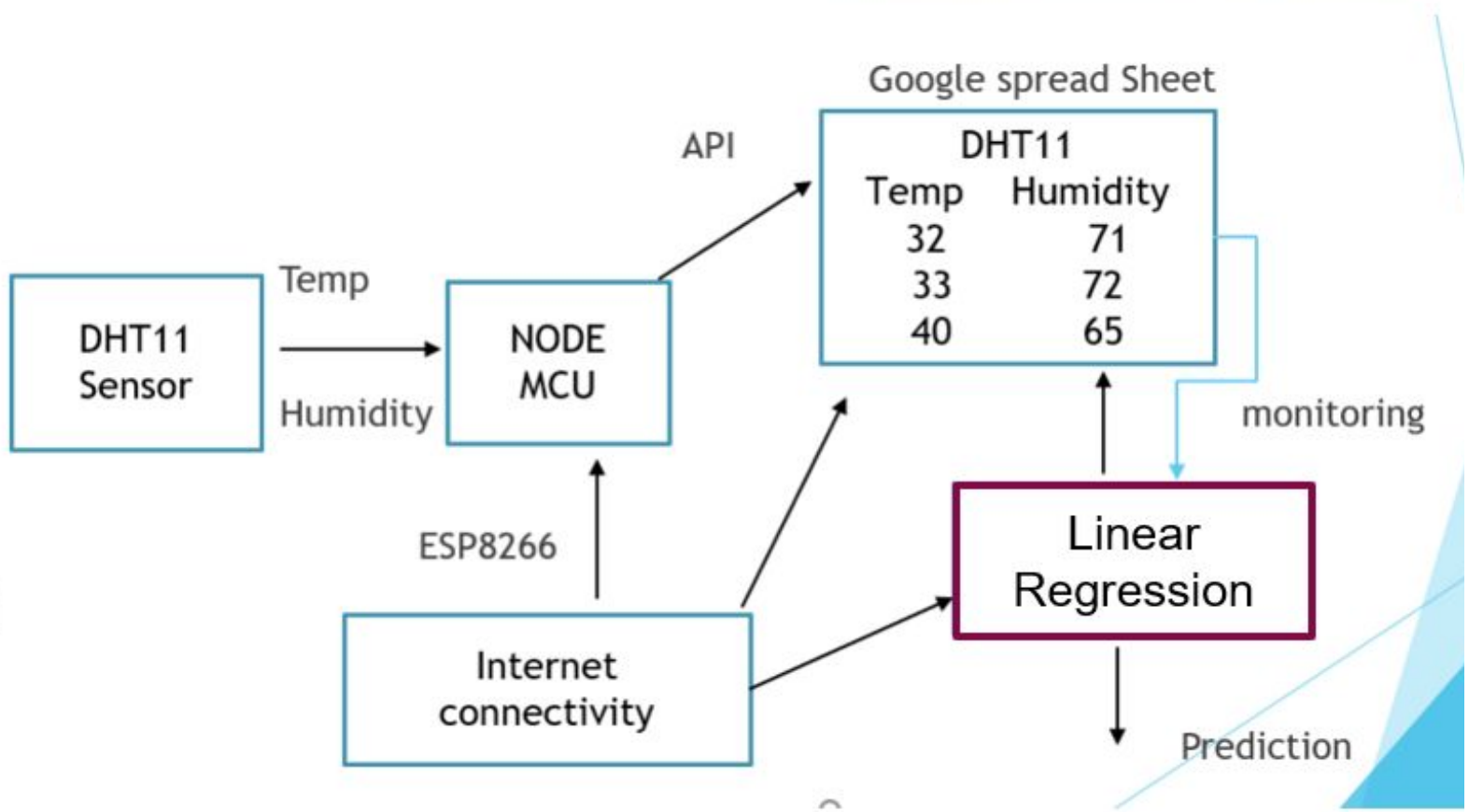
LITERATURE REVIEW

Sl No.	Year of Publication	Title	Author	Technology / Methodology used	Reference
2.	2020	A system to gather data for accurate weather predictions	S.Akter Nishe and Tahmina Aziz	• The design is based on a standalone Arduino board and the sensors for temperature, humidity along with GPS and Bluetooth modules are used. The communication between the mobile phone and Arduino board is wireless. ▪ The data collected is quantitative and continuous which is analyzed to show perfect correlation for accurate prediction	International Journal of Engineering and Advanced Technology (IJEAT)

LITERATURE REVIEW

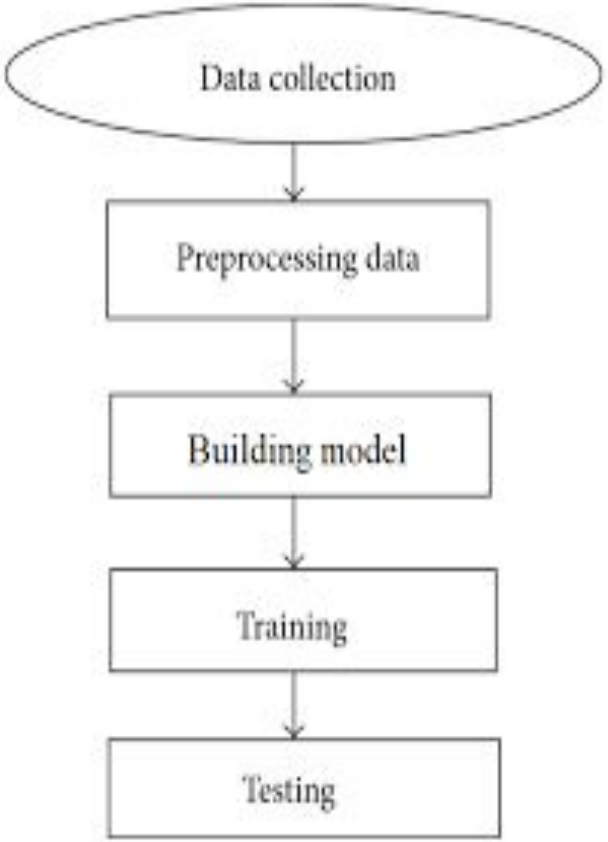
Sl No.	Year of Publication	Title	Author	Technology / Methodology used	Reference
3.	2020	Real Time weather Prediction using IOT and Machine Learning.	Gaurav Kumar , Pranjul Mittal and Shaista Farheen	• The paper can be utilized to foresee meteorological information that is climate expectation. • The paper gives a review of accessible writings of a few calculations utilized by various specialists to use different information mining methods for weather prediction	Institute of Electrical and Electronics Engineers(IEEE)

PROPOSED ARCHITECTURE

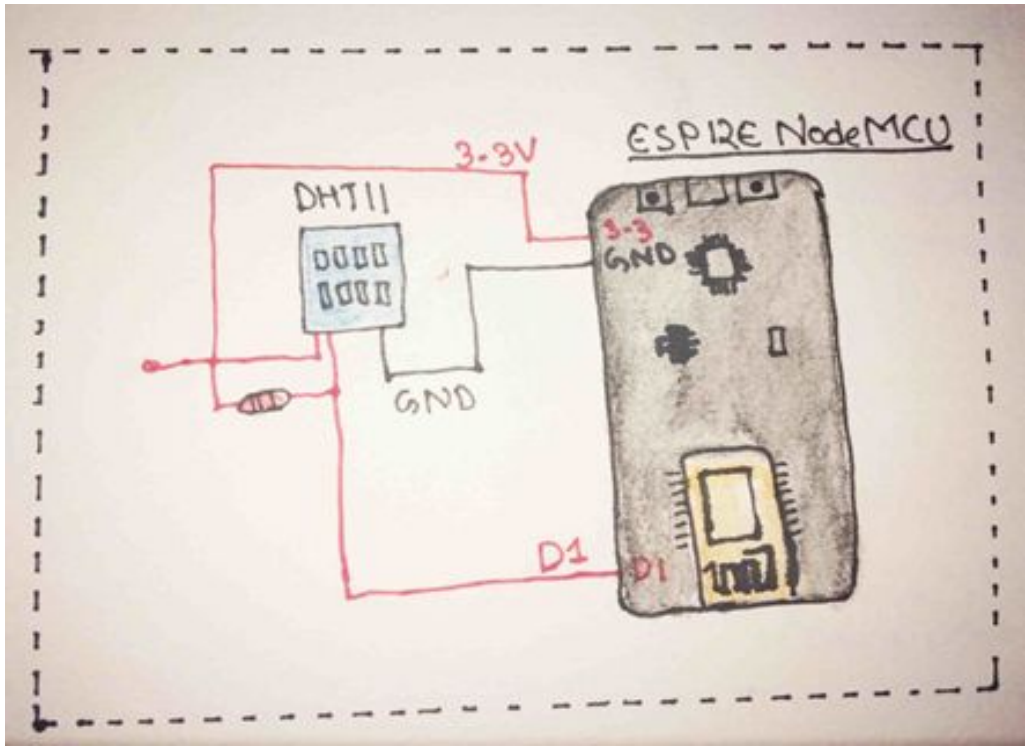


PROPOSED ARCHITECTURE

Data flow diagram



CIRCUIT DIAGRAM



MODULE WISE BREAKDOWN OF PROJECT

1. Collecting Data Module

- This model is trained using the pre- recorded values of sensor data. Further, NodeMCU records the data from sensors i.e. temperature, humidity and light intensity and then the values are transferred to the Jupyter notebook that utilizes a python environment.
- This real time data is used to test the model and prediction is done for a particular value by blinking the led connected to NodeMCU.
- DHT11 sensor is a module used for measuring temperature and humidity.
- It uses a capacitive humidity sensor and a thermistor to measure the surrounding air's humidity and temperature.
- NodeMCU is an updated version of Arduino with inbuilt wifi chip.
- It is cheaper than other modules performing the same function.

MODULE WISE BREAKDOWN OF PROJECT

2. Data Predicting Module

- Machine learning is the capacity of computer to learn without being expressly customized.
- It enables machines to find concealed patterns and insights.
- Linear Regression algorithm and Random forest has been used to predict the data for weather forecasting.
- Mostly, these models take the present weather conditions and process it to build a model for predicting the weather.

FUNCTIONAL REQUIREMENT

1. Project task management
   - Google Spreadsheet(To store data)
   - Google Script editor(to create an api which will be used to insert data in Google spreadsheet)
2. Weather forecast
   - Making circuit with Node MCU and sensor(DHT11)
   - Display 3 months Weather forecast
3. Prediction using supervised machine learning
   - Applying Linear Regression algorithm on the collected data.
   - Applying Random Forest algorithm on the collected data

IMPLEMENTATION OF ALGORITHMS

Selection criteria

1.Classification and regression capabilities

2.Data quality

3.Computational complexity

LINEAR REGRESSION ALGORITHM

1. CONCEPT -

The data is modelled using a straight line.

2. Used With -

Continuous Variable.

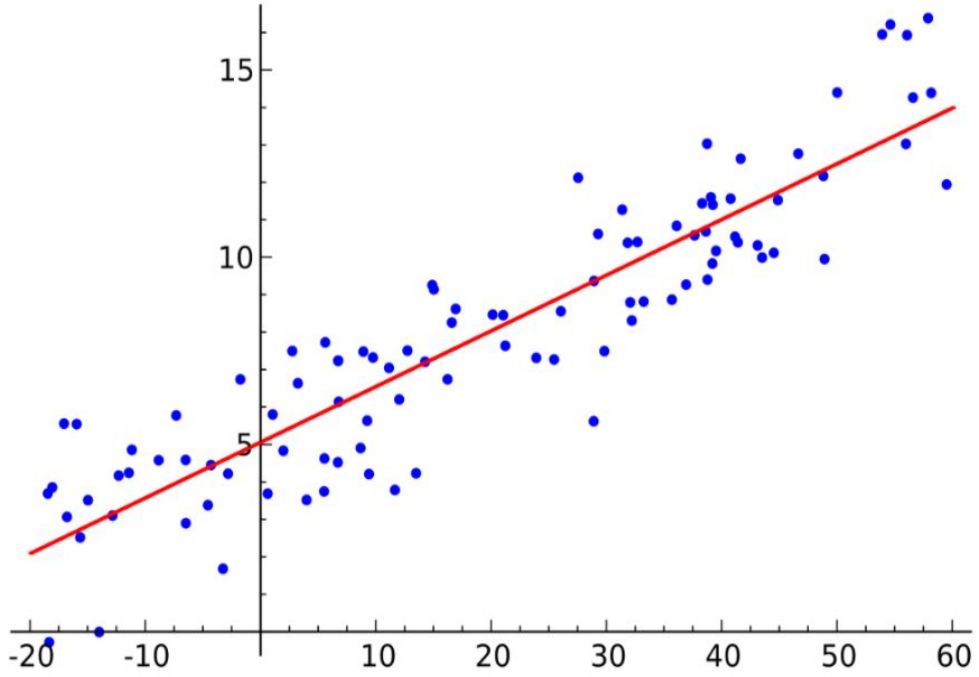
3. Predict –

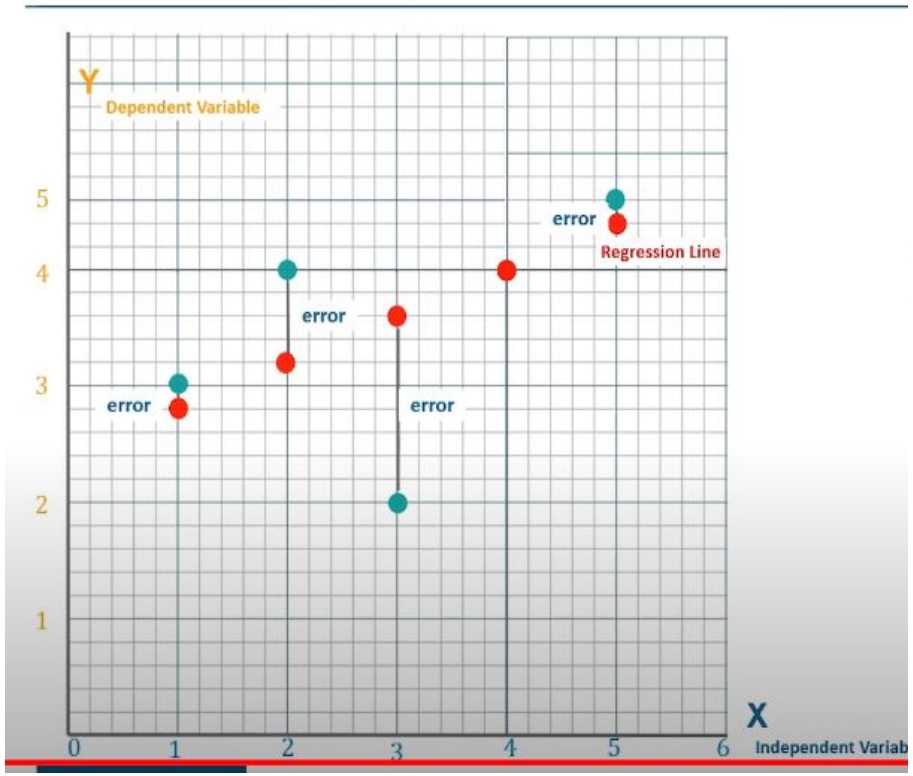
Value of The Variable

4. Accuracy and goodness of fit –

Measured by Loss, R Squared, Adjusted R Squared.

LINEAR REGRESSION ALGORITHM





RANDOM FOREST ALGORITHM

- Random forest is a supervised learning algorithm. The "forest" it builds, is an ensemble of decision trees, usually trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result.
- One big advantage of random forest is that it can be used for both classification and regression problems.

RANDOM FOREST ALGORITHM

- Pick at random K data points from the training set.
- Build the decision tree associated with those K data points
- Choose the number Ntree of trees you want to build and repeat step 1 and 2.
- For a new data point, make each one of your Ntrees trees predict the value of Y for the data point, and assign the new data point the average across all of the predicted Y values.

RANDOM FOREST ALGORITHM

No overfitting

Use of multiple trees reduce the risk of overfitting

Training time is less

High accuracy

Runs efficiently on large database

For large data, it produces highly accurate

Estimates missing data

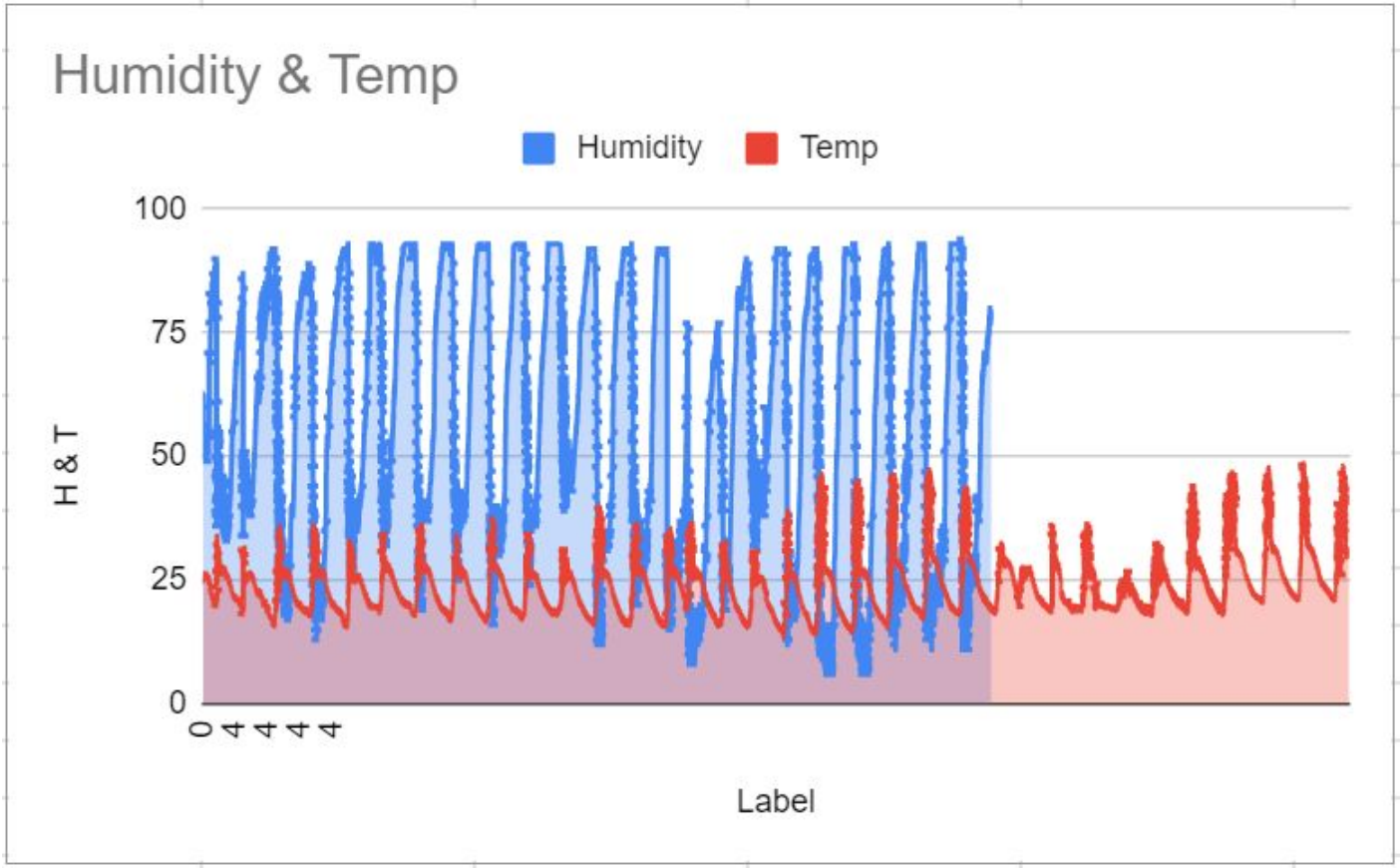
Random Forest can maintain accuracy when a large proportion of data is missing

RESULTS

1	id	Date	Time	Temp	Humidity	label
4147	4146	27-01-2021	05:50:49	18.2	87	4
4148	4147	27-01-2021	05:51:50	18.2	87	4
4149	4148	27-01-2021	05:52:51	18.2	87	4
4150	4149	27-01-2021	05:53:52	18.2	87	4
4151	4150	27-01-2021	05:54:53	18.1	88	4
4152	4151	27-01-2021	05:55:54	18	88	4
4153	4152	27-01-2021	05:56:55	18	89	4
4154	4153	27-01-2021	05:57:56	18.1	88	4
4155	4154	27-01-2021	05:58:57	18.1	88	4
4156	4155	27-01-2021	05:59:57	18.3	87	4
4157	4156	27-01-2021	06:00:59	18.4	86	4
4158	4157	27-01-2021	06:01:59	18.6	86	4
4159	4158	27-01-2021	06:03:00	18.5	86	6
4160	4159	27-01-2021	06:04:01	18.4	86	6
4161	4160	27-01-2021	06:05:02	18.5	86	6
4162	4161	27-01-2021	06:06:03	18.4	86	6
4163	4162	27-01-2021	06:07:05	18.5	86	6
4164	4163	27-01-2021	06:08:05	18.5	86	6
4165	4164	27-01-2021	06:09:06	18.6	86	6
4166	4165	27-01-2021	06:10:07	18.8	85	6

IoT Based Data Monitoring

RESULTS



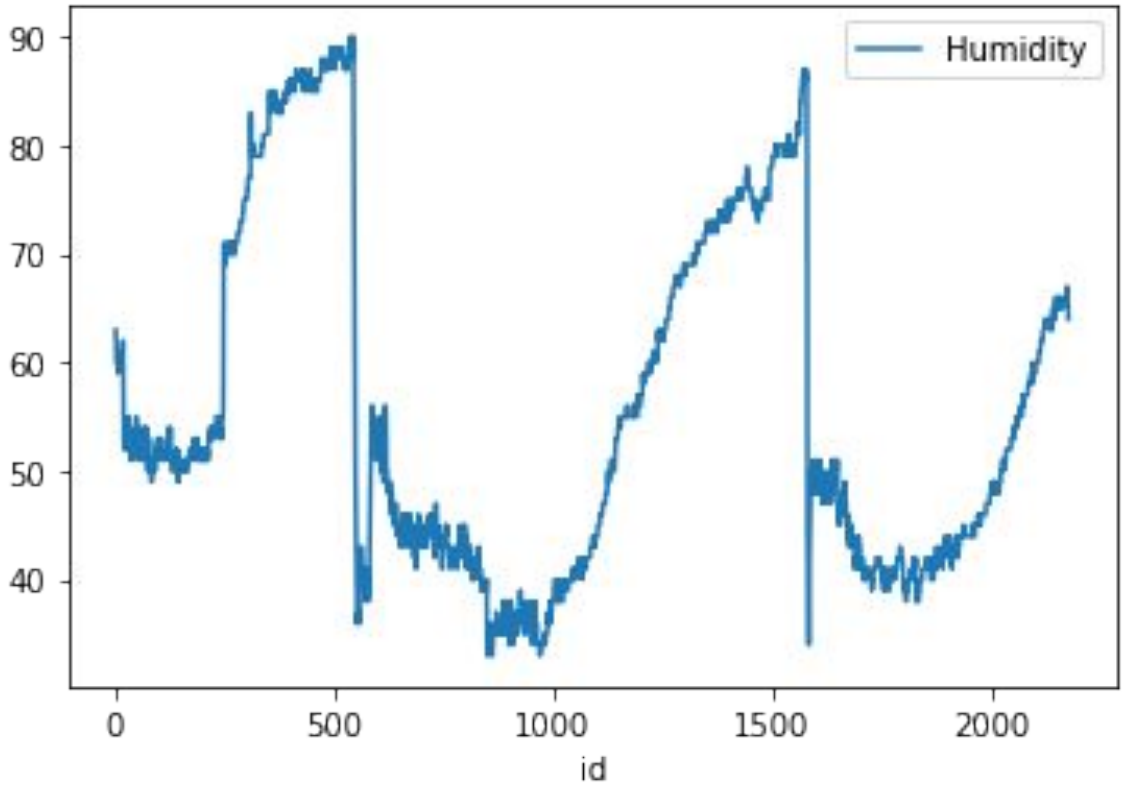
IoT Based Data Monitoring

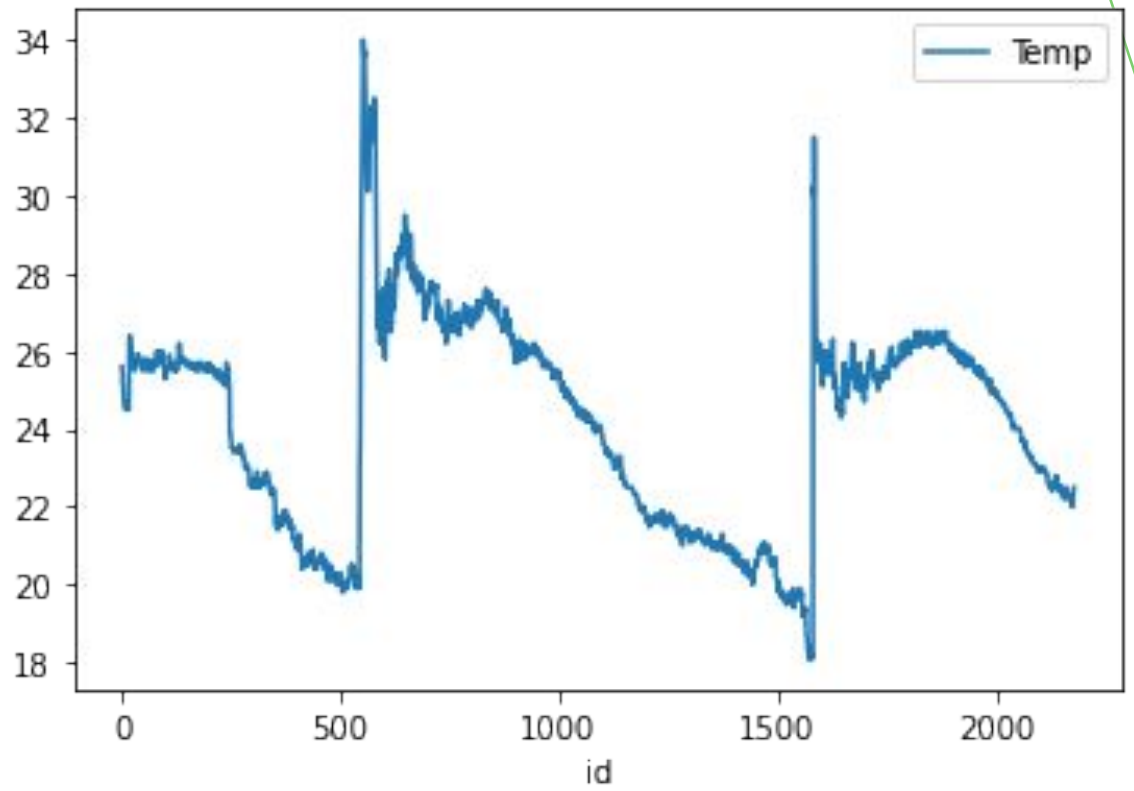
RESULTS

	id	Temp	Humidity	label
count	2178.000000	2178.000000	2178.000000	2178.000000
mean	1089.500000	24.185262	55.932048	1.820018
std	628.878764	2.639692	16.018248	1.992343
min	1.000000	18.100000	33.000000	0.000000
25%	545.250000	21.700000	43.000000	0.000000
50%	1089.500000	25.000000	51.000000	0.000000
75%	1633.750000	26.000000	70.000000	4.000000
max	2178.000000	34.000000	90.000000	4.000000

Data Analysis With Python

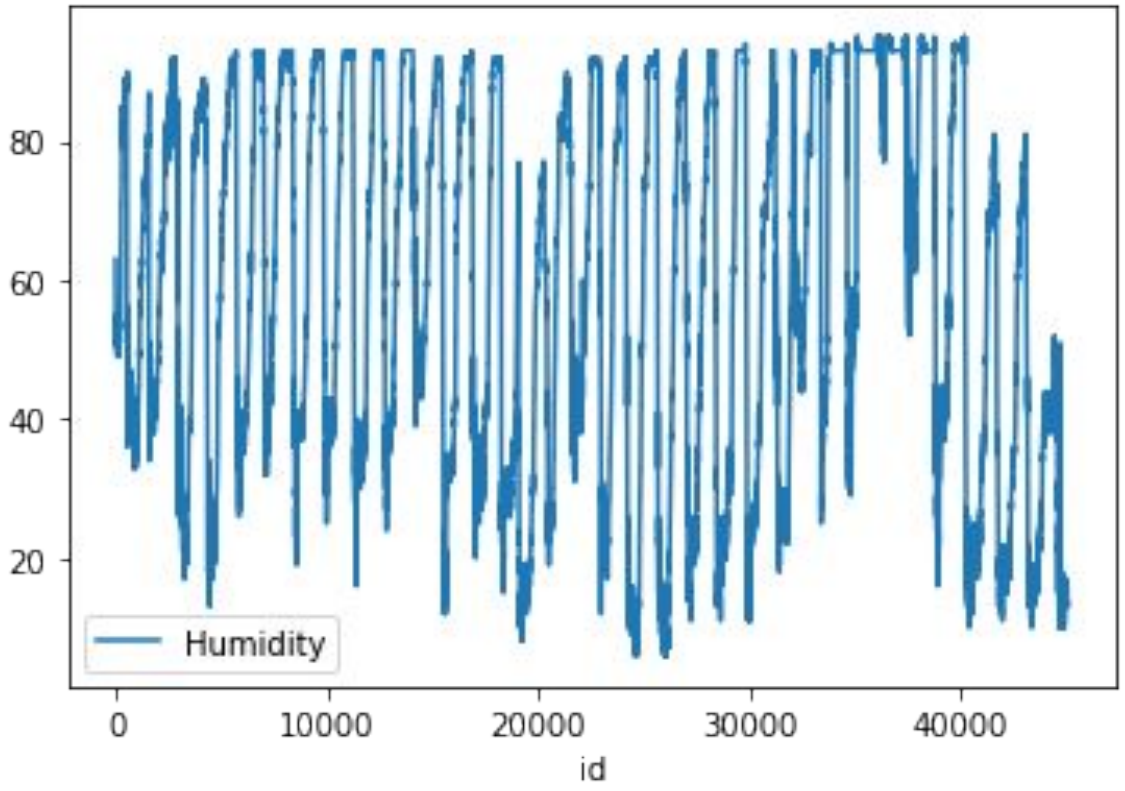
RESULTS

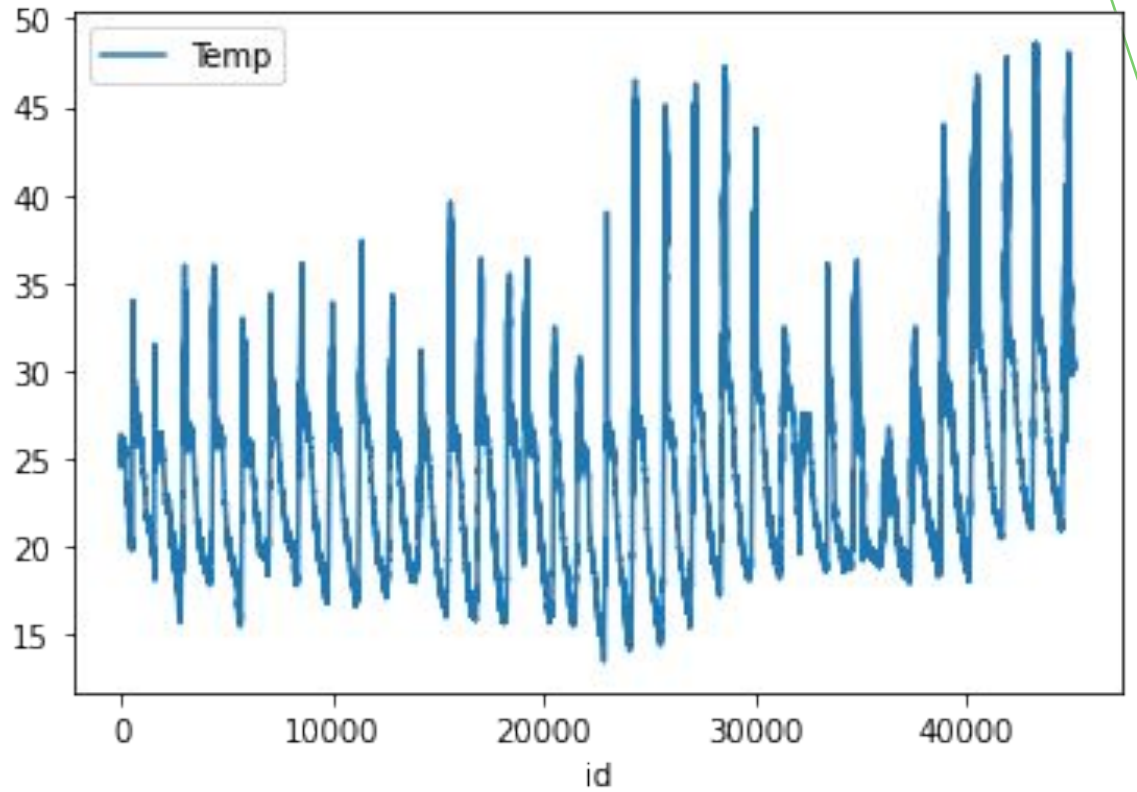




Data Visualisation With Python

RESULTS





Data Visualisation With Python

RESULTS

```
Epoch 40/50
52/52 [==============================] - 0s 1ms/step - loss: 0.0233 - acc: 0.9156
Epoch 41/50
52/52 [==============================] - 0s 1ms/step - loss: 0.0241 - acc: 0.9177
Epoch 42/50
52/52 [==============================] - 0s 1ms/step - loss: 0.0192 - acc: 0.9366
Epoch 43/50
52/52 [==============================] - 0s 1ms/step - loss: 0.0232 - acc: 0.9193
Epoch 44/50
52/52 [==============================] - 0s 1ms/step - loss: 0.0203 - acc: 0.9337
Epoch 45/50
52/52 [==============================] - 0s 1ms/step - loss: 0.0225 - acc: 0.9272
Epoch 46/50
52/52 [==============================] - 0s 1ms/step - loss: 0.0254 - acc: 0.9117
Epoch 47/50
52/52 [==============================] - 0s 1ms/step - loss: 0.0227 - acc: 0.9209
Epoch 48/50
52/52 [==============================] - 0s 1ms/step - loss: 0.0201 - acc: 0.9327
Epoch 49/50
52/52 [==============================] - 0s 1ms/step - loss: 0.0219 - acc: 0.9275
Epoch 50/50
52/52 [==============================] - 0s 1ms/step - loss: 0.0224 - acc: 0.9230
<tensorflow.python.keras.callbacks.History at 0x7ff6c736c2b0>
```

Model Training With Python TensorFlow

RESULTS

	Proposed	S. S. Bhatkande1 et al. [3]	Y. Radhika et al. [5]
Accuracy	78.8%	**82.62%**	**80%**
Parameters used	Temperature, Humidity and Light Intensity	Max and Min Temperature, Humidity and Wind Speed	Max and Min Temperature, Rainfall, Cloud Conditions And Wind Stream
Algorithm used	Logistic regression	Decision Tree	Artificial Neural Network

Comparison With Existing Systems

CHALLENGES


Privacy/Security/Safety
Complexity
Ethical and legal Issues
Compatibility
Challenges facing AI in IoT
Artificial Stupidity

APPLICATIONS

- Severe weather alerts and advisories.
- System can be modified to be used at commercial level.
- Have many applications in smart homes, buildings, sports, hospitals etc.

CONCLUSION

- The real time weather prediction system presented in this project has been developed around low cost IoT board and sensors.
- The system has been deployed in an indoor environment and values of the parameters have been recorded in Google spreadsheet.
- A Logistic regression model has been used in Jupyter notebook environment that is trained with prerecorded values of parameters and used to predict the weather parameters in real time environment.
- The result of the model is also compared with the other works available in literature and the proposed system is slightly better in terms of accuracy.

REFRENCES

1. Linear regression-

   https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a

2. Random Forest algorithm -

   https://builtin.com/data-science/random-forest-algorithm

3. Introduction to Machine Learning

   https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-eed6024fdb08

4. Study related to IOT sensors

   https://www.instructables.com/IoT-Temperature-Sensor-With-ESP8266/