UNIVERSITY of RWANDA

Website: www.aceiot.ur.ac.rw

Mail: aceiot@ur.ac.rw

College of Science and Technology

AFRICAN CENTER OF EXCELLENCE IN INTERNET OF THINGS

Research Thesis Title:

IoT-BASED CLIMATE CHANGE PREDICTION SYSTEM

Case Study: MASORO INDUSTRIAL ECONOMIC ZONE

A dissertation submitted in partial fulfillment of the requirements for the award of masters of science degree in the Internet of Things: Wireless Intelligent Sensor Networking

Submitted By:

Marie Louise NIRERE **(REF.NO:** 221001473**)**

DECEMBER 2022

Website: www.aceiot.ur.ac.rw

Mail: aceiot@ur.ac.rw

College of Science and Technology

AFRICAN CENTER OF EXCELLENCE IN INTERNET OF THINGS

Research Thesis Title:

IoT-BASED CLIMATE CHANGE PREDICTION SYSTEM

Case Study: MASORO INDUSTRIAL ECONOMIC ZONE

A dissertation submitted in partial fulfillment of the requirements for the award of masters of science degree in the Internet of Things: Wireless Intelligent Sensor Networking

Submitted by:

Marie Louise NIRERE **(REF.NO:** 221001473**)**

Supervised by:

Prof. Kayalvizhi Jayavel

Dr. NGENZI Alexander

DECEMBER 2022

DECLARATION

I, Marie Louise NIRERE, hereby declare that this research thesis titled "**IoT BASED CLIMATE CHANGE PREDICTION SYSTEM**" for the award of a Master's degree in Internet of Things-Wireless Sensor Networking is my original work and that it has not previously been submitted for any academic award in this or any other institution of higher learning for the academic award or any other purpose. The references to other journals or materials used in this document are listed in the references section.

Marie Louise NIRERE

REF. No: 221001473

Signature:

Date: 03/03/2023

Master Student in the Internet of Things-Wireless Intelligent Sensor Networking.

BONAFIDE CERTIFICATE

This is to certify that the project named "**IoT BASED CLIMATE CHANGE PREDICTION SYSTEM**" is a record of the original work of NIRERE Marie Louise (Ref. No. 221001473) done in partial fulfillment of the requirements for the award of a Master's Degree in the Internet of Things-Wireless Intelligent Sensor Networking from the University of Rwanda-College of Science and Technology through the African Center of Excellence in the Internet of Things, in the academic year 2020-2022.

Main Supervisor:

Prof. Kayalvizhi Jayavel

Date: 02/03/2023

Signature:

Co-Supervisor:

Dr. Alexander NGENZI

Date:

Signature:

Head of Masters and Training ACEIOT

Dr. James Rwigema

Date:

Signature:

DEDICATION

This research thesis is lovingly dedicated:

To Almighty God, who serves as my guide and source of strength and knowledge, for without His guidance, everything is futile.

To my lovely Husband, who has provided me with love, strength, support, patience, and motivation throughout this entire experience.

To my family, whose love and inspiration always steer me in the right direction.

To my friends and classmates for their cooperation in making this research project possible.

To the African Centre of Excellence in the Internet of Things (ACEIoT) management for the skills and knowledge, I gained from them during my master's studies.

ACKNOWLEDGEMENT

First of all, I would like to acknowledge the Almighty God for helping me from the beginning to the end of my studies.

The preparation of this research thesis required the effort, support, and guidance of many people whom I would like to thank.

My heartfelt gratitude goes to my supervisors, Prof Kayalvizhi Jayavel and Dr. Alexander NGENZI, for devoting their time to supervising and monitoring this project. Their leadership, duties, and numerous meetings for input and counsel to me all contributed to the effective completion of this project.

My deepest gratitude goes to the University of Rwanda / College of Science and Technology / African Centre of Excellence in the Internet of Things for their contributions to my education.

I extend my thanks to classmates who helped me morally and materially and shared their knowledge to improve my knowledge and for their good cooperation during our studies.

Finally, my deep appreciation goes to my husband, family and friends for their unforgettable contribution, help, guidance, care, prayer, and amazing love during this period.

May God bless you all!!

ABSTRACT

Climate change is one of the most significant challenges to every country's development, causing uncountable ravages in the life of all living around the globe. An increase in carbon dioxide emission, one of the main greenhouse gases has been gradually recorded from burning fuels and natural gases in generating power and heat, manufacturing processes, and deforestation in general. Numerous research and studies of strategies for tracking climate change have been raised by researchers. In Rwanda, the existing climate change tracking method uses a weather station model, in which numerous fixed weather stations are installed around the nation, however, due to its immobility; this process can not cover the whole country. With the lack of advanced methodologies and technology, the process of climate change tracking has become extremely expensive and suffered inaccuracies due to a lack of proper knowledge of which parameters to collect, knowledge of analyzing collected data, and the lack of specific accurate hardware.

Throughout this research, with the use of the MQ-135 sensor and DHT11 sensor, ESP8266 collects carbon dioxide gas and temperature/humidity respectively and other component include a push button for detecting the current season. Data from these sensors and push button are serially connected to it. With ESP built-in Wi-Fi capability, ESP8266 is programmed to send data over MQTT protocol which relies on Wi-Fi capability to send data to MQTT Broker which is hereby referred to as MQTT Box. With the Publish/Subscribe criteria of the MQTT protocol, node-red subscribes to the topics defined in MQTT broker to get data, which is sent to MongoDB for permanent storage, and also fed to the machine learning model for prediction of climate change/warming, this model is built by using Jupyter notebook which is a good tool for python users. This model is evaluated by using different machine learning classification algorithms for optimizing the prediction accuracy. Random Forest approves itself to be the best model in evaluating the built model with 99% of training accuracy and 84% of testing accuracy. The study shows that the increase in carbon dioxide gas leads to the gradual increase in the environmental temperature. Finally, the prediction clarifies that if no measures are taken presently, the climate change in Masoro's industrial zone will be dominated by warming periods in 2023.

Keywords: IoT, MQTT, Node Red, MongoDB, Machine Learning, climate change.

Table of Content

DECLARATION ii
BONAFIDE CERTIFICATE iii
DEDICATION iv
ACKNOWLEDGEMENT v
ABSTRACT vi
List of Figures ix
List of tables x
List of Symbols and Abbreviations xi
CHAPTER 1. INTRODUCTION 1
1.1 Background and Motivation 1
1.2 Problem Statement 2
1.3 Study Objectives 3
❖ General Objectives 3
❖ Specific Objectives 3
1.4 Hypothesis 3
1.5 Scope of the study 4
1.6 Significance of the study 4
1.7 Organization of the study 4
CHAPTER 2. LITERATURE REVIEW 5
2.1 Related Works 5
2.2 Gaps 9
2.3 Proposed solution 9
CHAPTER 3. RESEARCH METHODOLOGY 10
3.1 Overview 10
3.2 Data Collection 10
3.3 Visualization of collected data 10
3.4 System Architecture 12
3.4.1 The hardware Part 13
3.4.2. The Software Part 14
3.5 Data cycle 14

3.3.1 MQTT server ... 15
3.3.2 Node-red ... 15
3.3.3 MongoDB ... 15
3.6 Model Training and Evaluation ... 15
3.7 Introduction to building Model and explanation of algorithms ... 16
3.7.1 Python built model ... 16
3.7.2 Facebook Prophet Model ... 17
CHAPTER 4. SYSTEM DESIGN AND ANALYSIS ... 18
4.1 Overview ... 18
4.2 System Prototype ... 18
4.3 Design Description ... 19
4.3.1 Subscribing and publishing in the MQTT box ... 20
4.3.2 Building flows in Node-Red ... 20
4.3.3 MongoDB database ... 23
CHAPTER 5. RESULTS AND DISCUSSION ... 25
5.1 Overview ... 25
5.2 Data collection in Arduino and visualization in Serial Monitor ... 26
5.3 Study of correlation between parameters and the check for outliers ... 28
5.4 Model Building ... 29
5.4.1 Accuracy Measures ... 30
5.5 Training and evaluation implementation ... 30
5.6 Evaluation Metric for Predicted outputs ... 33
5.7 Comparative Analysis ... 34
5.8 Prophet Facebook model ... 35
5.9 Live data visualization ... 35
CHAPTER 6. CONCLUSION, RECOMMENDATION AND FUTURE WORK ... 36
REFERENCE ... 37

List of Figures

Figure 1 Design of the existing system based on data collection (MeteoRwanda, 2022) 2
Figure 2 Temperature measurements over time 11
Figure 3 Humidity measurements over time 11
Figure 4 CO2 measurements over time 12
Figure 5 System Architecture 12
Figure 6 DHT11 Sensor 13
Figure 7 MQ-135 13
Figure 8 Digital Push Button 13
Figure 9 NodeMCU / ESP8266 Microcontroller 14
Figure 10 GSM Module 14
Figure 11Model building process with Keras 16
Figure 12 System Prototype 18
Figure 13 System design 19
Figure 14 MQTT Overview 20
Figure 15 Nodes deployed in IoT based Climate change prediction 20
Figure 16 Nodes for visualization of climate change parameters 21
Figure 17 MQTT-in Node 21
Figure 18 function node 22
Figure 19 MongoDB out node1 22
Figure 20 MongoDB out node2 23
Figure 21 Message debugging node 23
Figure 22 MongoDB-Python data flow 23
Figure 23 MongoDB Compass interface 24
Figure 24 Adding subscription to MQTT topics 28
Figure 25 Study of the correlation between parameters and check for outliers 28
Figure 26 Model study with confusion matrix 29
Figure 27 Train-Test Model Evaluation 30
Figure 28 Model evaluation with Random Forest Classifier 31
Figure 29 Model evaluation with Multilayer Perceptron Classifier 31
Figure 30 Model evaluation with Logistic Regression 32
Figure 31 Model evaluation with Gradient Boost Classifier 32
Figure 32 Model evaluation with Decision Tree Classifier 33
Figure 33 Model evaluation with K Nearest Neighbor Classifier 33
Figure 34 Prophet Model-Based Prediction 35
Figure 35 Live Data Visualization in charts and gauges 35

List of tables

Table 1 Metrics of Algorithms 34
Table 2 Comparative analysis with different algorithms 34
Table 3 Parameters measurements 40
Table 4 Prediction vice Actual Value 40
Table 5 Logistic regression classification report 40
Table 6 MLP classification report 40
Table 7Gradient Boost classification report 41
Table 8 KNN Classification report 41
Table 9 Random Forest Classification report 41
Table 10 Decision Tree classification report 41

List of Symbols and Abbreviations

~: Approximation

ANN: Artificial Neural Network

CO2: Carbon dioxide gas

DB: Database

DC: Direct Current

DHT11: Digital Temperature and Humidity sensor

ESP8266: Espressif 8266

GB: Gradient Boost

GDP: Gross Domestic Product

GHG: GreenHouse Gas

GSM: Global System for Mobile Communications

IDE: Integrated Development Environment

IoT: Internet of Things

KNN: K Nearest Neighbor

LR: Logistic Regression

ML: Machine Learning

MLP: Multilayer Perceptron

MQTT: Message Queuing Telemetry Tool

NodeMCU: Node Microcontroller Unit

PDA: Personal Digital Assistant

PPM: Parts per million

RF: Random Forest

RH: Relative Humidity

SIM: Subscriber Identity Module

SMS: Short Message Service

SVM: Support Vector Machine

Wi-Fi: Wireless Fidelity

CHAPTER 1. INTRODUCTION

1.1 Background and Motivation

Weather forecasting involves predicting the weather conditions for a specified region or area. More importantly, accurate climate forecasting is required for day-to-day activities. Weather forecasts are important to both living and non-living things (1).Continued research in the areas of the Internet of things (IoT) and machine learning has allowed researchers to move on to various weather forecast models. Climate change is a major problem for people all around the world. However, the Internet of Things (IoT) provides a variety of resources and technologies to assist businesses and governments in reducing the negative impact of human activities on the environment (2). The Internet of Things has the potential to help us decarbonize our energy system, deliver modern energy systems to every human being, manage our infrastructure, and adapt to and solve climate change (3).IoT advancements helped companies, governments, and people to mitigate climate change without losing convenience. More consumers are expected to participate in activities that allow businesses and governments to collect and analyze Big Data for environmental purposes. Connected devices pave the way for the future of climate change and environmental health (3).

In Rwanda, the problem of accurately forecasting the weather persists. Meteo Rwanda's current method is the weather station model, which involves the installation of many static weather stations around the country. Massive observatories with powerful equipment (radar, satellite, etc.)(1) now do the work, although they are limited by technological advancements and are manufacturer-dependent; nonetheless, because of the use of manual strategies and the immobility of such stations, there is no real-time climate change information and the method of predicting climate change has low accuracy and is extremely-expensive.

Meteo Rwanda weather parameters are also collected from those stations by volunteers using the USSD application (see fig 1), where volunteers must look at and manually record the observed data, which is not real-time. The volunteer cloud application is used to store data sent from weather stations by USSD Volunteers, which is then reviewed and downloaded in Excel format before being uploaded to the Meteo Local Server for processing and data analysis.

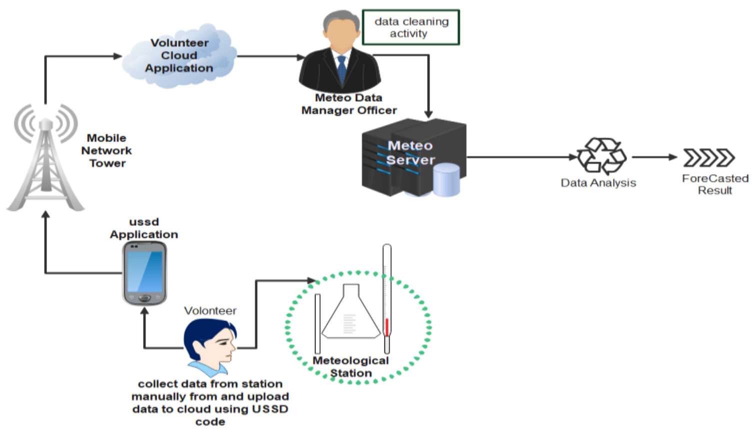


Figure 1 Design of the existing system based on data collection (MeteoRwanda, 2022)

With the help of IoT and models built on machine learning algorithms, it is possible to accurately monitor and predict climate change in any environment based on environmental parameters, which is why I was motivated to work on this project titled "IoT based climate change prediction system," which is a portable real-time IoT weather forecast system built using very low-cost components, with ease of implementation. This system model's implementation focuses on predicting whether it will be a warm or cold season by tracking environmental indicators such as CO2, temperature, and humidity parameters.

1.2 Problem Statement

Meteo Rwanda does not monitor environment with regards to climate change. Only weather monitoring is made considering only minimum temperature and maximum temperature of the environment. The data collected cannot fully contribute in climate change monitoring. There are no low-cost sophisticated methods or technologies for collecting and analyzing large amounts of data in order to effectively predict the weather.

Climate change is the term that scientists are currently using to describe the complex changes caused by greenhouse gas concentrations that may already be impacting the planet's weather and initiating global warming and climate systems (2).Carbon dioxide is the main greenhouse gas produced by burning organic materials such as coal, oil, gas, wood, and solid waste (3).Thus, predicting the level of CO2 in the air is important to make decisions based on analyses to decrease air pollution. Agriculture is one of the fields that is hugely affected by climate change, and the quality and quantity of the yield are negatively impacted too. Increased animal mortality rates due to fauna and flora destruction.

As a result, it's critical to fix those problems with the help of this proposed system" IoT-based climate change prediction system", which will also be a Rwandan IoT project. This project will address those challenges of collecting and analyzing relevant data from various areas using appropriate methods and technologies and minimizing costs to provide real-time climate change information while enabling useful and accurate predictions of future climate change with a focus on temperature and carbon dioxide (CO2) levels in the industrial economic zone.

1.3 Study Objectives

❖ General Objectives

The goal of this thesis is to design and implement an IoT-based climate change prediction system to monitor Masoro industrial economic zone, considering environmental parameters of temperature, humidity, and harmful gas concentrations in the air that lead and contribute to the warming of the environment; to provide real-time data, and to predict climate change through remote weather monitoring.

❖ Specific Objectives

- To study the existing weather monitoring system by analyzing how the climate change information is collected and how prediction is made.
- To develop an IoT weather station that would be installed in an industrial or manufacturing zone to collect data and save costs while improving accuracy.
- Implement a climate change prediction model, focusing on carbon dioxide levels (CO_2), temperature, and humidity.
- To provide real-time climate change information to the general public in Masoro industrial economic area

1.4 Hypothesis

With the help of IoT and models built upon machine learning algorithms, it is possible to monitor and predict climate change in any environment based on environmental parameters that include carbon dioxide, temperature, and humidity. With this process, it is possible to take preventive measures before any damage occurs, pursuing various strategies to reduce CO2 emissions in the environment and meeting the country's GDP by increasing clean productivity and clean and safe environments.

1.5 Scope of the study

This system is designed to be implemented in Rwanda, in the Masoro industrial economic zone located in the Gasabo district. This system is designed to collect weather data that includes carbon dioxide, humidity, and temperature from that area. To provide real-time information, an MQTT broker is built, that receives data from Esp8266. The collective sensors are connected in serial communication with the Esp8266 microcontroller that is enabled with Wi-Fi communication capability. From MQTT Broker, the data is sent and published to the Node-red which is installed locally as a local server. The MQTT broker also publishes data to the MongoDB for data permanent storage, and through MQTT publish/subscribe criteria, topics data are spread to different areas, which includes a predictive model built in python based Jupyter that keeps on learning from the data threads, a model that is evaluated by using different machine learning algorithms.

1.6 Significance of the study

This thesis aims to design and implement an IoT-based climate change prediction system that monitors environment parameters that include carbon dioxide gas, temperature, and humidity of the environment, parameters that hugely affect climate change and warming, as well as provide real-time climate change data and forecast warming of an environment as well. This project delivers a remote monitoring framework that relies on lower power and low-cost IoT paradigm to monitor climate change in industrial zones. The system is state-of-the-art, lightweight, real-time, modular, and versatile, and allows data exchange through state-of-the-art technologies.

1.7 Organization of the study

Chapter one (I): This chapter introduces the research thesis's background and describes the research thesis's motivation, description of the problem statement that set the objectives, study scope, significance, and organization study for this project.

Chapter two (II): Presents a review of related works.

Chapter three (III): Indicates the methodology used for this Research thesis.

Chapter four (IV): This chapter discusses the system analysis and design

Chapter five (V): Discusses obtained results and makes analysis based on graphs findings.

Chapter six (VI): Concludes the discussion and gives final remarks and give recommendations for future researchers.

CHAPTER 2. LITERATURE REVIEW

Various weather forecast models have risen as a result of ongoing research in the disciplines of the Internet of Things and Machine Learning. However, the problem of accurately forecasting or predicting the weather persists. Weather prediction has been a major challenge from the early days, new methodologies cluster every day replacing the old ones. Many research studies are conducted to improve the performance of prediction algorithms. A commonly used approach is to use several prediction algorithms in combination to get better accuracy.

2.1 Related Works

M.K. Nallakaruppan et al. conducted the study based on comprehending the weather prediction inconsistencies and in-proficiency in light of linear regression algorithms and time series models (1).The significant commitment of this research is to formulate a productive weather prediction model based on a Decision Tree and Time Series Analysis. This study did not mention the approach that would be utilized to collect big data to get the most accurate results.

The Applicability of Big Data in Climate Change research presented an overview of the interrelationship between data science and climate studies, as well as a description of how Big Data tools may be used to manage sustainable climate challenges. This research aimed to highlight the perspective of systems of systems (SoS) as the drivers and effects of climate as well as that their resilience and adaptation cannot be determined without the exploration of the synergies between new research trends and disciplines. It also discussed Big Data analytics technologies and their contributions to understanding the features of climate change, as well as climate action-related equivalents such as sustainability and social sciences, which are critical for the effective development and implementation of strategies (4).

IoT-based climate prediction using ANN for green networking is a proposed topology formulated an IoT-based weather monitoring using artificial neural network (ANN) for prediction. The main objective of this designed system is monitoring weather parameters such as temperature/humidity/ pressure/ rainfall etc and employing ANN for forecasting the occurrence of a flood using temperature and rainfall data using LM and GD learning algorithms (5).

Israr Ullah et al. presented a general conceptual model for learning and improving the accuracy of prediction algorithms in dynamic conditions. The proposed model evaluation and experimental analysis are conducted in a Greenhouse environment to accurately predict indoor climate conditions from noisy sensor readings and the ANN algorithm is used to improve the performance of the conventional Kalman filter algorithm in dynamically changing external conditions. This study suggested that in the future, deep learning algorithms (rather than ANN) be used to tune the performance of other prediction algorithms and that experimental analyses with big data be conducted in more complex real-world applications to further establish the validity of the proposed learning to prediction mechanism (6).

The study, titled "An Intelligent Weather Prediction System based on IoT" proposed an improved solution for weather monitoring using low-cost GPS-enabled IoT devices, connected with different sensors. The sensed information is stored in the server to predict the weather parameters such as temperature, humidity, air pressure, etc. of any defined application region of interest. This prediction can take place from the collected information to the predicted data set utilizing some soft computing tools like SVM, KNN, DNN, Linear Regression, etc. (7).

An integrated approach for estimating greenhouse gas emissions from 100 U.S. metropolitan areas found that Cities have become key players in climate change mitigation policy. To develop their climate policies, cities need good assessments of their current and future emissions. This study used publically available national datasets to develop an integrated approach for estimating GHG emissions at the metropolitan level over time, between multiple locations, and across sectors (8). According to the findings of this study, there is a lack of an appropriate method to address greenhouse gas emissions as well as sufficient large-scale data to measure progress.

Tsungnan Chou et al.presented the study with the intention of extracting useful features from cardholder data and applying artificial intelligence techniques to assist financial institutions in reducing their credit card default risk. The cardholder's personal information as well as repayment details were used to track consumer behavior for the predictive models. This study also aims to compare the prediction accuracy of newly developed deep learning methods to that of conventional machine learning based on credit card holders' financial information. The experiment results showed that the deep neural network outperformed the machine learning models in most evaluation metrics and achieved an impressively high accuracy of 0.93 (9).

Priya Singh et al. conducted research with the goal of developing an efficient and accurate system to inspect cardiovascular diseases, and the system is based on data mining techniques that can help to remedy such a situation. The system is built using classification algorithms such as Random Forest, Logistic Regression, Naive Bayes' and Support Vector Machine, as well as feature selection algorithms such as Pearson Correlation and Chi-Square, to improve classification system accuracy and reduce execution time. As a result, it was discovered that Logistic Regression had the highest accuracy of 84% when compared to the others (10).

Orlando P. Zacarias et al.in (11) used machine learning algorithms Support Vector Machine (SVM) and Random Forest (RF) to predict malaria cases in Mozambique.This study's dataset included records of malaria cases as well as climate data such as temperature, rain fall, and humidity. The study's findings revealed that the support vector machine algorithm outperforms the random forest and decision trees classifiers in terms of mean squared error (MSE).

Sumedh Kanind et al. presented a system for predicting stock prices for the next five years using Facebook Prophet, which can be used to make better investments. This makes it simple to choose which stock to invest in based on the predictions of the highest percentage of returns in a given period of time. Several other features of Facebook Prophet can be used to improve prediction accuracy while also making the application interactive and simple to use. Stock price predictive models have been developed and tested using publicly available stock data obtained from Yahoo Finance. By utilizing regression models, Prophet is capable of generating daily, weekly, and yearly seasonality, as well as holiday effects. The experimental results indicate that Facebook Prophet can be used to predict stock prices with reasonable accuracy over a long period of time (12).

Predicting Malarial Outbreak using Machine Learning and Deep Learning Approach is a review and analysis that focuses on determining which machine learning classification models, such as Support Vector Machine, K-Nearest Neighbors (KNN), Logistic regression, Random Forest, Gradient Boosting, Neural network, and Naive Bayes, were best suited for malaria case prediction using meteorological data malaria cases data collected over a six-year period. The study found that it is possible to use machine learning techniques to estimate the likelihood of future malaria cases and assist stakeholders in preventing future deaths from malaria. This study's dataset was divided into two independent sub-datasets, with 80% used for training and

20% used for testing. The overall outcomes of machine learning algorithms revealed that Gradient Boosting, Artificial Neural Networks, Random Forest, and Support Vector Machine achieved the best performance. Unfortunately, the researcher does not provide complete information on the model's performance on both training and testing data (13).

Ardvin Kester S. Ong et al. carried out the research with the goal of predicting and categorizing factors influencing the actual usage of the Morchana COVID-19 contact tracing mobile application. Furthermore, the use of machine learning algorithms (MLA) was considered to demonstrate how it could be used to assess factors influencing human behavior following the suggestion and considered limitations of traditional multivariate analysis such as SEM among several studies. MLA such as Random Forest Classifier (RFC) and Artificial Neural Network (ANN) have been used to assess significant factors and importance affecting the actual use of MorChana. According to the validated results, Hedonic Motivation and Facilitating Conditions were important factors that would lead to very high actual use. Furthermore, Habit and Understanding COVID-19 were important factors that would lead to widespread use of the MorChana COVID-19 contact tracing mobile application. It was discovered that when people understand the impact and causes of the COVID19 aftermath, its severity, and a way to mitigate it, they are more likely to use a system that caters to the benefits of health-related systems or technology (14).

B. Manjulatha et al. conducted the study using machine learning algorithms to help in the early prediction of diseases, which saves many lives around the world. Datasets such as the Pima Indian diabetes dataset (PIMA), the Indian liver patient data (ILPD), and the cardiovascular disease (CVD) are taken from the UCI repository and compared using various well-known algorithms. Each algorithm generates its own output, but determining the highest accuracy is difficult because each algorithm gives a different result. In this system, accuracy is increased by combining different algorithms such as decision trees, SVM, logistic regression, ANN, random forest classifiers, and KNN to create an ensemble hybrid model that provides more accurate, reliable results. The datasets for diabetes, liver disease, and heart disease are compared in this study using machine learning methods. According to the findings, the ensemble hybrid model's classification accuracy can reach 80%.The researchers did not, however, concentrate on applying deep learning algorithms to increase accuracy (15).

Rwanda is vulnerable to climate change as a whole. Climate change, if it occurs as predicted, would worsen Rwanda's current environmental, social, and economic concerns. Based on evaluations of 19 Global Circulation Models (GCMs) and 15 GCMs, Rwanda as a whole is anticipated to see a 0-10% rise in mean precipitation and a <2-4 degrees Celsius increase in mean temperature until 2080. The uncertainties are substantial, and they are projected to differ widely across the country (16). MeteoRwanda's main goal is to provide accurate, timely weather and climate information and products for the general welfare of Rwandans by improving weather and climate information reliability, increasing the number of observation stations from 334 to 340 by the end of 2024, and increasing the number of meteorological information and product users (2).

2.2 Gaps

According to the findings of the above existing systems, there is a lack of an appropriate method and technologies that can collect and analyze large amounts of data, give real-time information, and accurately predict the weather where powerful computations are required for complex analysis of large data sets, lack of useful strategy of addressing greenhouse gas (especially C02) emissions issue which affects our climate change. Meteo Rwanda's current strategy for weather prediction involves installing many static weather stations in the various districts of Rwanda, which requires a large number of components, most of which are extremely expensive, and due to the use of such pricey fixed weather stations, there is a problem that reliable data cannot be collected from all areas to provide real-time climate change information and an accurate result for the forecast.

2.3 Proposed solution

It is possible to accurately monitor and predict climate change in any environment using IoT and models built on machine learning algorithms, which is why I have been motivated to develop a proper, efficient, and reliable IoT-based system that will address the challenges of collecting and analyzing relevant big Data from various areas, particularly the industrial economic zone using appropriate methods and technologies while minimizing costs to provide real-time climate change information to the general public while enabling useful and accurate predictions of future climate change with a focus on temperature, humidity and carbon dioxide (CO_2) levels.

CHAPTER 3. RESEARCH METHODOLOGY

3.1 Overview

In this section, the research methodology techniques employed throughout the study are briefly explained. These include data collection, data visualization, system architecture of real-time IoT-based systems; training and evaluating machine learning models; and, finally, integrating IoT systems with machine learning predictive models to perform predictions on real-time data from field sensors.

3.2 Data Collection

The methods chosen were guided by the nature of the data to be collected, the time available, and the study's objectives, which are as follows:

Observation, interviews with concerned sectors and MeteoRwanda, as well as access internet resources for finding climatic change-related information and reports provided by organizations, institutions, and even Ministries.

The data acquisition method is also used to capture and gather environmental parameters using carbon dioxide sensors, humidity and temperature sensors, as well as a push button for detecting the current season, which will aid in the design and implementation of an effective climate change monitoring and prediction system.

The environmental data collected to model climate change include the current season status, carbon dioxide, humidity, and temperature recorded at Masoro Industrial Economical Zone in Gasabo district and aggregated on a daily basis.

3.3 Visualization of collected data

The environmental data collected to model climate change include temperature, relative humidity, and carbon dioxide levels recorded from Masoro industrial zone and aggregated on a daily basis. The figures below depict the distribution of those environmental variables based on the values of the data recorded with their corresponding time sequence.

With the DHT11 sensor, the temperature measurements over time are shown in figure 2, with a range mainly varying from 18 oC to 45 oC.

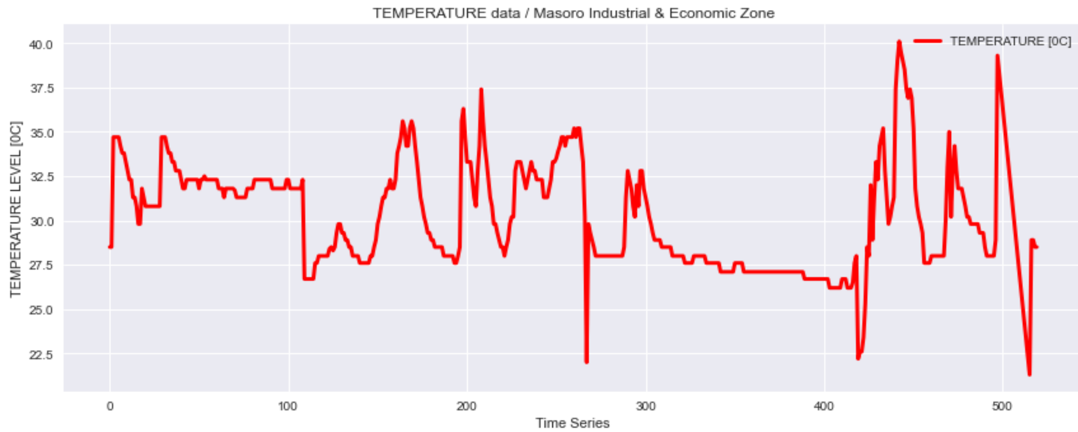


Figure 2 Temperature measurements over time

The corresponding humidity measurements over time are mapped as figure 3 shows, ranging from 20 % RH to 80 %RH.

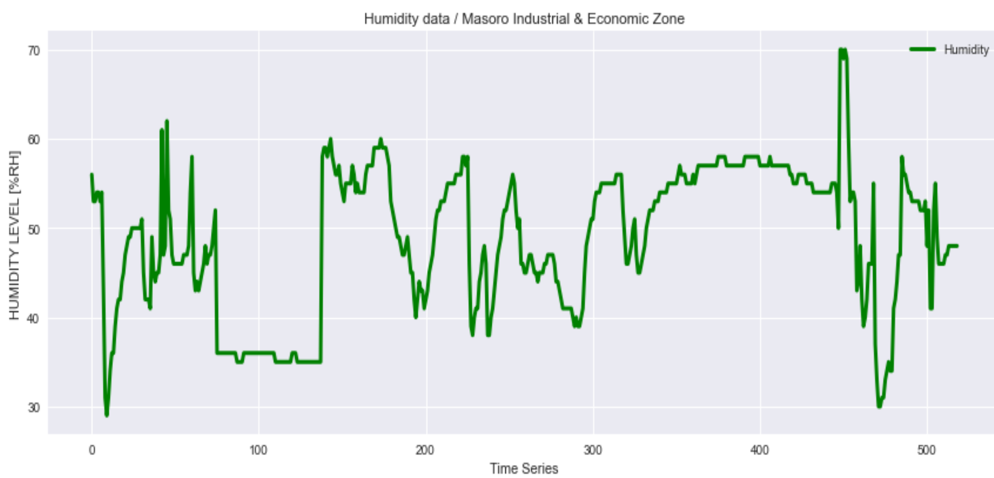


Figure 3 Humidity measurements over time

Figure 4 shows the readings of carbon dioxide over time. Throughout this research project, carbon dioxide is measured by an MQ-135 gas sensor, ranging from 0 to 600 PPM [Parts Per Million]. Other readings beyond that range are considered outliers and are removed during the data cleansing and wrangling process. For deployment, CO_2 normal range is between 200 PPM and 800 PPM. While cooking with a gas stove, this range may rise to 1600 PPM. In the dumb bins with rotten foods, this range rises to 2200 PPM which is considered the maximum.

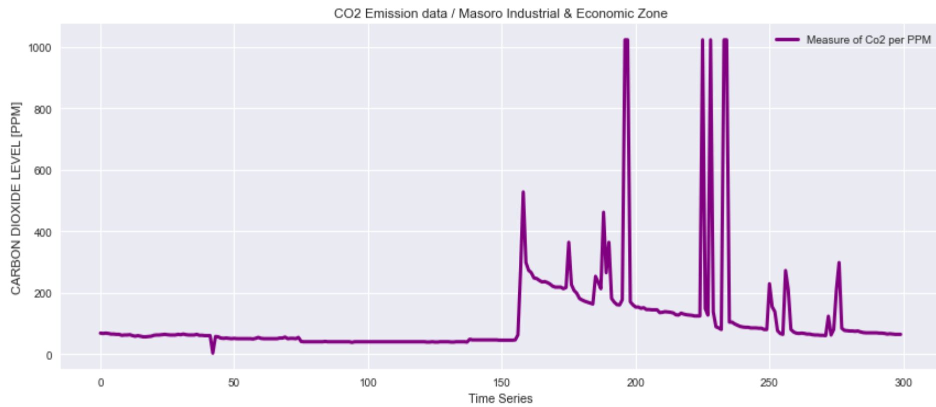


Figure 4 CO2 measurements over time

3.4 System Architecture

This system architecture is drawn in figure 5. The sensing component of the data collection system under consideration in this study includes three sensors for measuring ambient temperature, relative humidity, and carbon dioxide. Other components include a push button for detecting the current season and a power source to generate the needed power in the circuit, as well as Nodemcu (ESP8266), a microcontroller platform with Internet connection capability for transmitting processed sensor data to the IoT cloud via existing Internet-based network connectivity. The dataset in this instance is created and analyzed using a machine learning model to make predictions. A mobile and web application is also created to help a variety of users access real-time climate change data from cloud storage as well as the prediction outcomes whenever they want using any user interface device (phone, machine, etc.).

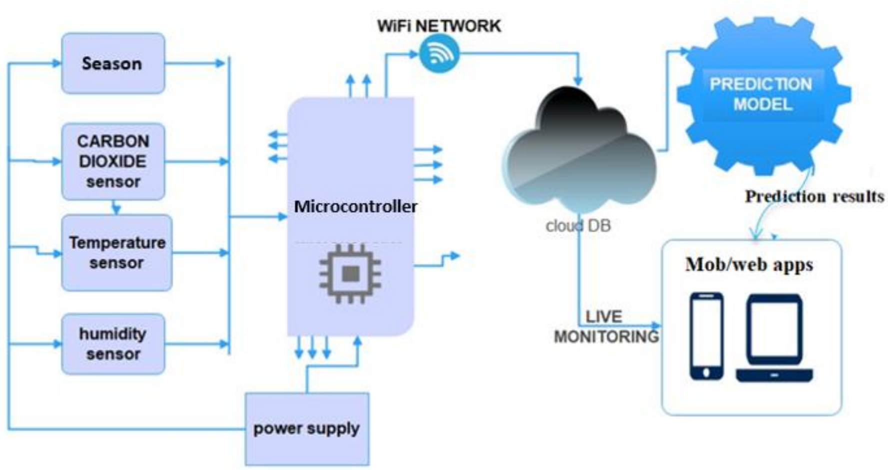


Figure 5 System Architecture

The system is made up of the following components, which include both hardware and software:

3.4.1 The hardware Part

This part includes environment monitoring sensors, a microcontroller, a push button, and connectors.

The sensors are as listed below:

- **DHT11 sensor** (17): is a basic, low-cost digital temperature and humidity sensor. It measures the surrounding air with a capacitive humidity sensor and a thermistor and outputs a digital signal on the data pin, as shown in figure 6.

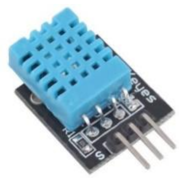


Figure 6 DHT11 Sensor

- **MQ-135**: is a popular gas sensor from the MQ series that is commonly used in air quality control equipment as shown in figure 7. It can detect various gases, including CO_2, and operates between 2.5V and 5.0V, consuming approximately 150mA (18). It requires some pre-heating before it can produce accurate results, and it can output both digital and analog signals.

Figure 7 MQ-135

- **Digital push button** (19): is employed to give a triggering pulse that sends a 1 to the microcontroller when the climate is warm and sends and 0 when the climate is cold as shown in figure 8.

Figure 8 Digital Push Button

- **NodeMCU(ESP8266)** (20): is a WiFi-capable microcontroller development board. It employs an ESP8266 microcontroller chip, as well as an integrated Wi-Fi communication protocol and multiple input/output pins functions as shown in figure 9.

Figure 9 NodeMCU / ESP8266 Microcontroller

- **GSM Module (21)**: stands for Global System for Mobile communication. This concept was developed by Bell Laboratories in 1970. In the entire world, it is a widely utilized mobile communication system. GSM is an open, digital cellular technology that uses the 850MHz, 900MHz, 1800MHz, and 1900MHz frequency bands to provide mobile voice and data services. GSM-based module has an antenna that it uses for transmission as shown in figure 10.

Figure 10 GSM Module

3.4.2. The Software Part

This part has C/C++ Arduino codes to command the microcontroller what to do and how to do it, and the Esp8266 module has a built-in Wi-Fi protocol.The data fetched and processed is sent to the cloud with the help of the MQTT Protocol. The MQTT protocol is built upon the act of subscribing and publishing (18).

3.5 Data cycle

Throughout this research project, the data moves from side to side, from collection to cleaning to visualization to storage. From storage to other usages that include training a predictive model built by using Python.

3.3.1 **MQTT server**

The server is deployed by using MQTT broker, hereby referred to as MQTTBox, a free and easy-to-use public tool that is downloaded as a browser extension. On the MQTT broker, the researcher has defined topics on which the client that runs code having these topics defined has to push payload. Once the payload is received, the MQTT broker sends it to all the end-side applications that subscribe to it.

3.3.2 **Node-red**

Node-red is a platform built by IBM to enable developers to run server services locally. It is employed in this project as a local server where the researcher relies on the data published by the MQTT broker to build instant data visualizations.

3.3.3 **MongoDB**

MongoDB is a NoSQL database that may run locally or as a cloud instance. As it is a NoSQL, it saves data not in tables but as collections or documents. Locally this database is accessed by using an IP address 127.0.0.1:1880

3.6 Model Training and Evaluation

Throughout this research project, a model is built upon Keras. The dataset used contains 521 samples for four different variables, including environmental temperature, relative humidity, carbon dioxide, and seasonal data that show the precise weather conditions of data collection. The dataset was initially split into two smaller datasets, a training dataset and a testing dataset as shown in figure 11. 25% of the original dataset is in the testing dataset, while 75% is in the training dataset. The training and testing processes were implemented through python programming codes and python libraries found in sklearn python package. These processes were started by importing the original dataset from CSV file by using dataframe implemented with panda’s python module.

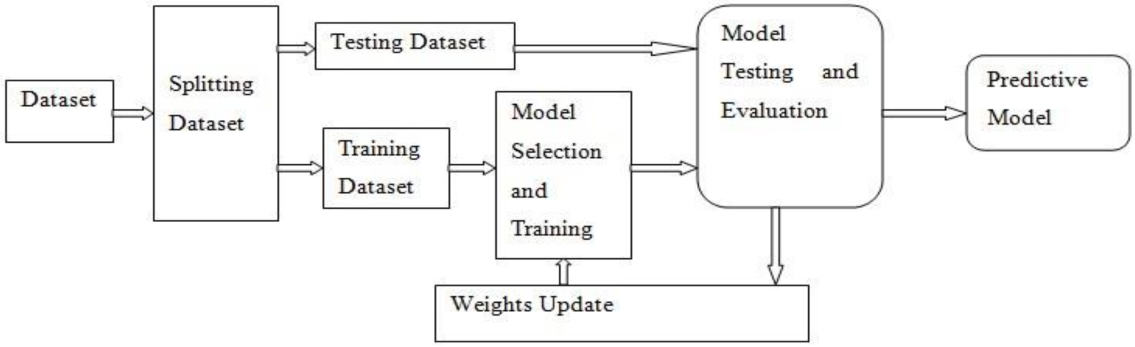


Figure 11Model building process with Keras

3.7 Introduction to building Model and explanation of algorithms

Predictive models are built in different ways and environments. This project employs 2 models. One is built upon Python and its core libraries, and another one is built upon the Facebook prophet algorithm.

A model built in python is evaluated by using other prebuilt python models, listed and described below:

- **3.7.1 Python built model**

- **Logistic regression**

A library from Scikit-learn linear model, a library that provides tools for statistical modeling that includes regression, classification, clustering, etc. Logistic regression is employed in Machine Learning based tasks when the input parameters are discrete. This is contrary to linear regression which is employed when the output parameter depends on the input parameter (22).

- **Gradient Boost classifier**

Gradient Boost is a library from the Scikit-learn ensemble model that combines or groups weak learning models to form one strong and good-performing predictive model, with gradient classification, the model is run through different weights to increase the accuracy of its performance (23).

- ❖ **K-Nearest Neighbor Classifier**

KNN classifier, a library from the Scikit-learn Neighbors model is employed when there is a study of where to place a new point, based on the study of its surrounding pre-located points,A new point is classified according to its nearest neighboring points (24).

- ❖ **Multilayer Perceptron Classifier**

MLP Classifier, a neural network-based model is employed when there is a need for deep analysis of a parameter. With MLP, there are three parameters, input layer, hidden layers, and output layer (25). The inputs are assigned weights and with the bias, the summation of products of weights with the bias is passed to the activation function that computes to give output.

- ❖ **Decision Tree Classifier**

The decision tree classifier is a Scikit-learn Tree-based model, with DT Classifier the branches are considered observations, and leaves are considered target values (26).

- ❖ **Random Forest Classifier**

A Random Forest Classifier is a Scikit-learn-based ensemble model. A classification system made up of several decision trees is called the random forest. It attempts to produce a randomization forest of trees whose forecast by committee is more accurate than that of any individual tree by using bagging and feature randomness when generating each individual tree (27).

3.7.2 Facebook Prophet Model

The prophet model is built by Facebook META. The prophet is utilized in numerous Facebook applications to generate accurate forecasts for planning and goal-setting. In most instances, we've discovered that it outperforms every other strategy. So that you can receive forecasts in only a few seconds, we fitted models in Stan. Prophet can predict the future based on past and current data (28).

CHAPTER 4. SYSTEM DESIGN AND ANALYSIS

4.1 Overview

This chapter covers all single details of the system design from hardware to software. System design is discussed in 2 parts, namely:

- Hardware Part made of : Sensors, Push button,GSM module,Microcontrollers,Power source, and Cables
- Software Part made of :Arduino IDE, MQTTBox, Node-Red, MongoDB, and Anaconda-based Jupyter Notebook for Python processing.

4.2 System Prototype

The prototype of the IoT-based climate change forecast can be seen in figure12 which shows various system components connected to the board. The used sensors are affordable and were created for research. The system can be powered entirely by a battery or cell with 5V (29)due to the minimal power requirements of all applied system components.

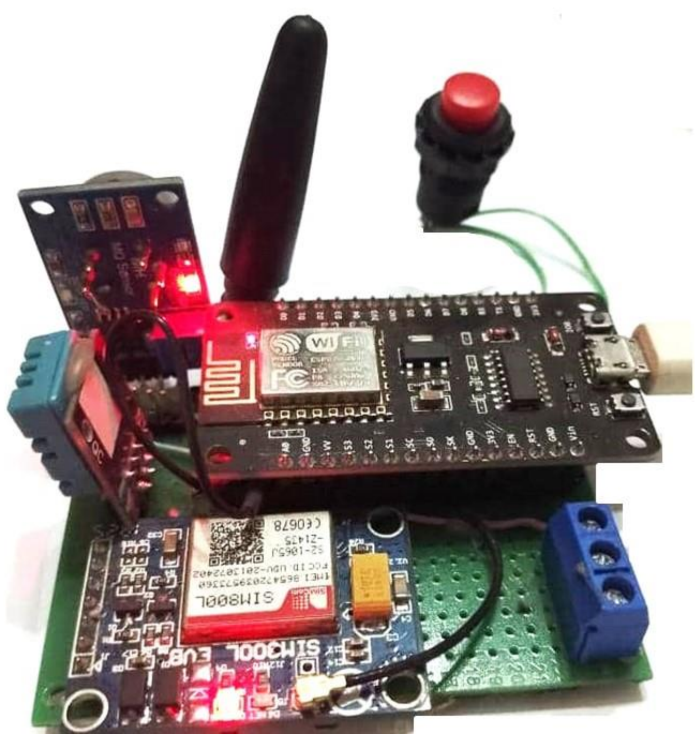


Figure 12 System Prototype

4.3 Design Description

The system designs of the "IoT based Climate Change prediction system" are all summarized in figure 13.

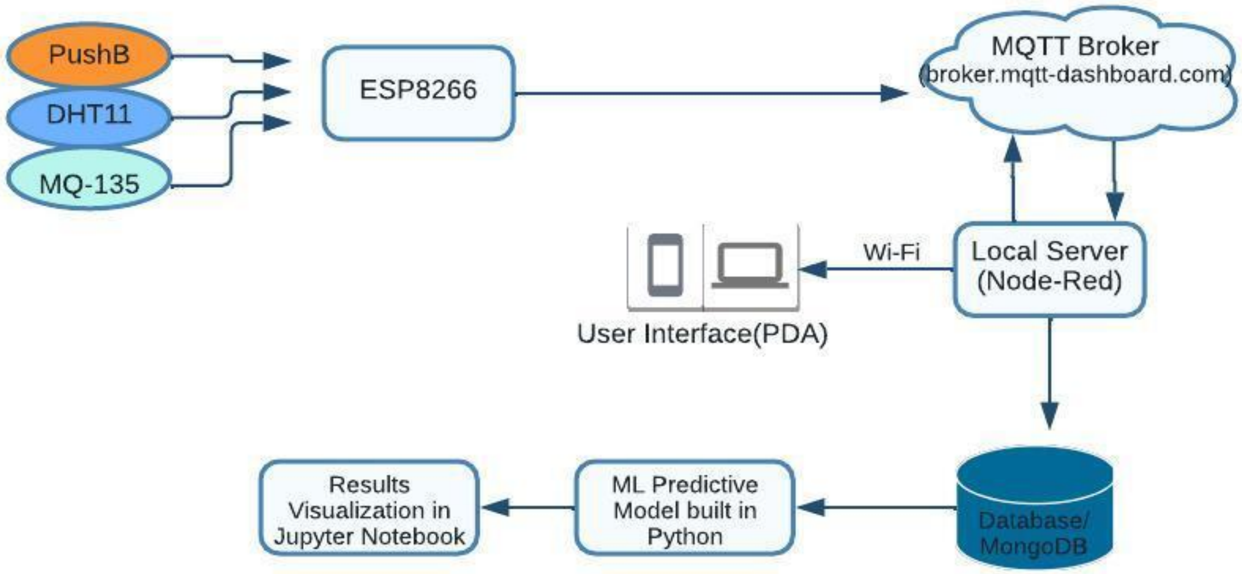


Figure 13 System design

According to the figure 13, the collected data will be published at the given topic to MQTT Broker via IoT Gateway specifically network router or access point which is connected directly to internet connectivity. At the other side of the MQTT broker, the client server(local server) was implemented by using Node-Red software tool, subscribes to the same topic and MQTT Broker as well as did by the publishing field sensor nodes(ground weather station). The received data by the Node-Red data collector nodes by the help of MQTT IoT communication protocol are visualized to the user interface and pushed into the database (MongoDB) to be stored permanently for later use.

Lastly, through python application software tools the data in the database are collected into the format needed to be used by the Machine Learning predictive model generated by model training. Finally, the data from the database are combined with the ML predictive model to generate new predictions and the resulted prediction is visualized in Jupyter Notebook.

4.3.1 Subscribing and publishing in the MQTT box

MQTT Broker in this study is employed as a cloud server that holds payload temporally to re-direct it to different far ends and gives the data once a client subscribes to it. MQTT Broker is built upon MQTT Protocol, a lightweight machine-to-machine communication protocol that transport messages from one node to another, with a tiny bandwidth utilization and code overhead (30).MQTT Architecture is briefly described in figure 14.

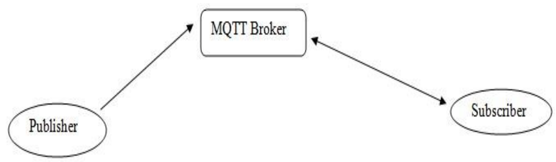


Figure 14 MQTT Overview

4.3.2 Building flows in Node-Red

Node-RED is a development tool for tying new and intriguing connections between physical components, APIs, and web services. With the help of the extensive selection of nodes in the palette that can be deployed to its runtime with a single click, it offers a browser-based editor that makes it simple to wire up flows. Figure 15 shows the deployed nodes that briefly include the mqtt-in node, nodes of functions to read payload from topics, messages debugger, gauges and charts for data visualization, and mongodb2 in and out nodes.

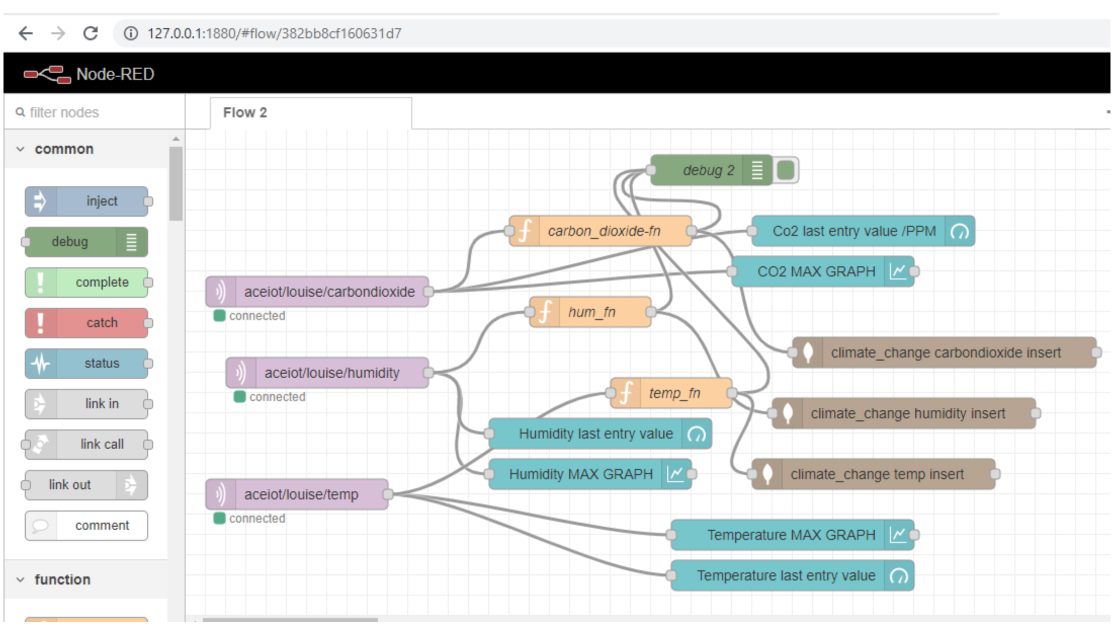


Figure 15 Nodes deployed in IoT based Climate change prediction

Figure 16 shows how the nodes climate change parameters are built. Hence the processes is reading data from MQTTBroker to display instant change in the payload by using the gauges and the charts.

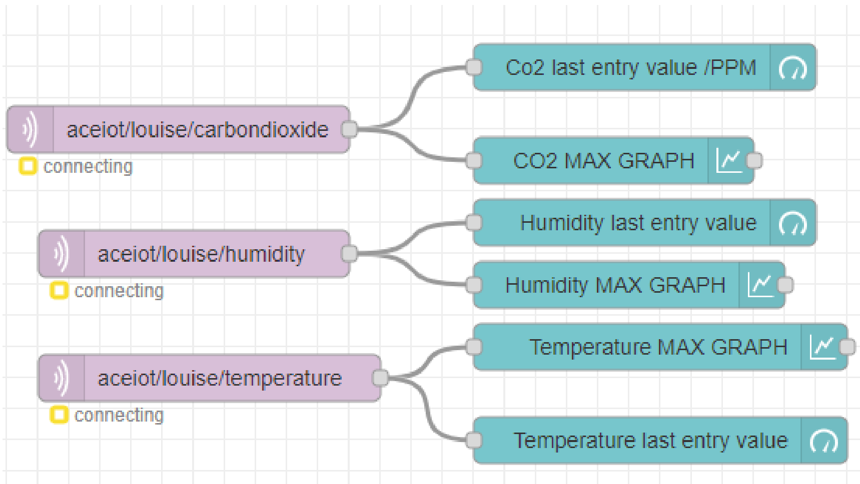


Figure 16 Nodes for visualization of climate change parameters

Figure 17 shows the customization of the MQTT-in nodes, as it subscribes to the MQTT broker, and receives the payload as a result of publishing function from the MQTT Broker.

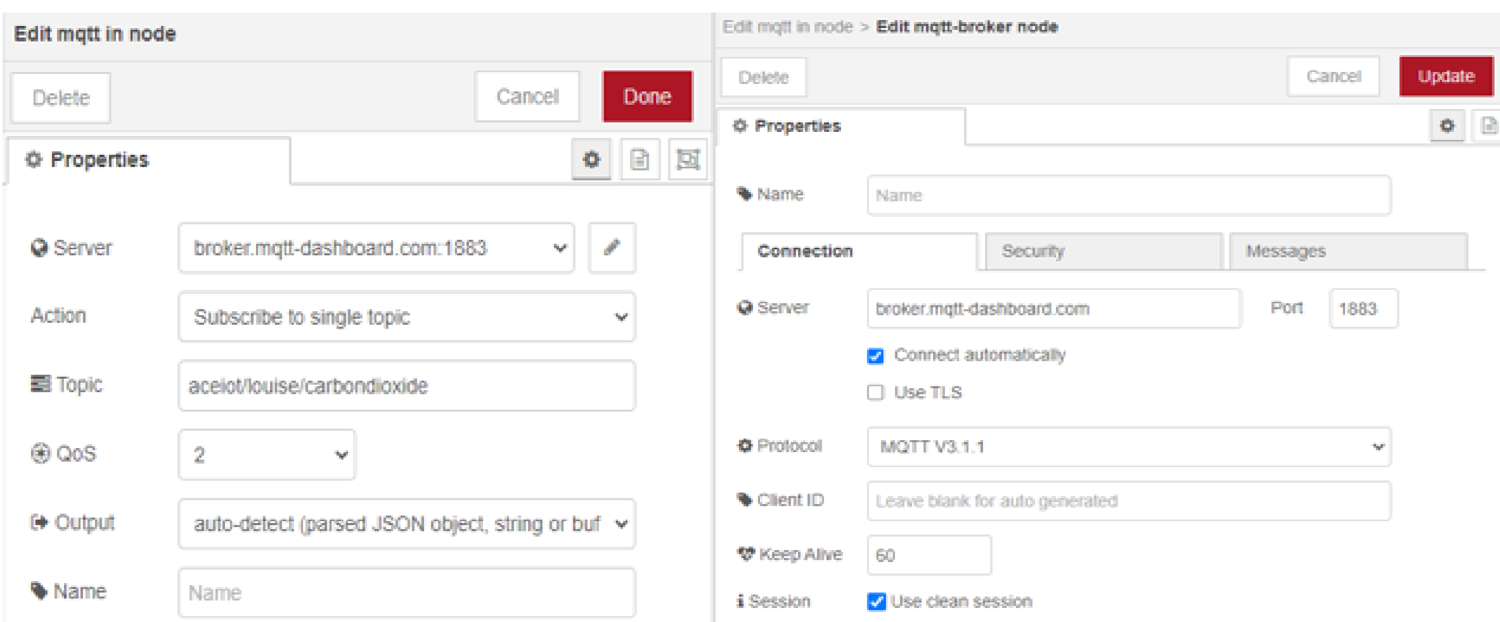


Figure 17 MQTT-in Node

As shown in figure 18, functions are written in JavaScript language and often define were to take the data or flows towards where, and how to assign values to variables. The codes are executed once the nodes are deployed.

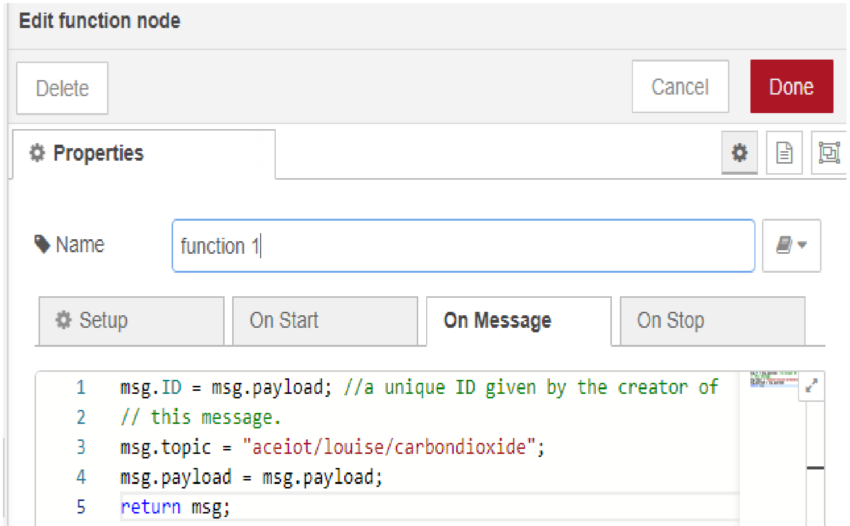


Figure 18 function node

The MongoDB is named 'climate_change'. Noticing that with Mongo, there are no tables, no columns, or rows. Every item is stored in a set of collections. As shown in figure 19 and figure 20, a collection named 'read' is defined and the insert function is chosen for the operation of inserting payload into the database to the defined collection. The database is running locally and is accessed on port **27017**.

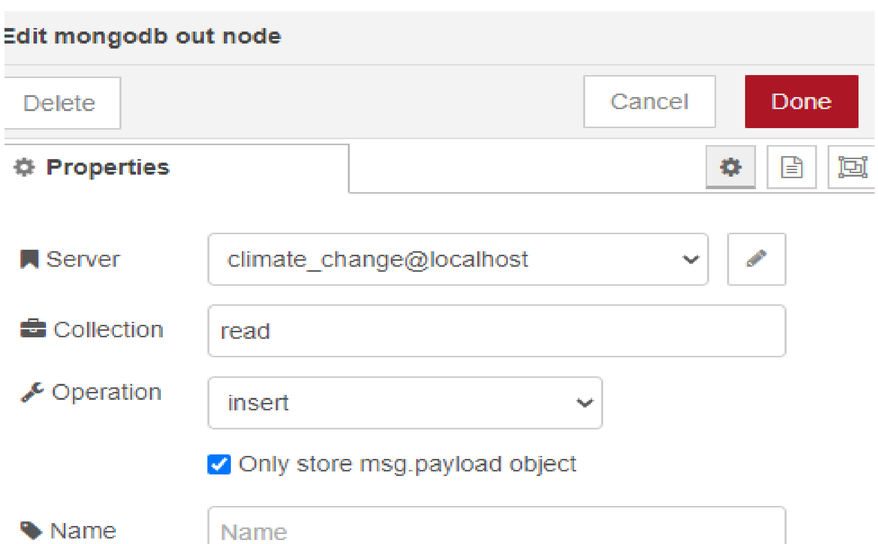


Figure 19 MongoDB out node1

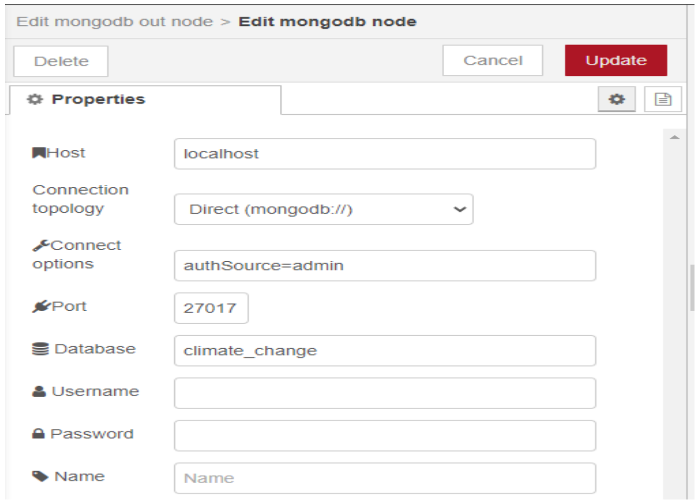


Figure 20 MongoDB out node2

To view the instant readings from the topics, a debugger node is used, and nothing to modify in it except its 'debug node' name and 'message name' as shown in figure 21.

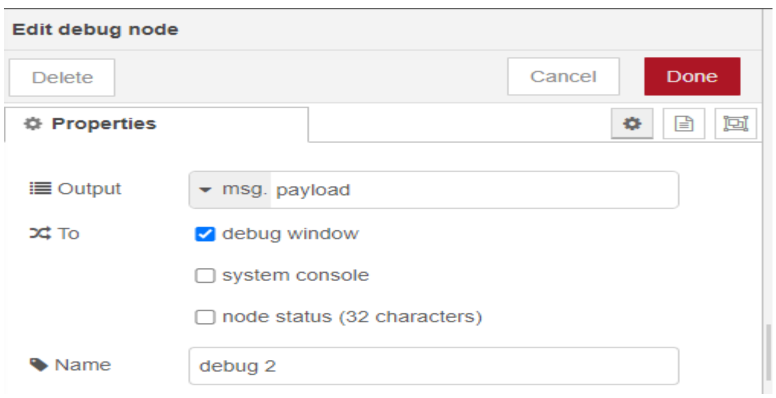


Figure 21 Message debugging node

4.3.3 MongoDB database

MongoDB is a No-SQL database that stores data in collections. With python script, the data is retrieved from the Mongo database to be fed to the ML, which continuously studies under the train-test-split function to give an accurate prediction/forecast as shown in figure 22.

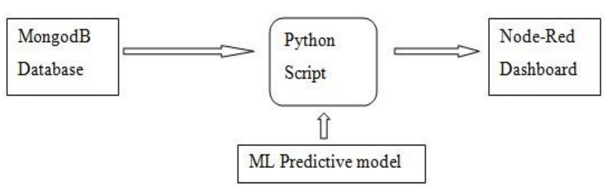


Figure 22 MongoDB-Python data flow

MongoDB as a no-SQL (31), saves data in a different format. Each row is called a document, and each document has a corresponding ID. The collection of documents forms a collection.

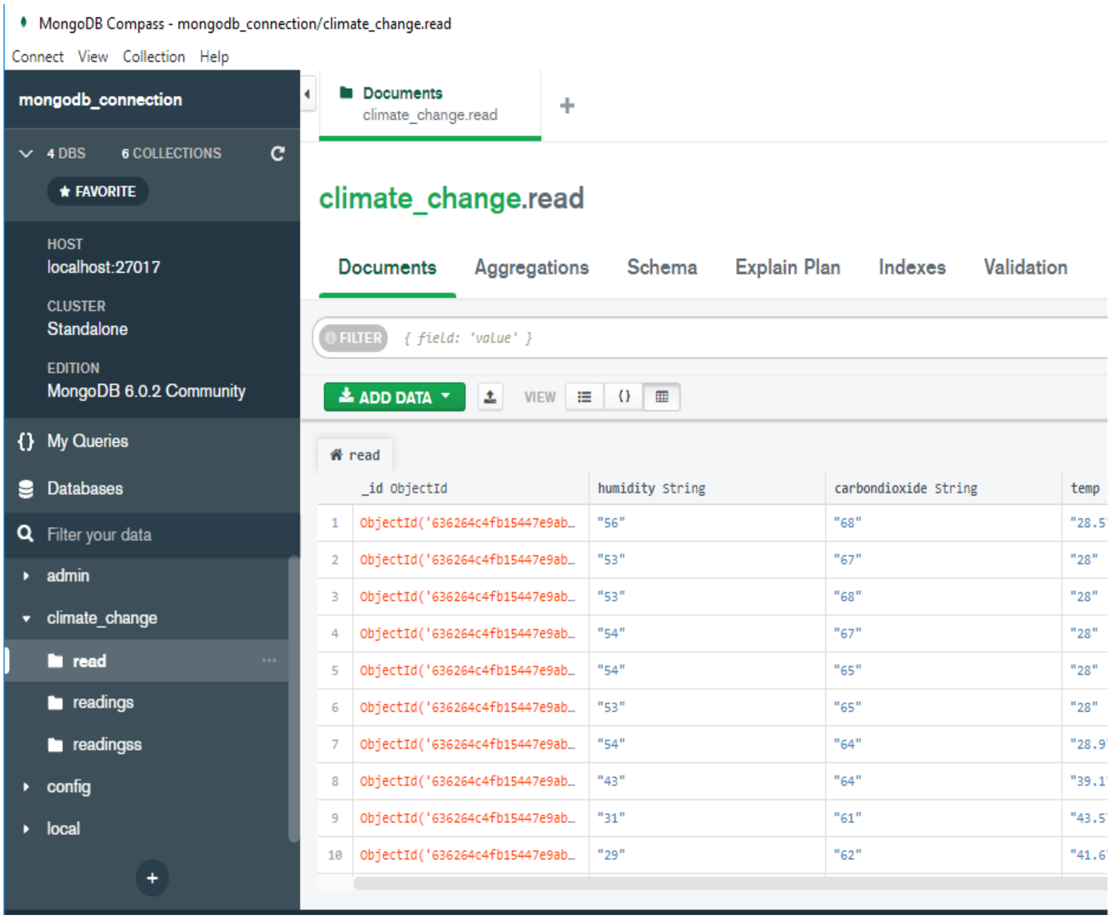


Figure 23 MongoDB Compass interface

With IoT, one of the big data generators, it is better to employ a generic database, as there is a change in aspects to measure, capture, to monitor or to analyze (32) MongoDB is employed in this study as it fulfills those requirements and moreover it enables the developer to build clusters, that make it accessed and used by clients over the cloud. It is not mandatory to be used locally.

MongoDB Compass is a querying, optimization, and analysis tool for your MongoDB data. Build pipelines using drag and drop and key insights (33). The saved data is of Temperature, Humidity, Carbon dioxide, and Season as shown in figure 23.

CHAPTER 5. RESULTS AND DISCUSSION

5.1 Overview

This chapter explains and gives insights into the study results, from analysis of collected data to data visualization. All of the work is done using Python.

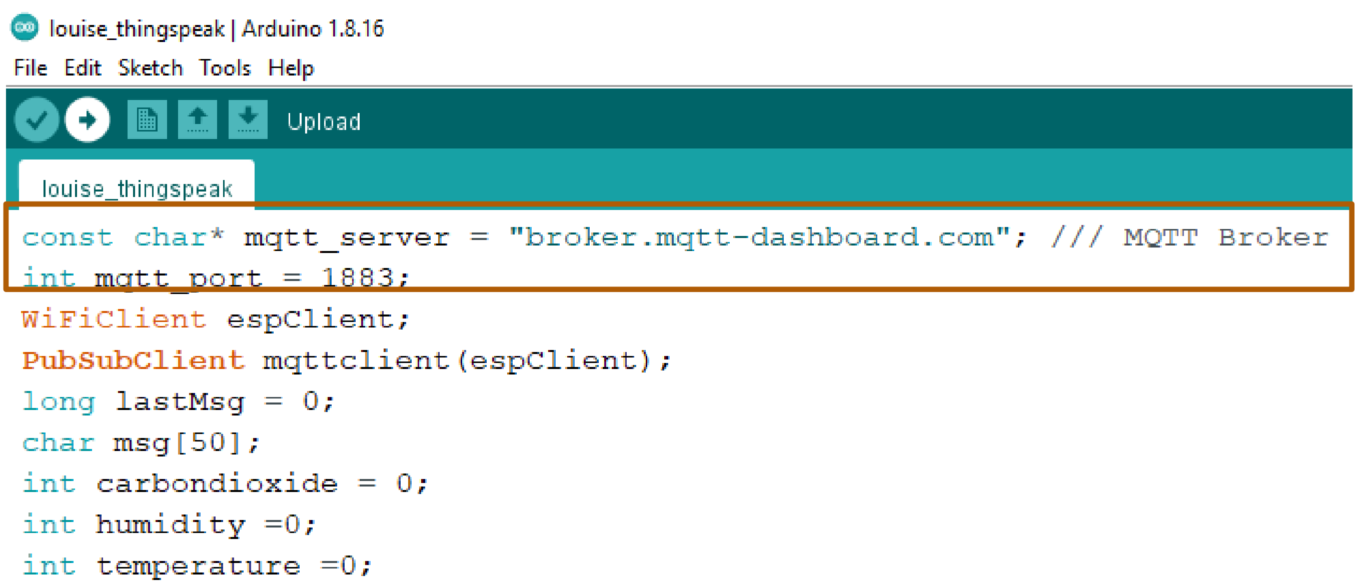
By using anaconda IDE, a Jupyter Notebook is built to assess and combine all the data processing and analytics in one form. The steps followed are as follows: visualization of collected data, data cleansing and wrangling building predictive model, and assessing the model accuracy by using different pre-built models. This study is generally a classification study (34) because, in the end, a forecast tells whether the climate is warm or cold. The employed pre-built models include Logistic regression (LR), Gradient Boos Classifier (GB), Random Forest Classifier (RF), Multilayer Perceptron classifier (MLP), Decision Tree Classifier (DT), K Nearest Neighbor Classifier (KNN) and Facebook Prophet Predictive model. Each model performs differently from others and gives its independent training and testing accuracy scores.

Several python libraries (35) are employed in this project and they include:

- Pandas: a python-based library used for analyzing and manipulating data. The data includes specific data structures and procedures for working with time series and mathematical tables.
- Numpy: Enormous, multi-dimensional arrays and matrices are supported by NumPy, a library for the Python programming language, with a substantial number of high-level mathematical computations that may be performed upon those arrays.
- Matplotlib is a plotting library for the Python computer language and its NumPy extension, which is used for numerical mathematics. With the help of general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK, it offers an object-oriented API for embedding plots into applications.
- Scipy: an open-source and free python library used for technical and scientific computing. SciPy includes modules for activities used often in science and engineering, including optimization, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, and ODE solvers.

- ❖ Keras: a python library that provides a python interface to artificial neural networks
  Scikit-learn: a python library that provides machine learning functions, such as Support-vector machines, random forests, gradient boosting, k-means, and DBSCAN, a few of the classification, regression, and clustering algorithms that are included. It is also made to work with Python's NumPy and SciPy libraries.
- ❖ Tensor Flow: a python library used for data automation, model tracking, performance monitoring, and model retraining. Success depends on the use of production-level technologies to automate and monitor model training throughout the lifespan of a good, service, or business procedure.
- ❖ Seaborn: a matplotlib-based Python data visualization package. It offers a sophisticated drawing tool for creating eye-catching and instructive statistical visuals.

5.2 Data collection in Arduino and visualization in Serial Monitor




```
const char* mqtt_server = "broker.mqtt-dashboard.com"; /// MQTT Broker
int mqtt_port = 1883;
WiFiClient espClient;
PubSubClient mqttclient(espClient);
long lastMsg = 0;
char msg[50];
int carbondioxide = 0;
int humidity =0;
int temperature =0;
```

The above lines of codes are written in c/c++, it is a sample taken from Arduino IDE to show how the MQTT server is defined and accessed, by using a call of “broker. mqtt-dashboard”. MQTT port number is 1883.

```
carbondioxide= analogRead(AOUTpin);
char carbondioxideString[8];
dtostrf(carbondioxide, 1, 2, carbondioxideString);
Serial.print("carbondioxide: ");
Serial.println(carbondioxideString);
mqttclient.publish("aceiot/louise/carbondioxide", carbondioxideString);

humidity = dht.readHumidity();
// Read Humidity as Celsius (the default)
char humString[8];
dtostrf(humidity, 1, 2, humString);
Serial.print("Humidity: ");
Serial.println(humString);
mqttclient.publish("aceiot/louise/humidity", humString);
```

The above set of codes shows how to define at which topics the MQTT client publishes the payload.

Once the codes are successfully compiled and uploaded to the ESP8266, the serial monitor output the following results, showing the status of the Wi-Fi connection, the corresponding IP address, and once the mqtt server is connected, displays the readings from the sensors as shown below:

```
WiFi connected
IP address:
192.168.137.227
Connected
MQTT Server broker.mqtt-dashboard.com:1883
ESP8266 IP 192.168.137.227
Attempting MQTT connection...connected
Attempting MQTT connection...connected

Cold period
Season: 0.00
Temperature: 25.00
Humidity: 63.00
carbondioxide: 21.00
Attempting MQTT connection...connected
Attempting MQTT connection...connected

Cold period
Season: 0.00
Temperature: 25.00
Humidity: 63.00
carbondioxide: 21.00
Attempting MQTT connection...
```

In MQTTBox happens a process of receiving data from the MQTT Server is defined in Arduino codes. It makes easy the process of adding publishers/subscribers to the topics as shown in figure 24.

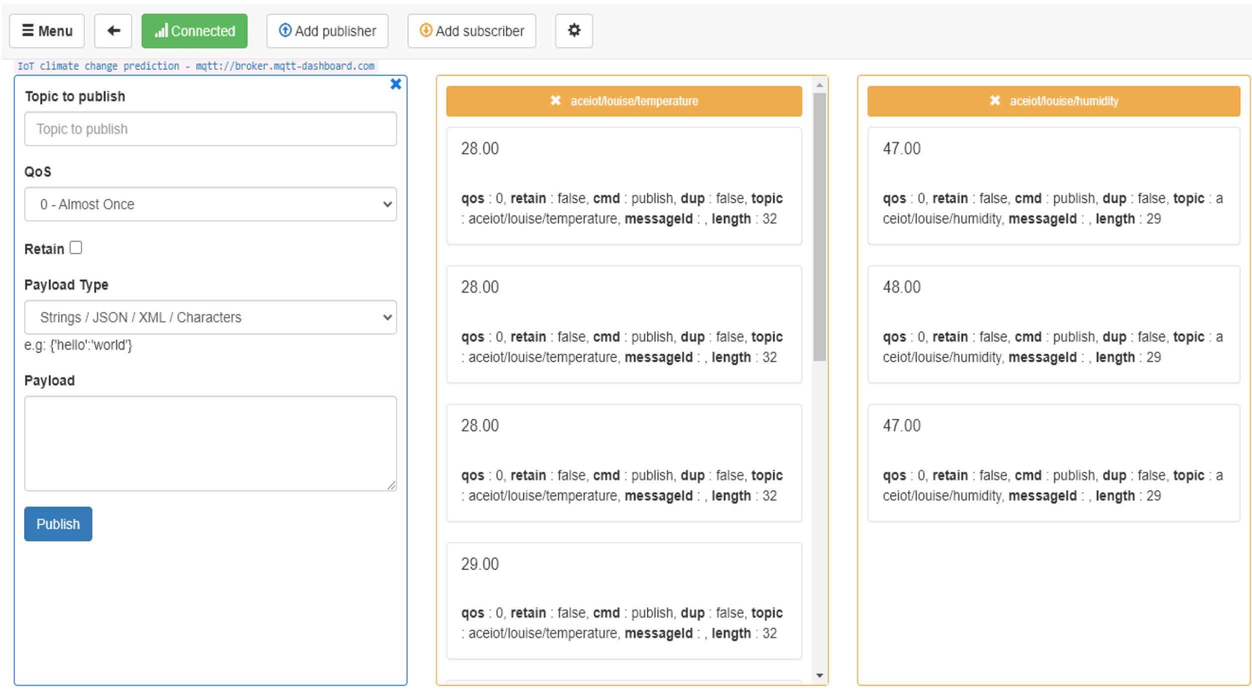


Figure 24 Adding subscription to MQTT topics

5.3 Study of correlation between parameters and the check for outliers

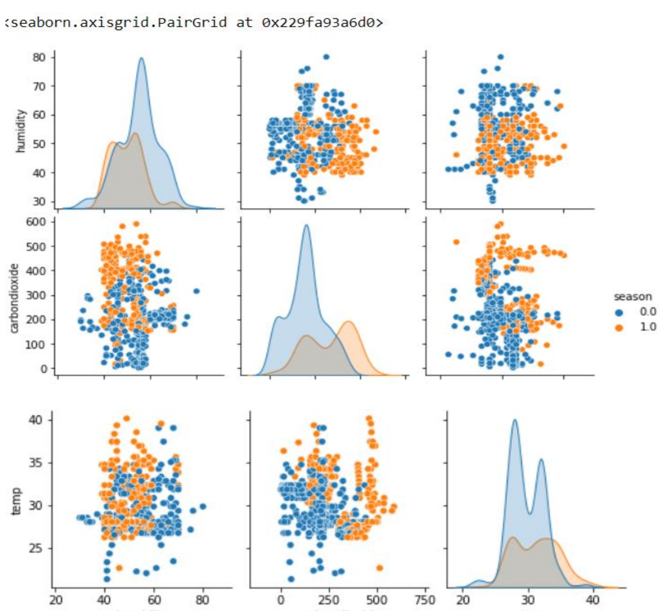


Figure 25 Study of the correlation between parameters and check for outliers

Figure 25 illustrates the correlation between monitored parameters and helps in identifying and checking for outliers. Once taken a clear look, it is visible that the increase in temperature and carbon dioxide measurements leads to warming, while the increase in humidity and decrease in temperature/carbon dioxide measurements leads to coldness.

An **outlier** is a data point in a data set that is far from all other observations is known as an outlier. a data point that deviates from the dataset's overall distribution. With the above illustration, it is clear that there are no outliers.

5.4 Model Building

The model training and testing processes were carried out using python programming codes and python libraries from the sklearn python package. These processes began with importing the original dataset from a CSV file into a dataframe created with Panda's Python module. Once the model is trained and tested, it is evaluated. Evaluating a model is the process of passing a model through other pre-built models to increase its training and testing accuracy. The selection of the specific evaluation metrics depends on the type of category of machine learning problem. The metrics techniques used during of model evaluation in this research study includes Accuracy, Recall, Precision and F1 score (36). These evaluation metrics are defined using confusion matrix and were selected because they are the best fit for machine learning classification problems. The confusion matrix is one of machine learning techniques for summarizing the performance of a machine learning classification models. As shown in the figure 26, the matrix contains four main elements that include True Positive (TP), False Positive (FP),True Negative and False Negative (37).In a confusion matrix, only the true values are considered to judge the performance of a model. As shown in figure 41, the True positive is 50, while the True negative is 42.

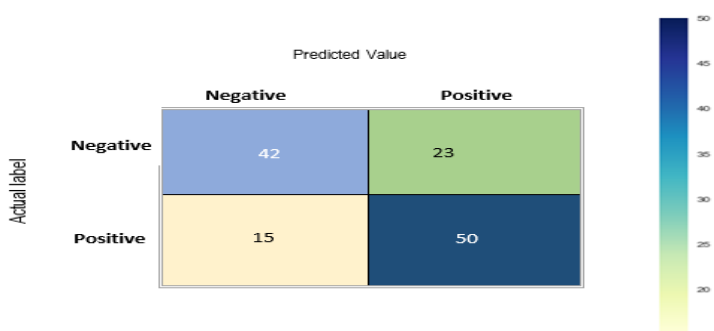


Figure 26 Model study with confusion matrix

5.4.1 Accuracy Measures

The accuracy measures are computed using four quantities. They are described as follows:

1. True-Positive (TP) 2. False-Positive (FP)

3. False-Negative (FN) 4. True-Negative (TN)

We can determine Accuracy, Precision, Recall, and F1 based on these factors.

- ❖ $Accuracy, A = TN+TP/TN+FN+TP+FN$
- ❖ Precision, $PV = TP/TP+FP$
- ❖ $Recall, R = TP/TP + FN$
- ❖ F1 = $2 * (PV * R / PV + \mathrm{R})$

The training accuracy of the built model is ~ 79% with the corresponding testing accuracy of ~78% as shown in figure 27.

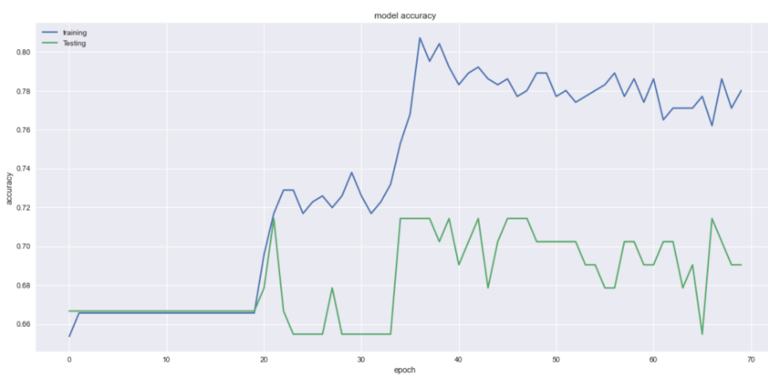


Figure 27 Train-Test Model Evaluation

5.5 Training and evaluation implementation

This study is generally a classification study because, in the end, a forecast tells whether the climate is warm or cold. The dataset for machine learning algorithms was prepared, and then the models were trained on training data using various machine learning classification models including Logistic regression (LR), Gradient Boost Classifier (GB), Random Forest Classifier (RF), Multilayer Perceptron classifier (MLP), Decision Tree Classifier (DT), K Nearest Neighbor Classifier (KNN) and Facebook Prophet Predictive model. These machine learning algorithms were used for training and testing data as well as for mapping the relationships between the dependent variables (model input) and independent variables (model output). Each model performs differently from others and gives its independent training and testing accuracy scores.

5.5.1. Random Forest classifier

The model built has a training accuracy of 79% and a testing accuracy of 78%. Trying to increase the accuracy by evaluating the model by using the Random Forest classifier model, the Training accuracy rises to 99% with the corresponding testing accuracy of 84% as figure 28 illustrates, and both accuracies continue to increase as the number of estimators increase.

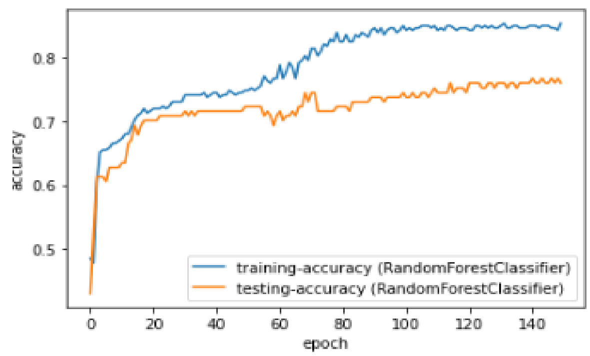


Figure 28 Model evaluation with Random Forest Classifier

5.5.2. Multilayer Perceptron Classifier

Once the model is evaluated by using Multilayer Perceptron, the accuracy improves as follows: both training and testing accuracies are 78% as illustrated in figure 29, and both accuracies continue to increase as the number of iterations increases.

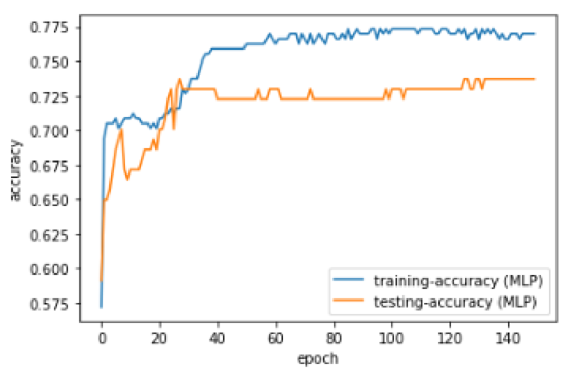


Figure 29 Model evaluation with Multilayer Perceptron Classifier

5.5.3. Logistic Regression

Employing Logistic Regression, the training accuracy rises to 82% with the corresponding testing accuracy of 80% as figure 30 shows, and both accuracies continue to increase as the number of iterations increases.

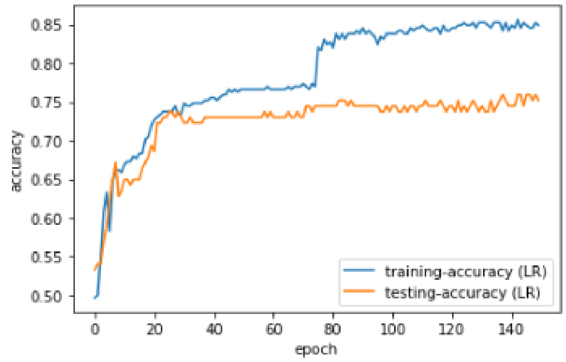


Figure 30 Model evaluation with Logistic Regression

5.5.4. Gradient Boost Classifier

Evaluating the built model with the Gradient Boost model, the training accuracy rise to 97% with a corresponding 77% as figure 31 illustrates, and both accuracies continue to increase as the number of estimators increase.

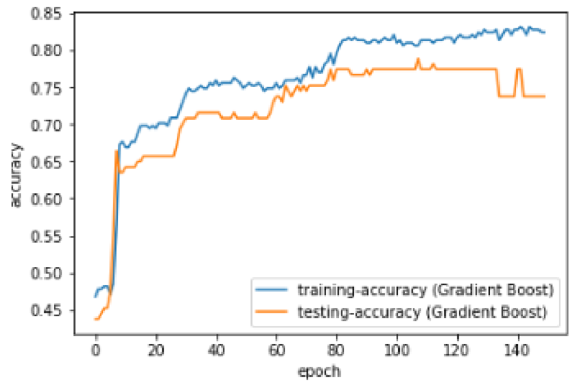


Figure 31 Model evaluation with Gradient Boost Classifier

5.5.5. Decision Tree Classifier

When the model is evaluated by employing the Decision Tree Classifier model, the accuracies become as follows: Training accuracy of 99% and 79% of testing accuracy as figure 32 shows. And both accuracies continue to increase as the maximum depth increase.

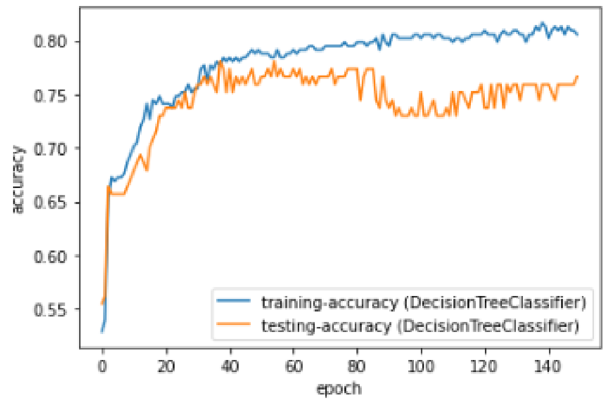


Figure 32 Model evaluation with Decision Tree Classifier

5.5.6. K Nearest Neighbor Classifier

When aggregating the built model with K Nearest Neighbor Classifier, the training accuracy rises to 84 % and the testing accuracy to 77% as illustrated in figure 33, and both accuracies continue to increase as the number of neighbors increases.

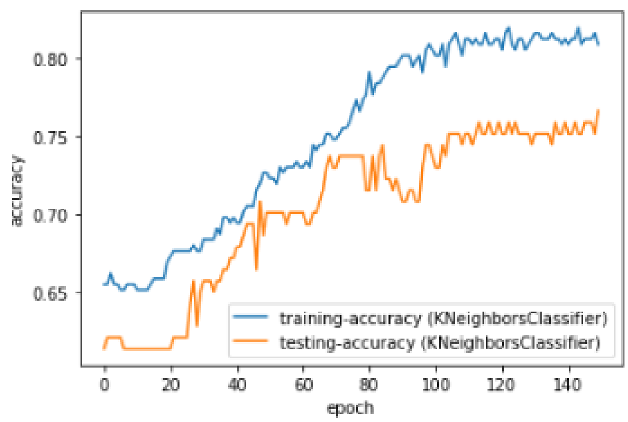


Figure 33 Model evaluation with K Nearest Neighbor Classifier

5.6 Evaluation Metric for Predicted outputs

The results of this prediction are shown in the table 1 along with relevant metrics like Precision, Recall, and F1 score for evaluating various algorithms utilized in this system as well as measuring algorithm's Performance.

Table 1 Metrics of Algorithms

Algorithm used	**Precision**	**Recall**	**F1 Score**
Logistic Regression (LR)	0.89	0.84	0.86
Gradient Boost Classifier (GB)	0.86	0.79	0.82
Random Forest Classifier (RF)	0.86	0.90	0.88
Multilayer Perceptron classifier (MLP)	0.83	0.89	0.86
Decision Tree Classifier (DT)	0.85	0.82	0.83
K Nearest Neighbor Classifier (KNN)	0.85	0.81	0.83

5.7 Comparative Analysis

When compared to other machine learning algorithms, Random Forest has a high prediction accuracy. This indicates that this system will perform better with Random Forest than with other machine learning algorithms, as shown in table 2.

Table 2 Comparative analysis with different algorithms

ALGORITHM	**ACCURACY**	
	TRAINING	**TESTING**
Logistic Regression (LR)	82%	80%
Gradient Boost Classifier (GB)	97%	77%
Random Forest Classifier (RF)	99%	84%
Multilayer Perceptron classifier (MLP)	78%	78%
Decision Tree Classifier (DT)	99%	79%
K Nearest Neighbor Classifier (KNN)	84 %	77%

5.8 Prophet Facebook model

With Prophet Facebook modeling, when the same dataset is fed to the prophet model to forecast the warming of the coming year, figure 34 shows that there will be a continuous increase of the warming periods throughout the year 2023, at Masoro Industrial Economic Zone. As the season is warm or cold, the measurements below Zero and the measurements above one are not considered during the interpretation of the graph. The general meaning of the illustrations is to let observers see that there will be a gradual increase in the periods of warming throughout the year.

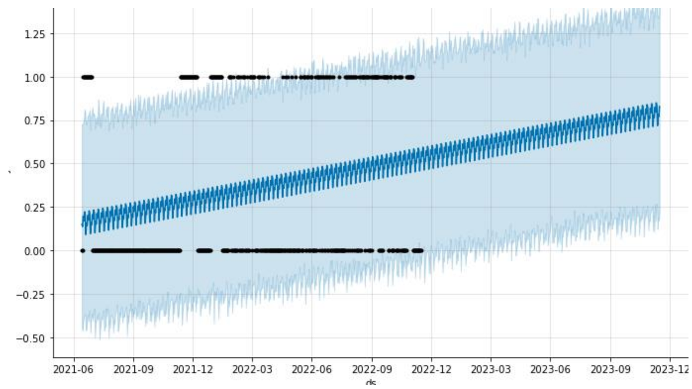


Figure 34 Prophet Model-Based Prediction

5.9 Live data visualization

The live data visualization from sensors by using Node-red is illustrated below in figure 35 with charts and gauges. Carbon Dioxide gas CO2 is ranging from zero to 300 PPM, Humidity is ranging from 0 to 100%RH, and Temperature is ranging from -10 to 50 ^{o}C

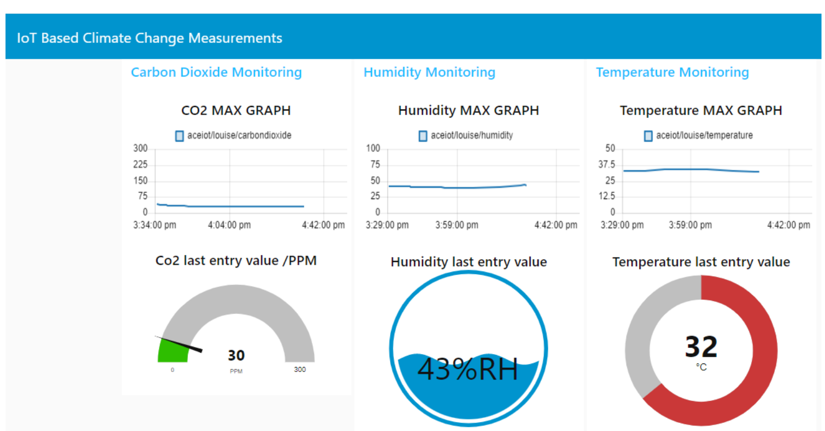


Figure 35 Live Data Visualization in charts and gauges

CHAPTER 6. CONCLUSION, RECOMMENDATION AND FUTURE WORK

During this study, I studied the use of the state-of-the-art technologies of IoT and Machine learning and deep learning aspects to predict climate change in the Masoro Industrial Economic Zone. It is proved that once the carbon dioxide gas emission goes higher, the temperature in the environment increases, which directly or indirectly affects the environment in general. Due to the python model built, the Random Forest approves itself to be the best model in evaluating the built model with 99% of training accuracy and 84% of testing accuracy. The study shows that once the carbon dioxide gas emission goes higher, the temperature in the environment increases, which directly or indirectly affects the environment in general. Finally, the prediction clarifies that if no measures are taken presently, the climate change in Masoro's industrial zone will be dominated by warming periods in the future. It is proved. The employed technologies are fully trustworthy and I can recommend them to other technologists. I recommend that Meteo Rwanda and other public/private institutions adopt this IoT system in order to improve the accuracy of their regular climate change monitoring and prediction tasks.

The future work will be to assess future climate change situations to approve the results of this research project in real situations. It was not easy to get a dataset as Meteo Rwanda does not capture CO_2 emissions and no other public or private institution was known to capture this parameter from the Rwandan environments. I am recommending to future researchers have enough hardware to collect data from different parts of the country throughout different periods of the year to predict the warming of the wide environment. Also, I am recommending them have a bit larger dataset so that the model to build becomes more accurate.

REFERENCE

1. **Kumaran, M.K. Nallakaruppan and U. Senthil**,"IoT based Machine Learning Techniques for Climate Predictive Analysis," International Journal of Recent Technology and Engineering (IJRTE), January 2019., Vol. 7 . ISSN: 2277-3878,.

2. **Telenor.** IoT Predictions : NorthStream Part of Accenture, 2020. Prediction Report.

3. **N. Abd El-Mawla, M. Badawy, and H. Arafat,"**IoT for the Failure of Climate-Change Mitigation and Adaptation and IIoT as a Future Solution," World Journal of Environmental Engineering, Apr. 2019, Vol. 6. pp. 7–16,doi: 10.12691/wjee-6-1-2..

4. **Viktor Sebestyén, Tímea Czvetkó and János Abonyi,"**The Applicability of Big Data in Climate Change Research:The Importance of System of Systems Thinking," Frontiers in Environmental Science, 17 March 2021. Sci.

5. **S Sheeba Rani, J Janet ,KC Ramya and V Gomathy,"**IoT based climate prediction using ANN for green networking," 993,: IOP Conf. Series: Materials Science and Engineering , 2020. 012090.

6. **Israr Ullah, Muhammad Fayaz, Nasir Naveed, And Dohyeun Kim**,"ANN Based Learning to Kalman Filter Algorithm for Indoor Environment Prediction in Smart Greenhouse," Digital Object Identifier 10.1109/IEEE ACCESS , August 13, 2020. 3016277.

7. **Sadhukhan, Mrinmoy and Bhattacharya, Sudakshina Dasgupta and Indrajit,"**An Intelligent Weather Prediction System Based on IOT," Devices for Integrated Circuit (DevIC) 978-1-7281-9955-9/20/$31.00 , 20 May 2021. ©2021 IEEE.

8. **SA Markolf, HS Matthews, IL Azevedo**,"An integrated approach for estimating greenhouse gas emissions from 100 US metropolitan areas," Environmental Research Letters , 25 January 2017, Vol. 12.

9. **Lo, Tsungnan Chou and Mingmin**,"Predicting Credit Card Defaults with Deep Learning and Other Machine Learning Models," International Journal of Computer Theory and Engineering , 2018, Vol. 10. 105-110.

10. **Priya Singh, Gyanendra Kumar Pal, and Sanjeev Gangwar** Prediction of Cardiovascular Disease Using Feature Selection," International Journal of Computer Theory and Engineering, 2022, Vol. 14. 97-103.

11. **Bostrom, Orlando P. Zacarias and Henrik,"**Comparing support vector regression and random forests," Int. Conf. Adv. ICT Emerg. Reg. ICTer, 2013. pp. 217–221,doi: 10.1109/ICTer.2013.6761181..

12. **Sumedh Kaninde, Manish Mahajan , Aditya Janghale and Bharti Joshi,"**Stock Price Prediction using Facebook Prophet," ITM Web of Conferences, 2022, Vol. 44. doi https://doi.org/10.1051/itmconf/20224403060.

13. **Kalipe, Godson, Gautham, Vikas and Behera, Rajat Kumar**,"Predicting Malarial Outbreak using Machine and Deep Learning Approach: A Review and Analysis," Bhubaneswar, India : 2018 International Conference on Information Technology (ICIT), 2018. pp. 33-38, doi: 10.1109/ICIT.2018.00019..

14. **Ardvin Kester S. Ong, Yogi Tri Prasetyo , Nattakit Yuduang , Reny Nadlifatin, Satria Fadil Persada,Kirstien Paola E. Robas,Thanatorn Chuenyindee,Thapanat Buaphiban** ,"Utilization of Random Forest Classifier and Artificial Neural Network for Predicting Factors Influencing the Perceived," International Journal of Environmental Research and Public Health, 29 June 2022, Vol. 19. doi.org/10.3390/ijerph19137979.

15. **Pabboju, B. Manjulatha and Suresh**,"An Ensemble Model for Predicting Chronic Diseases Using Machine Learning Algorithms," Smart Computing Techniques and Applications, 2021, Vol. 224. 978-981-16-1501-6.

16. **Anders Ekbom, Emelie César.** Rwanda Environment and Climate Change Analysis. RWANDA : Sida's Helpdesk, 2019. desk study.

17. **Ejodamen Pius Uagbae, Ekong, Victor Eshiet , Inyang, Udoinyang Godwin,"**Arduino-Based Weather Monitoring System," Delta State, Nigeria ,SMART- SMART-iSTEAMS Multidisciplinary Conference Ogwuashi-uku, February 2018.

18. **K. B. K. Sai, S. Ramasubbareddy, and A. K. Luhach,"**IOT based air quality monitoring system using MQ135 and MQ7 with machine learning analysis," Scalable Computing, 2019, Vol. 20. pp. 599–606, , doi: 10.12694/scpe.v20i4.1561.S..

19. **F. Wu, H. Chen, K. Hou, Z. Cheng, and T. Z. Qiu,"**Adaptive Pushbutton Control for Signalized Pedestrian Midblock Crossings," J Transp Eng A Syst,, Apr. 2022, Vol. 148. doi: 10.1061/jtepbs.0000659..

20. nodemcu-details-specifications. Make-It.ca. [Online] PCBoard.ca. https://www.make-it.ca/nodemcu-details-specifications/.

21. **I. O. Ewona, U. J. Ekah, A. O. Ikoi, and E. Obi.** "Measurement and performance assessment of GSM networks using received signal level.". [Online] www.unicrossjournals.com.

22. **E. Dumitrescu, S. Hué, C. Hurlin, and S. Tokpavi**,"Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects," Eur J Oper Res, Mar. 2022, Vol. 297. pp. 1178–1192, doi: 10.1016/j.ejor.2021.06.053..

23. **P. Datta, P. Das, and A. Kumar,"**Hyper parameter tuning based gradient boosting algorithm for detection of diabetic retinopathy: an analytical review," Bulletin of Electrical Engineering and Informatics, Apr. 2022, Vol. 11. pp. 814–824, doi:10.11591/eei.v11i2.3559..

24. **J. Gou, H. Ma, W. Ou, S. Zeng, Y. Rao, and H. Yang,"**A generalized mean distance-based k-nearest neighbor classifier," Expert Syst Appl, Jan. 2019, Vol. 115. pp. 356–372,doi: 10.1016/j.eswa.2018.08.021.

25. **C. Zhao, Y. Gao, J. He, and J. Lian**,"Recognition of driving postures by multiwavelet transform and multilayer perceptron classifier," Eng Appl Artif Intell, Dec. 2012, Vol. 25. pp. 1677–1686,doi: 10.1016/j.engappai.2012.09.018.

26. **Abdulazeez, B. Charbuty and A,"**Classification Based on Decision Tree Algorithm for Machine Learning," Journal of Applied Science and Technology Trends, Mar. 2021, Vol. 2. pp.20–28, doi: 10.38094/jastt20165..

27. **Ong AKS, Prasetyo YT, Kusonwattana P, Mariñas KA, Yuduang N, Chuenyindee T, Robas KPE, Persada SF, Nadlifatin R'**"Utilization of Random Forest Classifier and Artificial Neural Network for Predicting Factors Influencing the Perceived Usability of COVID-19 Contact," 13, MorChana in Thailand : Int J Environ Res Public Health, Jul.2022, Vol. 19. doi: 10.3390/ijerph19137979.

28. **S. Kaninde, M. Mahajan, A. Janghale, and B. Joshi,"**Stock Price Prediction using Facebook Prophet," ITM Web of Conferences, 2022, Vol. 44. p. 03060, doi: 10.1051/itmconf/20224403060.

29. **E. Garcia-Espinosa, O. Longoria-Gandara, I. Pegueros-Lepe, and A. Veloz-Guerrero,"**Power consumption analysis of bluetooth low energy commercial products and their implications for IoT applications,"12, Switzerland : Electronics, MDPI AG, Dec. 01, 2018, Vol. 7. doi: 10.3390/electronics7120386.

30. **K. Hwang, J. M. Lee, and I. H. Jung,**"Performance Monitoring of MQTT-based Messaging Server and System," Journal of Logistics, Informatics and Service Science, 2022, Vol. 9. pp. 85–96,doi: 10.33168/LISS.2022.0107..

31. **E. Gupta, S. Sural, J. Vaidya, and V. Atluri,"**Attribute-Based Access Control for NoSQL Databases," CODASPY : Proceedings of the 11th ACM Conference on Data and Application Security and Privacy, Apr.2021. pp. 317–319. doi: 10.1145/3422337.3450323..

32. **. S. Kang, I. H. Park, J. Rhee, and Y. H. Lee,"**MongoDB-Based Repository Design for IoT-Generated RFID/Sensor Big Data," IEEE Sens J, Jan. 2016, Vol. 16. pp. 485–497,doi: 10.1109/JSEN.2015.2483499..

33. **Ganesan, S. Chellappan and D**,"MongoDB Recipes," Apress, 2020. doi: 10.1007/978-1-4842-4891-1..

34. **Abdulazeez, B. Charbuty and A,"**Classification Based on Decision Tree Algorithm for Machine Learning," Journal of Applied Science and Technology Trends, Mar. 2021, Vol. 2. pp.20–28, doi: 10.38094/jastt20165..

35. **Unpingco, J.,"**Python Programming for Data Analysis," Springer International Publishing, 2021. doi: 10.1007/978-3-030-68952-0..

36. **Hossin, M. and Sulaiman, M.N,"**A Review on Evaluation Metrics for Data Classification Evaluations," International Journal of Data Mining & Knowledge Management Process (IJDKP) , March 2015, Vol. 5. doi: 10.5121/ijdkp.2015.5201 .

37. **Bhandari, Aniruddha.** Everything you Should Know about Confusion Matrix for Machine Learning. analytics vidhya. [Online] April 17, 2020. https://www.analyticsvidhya.com/terms/.

Appendix 1

Table 3 Parameters measurements

	humidity	carbondioxide	temp	season
0	40	23	28.5	0
1	39	135	28.5	0
2	40	400	34.7	1
3	39	401	34.7	1
4	39	402	34.7	1
5	40	403	34.7	1
6	41	404	34.2	1
7	45	405	33.8	1
8	48	406	33.8	1
9	49	407	33.3	1

Table 4 Prediction vice Actual Value

	Season	Predicted
275	0	0
93	0	0
6	0	0
167	0	0
90	0	0
...	...	...
11	1	1
338	0	1
22	0	0
462	0	1
508	1	1

78 rows × 2 columns

Table 5 Logistic regression classification report

	precision	recall	f1-score	support
0	0.89	0.84	0.86	56
1	0.64	0.73	0.68	22
accuracy			0.81	78
macro avg	0.76	0.78	0.77	78
weighted avg	0.82	0.81	0.81	78

Table 6 MLP classification report

	precision	recall	f1-score	support
0	0.83	0.89	0.86	56
1	0.67	0.55	0.60	22
accuracy			0.79	78
macro avg	0.75	0.72	0.73	78
weighted avg	0.79	0.79	0.79	78

Table 7Gradient Boost classification report

	precision	recall	f1-score	support
0	0.86	0.79	0.82	68
1	0.66	0.75	0.70	36
accuracy			0.78	104
macro avg	0.76	0.77	0.76	104
weighted avg	0.79	0.78	0.78	104

Table 8 KNN Classification report

	precision	recall	f1-score	support
0	0.85	0.81	0.83	68
1	0.67	0.72	0.69	36
accuracy			0.78	104
macro avg	0.76	0.77	0.76	104
weighted avg	0.78	0.78	0.78	104

Table 9 Random Forest Classification report

	precision	recall	f1-score	support
0	0.86	0.90	0.88	61
1	0.85	0.79	0.82	43
accuracy			0.86	104
macro avg	0.85	0.85	0.85	104
weighted avg	0.86	0.86	0.85	104

Table 10 Decision Tree classification report

	precision	recall	f1-score	support
0	0.85	0.82	0.83	61
1	0.76	0.79	0.77	43
accuracy			0.81	104
macro avg	0.80	0.81	0.80	104
weighted avg	0.81	0.81	0.81	104