**Software Requirements Specification**

**for**

***Predictive Analysis on Climate Change***

Version 1.0 approved

Prepared by Abdul Wasay (1812254)

Muhammad Talha Siddiqui (1812267)

SZABIST

June 18, 2022

Table of Contents

1.Introduction…….………………………………………………………………………………3

1.1 Purpose....………………………………………………………………………………...…………….3

1.2 Document Convention.………………………………………………………………………………...3

1.3 Intended Audience and Reading Suggestion.………………………………………………………….3

1.4 Product Scope………………………………………………………………………………………….4

1.5 Feasibility Study……………..…………………………………………………………………………4

1.6 Expertise of the Team Members………………………………………………………………………..4

1.7 Milestones………………………………………………………………………………………………4

1.8 References……………………………………………………………………………………………..5

2.Overall Description……………………………………………………………………………6

2.1 Product Perspective……………………………………………………………………………………6

2.2 Product Function………………………………………………………………………………………6

2.3 Operating Environment………………………………………………………………………………..6

2.4 Design and Implementation Constraints………………………………………………………………6

3. External Interface Requirement……………………………………………………………..7

3.1 User Interfaces…………………………………………………………………………………………7

3.2 Hardware Interfaces……………………………………………………………………………………7

3.3 Software Interfaces…………………………………………………………………………………….8

4. System Features……………………………………………………………………………….9

4.1 System Feature 1………………………………………………………………………………………9

4.2 System Feature 2………………………………………………………………………………………9

5. Other Nonfunctional Requirements……………………………………………………......10

5.1 Performance Requirements………………………………………………………………………......10

5.2 Safety Requirements………………………………………………………………………………....10

5.3 Security Requirements….……………………………………………………………………………10

5.4 Business Rules……………………………………………………………………………………….10

6. UML Diagrams……………………………………………………………………………...11

6.1 Use Case Diagram…………………………………………………………………………………....11

6.2 Sequence Diagram……………………………………………………………………………………12

6.3 Activity Diagram……………………………………………………………………………………..13

6.4 Context Diagram.. ……………………………………………………………………………………14

7. Datasets………………………………………………………………………………………15

7.1 Green House Gases Dataset…………………………………………………………………………..15

7.2 AQI Dataset ……………………………………………………………………….…………………15

**7**.3 Climate Dataset………………………………………………………………………………………1**6**

7.4 Rainfall Dataset………………………………………………………………… ..…………………..16

8. Use Case………………………………………………………………………………………17

9. Other Requirement………………………………………………………………………….20

9. Appendix A: Glossary……………………………………………………………………….20

10. Appendix B: Analysis Models……………………………………………………………...21

10.1 LSTM Model………………………………………………………………………………………..21

10.2 Linear Regression Model……………………………………………………………………………22

**11.Test Cases……………………………………………………………………………………………24**

**1.Introduction**

**1.1 Purpose**

The purpose of this SRS is to collect and analyze all the ideas we have come up with in order to define the system we will create. We have defined each and every aspect of our project for the better understanding of it and outline the concepts which may or may not help in the development of the product later.

The aim of this document is to provide an explicit look of our desired project, its requirements and the objective of the project. It describes the software requirements, Datasets requirements, how the models will be displayed. It will give a detailed view of how researchers will see our product and Meteorologists can use it and can come up with future predictions.

**1.2 Document Conventions**

Main Headings

Font: Times New Roman, Size: 16, Bold

Subheadings within main headings

Font: Times New Roman, Size: 12. Bold.

Normal Text

Font: Calibri, Size 11.

**1.3 Intended Audience and Reading Suggestions**

This document is intended for:

1- Researchers: For the ease in understanding the requirements of the project according to the desired outcome and how to develop the system according to it. It will also help them in understanding the main focus point of the Project and where the extra effort is required.

2- Observation Team: The Observation team designed to pounder upon the outputs of the system can use this document as the evaluation criterion to make sure all the requirements are fulfilled.

3- Scientists: Meteorologists can use the models to predict the future precipitation in a region and prevent any hazardous situation from happening in the region.

**1.4 Product Scope**

Climate had been changing negatively throughout the world due to excessive Global warming and weakening of ozone layer and Pakistan situated in Southern Asia is facing its effects immensely in form of extreme temperatures, less rainfall and poorer air quality. The knowledge of how climate change adaptation is an important aspect of existing policy sectors and operations, it is crucial to be sure of timely climatic actions across different levels. The study of climate change has been made difficult due to the lack of relevant data available, the accuracy of the existing methods proving to be ineffective. Introduction of machine learning algorithms to interpret the data and predicting the change in future will provide the researchers with the information required either to prevent it from happening or to create adaptation policies for it. We will be using various models defined in our research onto our data scrapped from different sources and illustrate its usefulness with quantitative analysis. Our interpretations will be providing a predictive analysis for the researchers in analyzing the trends and patterns from the historical patterns to the predicted patterns and take precautionary measures in how to avoid any disastrous situation.

This system would be used to predict climate change over the years and its effects on the environment and the world. These predictions would be used to take precautions in-order to mitigate climate change and further preserve the environment.

**1.5 Feasibility Study**

With above defined scope, we’ll be able to meet our project schedule and target by overcoming following aspects:

1.5.1 Risks involved:

A. Faulty Data

1.5.2 Resource requirement:

A. Server for data processing.

B. Powerful PCs

C. Investment

**1.6 Expertise of the Team Members**

All the students involved in this respective project shall be following the proper guidelines and steps included to complete this particular project using the comprehensive and detailed guide over the internet along with other free open resources that would help us in handling and working with our Models on particular platforms.

**1.7 Milestones**

Our goal is to complete this project with all of the above-mentioned details within the time span of one year, along with:

* Gathering Data
* Analyzing the data
* Extracting useful parameters from the gathered data
* Forecasting the data
* Using extracted data to model forecasting
* Testing various models and choosing the best model for predictions
* Checking accuracy of the finalized model using MSE and MAE
* Creation of a dashboard for visualization of data.

# **1.6 References**

[1]  Khan, Z. &. (2014). *Hourly Based Climate Prediction Using Data Mining Techniques by Comprising Entity Demean Algorithm.*

[2] Liu Z L, P. C. (2010). Application of artificial neural networks in global climate change and ecological research. An overview. . *ChineseSci Bull, 2010, 55: 3853*.

[3] Mulomba Mukadi, P. G.-G. (2021). *Time Series Analysis of Climatic Variables in Peninsular Spain. Trends and Forecasting Models for Data between 20th and 21st Centuries.*

[4] S. Kareem, Z. J. (2021). An evaluation of CNN and ANN in prediction weather forecasting: A review. *Sustainable Engineering and Innovation, vol. 3, no. 2, pp. 148-159.*

[5] Sreehari, E. &. (2019). Climate Changes Prediction Using Simple Linear Regression. *Journal of Computational and Theoretical Nanoscience.*

[6] Yousif. (2013). *Weather Prediction System Using KNN Classification Algorithm.*

[7] Muhuri, P. &. (2020). *Using a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) to Classify Network Attacks.*

[8] Jitendra Kumar, R. G. (2018). *Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenters.*

[9] <https://www.logianalytics.com/predictive-analytics/predictive-algorithms-and-models/>

[10] <https://towardsai.net/p/l/the-future-of-artificial-intelligence-in-weather-forecasting>

[11] [https://scied.ucar.edu/learning-zone/climate-change-impacts/predictions-future-global climate](https://scied.ucar.edu/learning-zone/climate-change-impacts/predictions-future-global%20climate)

[12] <https://www.kaggle.com/maryamalizadeh/international-ghg-emission-data-analysis>

[13] <https://towardsai.net/p/l/the-future-of-artificial-intelligence-in-weather-forecasting>

[14] <https://www.springeropen.com/collections/AICC>

[15] <https://earth.org/data_visualization/ai-can-it-help-achieve-environmental-sustainable/>

[16] <https://www.sciencedirect.com/science/article/abs/pii/S0952197603000629>

[17] <https://en.tutiempo.net/climate/pakistan.html>

[18] <https://www.researchgate.net/publication/224082868_Engineering_the_Software.>

[19] <http://berkeleyearth.org/simmod-a-simple-python-based-climate-model-new/>

[20] <https://www.logianalytics.com/predictive-analytics/predictive-algorithms-and-models/>

[21] <https://climateknowledgeportal.worldbank.org/>

[22] https://www.hindawi.com/journals/complexity/2020/6622927/

**2. Overall Description**

**2.1 Product Perspective**

The perspective behind creating this system is that our system will be processing the variety of data observed in playing its role in climate change. We will process different datasets, we will further train models for each factor affecting climate change. Our systems would be tested for efficiency & accuracy in-order to find prediction patterns for the near future.

**2.2 Product Functions**

The main function of our system will be to showcase the models on a web page. It will show plotted climatic parameters such as precipitation, temperature, humidity and the natural disasters it follows.

The System will showcase the models of the last few years and it will also show the future prediction of those models along with any natural hazard that follows a particular region.

**2.3 Operating Environment**

This software will be deployed in a controlled environment with specific system hardware.

The minimum hardware requirement for the application to work are as follows:

OS - 64-bit Windows 8.1, Windows 10

Intel(R) Core (TM) i5-4300M CPU, 2.60GHz-3.00GHz (i5 or i7 Intel processor or equivalent AMD).

RAM – 16 (minimum) GB

Hard Drive - 20 GB 5400 RPM hard drive.

**2.4 Design and Implementation Constraints**

The Design constraints that apply on the development of the system are:

We will be using different machine learning models to train and evaluate our data such as scikit-learn which has different algorithms for analyzing the data.

Different libraries will be used to customize and plot the data such as Matplotlib, Pandas, Seaborn

For our prediction analysis, we will be using different algorithms to find the best accuracy among them.

**3. External Interface Requirements**

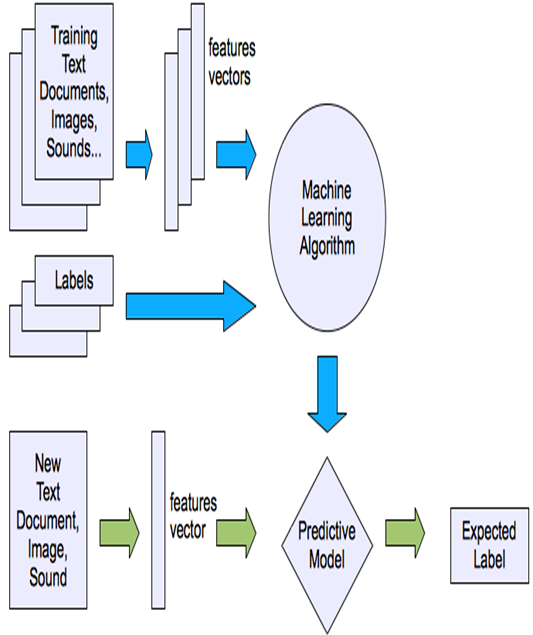
**3.1 Hardware Interfaces**

Since our system will be displaying details on a website page, there are no such special hardware interface requirements. A device with an internet connection and basic system requirements met will be able to access it.

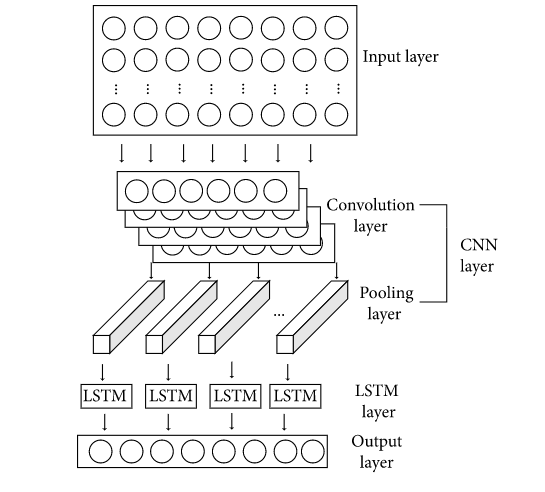
**3.3 Software Interfaces**

There is no such software in our project except for that the results will be displayed on a website page. The process to build all those models will make use of different algorithms such as Numpy, Scikit learn, and different models such as Linear regression, ARIMA, Clustering models, Classification model and K-means.

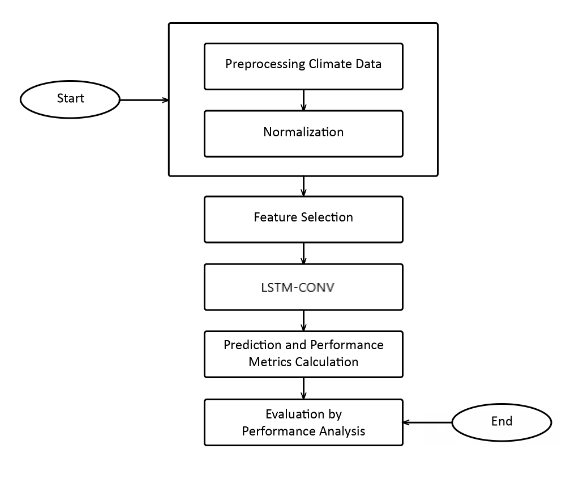
**3.3.1 General Model**



**3.3.2 CNN LSTM Memory Cell**



**3.3.3 LSTM Model**

****

**4. System Features**

**4.1 System Feature 1**

**4.1.1** - Data Management which involves Data gathering through Website Scraping, Data cleaning and data preparation for further pre-processing. (High Priority).

**4.1.2** - R&D on Algorithms (Linear Regression, Time series Forecast, PMD ARIMA, LSTM, RNN). (High Priority).

**4.1.3** - Training and Testing Model. (High Priority).

**4.1.4** - Precision and Optimization. (High Priority).

**4.1.5** - Data Visualization. (High Priority).

**4.1.6** – Choosing the Best algorithm for Yearly Prediction. (High Priority).

**4.2 System Feature 2**

**4.2.1** – Data Collection and pre-processing as per new requirements. (High Priority).

**4.2.2** - 100% of our documentation and 80% of Research paper. (High Priority).

**4.2.3** - Training and testing of our model. (High Priority).

**4.2.4** - Designing a Final System. (High Priority).

**4.2.5** - Precision and Optimization of Algorithms. (High Priority).

**4.2.6** – Checking Accuracy of model using MAE, MSE. (Medium Priority).

**4.2.7** - Website dashboard for visualization. (Low Priority).

**5. Other Nonfunctional Requirements**

**5.1 Performance Requirements**

The accuracy of our system should be above 80%. The accuracy of our model is a high priority since our system is based upon future prediction using trained models and if there is a lack of accuracy within the system, the future predictions will be unjustifiable.

**5.2 Security Requirements**

Data acquired by the Pakistan Meteorological Department is on certain rules and regulations set by them and according to which, the confidentiality of the data has to be maintained and cannot be accessed by any unauthorized personnel.

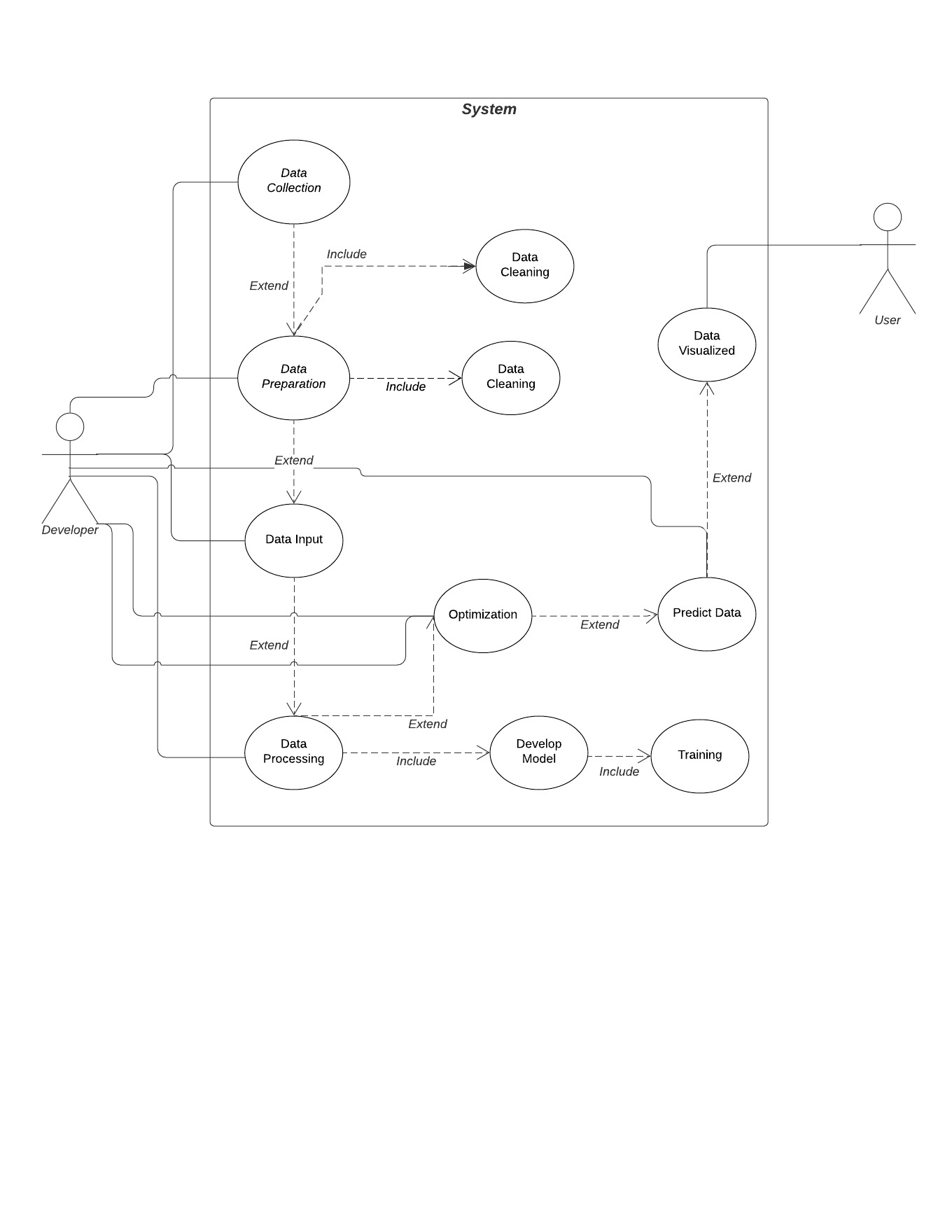
Data being used can be a target of data poisoning which will interrupt the model validation, model accuracy and hence failure in presenting the accurate predicted values.

We will be doing regression testing, manual moderation, and using various statistical techniques to detect anomalies for prevention from any sort of data poisoning.

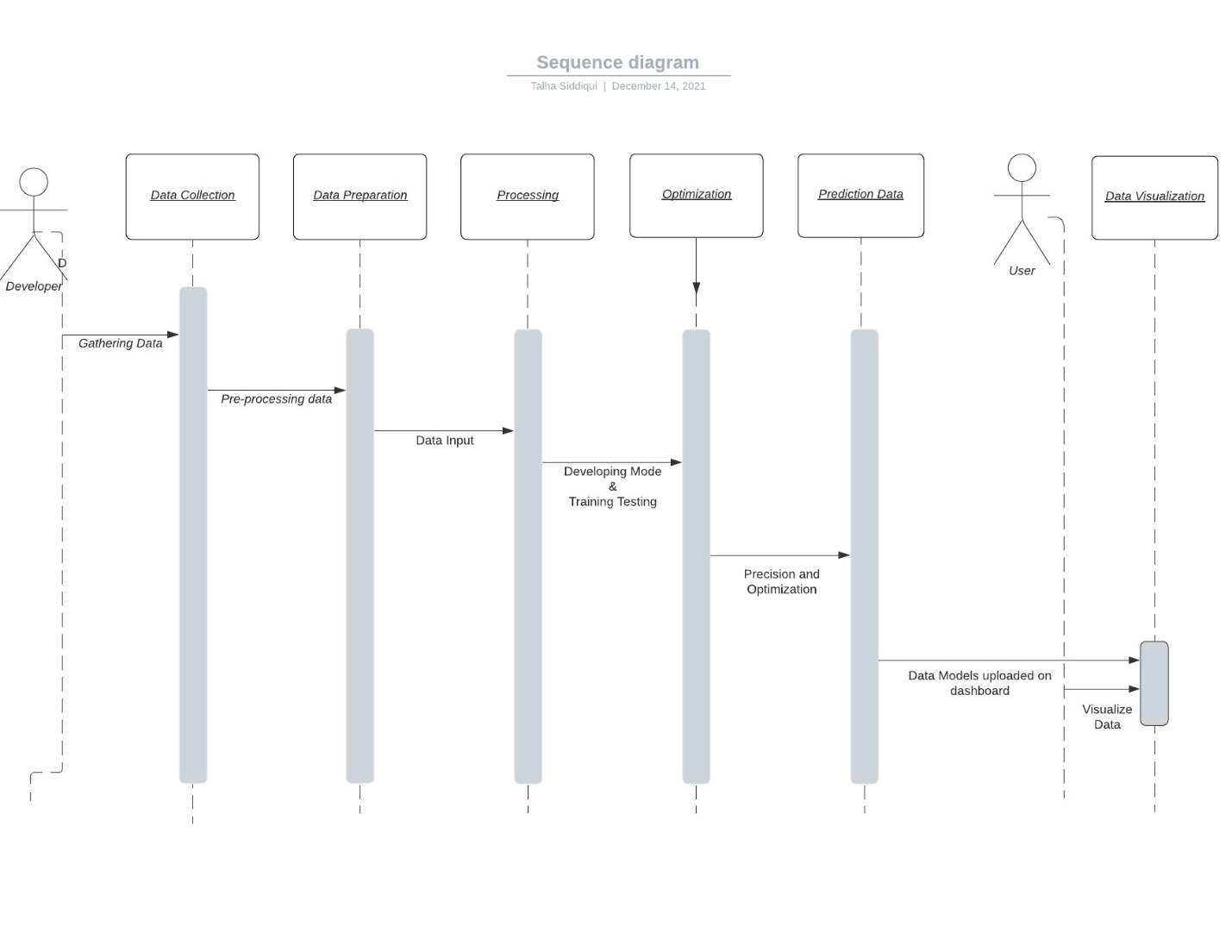
Systems being used for training and testing of the model should be kept secure from any sort of access from unauthorized user to prevent any sort of mishap and deleting of the complete progress.

**6. UML Diagrams**

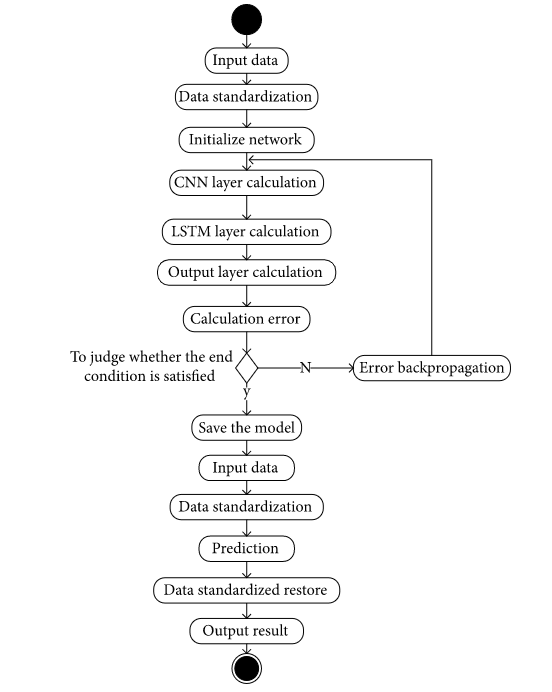
**6.1 Use case Diagram**

****

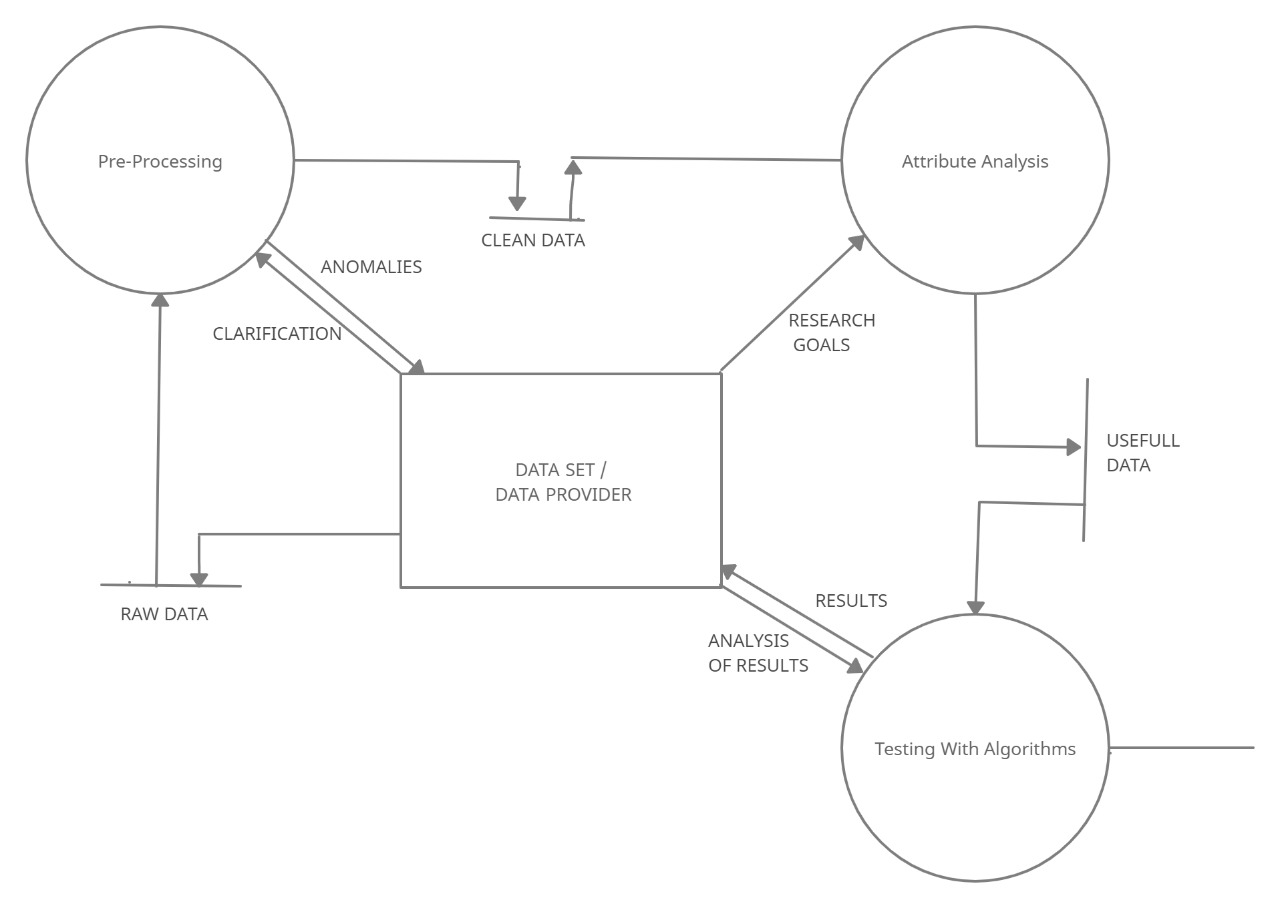
**6.2 Sequence Diagram**

****

**6.3 Activity Diagram**

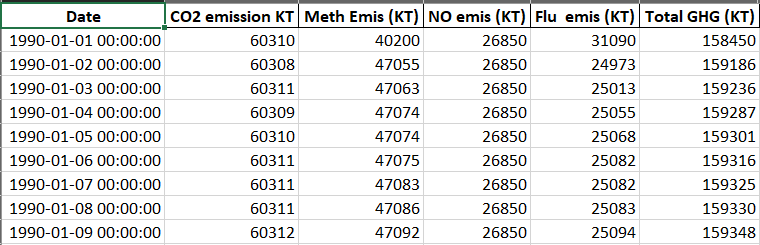
****

**6.4 Context Diagram**

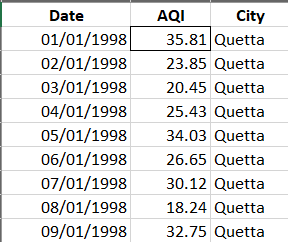
****

**7. Datasets**

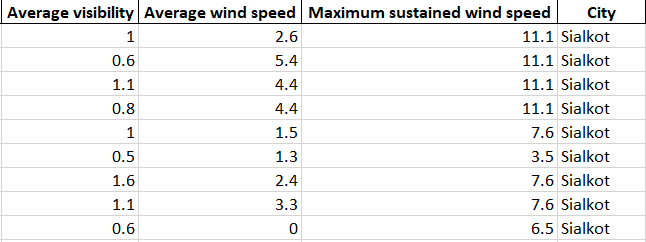
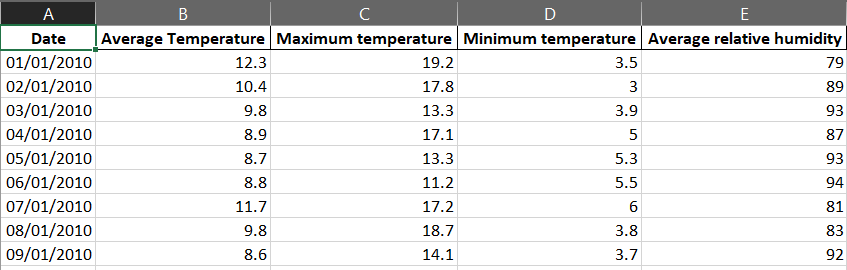
**7.1 Green House Gases**

****

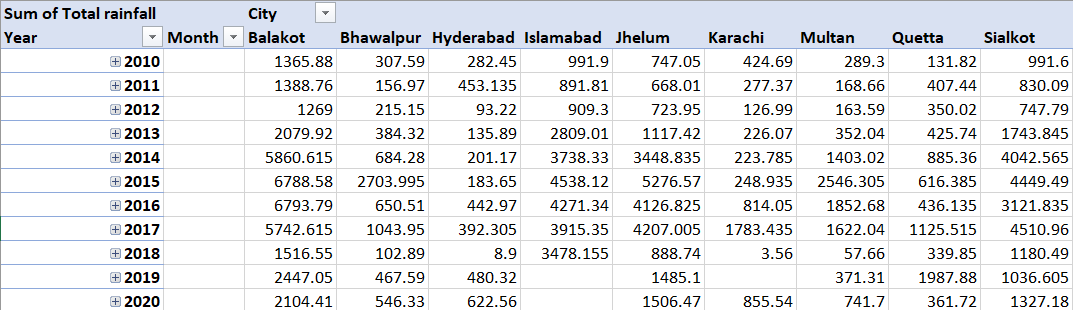
**7.2 Air Quality**

****

**7.3 Climate**

****

**7.4 Rainfall**

****

**8. USE CASE**

**8.1 Use Case**

|  |  |  |
| --- | --- | --- |
| **Use Case Name:** | Data Analytics | |
| **ID:** | AC 1 | |
| **Actors Involved:** | Developers | |
| **Brief Description** | Performing data analysis on the available datasets. | |
| **Pre-Conditions** | --- | |
| **Post-Conditions** | Will perform data collection, cleansing and analysis  Will send the data in the system | |
| **Normal Flow of Events:** | **Actor Action** | **System Response** |
| 1. The Actor will enter data into the system. | 1. System will perform operations on the data through the algorithms |

**8.2**

|  |  |  |
| --- | --- | --- |
| **Use Case Name:** | System | |
| **ID:** | AC 2 | |
| **Actors Involved:** | System (Spyder Notebook) | |
| **Brief Description** | Converting the data into useful information using Python Language | |
| **Pre-Conditions** | --- | |
| **Post-Conditions** | Will perform data collection, cleansing and analysis  Will send the data in the system | |
| **Normal Flow of Events:** | **Actor Action** | **System Response** |
| 1. Will Covert the data | 1. Convert the data into useful information using python script. |

**8.3**

|  |  |  |
| --- | --- | --- |
| **Use Case Name:** | Data Normalization | |
| **ID:** | AC 3 | |
| **Actors Involved:** | Developer and System (Spyder Notebook) | |
| **Brief Description** | Doing Normalization onto the data | |
| **Pre-Conditions** | --- | |
| **Post-Conditions** | --- | |
| **Normal Flow of Events:** | **Actor Action** | **System Response** |
| 1. Will provide the data | 1. Will Normalize the data for further processing |

**8.4**

|  |  |  |
| --- | --- | --- |
| **Use Case Name:** | Over Sampling and Denormalization | |
| **ID:** | AC 4 | |
| **Actors Involved:** | Developer and System (Spyder Notebook) | |
| **Brief Description** | Doing oversampling on the limited data for algorithm training and performing Denormalization. | |
| **Pre-Conditions** | AC 3 | |
| **Post-Conditions** | --- | |
| **Normal Flow of Events:** | **Actor Action** | **System Response** |
| 1. Will provide the normalized data | 1. Will process oversampling using python script and denormalize it for further processing. |

**8.5**

|  |  |  |
| --- | --- | --- |
| **Use Case Name:** | Model Training | |
| **ID:** | AC 5 | |
| **Actors Involved:** | Developer and System (Spyder Notebook/Google Colab) | |
| **Brief Description** | Providing data to the algorithm and training the model for prediction. | |
| **Pre-Conditions** | --- | |
| **Post-Conditions** | --- | |
| **Normal Flow of Events:** | **Actor Action** | **System Response** |
| 1. Will input the data and set training and testing values. | 1. Will process the algorithm and train the model according to the inputted epochs and predict the future. |

**8.6**

|  |  |  |
| --- | --- | --- |
| **Use Case Name:** | Visualization dashboard | |
| **ID:** | AC 6 | |
| **Actors Involved:** | Developer and System (Spyder Notebook) | |
| **Brief Description** | Creating excel files of the predicted data and loading them onto dashboard. | |
| **Pre-Conditions** | AC5 | |
| **Post-Conditions** | --- | |
| **Normal Flow of Events:** | **Actor Action** | **System Response** |
| 1. Will write the python script for creation of dashboard and input files. | 1. Will process the script, create the tabs and load the predicted data onto the graphs. |

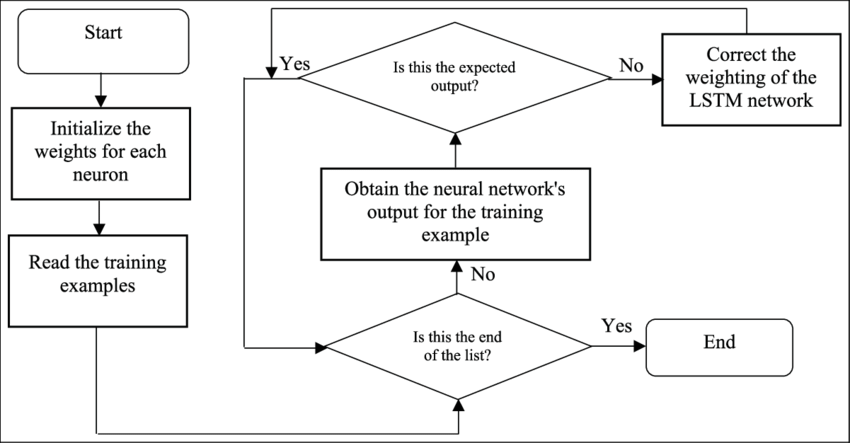
**9. Other Requirements**

**Appendix A: Glossary**

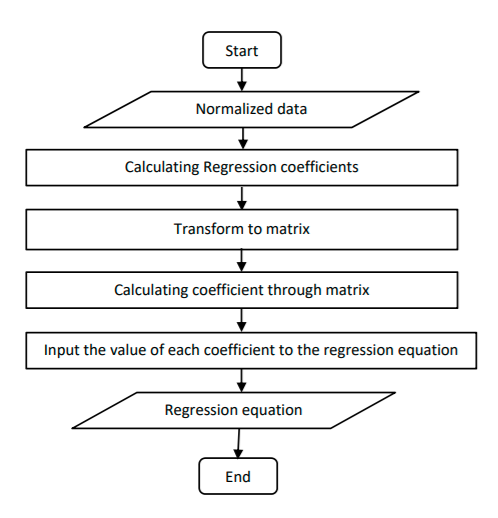
|  |  |
| --- | --- |
| Term | Definitions |
| LSTM | Long short-term memory (LSTM) is an artificial recurrent neural network architecture used in the field of deep learning |
| CNN | A **Convolutional Neural Network** is a Deep Learning algorithm. They are composed of multiple layers of artificial neurons. |
| PMD ARIMA | PMD ARIMA is a statistical library designed to fill the void in Python's time series analysis capabilities. |
| MSE | The Mean Squared Error (MSE) is a measure of how close a fitted line is to data points. |
| MAE | The Mean Absolute Error of a model with respect to a test set is the mean of the absolute values of the individual prediction errors on over all instances in the test set. |
| RNN | Recurrent Neural Networks enable you to model time-dependent and sequential data problems |

**10. Appendix B: Analysis Models**

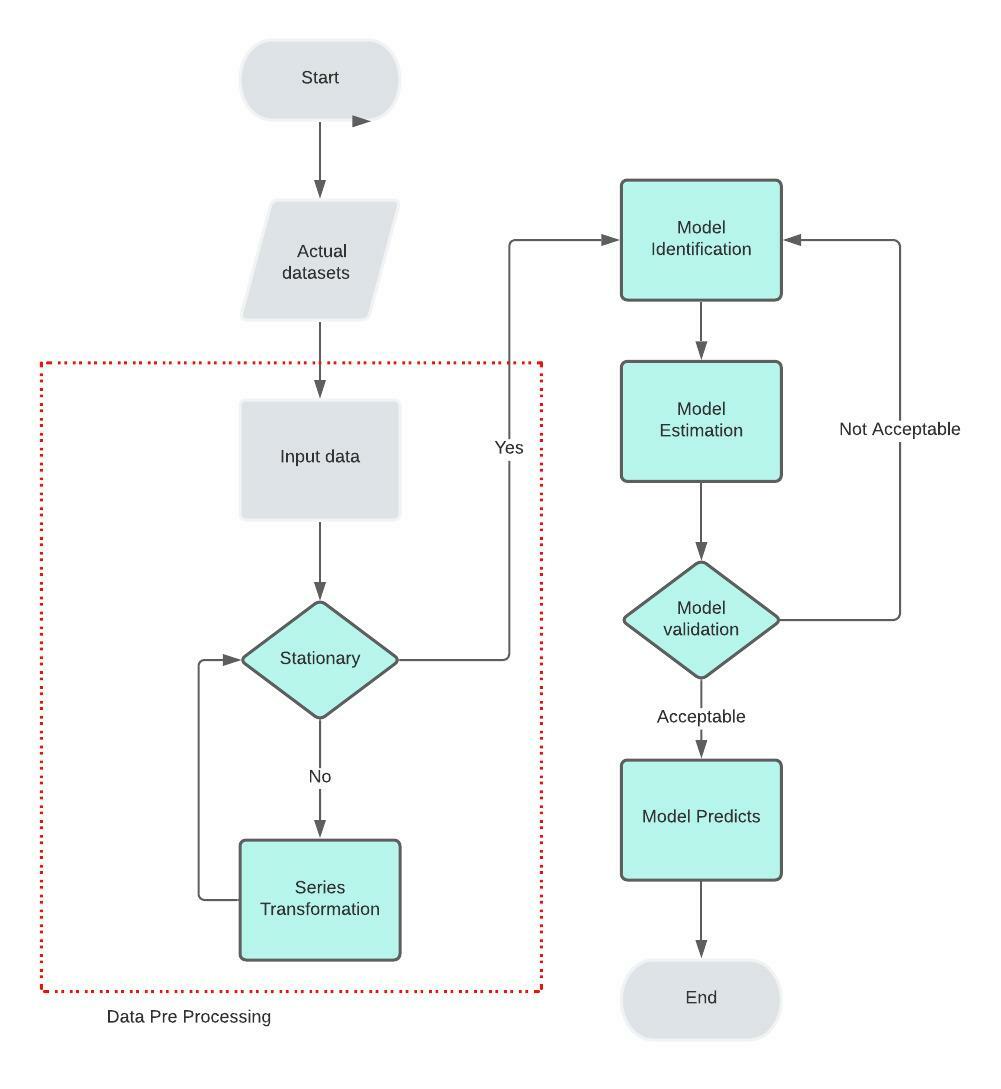
**10.1 LSTM Model**

****

* 1. **Linear Regression Model**

****

* 1. **Time series forecast**

****

**11. TEST CASES**

**11.1 Test Case: Max Temperature**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test** **Case**  **ID** | **Test Case**  **Name** | **Test Case**  **Summary** | **EPOCH TIME** | **EPOCH TIME LOSS** | **MSE** | **MAE** |
| 1. | Max Temp Forecasting | Forecasting maximum temperature using CONV1D LSTM | 232s | 0.0127 | 0.01269 | 0.08285 |
| 2. | Max Temp Forecasting (2) | Forecasting maximum temperature using LSTM | 232s | 0.0463 | 0.04632 | 0.16820 |

**11.2 Test Case: Min Temperature**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test**  **Case**  **ID** | **Test Case**  **Name** | **Test Case**  **Summary** | **EPOCH TIME** | **EPOCH TIME LOSS** | **MSE** | **MAE** |
| 1. | Min Temp Forecasting | Forecasting Minimum temperature using CONV1D LSTM | 244s | 0.0111 | 0.01108 | 0.07907 |
| 2. | Min Temp Forecasting (2) | Forecasting Minimum temperature using LSTM | 235s | 0.0488 | 0.04875 | 0.17182 |

**11.3 Test Case: Max Temperature**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test**  **Case**  **ID** | **Test Case**  **Name** | **Test Case**  **Summary** | **EPOCH TIME** | **RMSE** | **MSE** | **Residual Sum** |
| 1. | Max Temp Forecasting | Forecasting maximum temperature using Linear Regression | 168 | 1.74697404157 | 3.0320551079734 | 9.713125 |
| 2. | Max Temp forecasting | Forecasting maximum temperature using Multiple Linear Regression | 142 | 2.64860578104 | 7.26230252865065 | 6.1546 |

**11.4 Test Case: Bahawalpur Max/Min Temperature**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test** **Case**  **ID** | **Test Case**  **Name** | **Test Case**  **Summary** | **EPOCH TIME** | **EPOCH TIME LOSS** | **MSE** | **MAE** |
| 1. | Max Temp Prediction  (5years) | Predicting maximum temperature using CONV1D LSTM | 250s | 3.8610e-04 | 2.4048e-04 | 0.0125 |
| 2. | Min Temp Prediction  (5years) | Predicting maximum temperature using CONVID LSTM | 260s | 1.9697e-04 | 2.3751e-04 | 0.0109 |

**11.5 Test Case: Bahawalpur Relative Humidity and AQI**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test** **Case**  **ID** | **Test Case**  **Name** | **Test Case**  **Summary** | **EPOCH TIME** | **EPOCH TIME LOSS** | **MSE** | **MAE** |
| 1. | Relative Humidity Prediction (5years) | Predicting Relative Humidity using CONV1D LSTM | 270s | 5.1030e-04 | 3.8340e-04 | 0.0164 |
| 2. | AQI Prediction (5years) | Predicting AQI using CONVID LSTM | 241s | 9.0697e-03 | 6.605e-04 | 0.0217 |

**11.6 Test Case: Multan Max/Min Temperature and Humidity**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test** **Case**  **ID** | **Test Case**  **Name** | **Test Case**  **Summary** | **EPOCH TIME** | **EPOCH TIME LOSS** | **MSE** | **MAE** |
| 1. | Max Temp Prediction  (5years) | Predicting maximum temperature using CONV1D LSTM | 285s | 8.8354e-04 | 7.6510e-04 | 0.0067 |
| 2. | Min Temp Prediction  (5years) | Predicting minimum temperature using CONV1D LSTM | 248 | 6.0826e-04 | 7.746e-04 | 0.0192 |
| 3. | Relative Humidity Prediction (5years) | Predicting Relative Humidity using CONV1D LSTM | 241s | 2.9797e-03 | 2.0298e-04 | 0.0105 |

**11.7 Test Case: Quetta Max/Min Temperature and Rainfall**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test** **Case**  **ID** | **Test Case**  **Name** | **Test Case**  **Summary** | **EPOCH TIME** | **EPOCH TIME LOSS** | **MSE** | **MAE** |
| 1. | Max Temp Prediction  (5years) | Predicting maximum temperature using CONV1D LSTM | 285s | 3.2154e-04 | 2.7550e-04 | 0.00184 |
| 2. | Min Temp Prediction  (5years) | Predicting minimum temperature using CONV1D LSTM | 248 | 2.5681e-04 | 2.5681e-04 | 0.0122 |
| 3. | Rainfall Prediction (5years) | Predicting Rainfall using CONV1D LSTM | 241s | 5.8463e-07 | 4.7989e-07 | 6.2086e-04 |

**11.8 Test Case: Karachi Max/Min Temperature, Rainfall and AQI**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test** **Case**  **ID** | **Test Case**  **Name** | **Test Case**  **Summary** | **EPOCH TIME** | **EPOCH TIME LOSS** | **MSE** | **MAE** |
| 1. | Max Temp Prediction  (5years) | Predicting maximum temperature using CONV1D LSTM | 285s | 6.0264-04 | 1.5261e-04 | 0.0421 |
| 2. | Min Temp Prediction  (5years) | Predicting minimum temperature using CONV1D LSTM | 248 | 5.2155e-04 | 2.4556e-04 | 0.0512 |
| 3. | Rainfall Prediction (5years) | Predicting Rainfall using CONV1D LSTM | 241s | 5.1297e-07 | 3.0754e-04 | 4.2123e-04 |
| 4. | AQI Prediction (5years) | Predicting AQI using CONV1D LSTM | 352s | 0.3472 | 0.1215 | 0.4775 |

**11.9 Test Case: CO2 Emission**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test** **Case**  **ID** | **Test Case**  **Name** | **Test Case**  **Summary** | **EPOCH TIME** | **EPOCH TIME LOSS** | **MSE** | **MAE** |
| 1. | CO2 Prediction  (5years) | Predicting CO2 Emission using CONV1D LSTM | 400s | 9.001e-04 | 6.1210e-04 | 0.9867 |