**Software Design Specification**

**for**

***Predictive Analysis on Climate Change***

Version 1.0 approved

Prepared by Abdul Wasay

Muhammad Talha Siddiqui

SZABIST

Feb 01, 2021

Table of Contents

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

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

1.2 Project Scope….…………………………………………………………………………….3

1.3 Definitions, acronyms and abbreviations……………………………………………………..3

1.4 References…..……………………………………………………………………………....4

1.5 Overview of document………..……………………………………………………………...4

2. System Architecture Description…………………………………………………………5

2.1 Section Overview…….…..………………………………………………………………….5

2.2 General Constraints …..…..………………………………………………………………….5

2.3 Program Structure……..…………………………………………………………………….6

2.4 Alternatives Considered……..………………………………………………………………6

3. Detailed description of components……………………………………………………...7

3.1 Section Overview…………..………………………………………………………………..7

3.2 Component and Detail…..……..…………………………………………………………….7

3.2.1 Description….…..……..……………………………………………………………………7

3.2.1 Data Members………………………………………………………………………………7

3.3 SDS Components…………………………………………………………………………………11

3.2.1 Component 1. ….…..……..……………………………………………………………….11

3.3.2 Component 2….…..……..…………………………………………………………...11

3.3.3 Component 3….…..……..………………………………………………………………...12

3.3.4 Component 4….…..……..………………………………………………………………...12

4. User Interface Design…………………………………………………………………….13

4.1 Section Overview….…..……..……………………………………………………………..13

4.2 Detailed Description….…..……..…………………………………………………………..14

5. Reuse and Relationships….…..……..……………………………………………………18

6. Design decisions and tradeoff….…..……..……………………………………………...18

7. Pseudocode for components….…..……..……………………………………………….18

8. Appendices….…..……..………………………………………………………………….19

8.1 Activity Diagram….…..……..……………………………………………………………..20

8.2 Sequence Diagram….…..……..……………………………………………………………21

8.3 Use Case Diagram….…..……..…………………………………………………………….22

8.4 Context Diagram….…..……..……………………………………………………………..23

8.5 Deployment Diagram….…..……..…………………………………………………………24

8.6 System Block Diagram….…..……..………………………………………………………..25

8.7 Work Flow Diagram….…..……..……………………………………………………...…...26

8.8 System Architecture Diagram….…..……..…………………………………………………27

1. **Introduction**
   1. **Purpose**

The Purpose of this Software Design Document is to provide a detailed description of the design of the models. It will help a developer to understand which Machine learning model is used, why is it used and what is the expected outcome from those models.

Each and every aspect of the project is thoroughly defined for the better understanding of it and outline the concepts which may or may not help in the development of the product later.

**1.2 Scope of the development project**

As we know Climate change not only leads to extreme weather events and dangerous health effects the world over, but, as a threat multiplier, it poses a direct risk to human survival for many of the world’s most vulnerable communities.

In this project we examine and assess the patterns of change in these conditions and the magnitude of their effects on the global climate. We proposed a system which provides the overall analysis of climate change over the years and the prediction for the future. We believe that mitigating climate change is not only important for global peace and security, but key to fulfilling our moral imperative to protect those communities most at risk. Our work is time sensitive and urgent.

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. **Definitions, acronyms, and abbreviations**

|  |  |
| --- | --- |
| **Term** | **Definition** |
| **LSTM** | Long short-term memory (LSTM) is an artificial [recurrent neural network](https://en.wikipedia.org/wiki/Recurrent_neural_network) architectureused in the field of [deep learning](https://en.wikipedia.org/wiki/Deep_learning). |
| **CNN** | A **Convolutional Neural Network** is a Deep Learning algorithm. They are composed of multiple layers of artificial neurons |
| **MSE** | Mean Square Error |
| **MAE** | Mean Absolute Error |
| **Linear Regression** | Linear regression models use a straight line, while logistic and nonlinear regression models use a curved line |
| **RSS** | The residual sum of squares (RSS) measures the level of variance in the error term, or residuals, of a regression model. |

# **1.4 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/>

**1.5 Overview of document**

The rest of the document will be briefly describing all the aspects of climate and how they are affecting in the climate change. Machine learning models forecasting the trends in how each factor will change in the future and its effect on the climate.

Lastly the user will find all the relevant relations and UML diagrams at the end.

1. **System architecture description**

This section provides an overview and rationale for the program's data and architectural design decisions.

* 1. **Section Overview**  
     This section explains in detail regarding the constraints we had in acquiring the data, the structure of the data we acquired.
  2. **General Constraints**

There are certain limitations that we faced to get the results we wanted. These limitations are as follows:

* Missing values in our data sets and outliers.
* **Shortage of massive data sets needed to train machine learning**
* **There were times when we faced trouble in data labeling.**

To get the desired results, we overcame by using linear interpolation and eliminating the outliers with the help of data visualization tools which made us understand the nature of the data.

We used data interpolation to increase our data. We used Quadratic Sum Formula which is:

Yt1=1/4{Yt-4,5/12(Yt-Yt-1)}

Yt2=1/4{Yt-1,5/12(Yt-Yt-1)}

Yt3=1/4{Yt+1,5/12(Yt-Yt-1)}

Yt4=1/4{Yt+4,5/12(Yt-Yt-1)}

We also used Data oversampling to increase the size of training data so that our model would be trained in an appropriate manner.

* 1. **Pseudocode**

(1) Input data: input the data required for CNN-LSTM training.

(2) Initialize network: initialize the weights and biases of each layer of the CNN-LSTM.

(3) CNN layer calculation: the input data are successively passed through the convolution layer and pooling layer in the CNN layer, the feature extraction of the input data is carried out, and the output value is obtained.

(4) LSTM layer calculation: the output data of the CNN layer are calculated through the LSTM layer, and the output value is obtained.

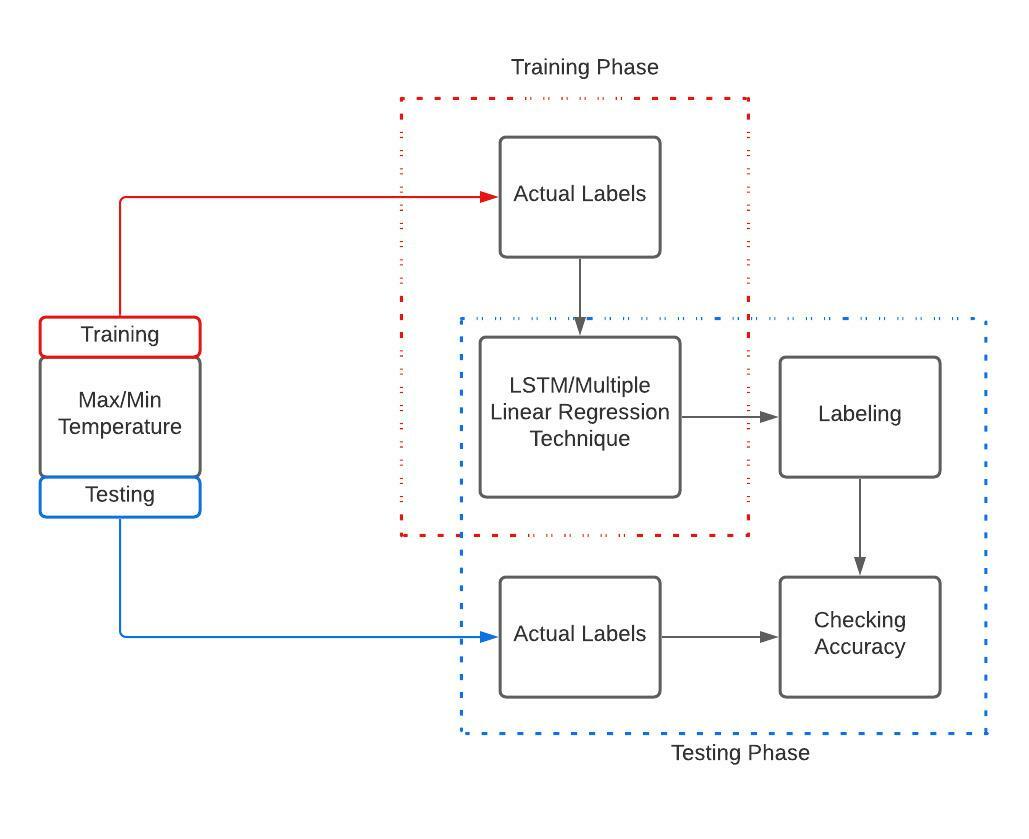
(5) Output layer calculation: the output value of the LSTM layer is input into the full connection layer to get the output value.

(6) Save the model: save the trained model for forecasting

(7) Input data: input the input data required for the forecasting.

(8) Output result: output the restored results to complete the forecasting process.

* 1. **Program Structure**

****

* 1. **Accuracy Calculation**

The (MSE) mean square error and (MAE) mean absolute error is a method of measuring how accurate a predicted system is.

Mean Square Error

* *Yi* = Actual value
* *Yi* = Predicted value
* *n* = Total number of samples.
  1. **Alternative Considered**There are no alternatives that we had to consider. The architecture of our data model is at it is required.

1. **Detailed description of components**
   1. **Section Overview**  
      This section will be providing the thorough details of the components used in our project.
   2. **Component and Detail (include a sub-section for each component)**  
      A structured description usually works. For example, if your components are classes you may wish to include the following subsections
      1. Description:

The data sets we acquired were from web scraping. Our first dataset consists of Avg temperature, Min temperature, Max temperature, Relative humidity and Rainfall. The data is from 1st of January 2010 to December 2020. We have 11 cities data from Pakistan and it is measured on daily basis.

Our second dataset consists of Greenhouse Gases factors which consists of Fossil Fuel emission, CO2 emission, Methane Emission and Total Green House gas emission which is measured in kilo tones. We have monthly based data from 1969 to 2020. We Intend to interpolate this data to daily basis emissions on each factor.

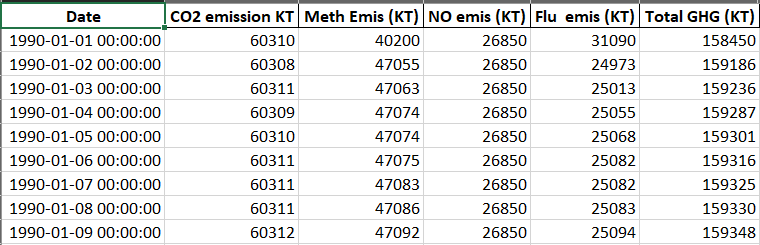
Our third dataset is in its initial basis which is a sea level pressure dataset. This data is recorded from Karachi port and is dated from 1916 to 2016.

All the data is acquired from reliable sources and cross checked for their validity.

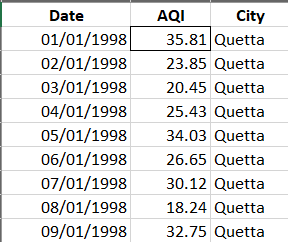
* + 1. Data Members

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO | Attribute | Type | Description |
| 1. | DATE | Datetime | Year, month, day. |
| 2. | AVG | Float | Average temperature |
| 3. | MAX | Float | Maximum temperature |
| 4. | MIN | Float | Minimum temperature |
| 5. | R.H | Float | Relative Humidity |
| 6. | PRCP | Float | Total Precipitation |
| 7. | WS | Numeric | Maximum Wind Speed |
| 8. | NO | Float | Nitrogen Oxide emission |
| 9. | CO2 | Float | Carbon Dioxide emission |
| 10. | Meth | Float | Methane emission |
| 11. | Flu Gases | Float | Fluorinated gases |
| 12. | GHG | Float | Total Green House |

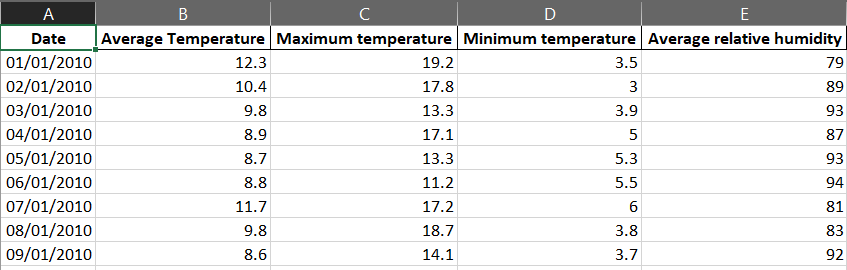
1. **Green House Gases Dataset**

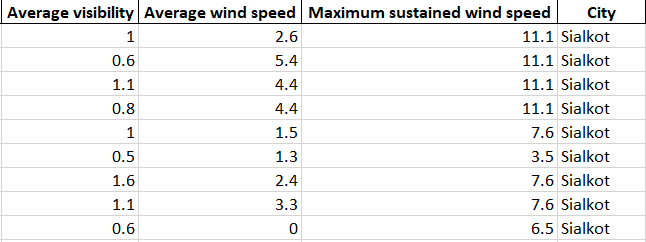
****

1. **Air Quality Dataset**

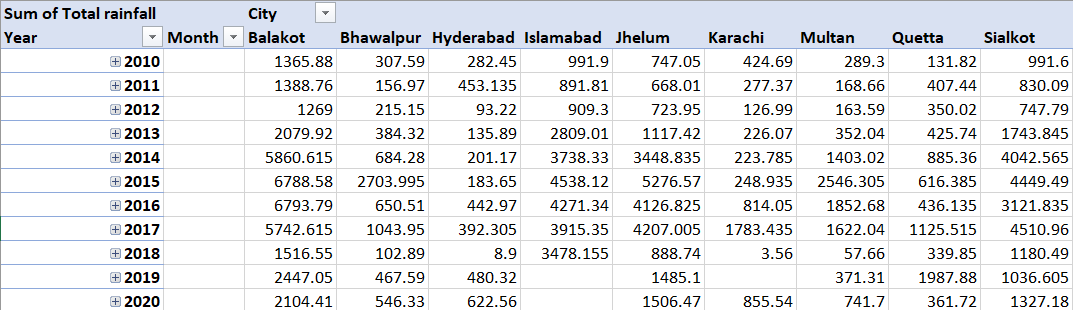
****

1. **Climate Dataset**

****

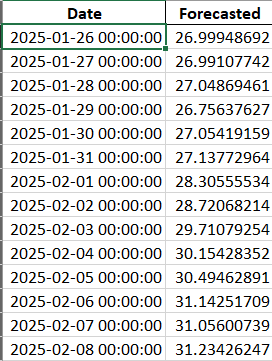
****

1. **Rainfall**

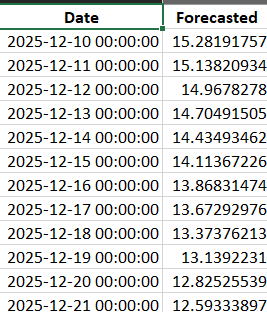


**Predicted Data:**

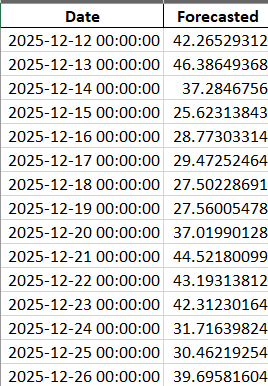
* **Karachi Max Temperature**



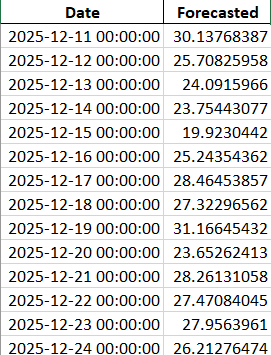
* **Karachi Min Temperature**



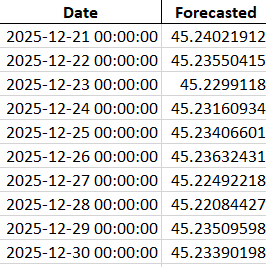
* **Karachi Humidity**



* **Karachi Wind Speed**



* **Karachi AQI**



* 1. SDS Components
     1. Component 1

|  |  |
| --- | --- |
| Identification | **Data Set** |
| Type | Xlsx, csv |
| Purpose | Dataset is used for training, validating, and testing the ML model. Bigger the set of data, better are the learning opportunities for the model, higher are the chances of achieving accuracy in results. |
| Function | The main function of using dataset is to predict the future values using the historical ones. |
| Dependencies | The Data is dependent upon the recorded values via the sensors. |
| Resources | Data was acquired by web scraping the tutiempo website, accuweather and data world bank.  The operating system requirements are Windows 7,8 or 10. |
| Processing | The system processes whenever the data is being processed through testing or training. |

* + 1. Component 2

|  |  |
| --- | --- |
| Identification | Predictive Algorithms |
| Type | Classification or Regression. |
| Purpose | an algorithm as a mathematical or logical program that turns a data set into a model. |
| Function | Machine learning algorithms use computational methods to “learn” information directly from data without relying on a predetermined equation as a model. |
| Dependencies | The efficiency of an algorithm is dependent upon the size of the data. |
| Processing | It works by analyzing current and historical data and projecting what it learns on a model generated to forecast likely outcomes. |

* + 1. Component 3

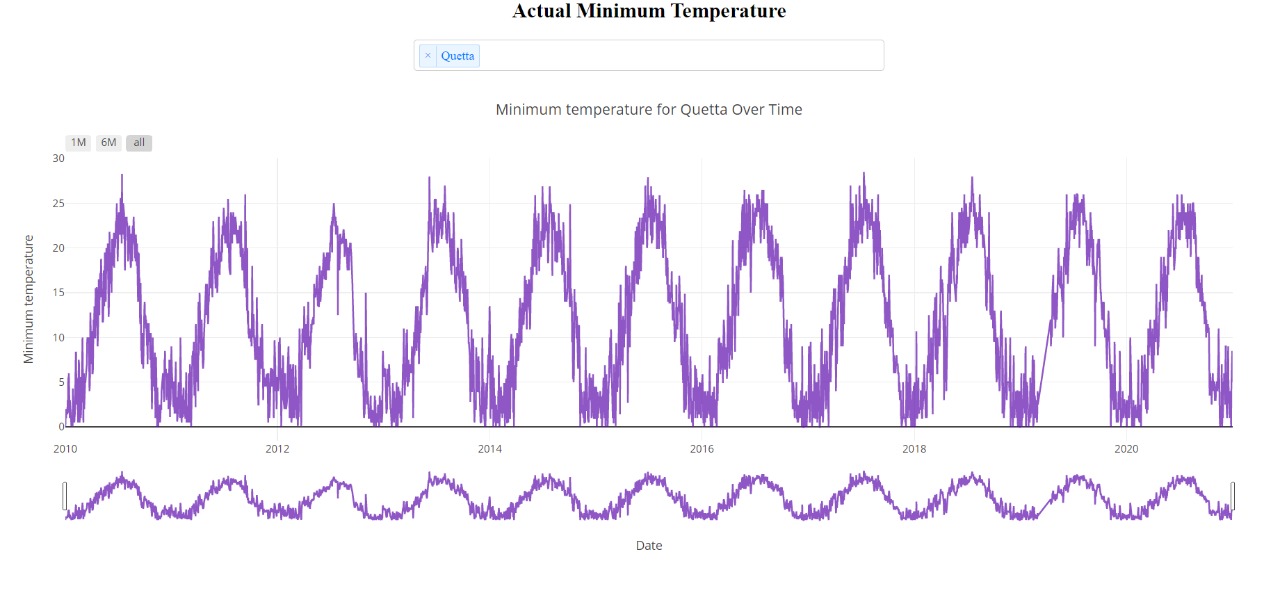
|  |  |
| --- | --- |
| Identification | Predictive Models |
| Type | Convolutional LSTM |
| Purpose | A model is a computational representation of real-world processes. |
| Function | An ML model is trained to recognize certain types of patterns by training it over a set of data using relevant algorithms. Once a model is trained, it is used to make predictions. |
| Dependencies | It is dependent upon the current and historical data, analyzing it. |
| Processing | It works by analyzing current and historical data and projecting the outcomes. |

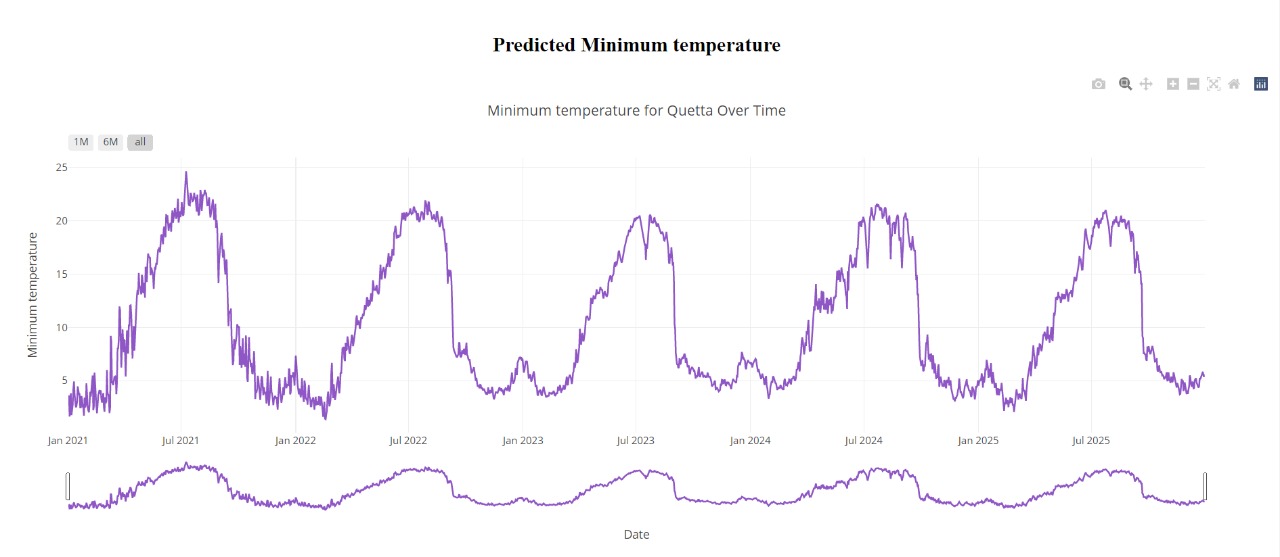
* + 1. Component 4

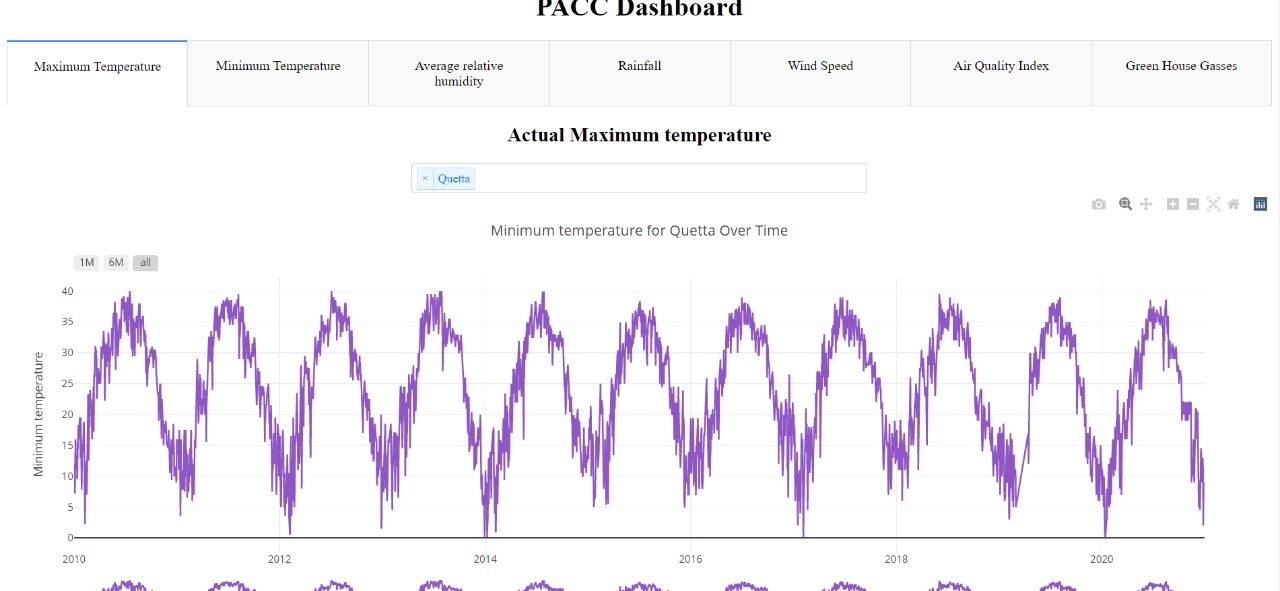
|  |  |
| --- | --- |
| Identification | Data Visualization tools |
| Type | Plotly, matplotlib, pandas, |
| Purpose | These libraries are used for the visualization of the dataframe that is being used in the model. |
| Function | The main function of these libraries is to visualize the data in an attractive manner. It allows you to turn analyses into interactive web apps using only Python scripts. |
| Dependencies | These libraries are only dependent on the available environment in the system. |
| Processing | Once the plots are created, it can build fields on top of it so users can filter and sort data. |

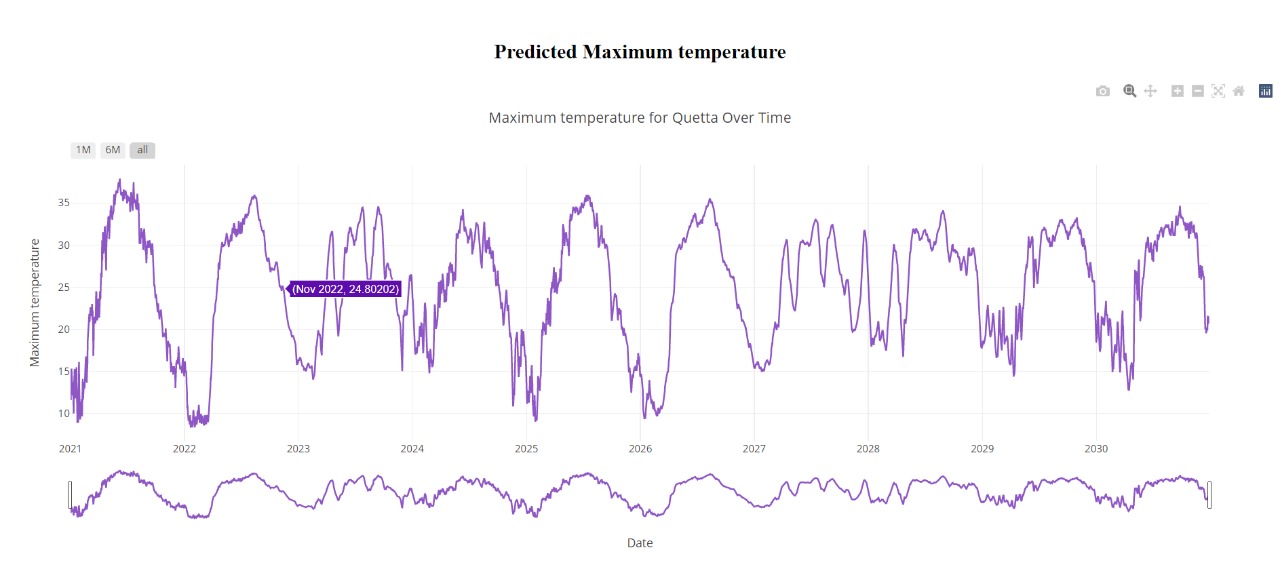
1. **User Interface Design**
   1. **Section Overview**  
      This section is providing with the actual values plotted from our dataset on the webpage using dash library of python. We created basic html page, created tabs and division to input the graphs. We created graphs of the actual dataset as well as the predicted values.
   2. **Detailed Description**

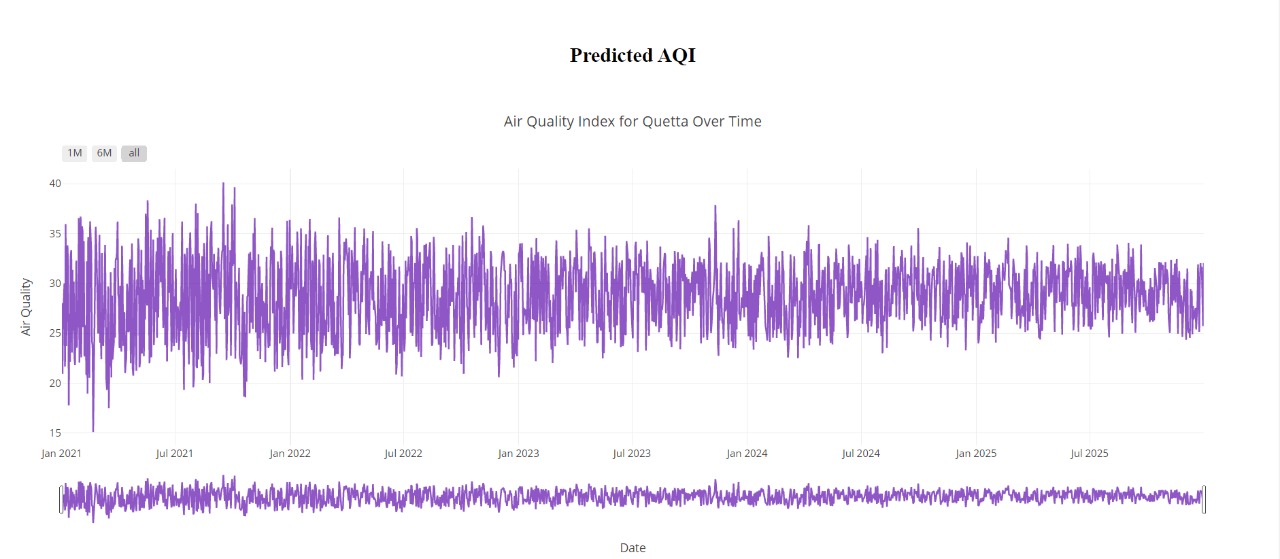
We have created a dashboard on the webpage visualizing the Actual data and also the Forecasted data Using Convolutional LSTM.

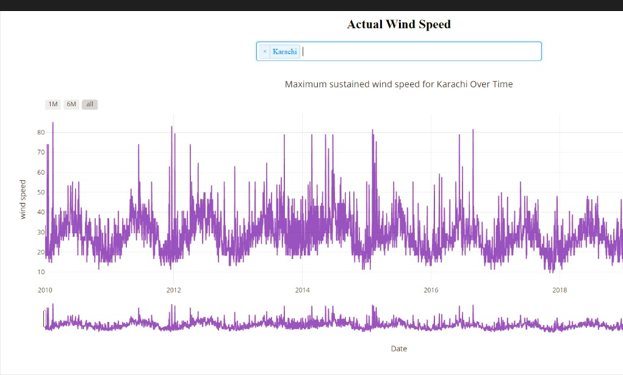


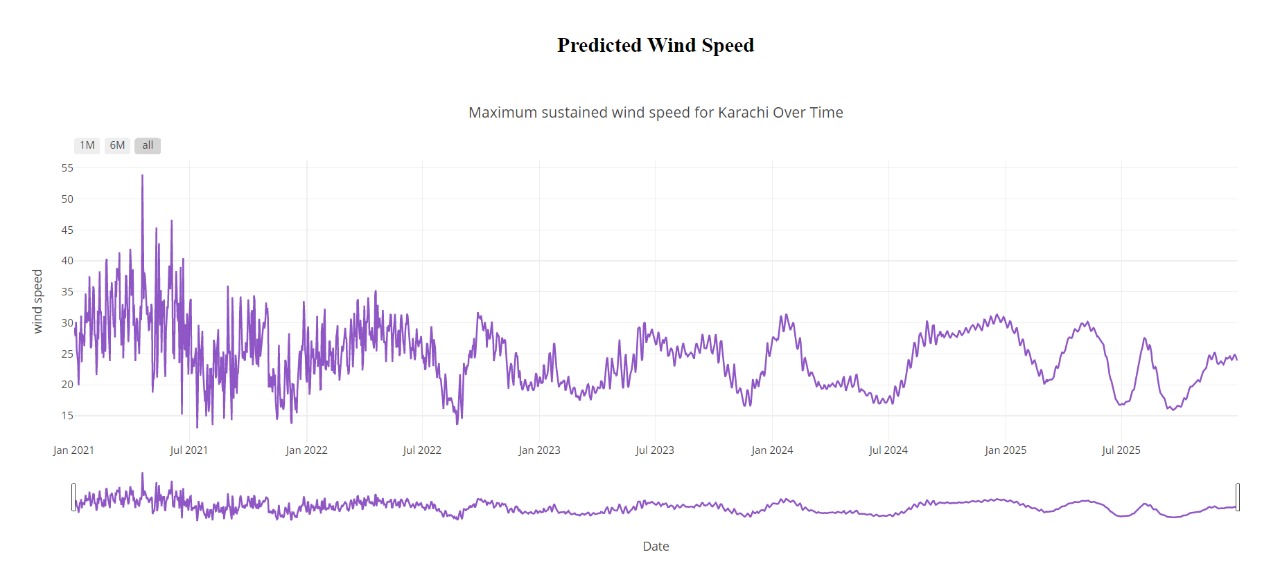


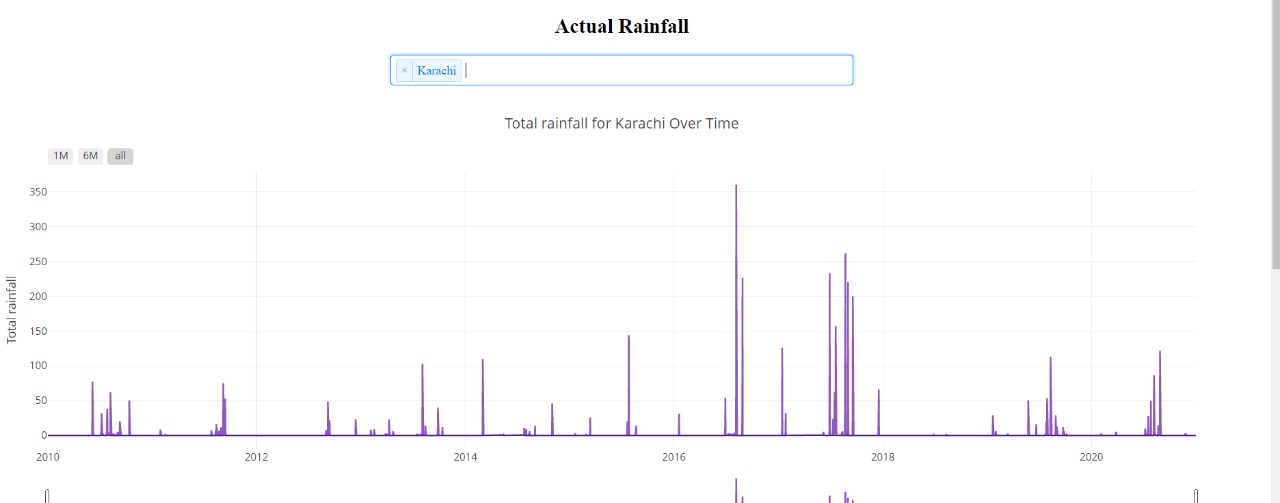


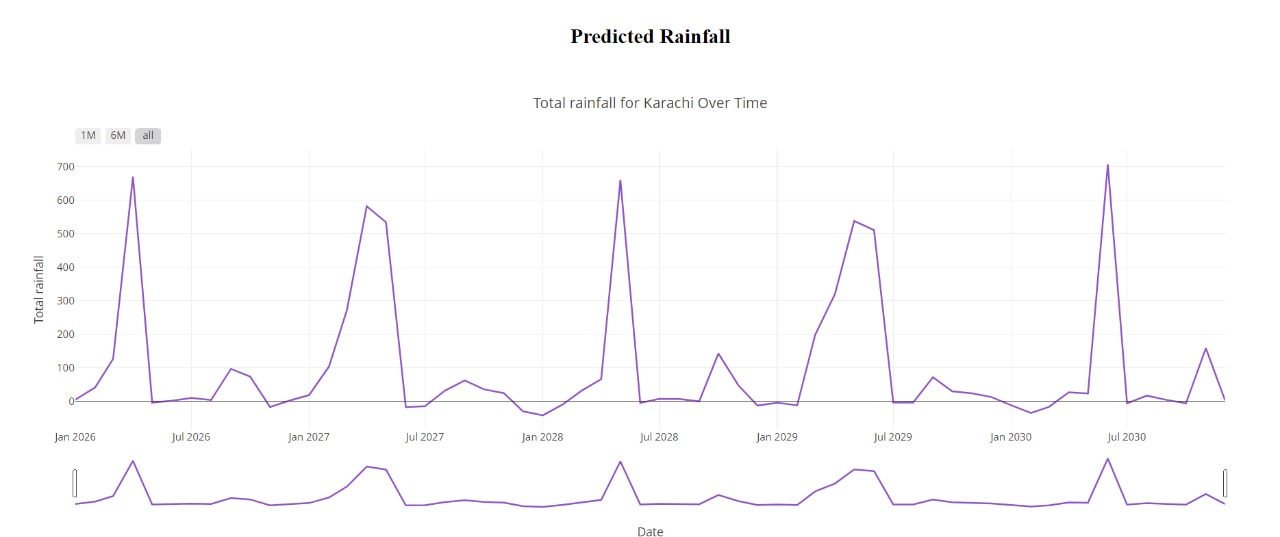


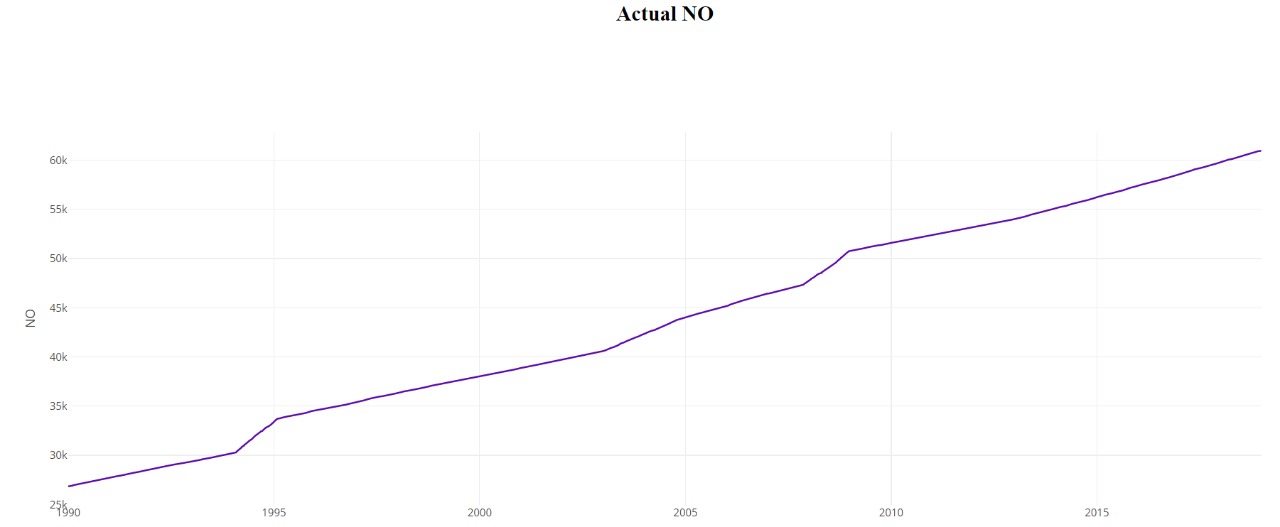


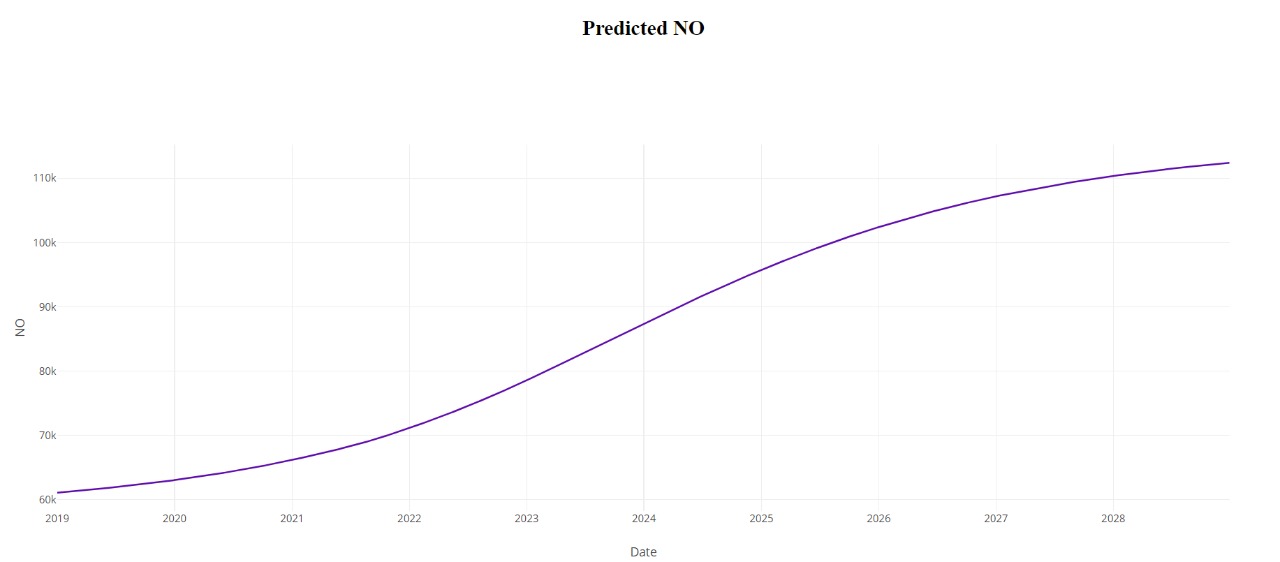












1. **Reuse and relationships to other products**

Since our system is in the early phase of its development therefore the reusable components haven’t been decided. There might be the reuse of LSTM model for predicting the temperature (max and min) and precipitation.

We might be able to create a relationship between the predicted temperature values and predicted Green House Gas emission since it contributes in the change in temperature.

1. **Design decisions and tradeoffs**

Since there are different models being tested to find the one providing with the best accuracy and predicted values, this will help in predicting the actual change in climate in the upcoming years.

There are certain factors that we could not include which could have provided us with an even better results but since there are certain limitations with the amount of data we require, and few factors upon which there is no data being recorded.

1. **Pseudocode for components**

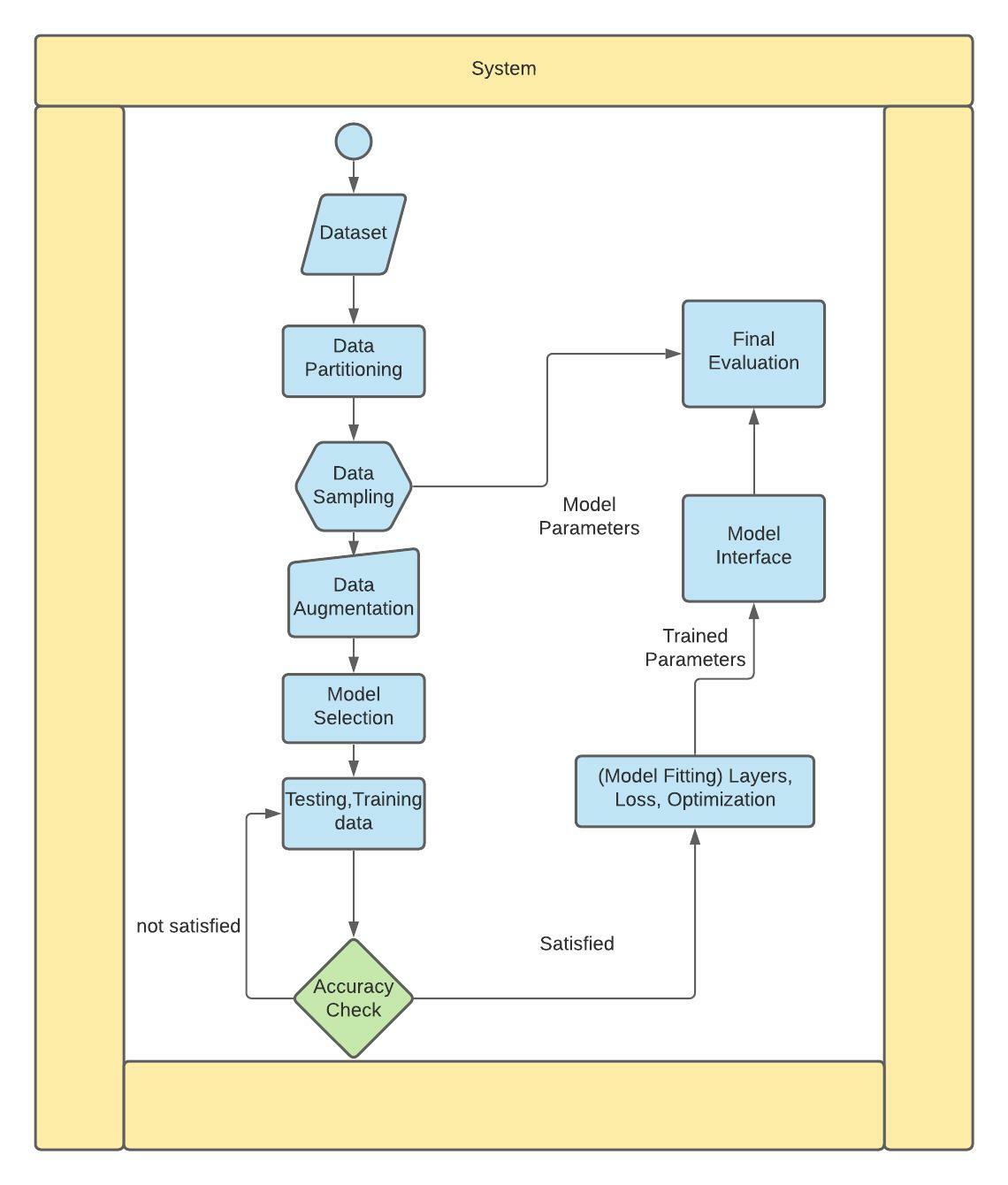
The main components applied and their pseudocode are as follows:

* The dataset component is the collection of the entire data that will be used for the training of the model so that we can predict the future values.
* The algorithm applied is the one with the most accuracy so that the predicted values are accurate and the system is legit with minimum drawbacks.
* The model used is a Convolutional LSTM model which is analyzing the training and testing data and provide with the predictive values.
* Created an Html page and used the python Dash library for the visualization of the data. Different components of Dash library are used such as tables and graphs.
* Different python libraries are used for the plotting of the graph such as seaborn, Plotlnine and matplotlib.

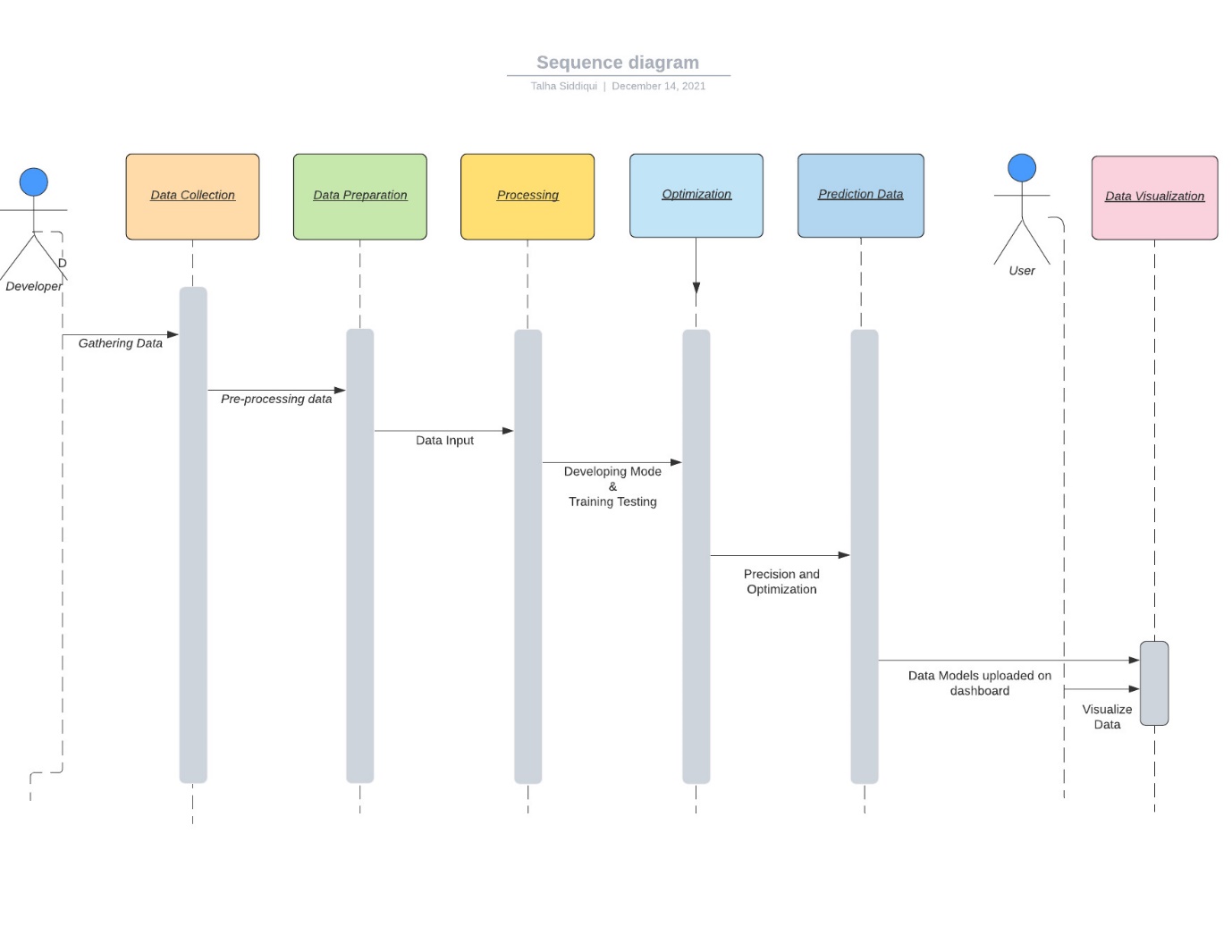
1. **Appendices**

|  |  |
| --- | --- |
| Activity Diagram | Describes activities and actions taking place in a system |
| Sequence Diagram | Shows one or several sequences of messages sent among a set of objects |
| Use-case Diagrams | Illustrates the relationships between use cases |
| Context Diagram | It shows the interactions between a system and other actors |
| Deployment Diagram | A special case of a Class Diagram used to describe hardware within the overall system architecture |
| System Block diagram | A diagram showing the major components of the system with its interconnections and external interfaces |
| Work Flow Diagram | It provides an overview of the process |
| System Architecture | It is the conceptual model that defines the structure, behavior, and more views of a system. |

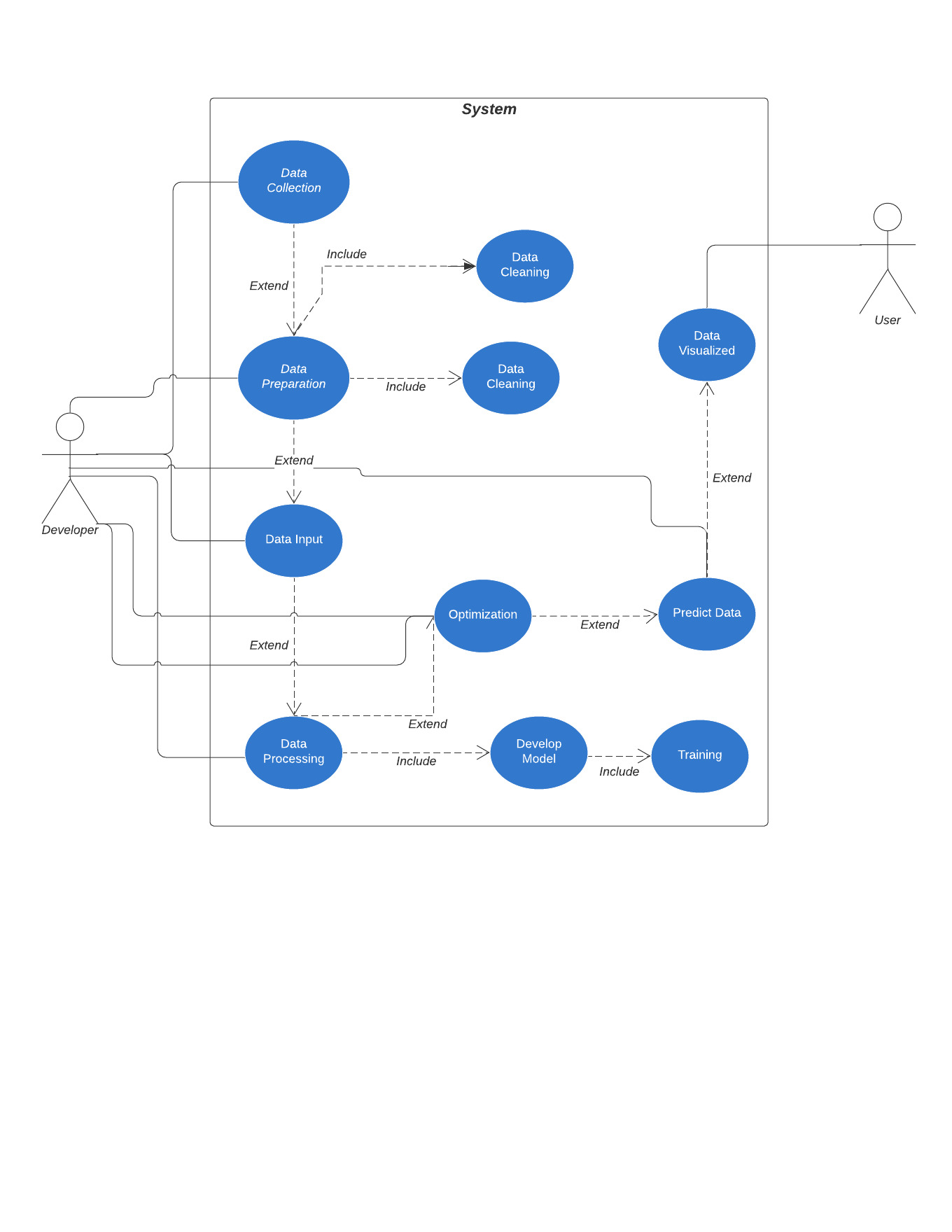
* 1. **Activity Diagram**

****

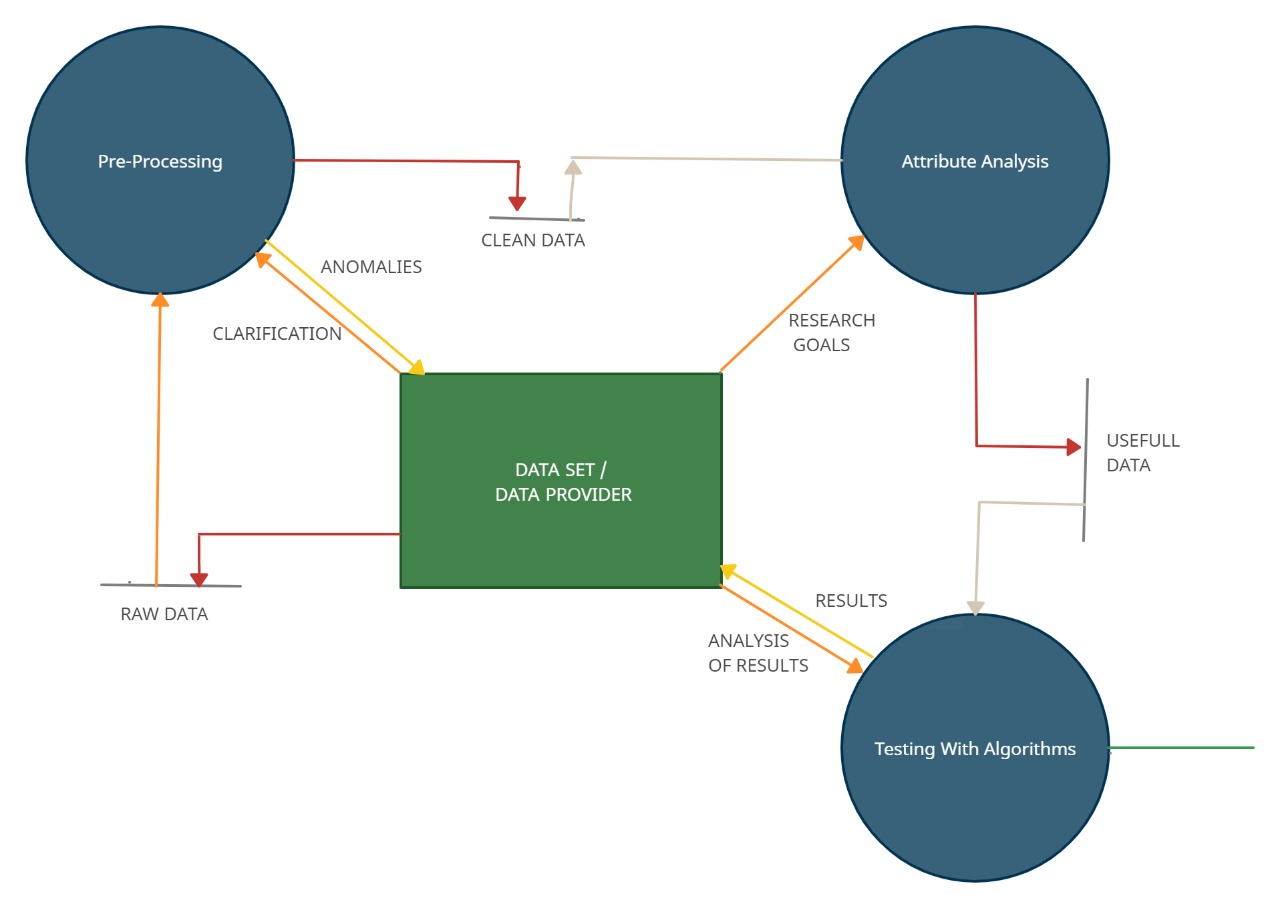
* 1. **Sequence Diagram**



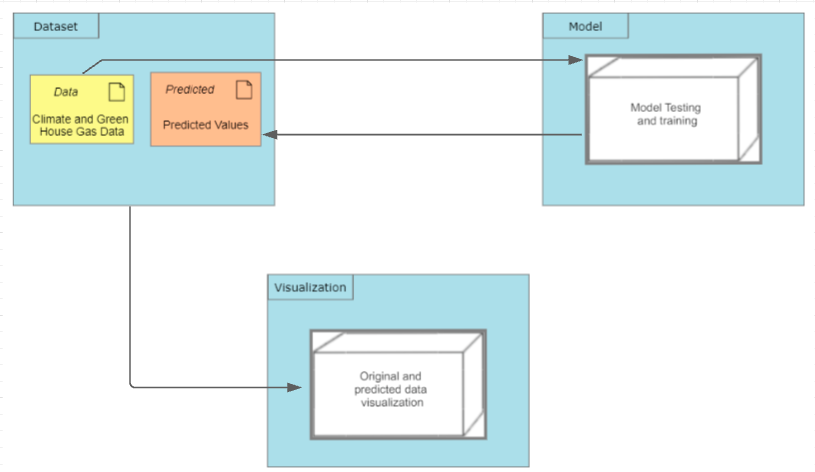
* 1. **Use Case Diagram**

****

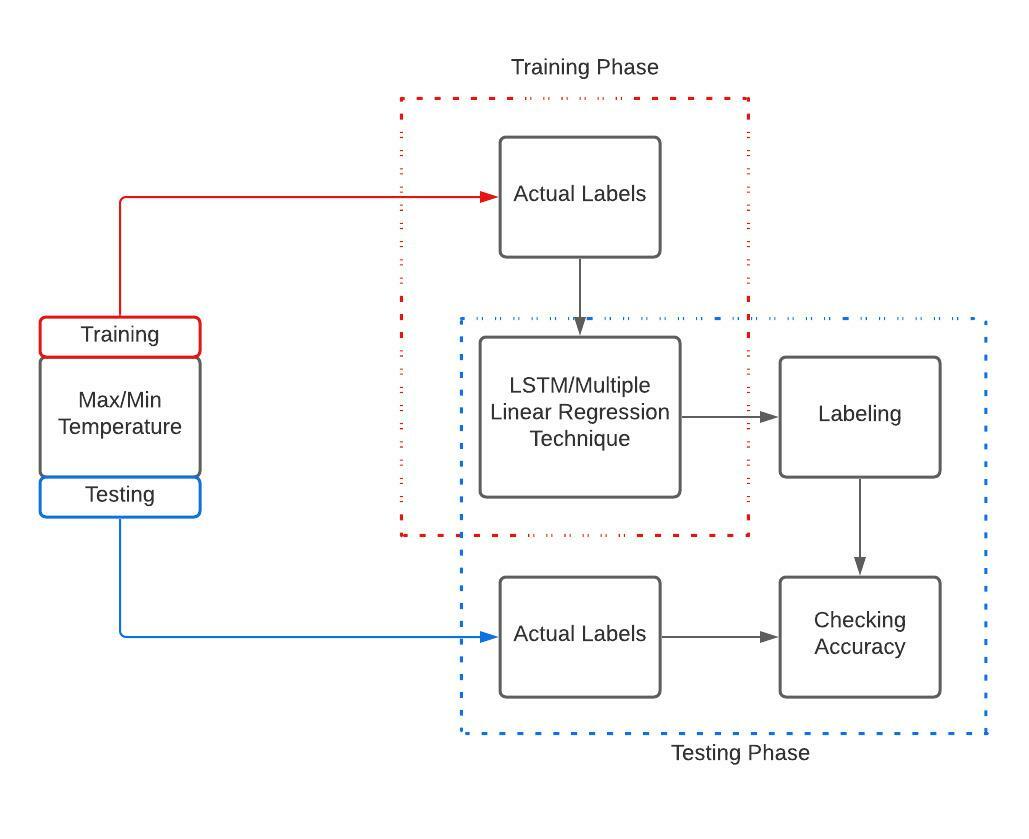
* 1. **Context Diagram**



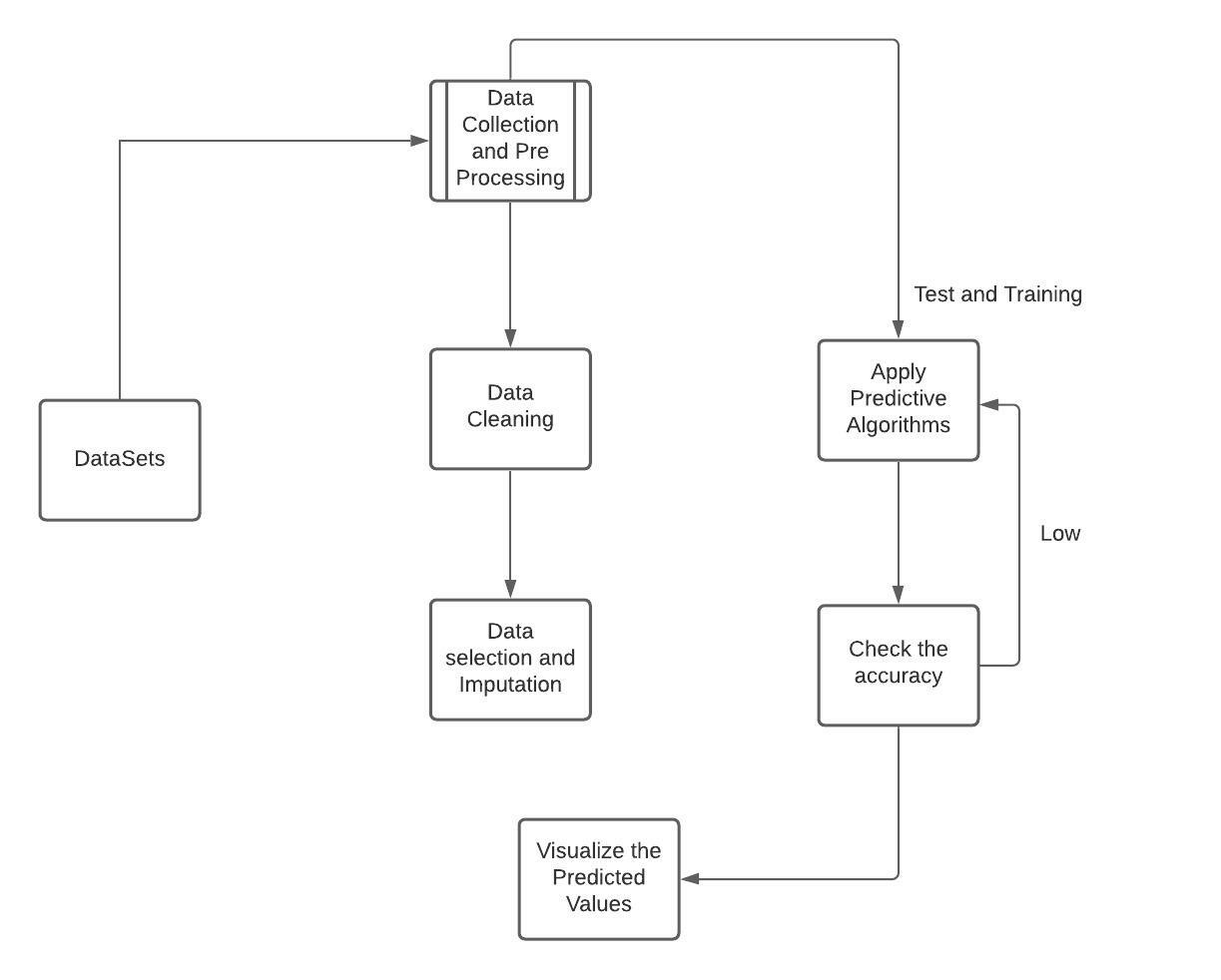
* 1. **Deployment Diagram**

****

* 1. **System Block Diagram**



* 1. **Work Flow Diagram**

****

* 1. **System Architecture Diagram**

