**Final Year Project Report**

**Predictive Analysis On Climate Change**

**Project Team:**

**Abdul Wasay 1812254**

**Muhammad Talha Siddiqui 1812267**

**Project Supervisor:**

**Maira Sami**

**Date: 20th June, 2022**

**Submitted in partial fulfillment of the requirements for the degree of**

**Bachelors of Science in Computer Science in the**

**Faculty of Computing and Engineering Sciences**

**Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology (SZABIST) Karachi Campus**

**Table of Contents**

[Project Proposal 1](#_Toc110733870)

[1. Introduction 2](#_Toc110733871)

[2. Objective 2](#_Toc110733872)

[3. Problem Description 2](#_Toc110733873)

[4. Methodology 3](#_Toc110733874)

[5. Project Scope 3](#_Toc110733875)

[6. Feasibility Study 4](#_Toc110733876)

[7. Tools/Technology 4](#_Toc110733878)

[8. Expertise of the Team Members 4](#_Toc110733879)

[9. Milestones 5](#_Toc110733880)

[10. Project Schedule 6](#_Toc110733881)

[Software Requirements Specification 8](#_Toc110733882)

[1. Introduction 9](#_Toc110733883)

[1.1 Purpose 9](#_Toc110733884)

[1.2 Document Conventions 9](#_Toc110733885)

[1.3 Intended Audience and Reading Suggestions 9](#_Toc110733886)

[1.4 Product Scope 9](#_Toc110733887)

[1.5 References 11](#_Toc110733888)

[2. Overall Description 12](#_Toc110733889)

[2.1 Product Perspective 11](#_Toc110733890)

[2.2 Product Functions 11](#_Toc110733891)

[2.3 Operating Environment 14](#_Toc110733893)

[2.4 Design and Implementation Constraints 15](#_Toc110733894)

[3. External Interface Requirements 16](#_Toc110733897)

[3.1 User Interfaces 16](#_Toc110733898)

[3.2 Hardware Interfaces 16](#_Toc110733899)

[3.3 Software Interfaces 17](#_Toc110733900)

[4. System Features 18](#_Toc110733902)

[4.1 System Features 1 18](#_Toc110733903)

[4.2 System Features 2 18](#_Toc110733904)

[5. Other Non-Functional Requirements 19](#_Toc110733905)

[5.1 Performance Requirements 20](#_Toc110733906)

[5.2 Safety Requirements 20](#_Toc110733907)

[5.3 Security Requirements 21](#_Toc110733908)

[5.4 Business Rules 21](#_Toc110733910)

[6. UML DIagrams 22](#_Toc110733911)

[6.1 Use Case Diagram 22](#_Toc110733912)

[6.2 Sequence Diagram 23](#_Toc110733913)

[6.2 Activity Diagram 24](#_Toc110733913)

[6.2 Context Diagram 25](#_Toc110733913)

[7. Datasets 26](#_Toc110733911)

[Appendix A: Glossary 26](#_Toc110733914)

[Appendix B: Analysis Models 27](#_Toc110733915)

[ LSTM Model 27](#_Toc110733916)

[ Linear Regression Model 28](#_Toc110733917)

[ Test Cases 29](#_Toc110733918)

[Software Design Specification 32](#_Toc110733920)

[1. Introduction 32](#_Toc110733921)

[1.1 Purpose of this document 32](#_Toc110733922)

[1.2 Scope of the development project 33](#_Toc110733923)

[1.3 Definitions, acronyms, and abbreviations 34](#_Toc110733924)

[1.4 References 35](#_Toc110733925)

[1.5 Overview of document 35](#_Toc110733926)

[2. System architecture description 36](#_Toc110733927)

[2.1 Section Overview 36](#_Toc110733928)

[2.2 General Constraints 37](#_Toc110733929)

[2.3 Program Structure 38](#_Toc110733931)

[2.4 Alternatives Considered 39](#_Toc110733932)

[3. Detailed description of components 40](#_Toc110733933)

[3.1 Section Overview 41](#_Toc110733934)

[3.2 Component and Detail 42](#_Toc110733934)

[3.3 SDS Components 46](#_Toc110733934)

[4. User Interface Design 48](#_Toc110733935)

[4.1 Section Overview 52](#_Toc110733936)

[4.2 Detailed description 53](#_Toc110733939)

[5. Reuse and relationships to other products 57](#_Toc110733940)

[6. Pseudocode for components 57](#_Toc110733941)

[7. Appendices 58](#_Toc110733942)

7.1 [Activity Diagram 59](#_Toc110733957)

7.2 [Sequence Diagram 60](#_Toc110733957)

7.3 [Use Case Diagram 61](#_Toc110733957)

7.4 [Context Diagram 62](#_Toc110733957)

7.5 [Deployment Diagram 63](#_Toc110733957)

7.6 [System Block Diagram 64](#_Toc110733957)

7.7 [Work Flow Diagram 65](#_Toc110733957)

7.8 [System Architeture Diagram 66](#_Toc110733957)

[PROJECT MEETING LOG SHEET 67](#_Toc110733958)

[ITERATION PLAN 68](#_Toc110733959)

[GANTT CHART 69](#_Toc110733960)

[Plagiarism Free Certificate 70](#_Toc110733965)

# Project Proposal

**Predictive Analysis on Climate Change**

**Version 1.0 approved**

**Supervisor:**

**Ms. Maira Sami**

**Prepared by:**

**Muhammad Talha Siddiqui 1812267**

**Abdul Wasay 1812254**

**SZABIST**

**20th June, 2022**

# Introduction

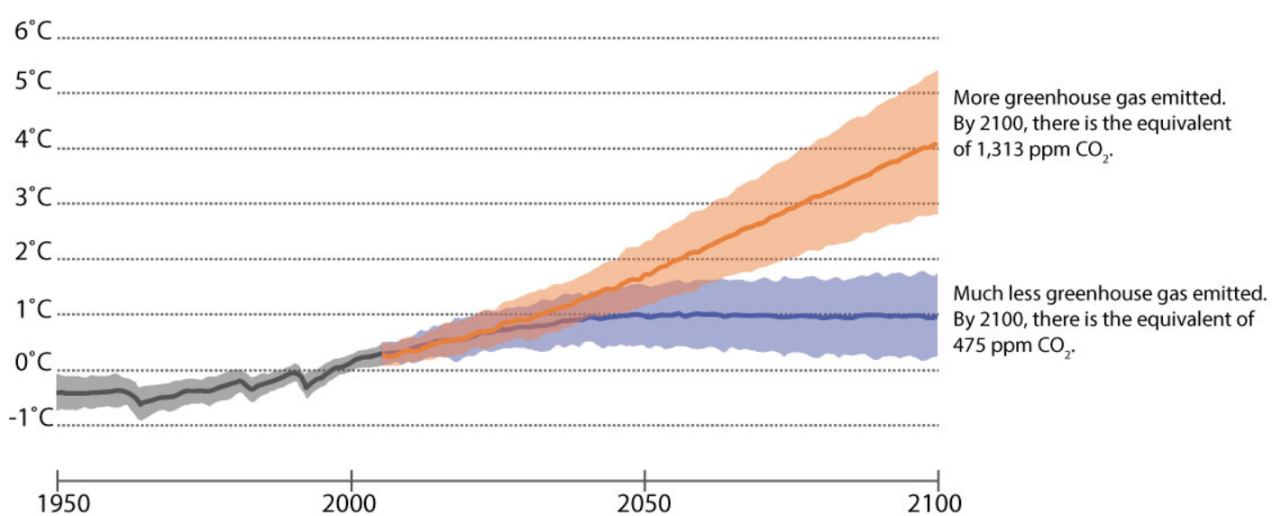
Climate change refers to changes in the Earth’s climates, at local, regional, or global scales, and can also refer to the effects of these change. As it can be seen in the past 100 years, the world's surface air temperature increases gradually due to burning fossil fuels that releases carbon dioxide and other greenhouse gases into the atmosphere. PACC models will predict the change in Earth’s climate and its effect on the environment.

# Objective

The objective of this project is to asses global climate change factors and further predict how these factors will change within the future effecting the climate. during this project different models are getting used to predict the climate changing factors. Furthermore, using these predictions to accurately predict global temperature change and its effects on the environment and also the world.

# Problem Description

Climate Change is that the defining issue of our time and that we are at a defining moment. From shifting weather patterns that threaten food production, to rising sea levels that increase the danger of catastrophic flooding, the impacts of global climate change are global in scope and unprecedented in scale. Without drastic action today, adapting to those impacts within the future are tougher and expensive.

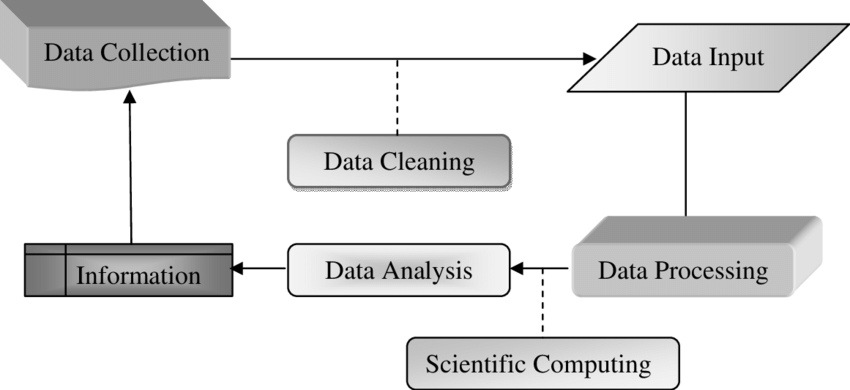


Climate change is causing many other aspects of Earth to vary, including the examples noted during this graphic below: Using models, we will project how these aspects of Earth are likely to vary within the future because the climate continues to vary.

# Methodology

Our system is going to be processing the variability of information observed in playing its role in global climate change. we'll process different datasets, we'll further train models for every factor effecting global climate change. Our systems would be tested for efficiency & accuracy in-order to search out prediction pattern for the near future. Once we've the predictive analysis of these factors, the expected data are accustomed integrate all the factors into one main predictive analysis model which will provide the general analysis in change of climate over the years and therefore the prediction for the long run.

First step of the project would be data collection, after collection data must be processed because of which data goes through several steps usually referred to as processing cycle. Below are the steps included in processing cycle.



# Project Scope

During the subsequent decades and beyond, heating is expected to cause further increases in atmospheric moisture content, more extreme heatwaves, fewer frosts, further decreases within the extent and thickness of sea ice, further melting of mountain glaciers and ice sheets, shifts in rainfall, ocean warming and rises in sea levels. The magnitude of expected change depends on future greenhouse emission and climate feedbacks.  
  
As we all know Climate change not only leads to extreme weather events and dangerous health effects the world over, but, as a threat multiplier, it poses an immediate 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 give the overall analysis in change of climate over the years and the prediction for the future. We believe that mitigating global climate change is not only important for global peace and security, but key to fulfilling our moral imperative to guard those communities most at risk. Our work is time sensitive and urgent.  
  
This system would be used to predict the climate change over the years and its effects on the environment and the world. These predictions would be to take extra precautions in-order to mitigate climate change and further preserve the environment.

1. **Feasibility Study**

Climate predictions suffer from errors because of unavoidable uncertainties, which prevent forecast systems from taking full advantage of the massive range of predictability sources. There are three main sources of uncertainty in climate prediction.  
• The first source arises from natural internal variability, intrinsic to the climate system. Internal variability can be initialized in an exceedingly predictions, but the uncertainty within the initial conditions because of our inability to perfectly know the state of the climate system is nonlinearly amplified.  
• The second source is that the uncertainty within the past, present and future changes within the forcing of the climate system (anthropogenic emissions, land use and natural forcing’s like volcanic eruptions and solar activity) arising from a scarcity of observations and therefore the limitations to grasp their future evolution.

# Tools/Technology

Python, Spyder, Anaconda, TensorFlow, Pandas, NumPy, Matplotlib, Keras, Pytorch, Sckeit-learn and as for our requirement.

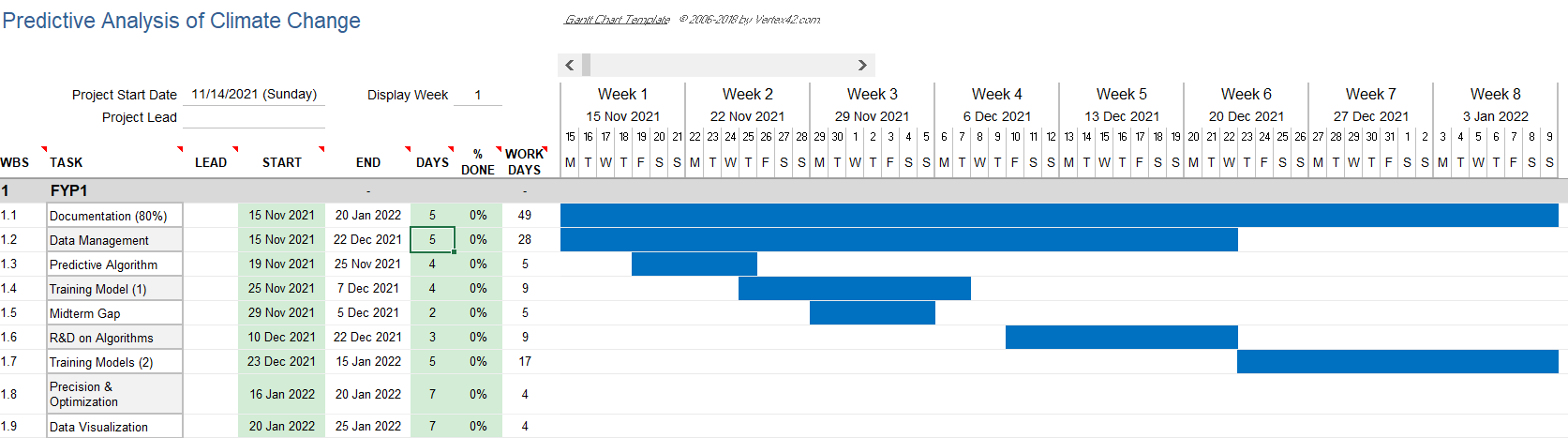
# Expertise of the Team Members

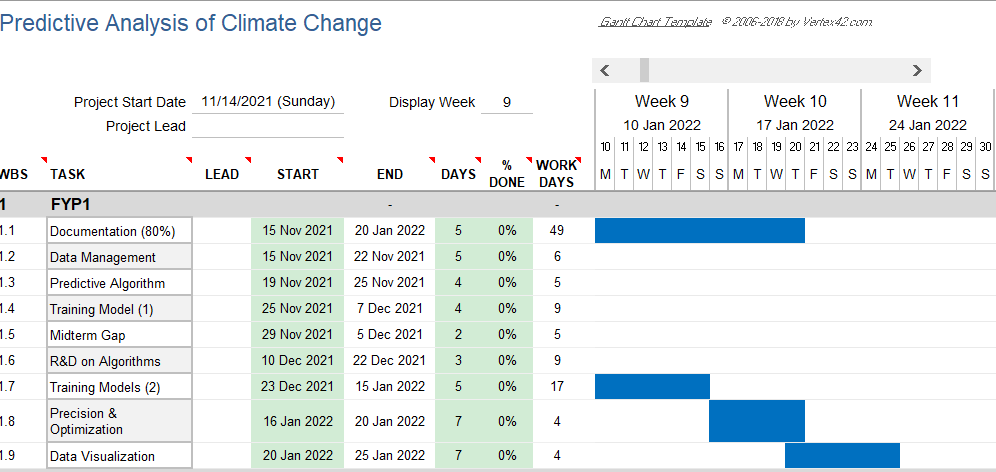
All the process required and implementation of the tasks will be done by both the team members Abdul Wasay and Talha Siddiqui.

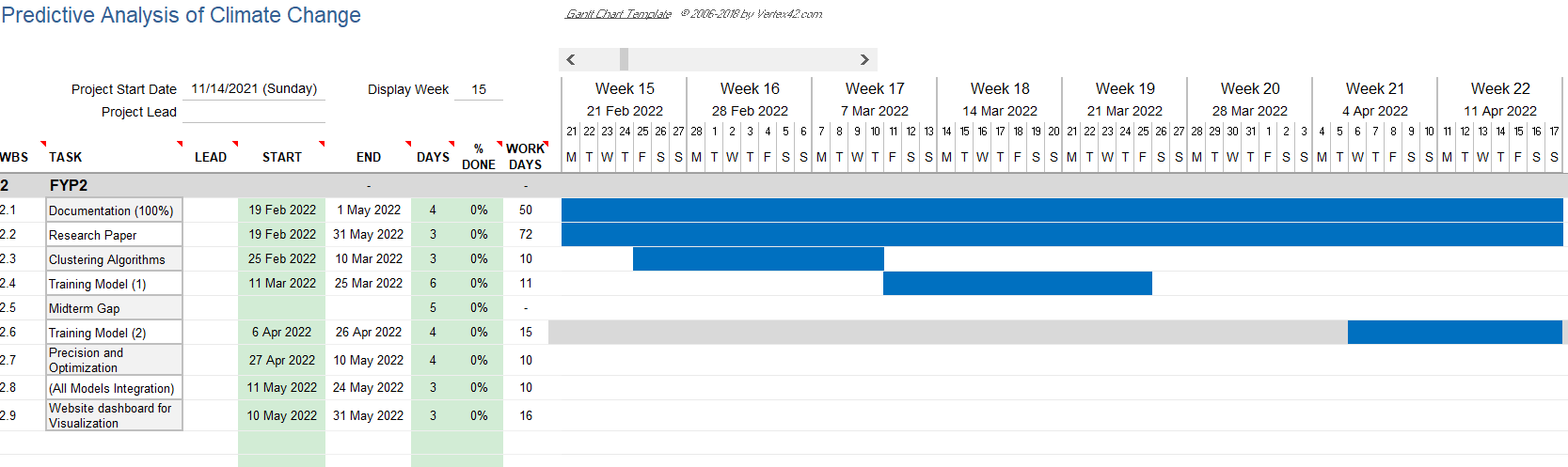
# Milestones

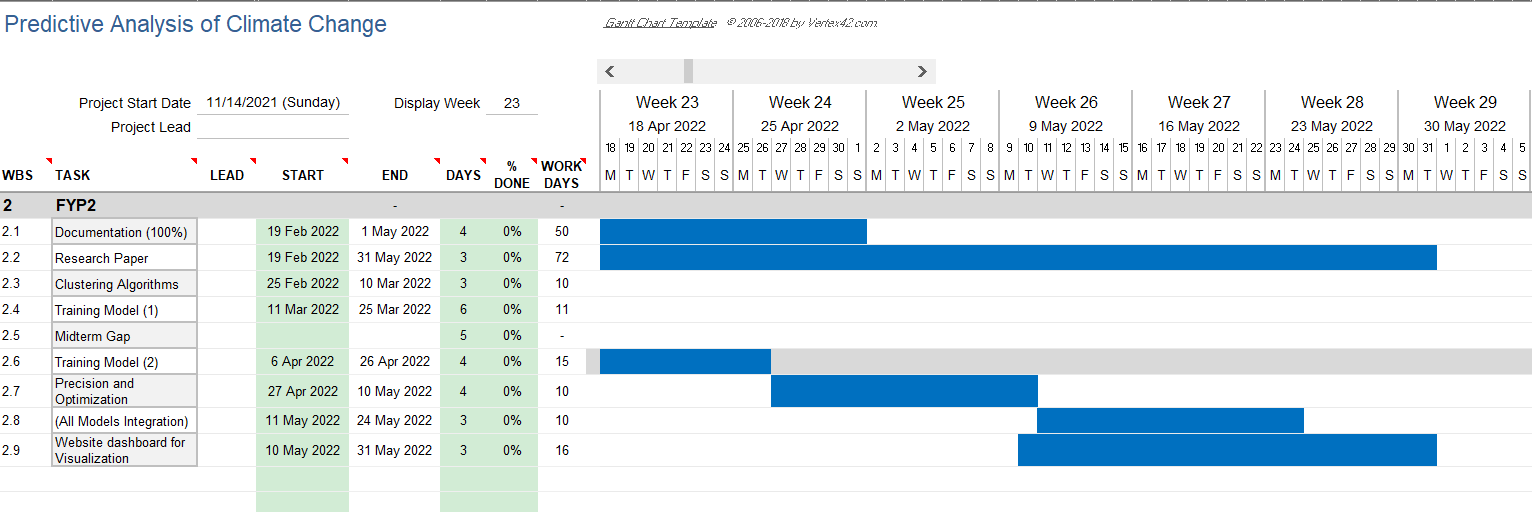
|  |  |  |
| --- | --- | --- |
|  | **Milestones** |  |
|  | **FYP1** |  |
|  | Documentation (80%) | 15-11-21 to 20-01-22 |
|  | Data Management | 15-11-21 to 22-11-21 |
|  | Predictive Algorithm | 19-11-21 to 25-11-21 |
|  | Training Model (1) | 25-11-21 to 07-12-21 |
|  | Midterm Gap | 29-11-21 to 05-12-21 |
|  | R&D on Algorithms | 10-12-21 to 22-12-21 |
|  | Training Models (2) | 23-12-21 to 15-01-22 |
|  | Precision & Optimization | 16-01-22 to 20-01-22 |
|  | Data Visualization | 20-01-22 to 25-01-22 |
|  | Finals | 31-01-22 to 18-02-22 |
|  | **FYP2** |  |
|  | Documentation (100%) | 19-02-22 to 01-05-22 |
|  | Research Paper | 19-02-22 to 31-05-22 |
|  | Clustering Algorithms | 25-02-22 to 10-03-22 |
|  | Training Model (1) | 11-03-22 to 25-03-22 |
|  | Midterm Gap | No date |
|  | Training Model (2) | 06-04-22 to 26-04-22 |
|  | Precision and Optimization | 27-04-22 to 10-05-22 |
|  | (All Models Integration) | 11-05-22 to 24-05-22 |
|  | Website dashboard for Visualization | 10-05-22 to 31-05-22 |
|  | Finals | No Date |

# Project Schedule



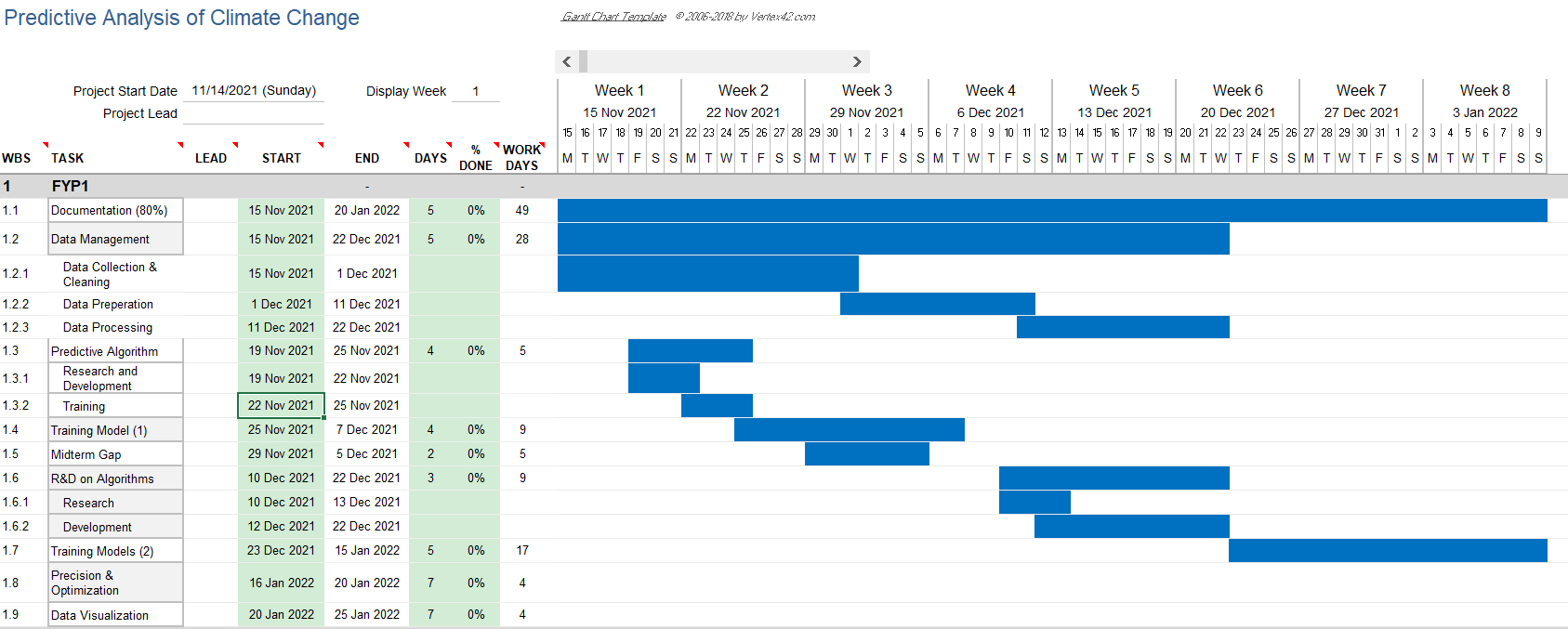




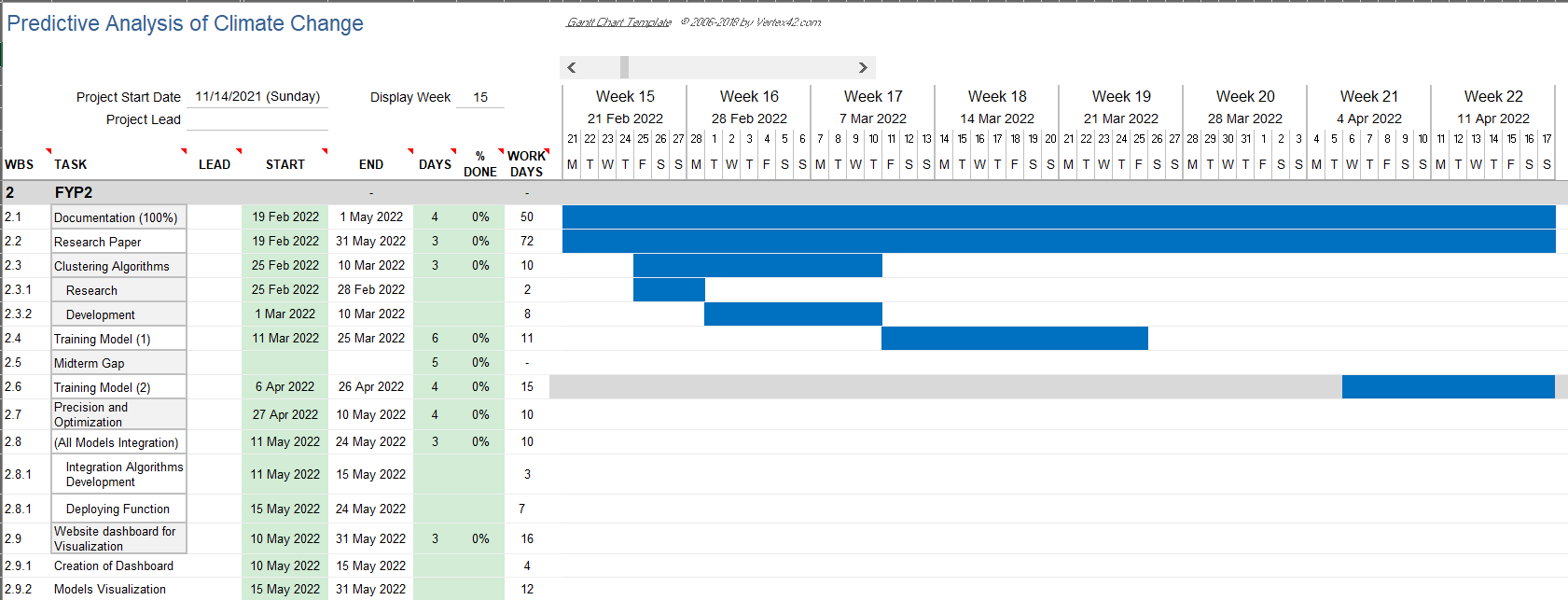


# Work Breakdown Structure

# Final Year Project 1



**Final Year Project 2**



**Software Requirements Specification**

**for**

***Predictive Analysis on Climate Change***

Version 1.0 approved

Prepared by Abdul Wasay (1812254)

Muhammad Talha Siddiqui (1812267)

SZABIST

June 20th, 2022

## 

## 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: Times New Roman, 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/>

**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**

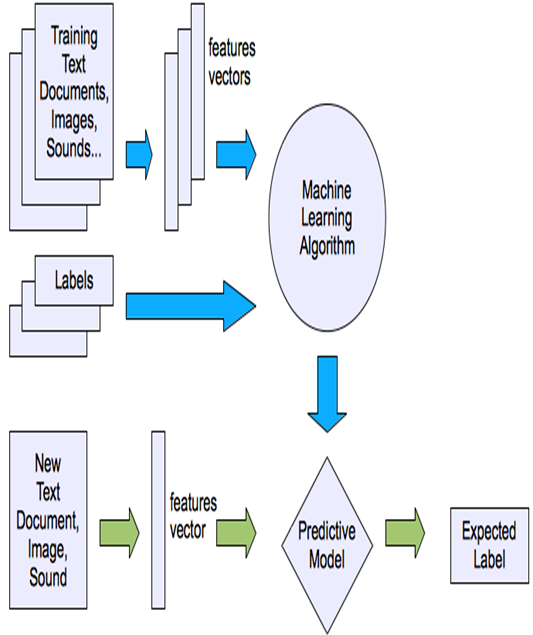
**3.1 Hardware Interfaces**

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.

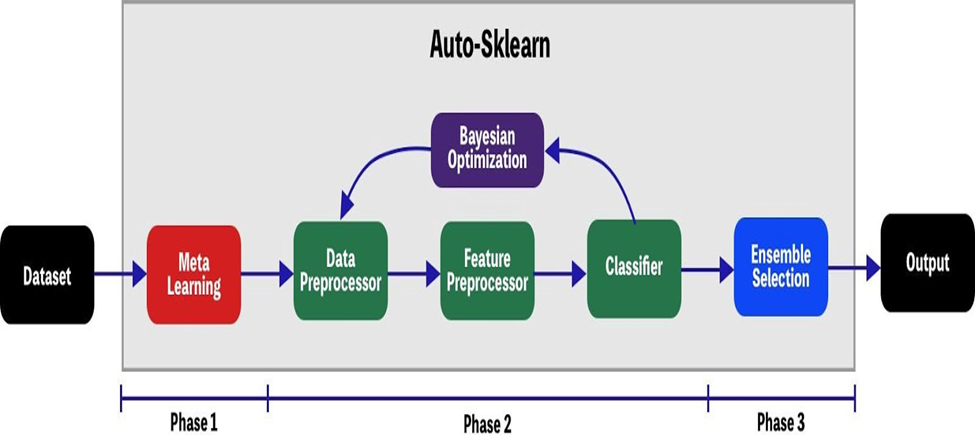
**3.3 Software Interfaces**

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.

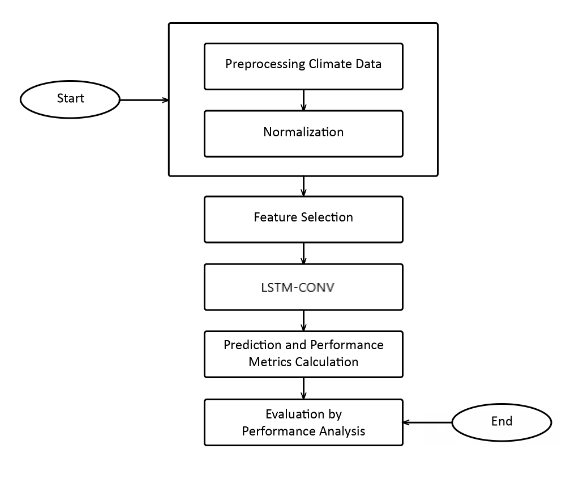
**3.3.1 General Model**



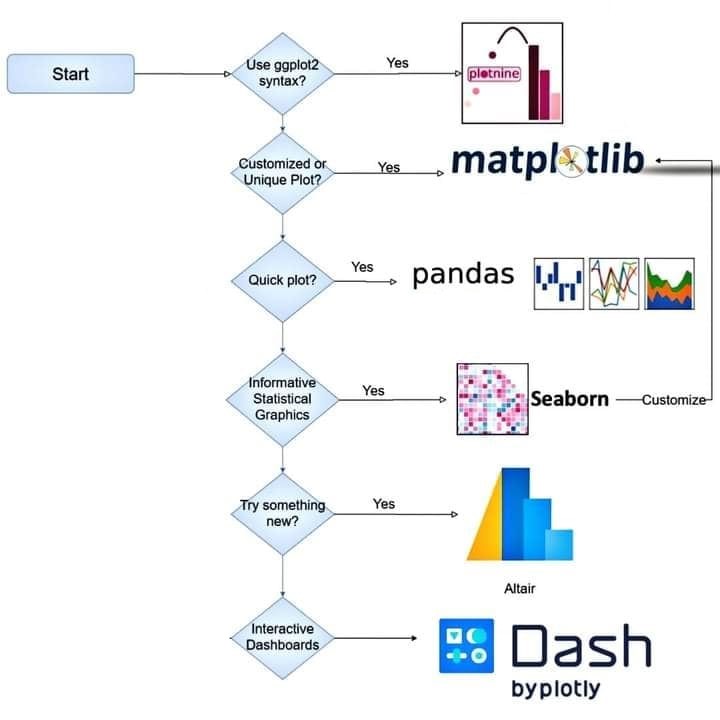
**3.3.2 Scikit-Learn Model**



**3.3.3 LSTM Model**

****

**3.3.4 Tools for Data Visualization**



**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.

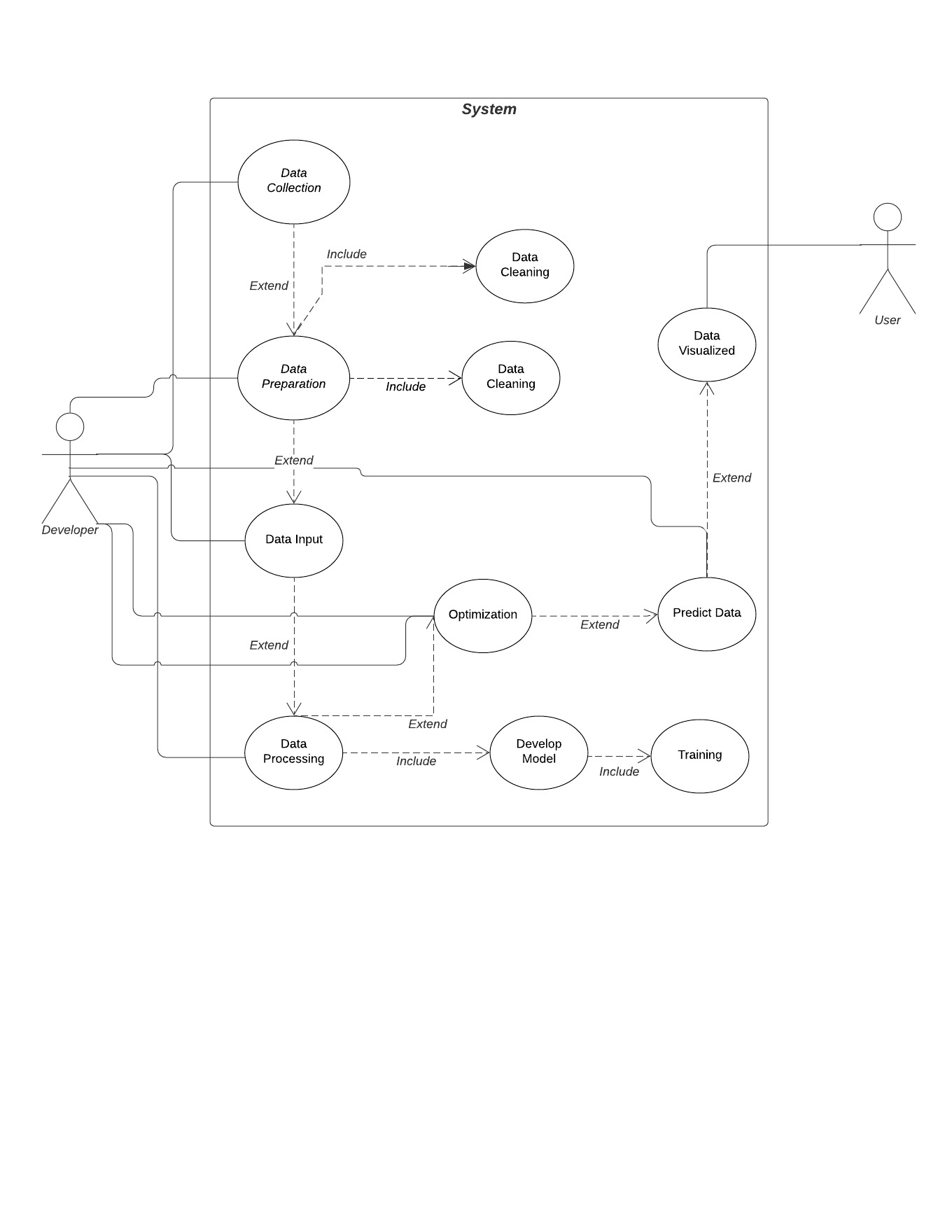
Data being used can be a target of data poisoning which will interrupt the model validation, model accuracy and hence failure in presenting the accurate predicted values.

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

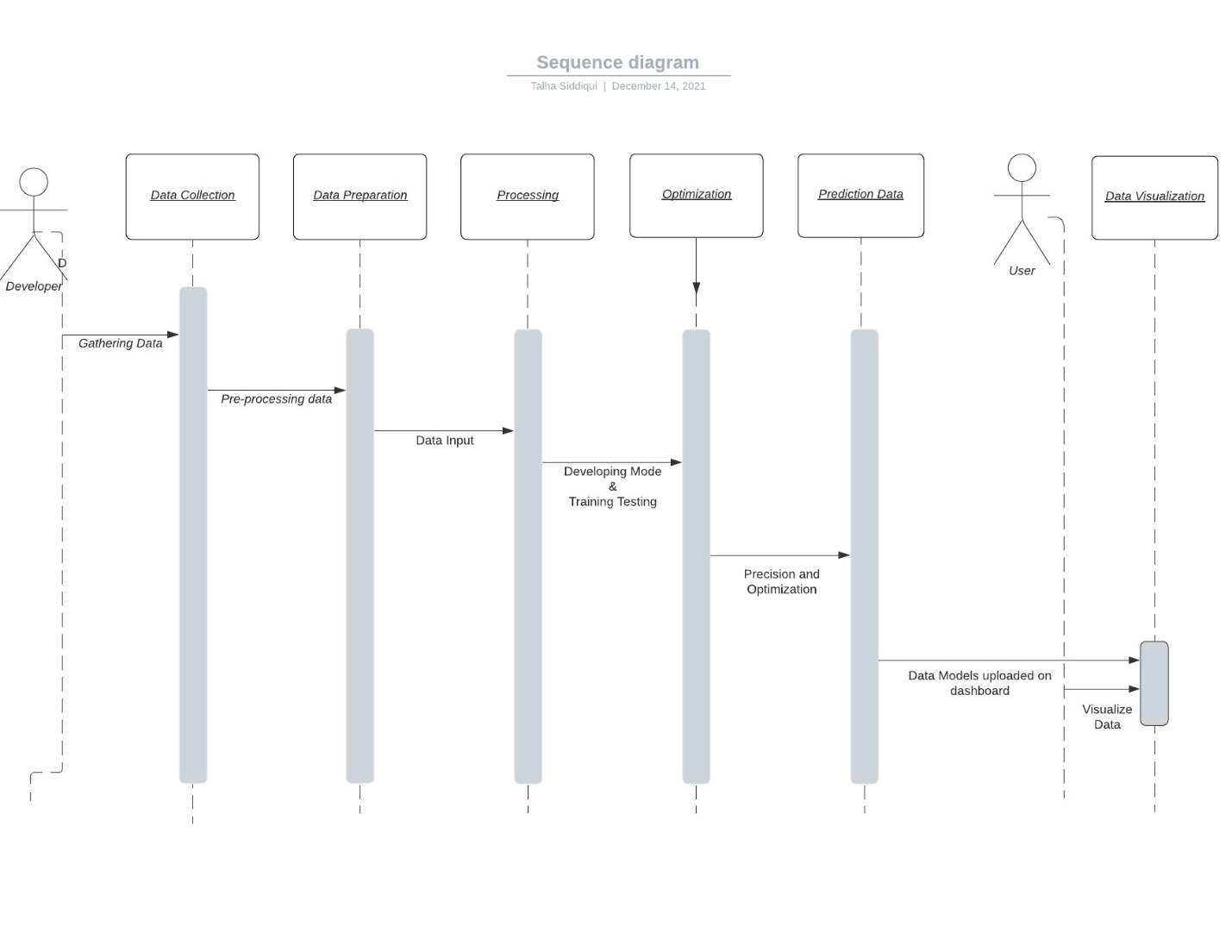
Systems being used for training and testing of the model should be kept secure from any sort of access from unauthorized user to prevent any sort of mishap and deleting of the complete progress.

**6. UML Diagrams**

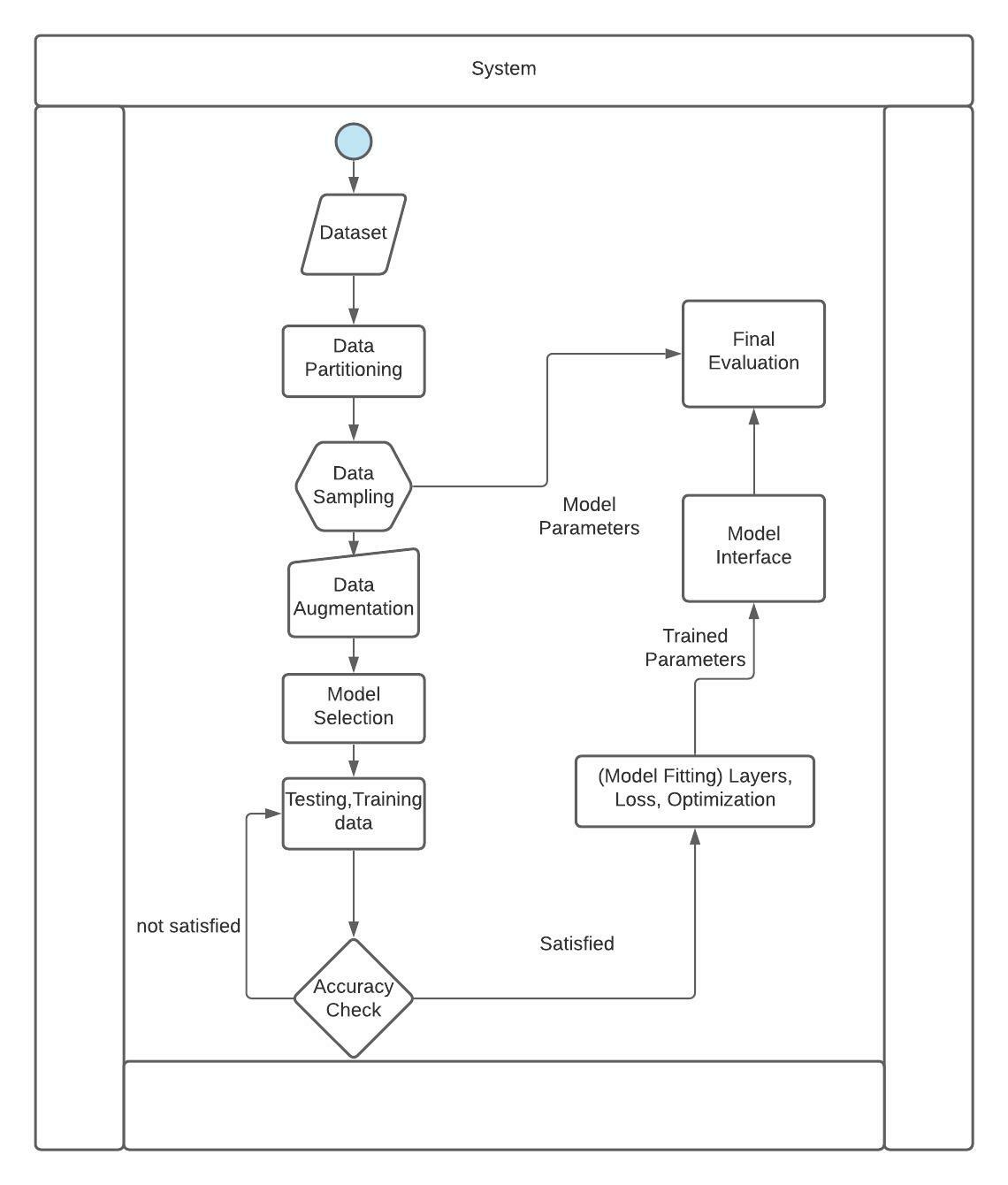
**6.1 Use case Diagram**

****

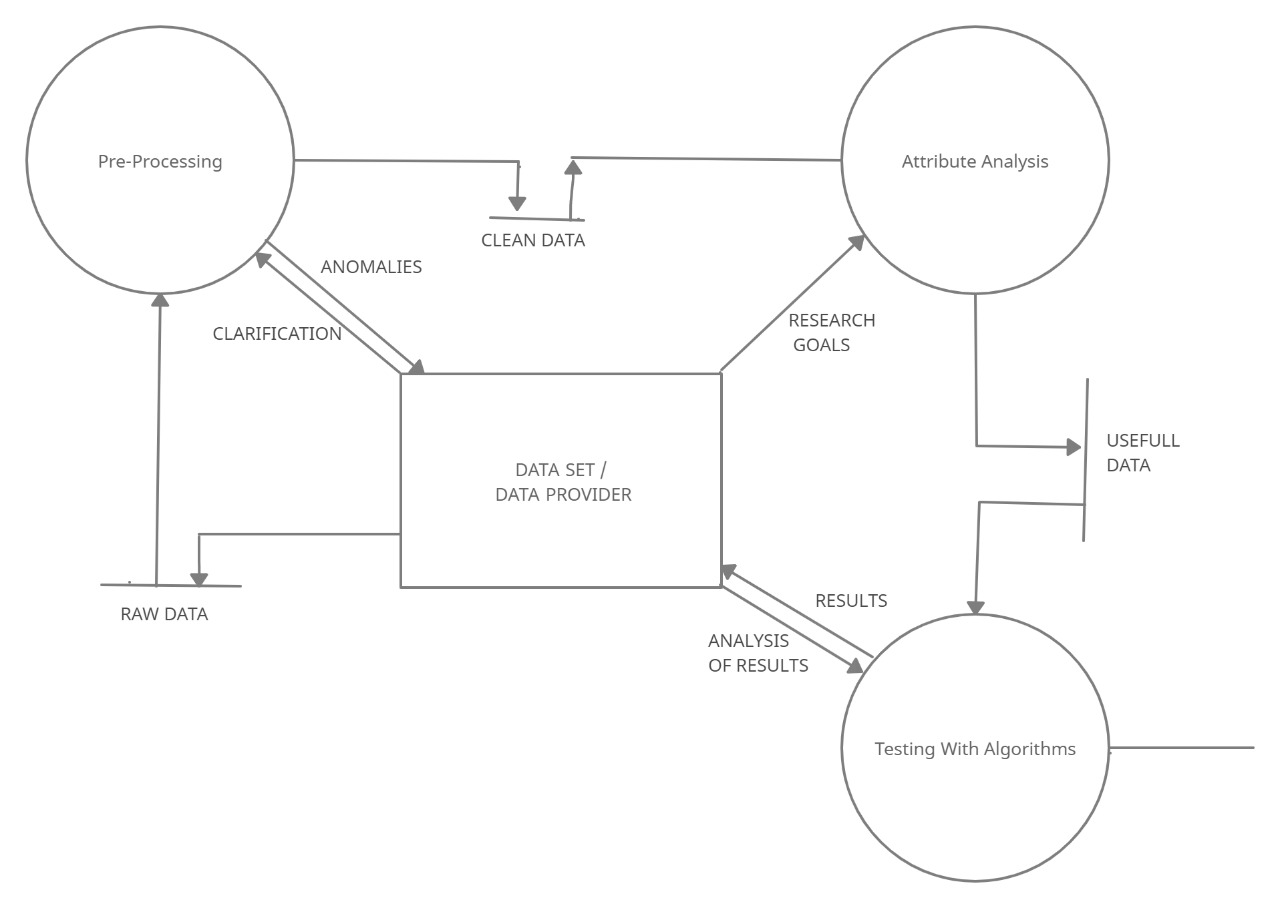
**6.2 Sequence Diagram**

****

**6.3 Activity Diagram**

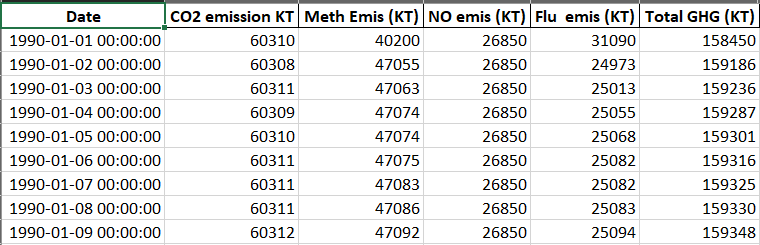
****

**6.4 Context Diagram**

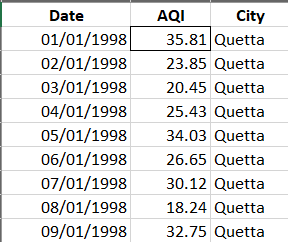
****

**7. Datasets**

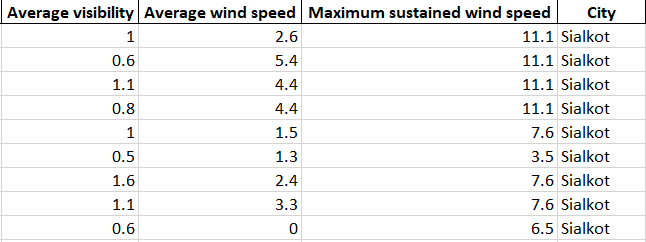
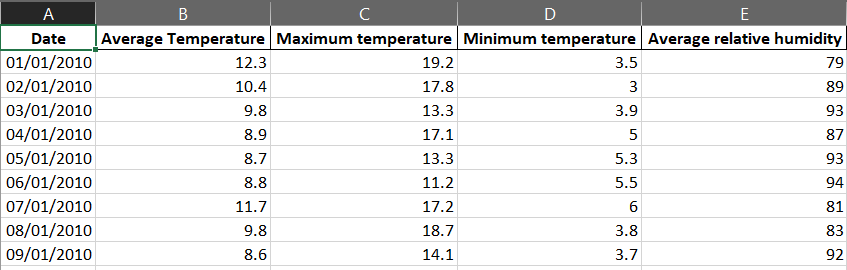
**7.1 Green House Gases**

****

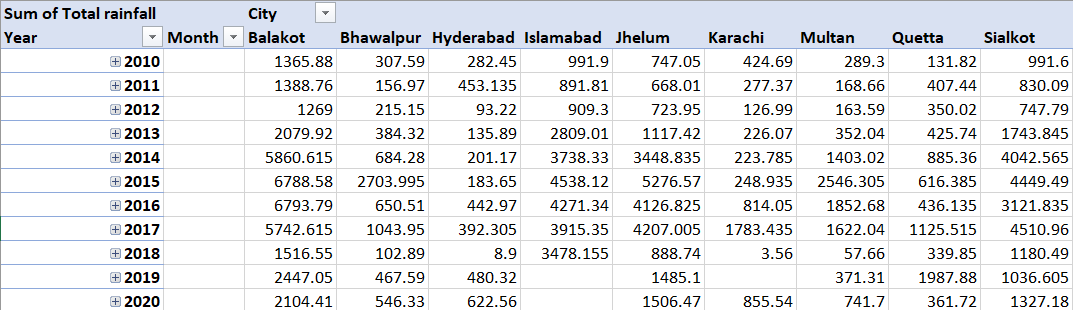
**7.2 Air Quality**

****

**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. |

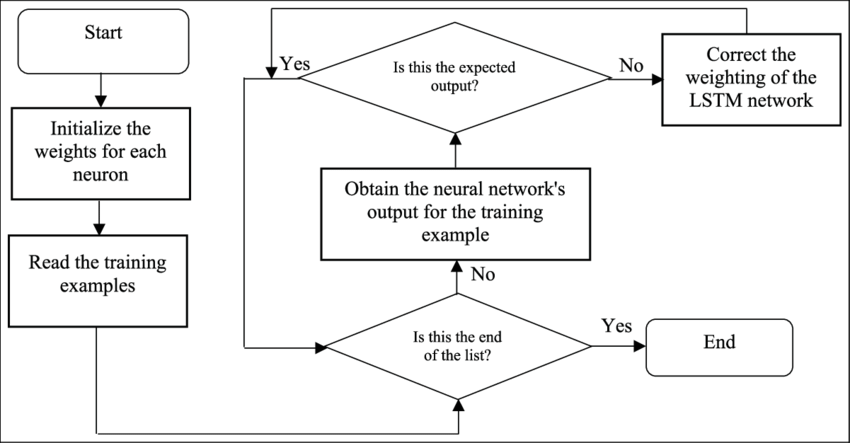
**9. Other Requirements**

**Appendix A: Glossary**

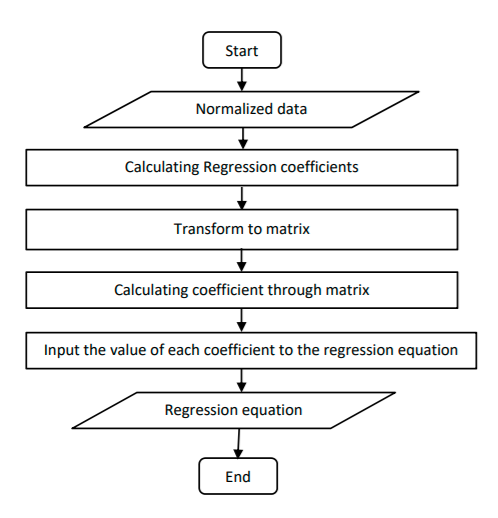
|  |  |
| --- | --- |
| Term | Definitions |
| LSTM | Long short-term memory (LSTM) is a recurrent neural network. |
| CNN | **CNN stands for Convolutional Neural Network which** is basically a Deep Learning algorithm. |
| PMD ARIMA | PMD ARIMA is a statistical method. |
| MAE | The Mean Absolute Error of a model is the mean of the absolute values of the prediction errors on over all instances in the test set. |

**10. Appendix B: Analysis Models**

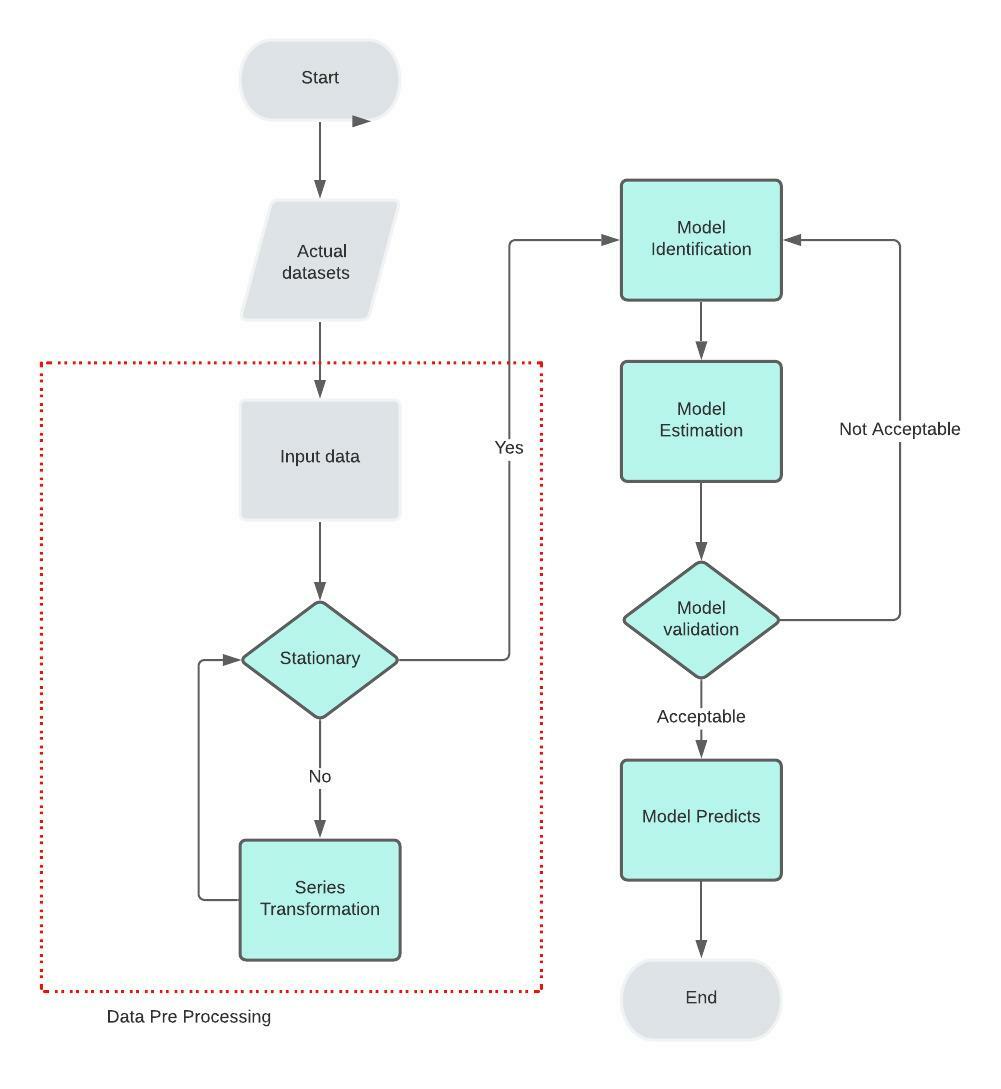
**10.1 LSTM Model**

****

* 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 |

**Software Design Specification**

**for**

***Predictive Analysis on Climate Change***

Version 1.0 approved

Prepared by Abdul Wasay (1812254)

Muhammad Talha Siddiqui (1812267)

SZABIST

June 20, 2022

## 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.

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 propose a system which provides the overall analysis of climate change. 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). |
| **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 of system architecture

* 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)}

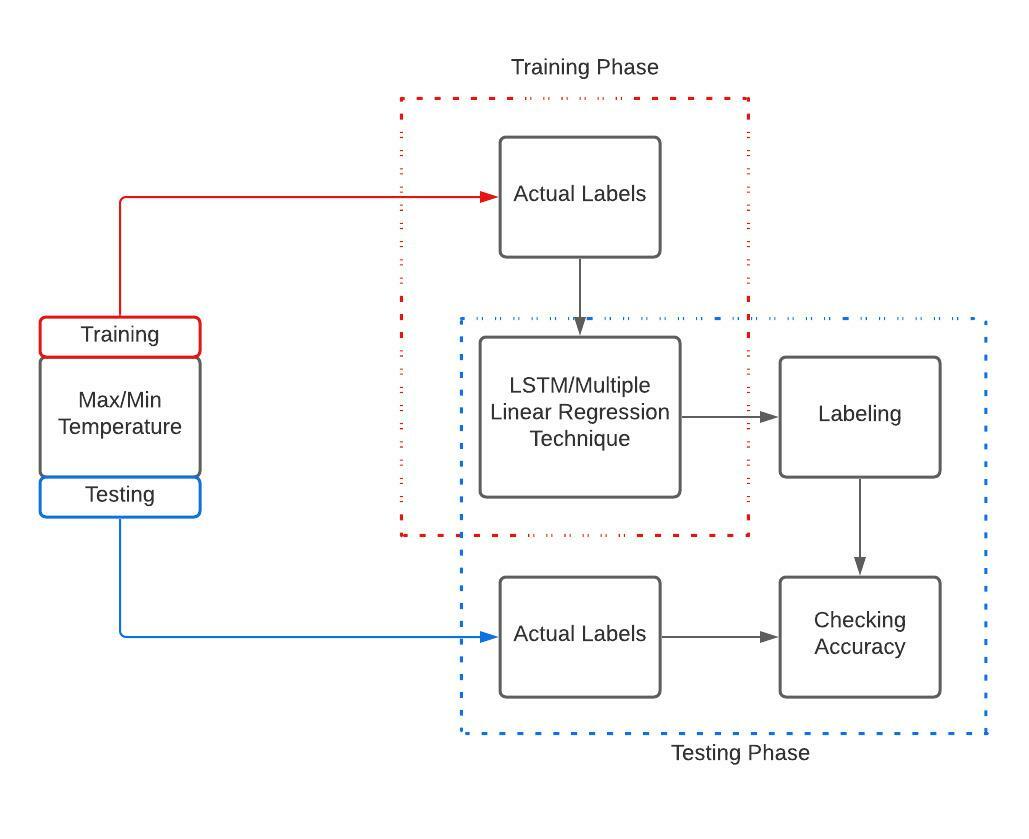
Yt2=1/4{Yt-1,5/12(Yt-Yt-1)}

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. **Program Structure**

****

* 1. **Alternatives 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.

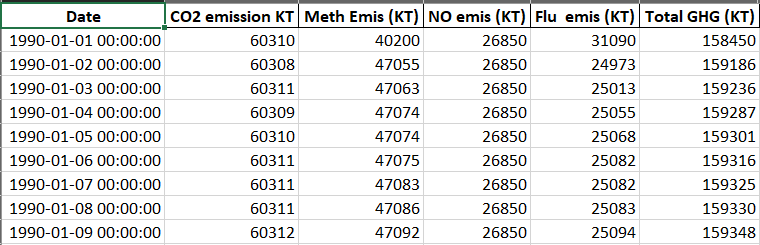
Our third dataset is in its initial basis which is a sea level pressure dataset. This data is recorded from Karachi port and is dated from 1916 to 2016.

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

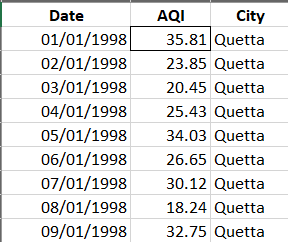
* + 1. Data Members

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |

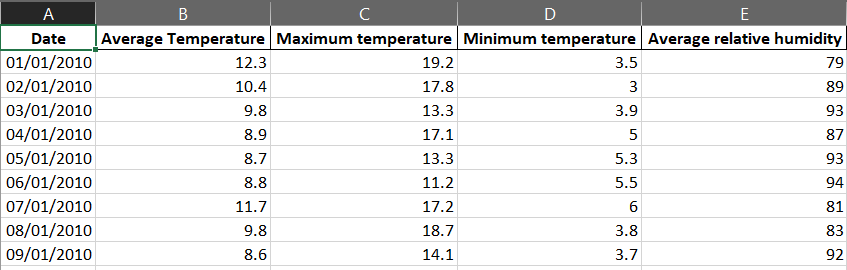
1. **Green House Gases Dataset**

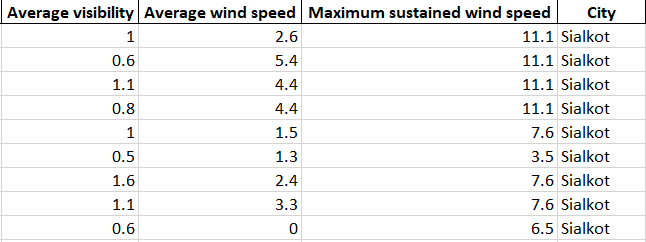
****

1. **Air Quality Dataset**

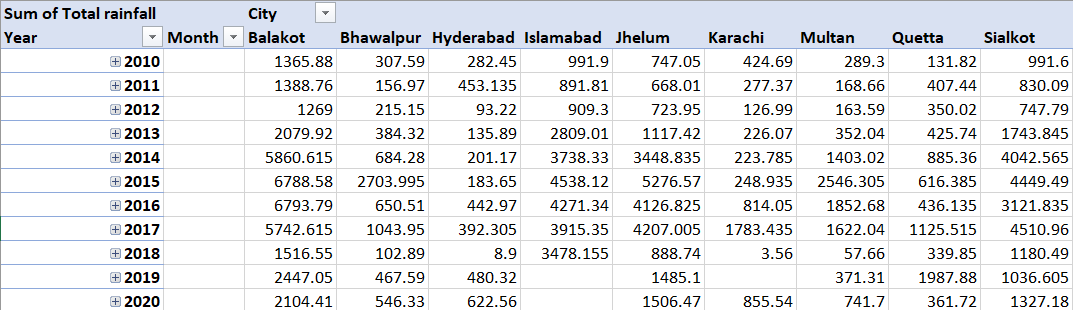
****

1. **Climate Dataset**

****

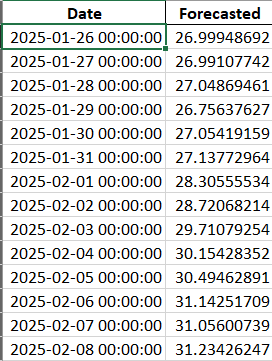
****

1. **Rainfall**

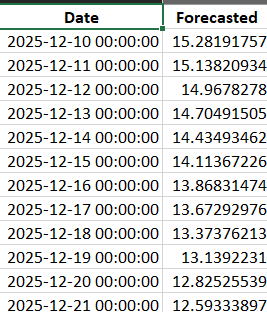


**Predicted Data:**

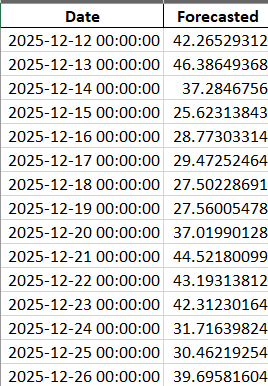
* **Karachi Max Temperature**



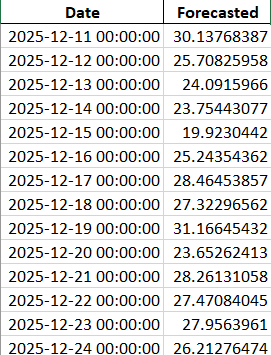
* **Karachi Min Temperature**



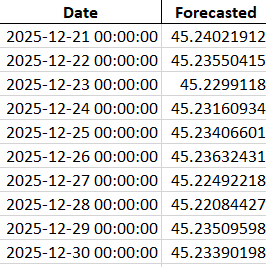
* **Karachi Humidity**



* **Karachi Wind Speed**



* **Karachi AQI**



* 1. SDS Components
     1. Component 1

|  |  |
| --- | --- |
| Identification | **Data Set** |
| Type | Xlsx, csv |
| Purpose | Dataset used for training, validating, and testing the ML model. Greater the set of data, model will learn better. |
| Function | The main reason of using dataset is to predict the future values. |
| 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

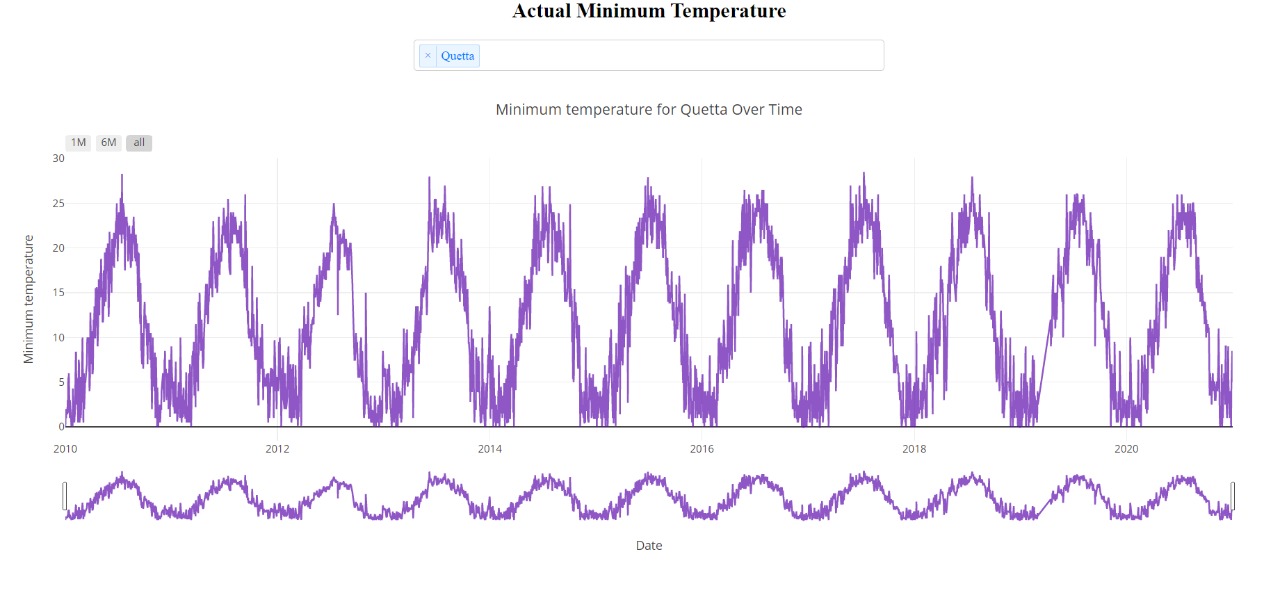
|  |  |
| --- | --- |
| Identification | Predictive Models |
| Type | Convolutional LSTM |
| Purpose | A model is a computational representation of real-world processes. |
| Function | An ML model is trained to recognize certain types of patterns by training it over a set of data using relevant algorithms. 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. |

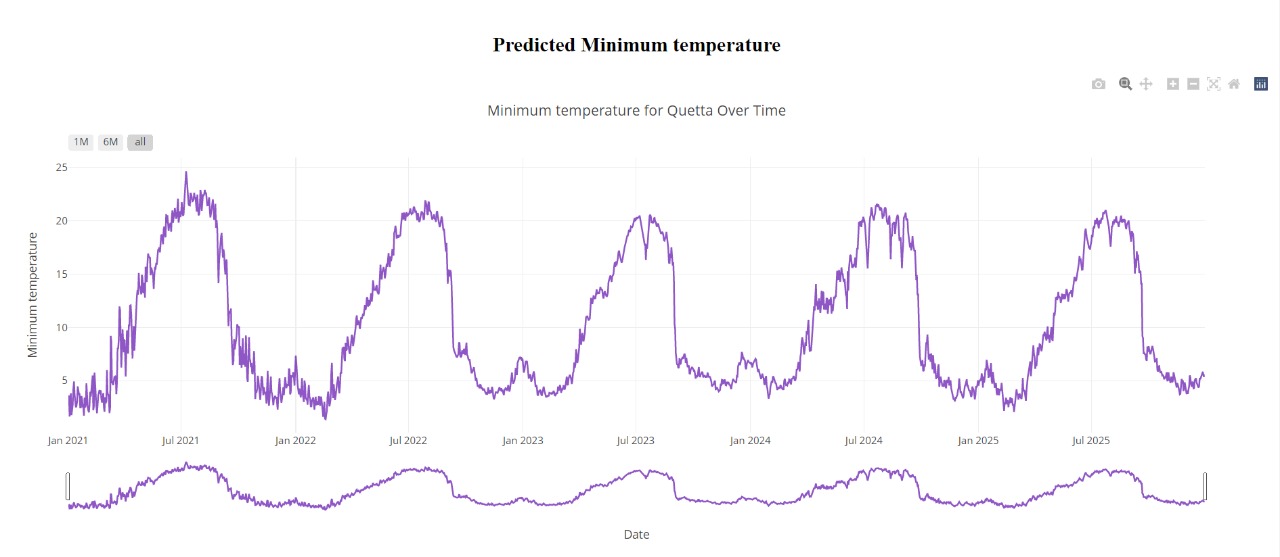
* + 1. Component 4

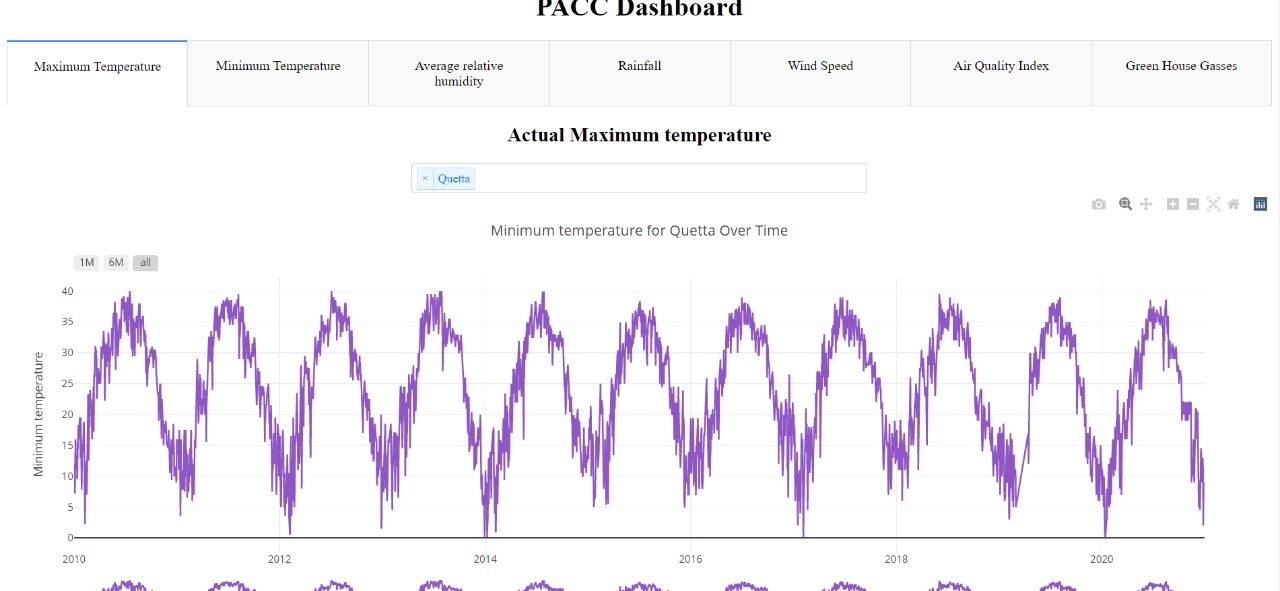
|  |  |
| --- | --- |
| Identification | Data Visualization tools |
| Type | Plotly, matplotlib, pandas, |
| Purpose | These libraries are used for the visualization of the dataframe that is being used in the model. |
| Function | The main function of these libraries is to visualize the data in an attractive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

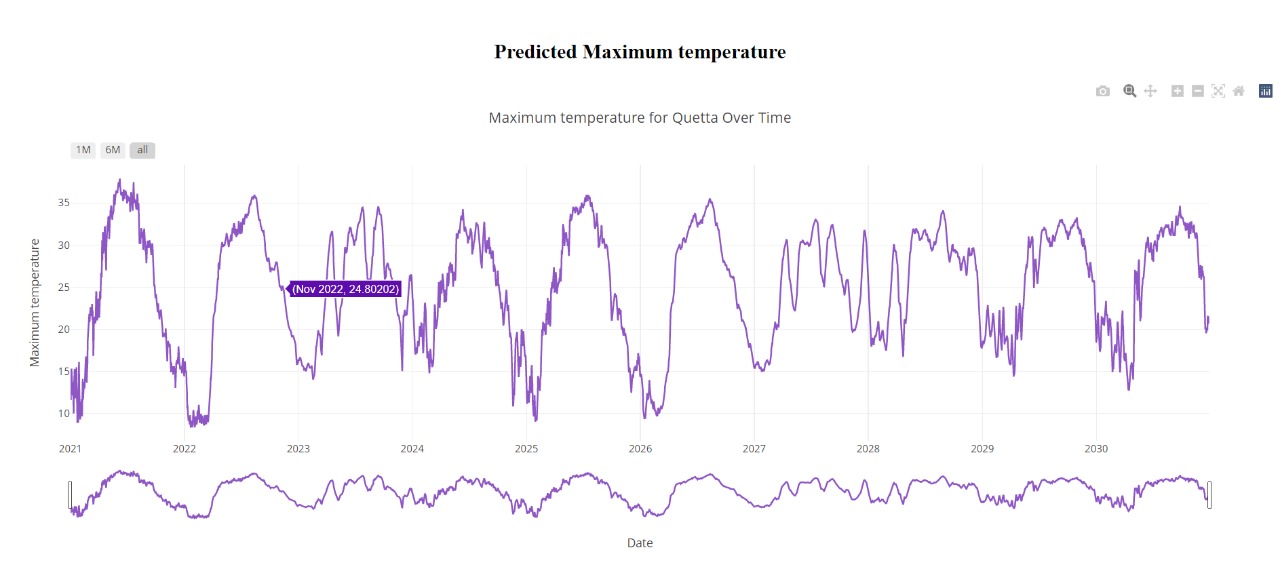
1. **User Interface Design**
   1. **Section Overview**  
      This section is providing with the actual values plotted from our dataset on the webpage using dash library of python. We created basic html page, created tabs and division to input the graphs. We created graphs of the actual dataset as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

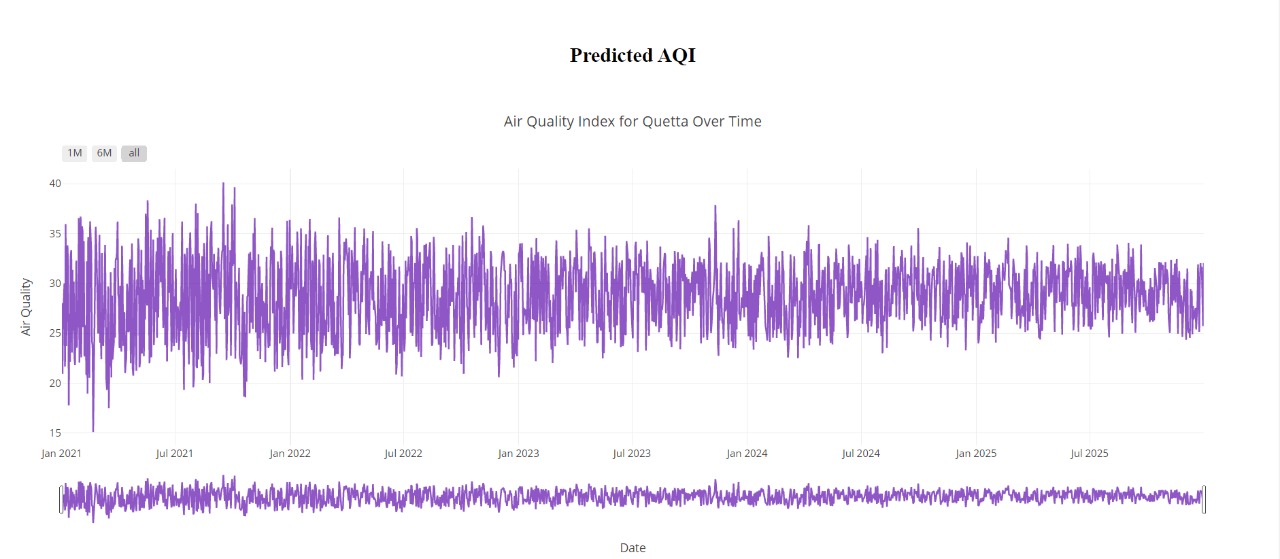
We have created a dashboard on the webpage visualizing the Actual data and also the Forecasted data Using Convolutional LSTM.

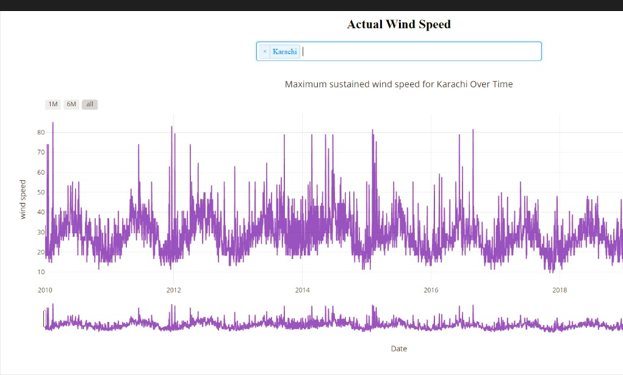


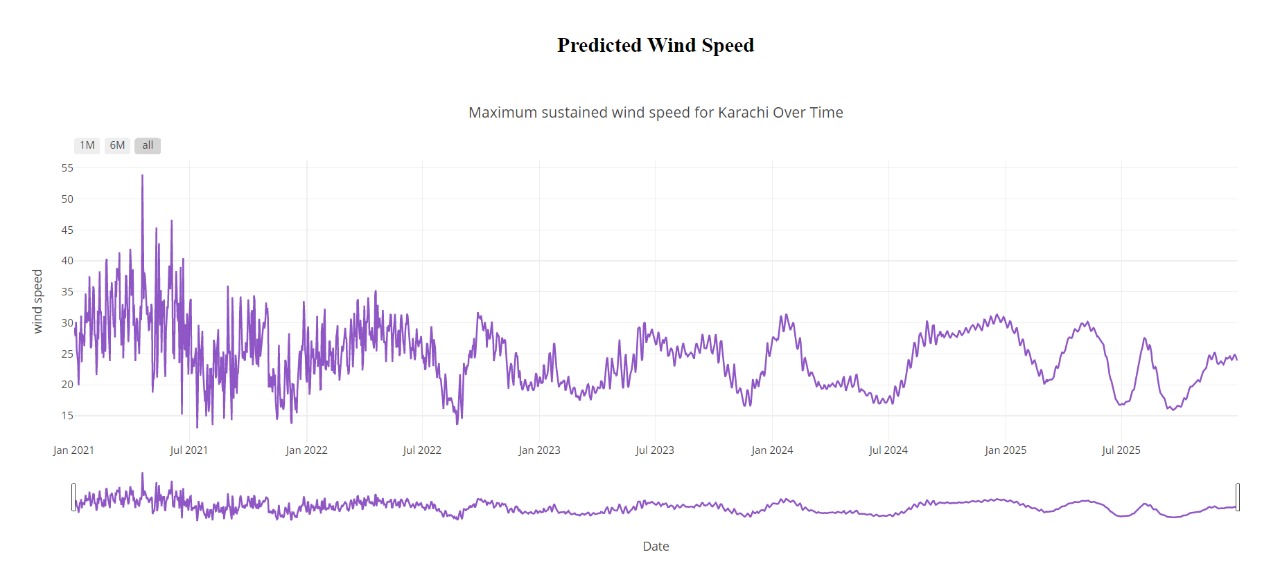


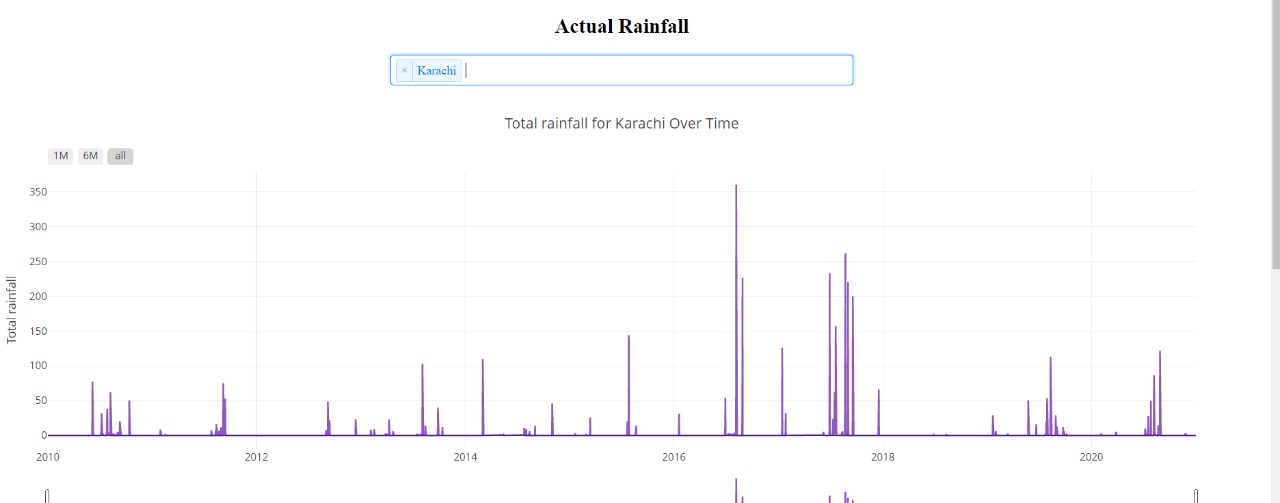


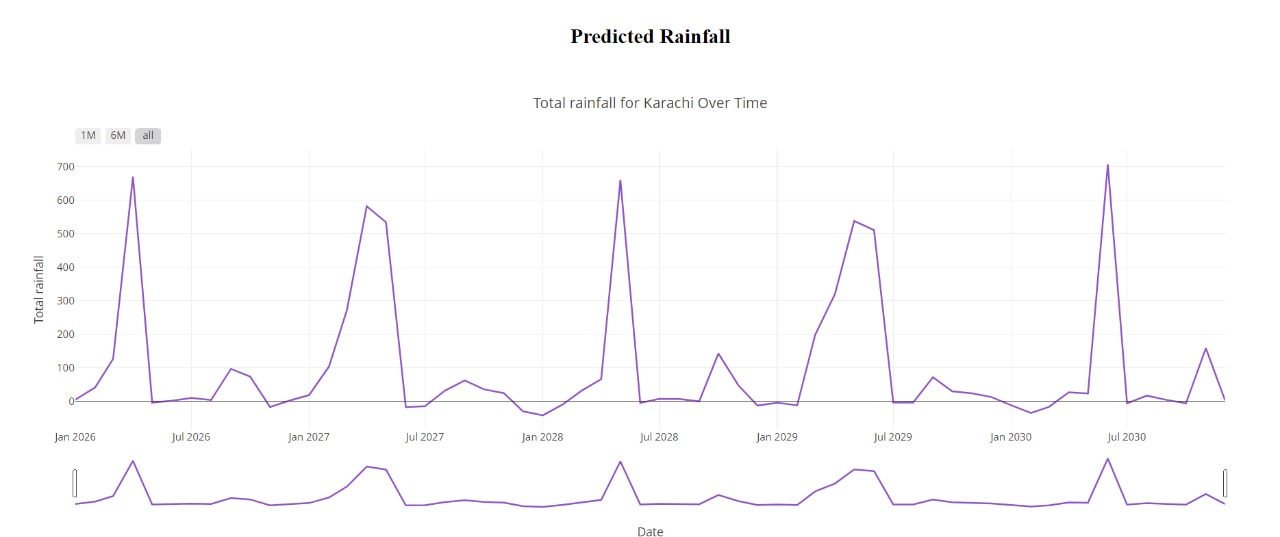


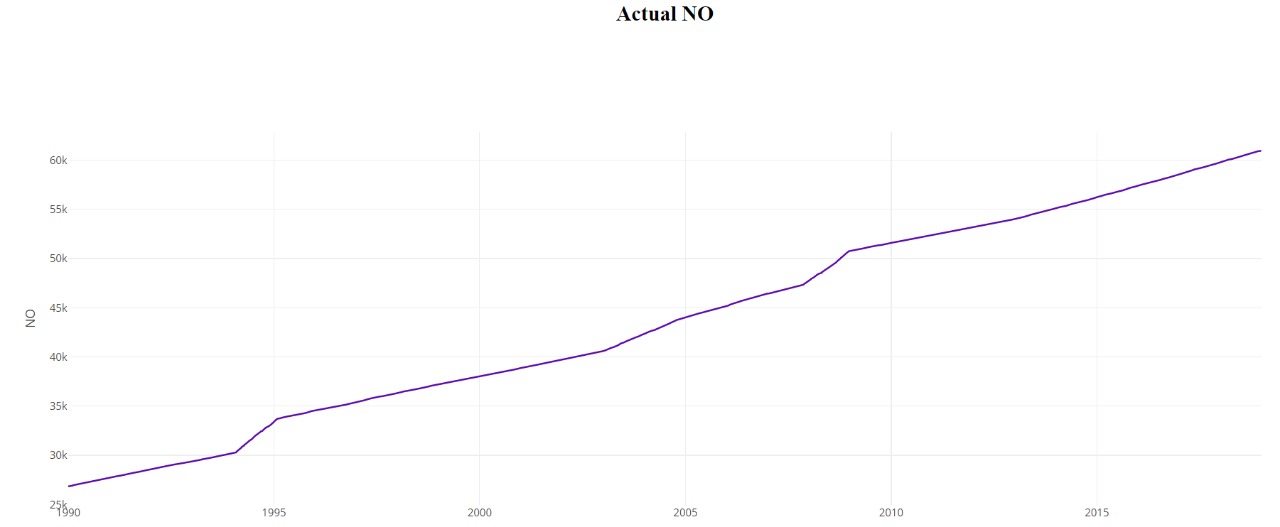


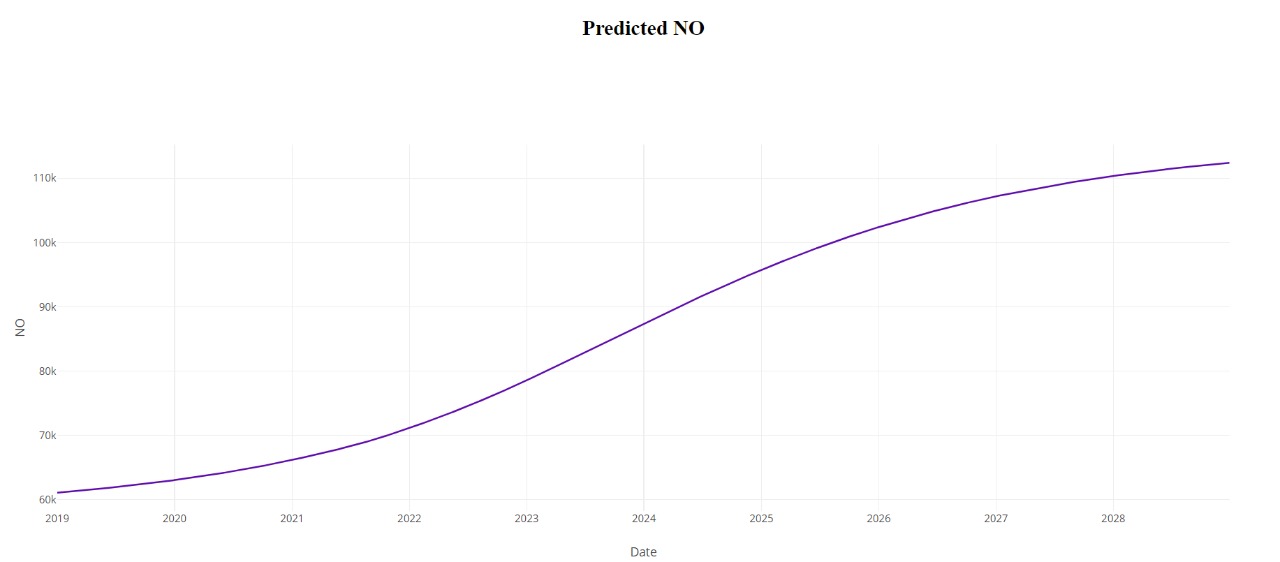












1. **Reuse and relationships to other products**

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**

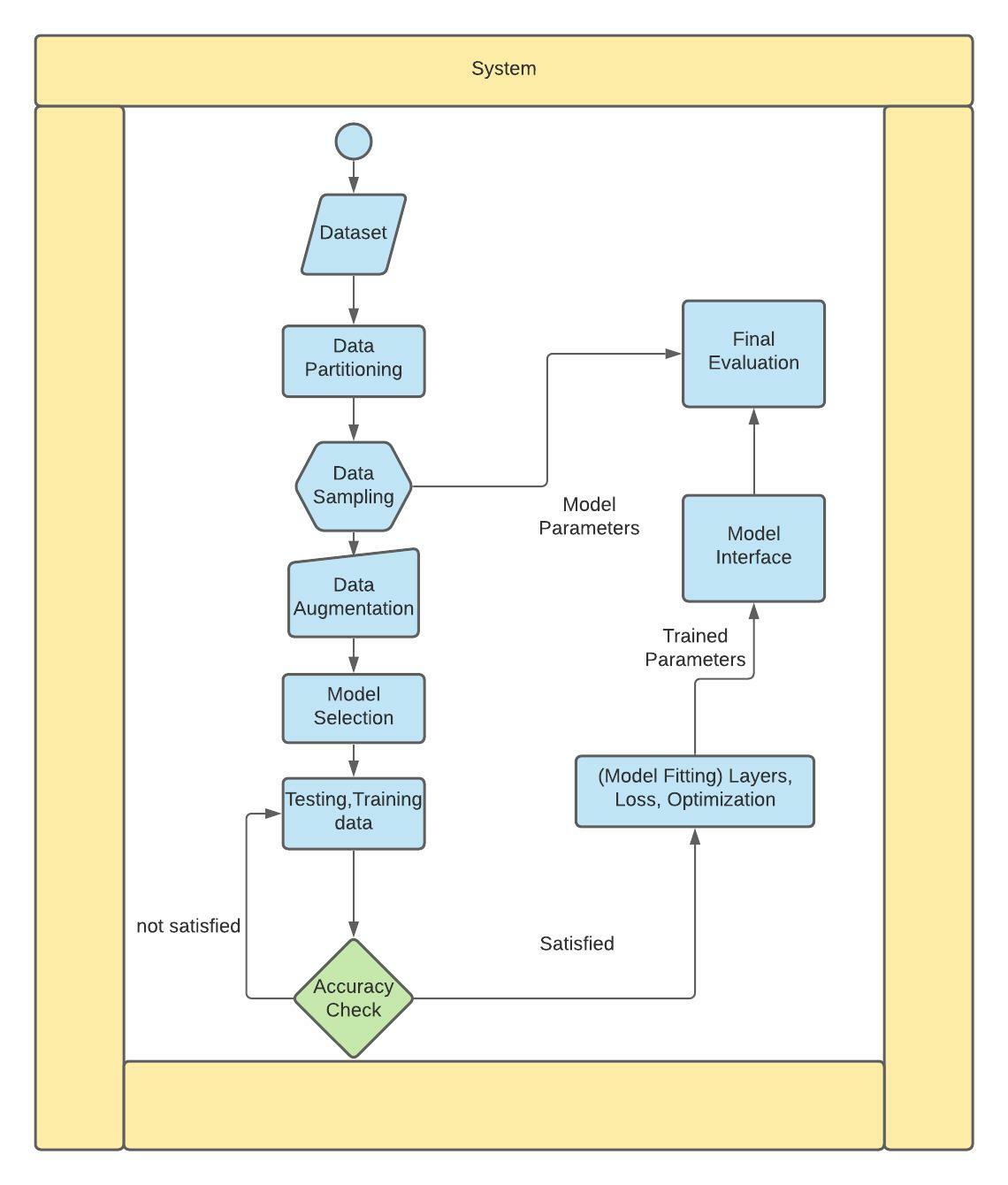
The main components applied and their pseudocode are as follows:

* 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.

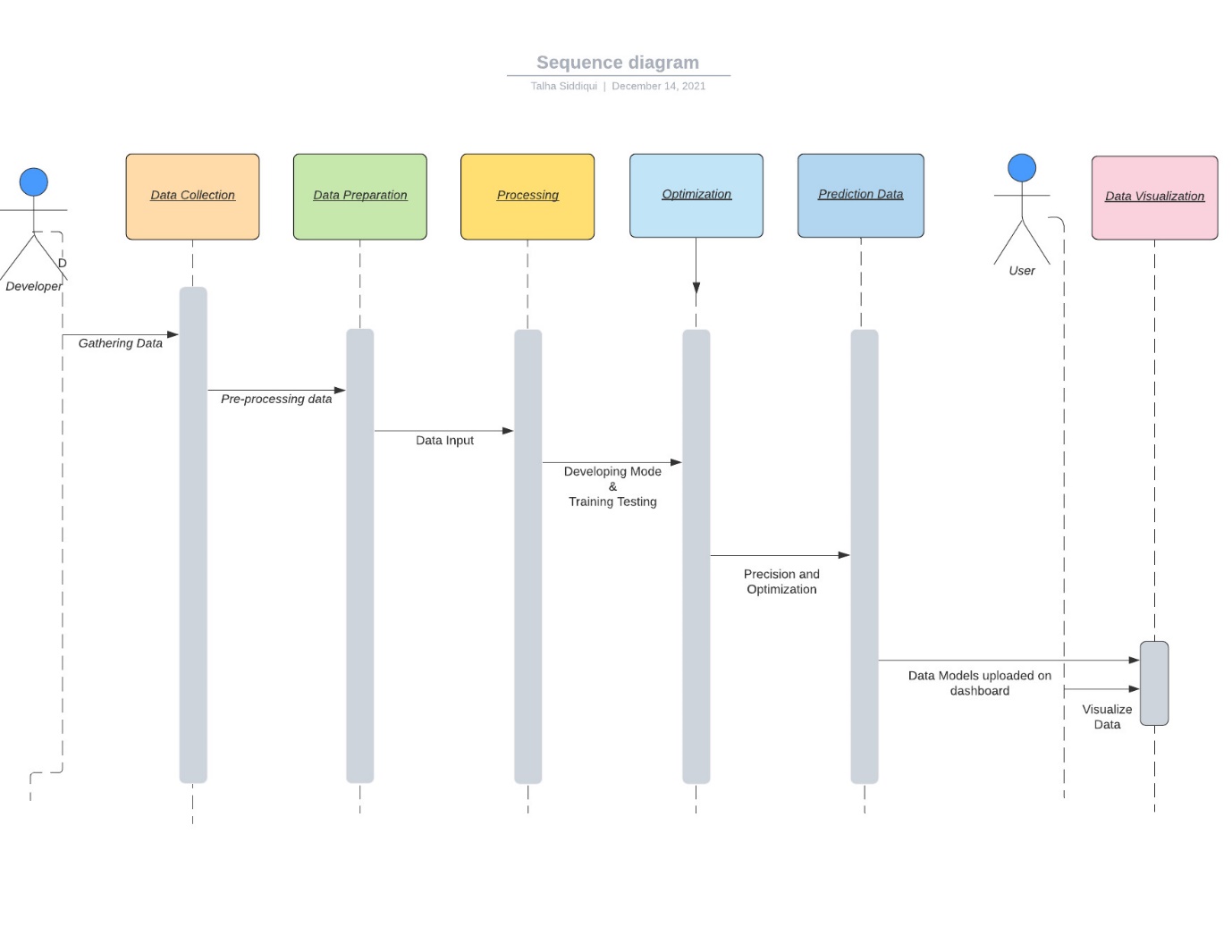
1. **Appendices**

|  |  |
| --- | --- |
| Activity Diagram | Describes activities. |
| Sequence Diagram | Shows one or several sequences. |
| Use-case Diagrams | Illustrating relationships between use cases |
| Context Diagram | Interactions between a system and other actors |
| Deployment Diagram | A special case of a Class Diagram. |
| System Block diagram | A diagram showing the major components of the system. |
| Work Flow Diagram | It provides an overview of the process |
| System Architecture | It is the conceptual model that defines the structure of a system. |

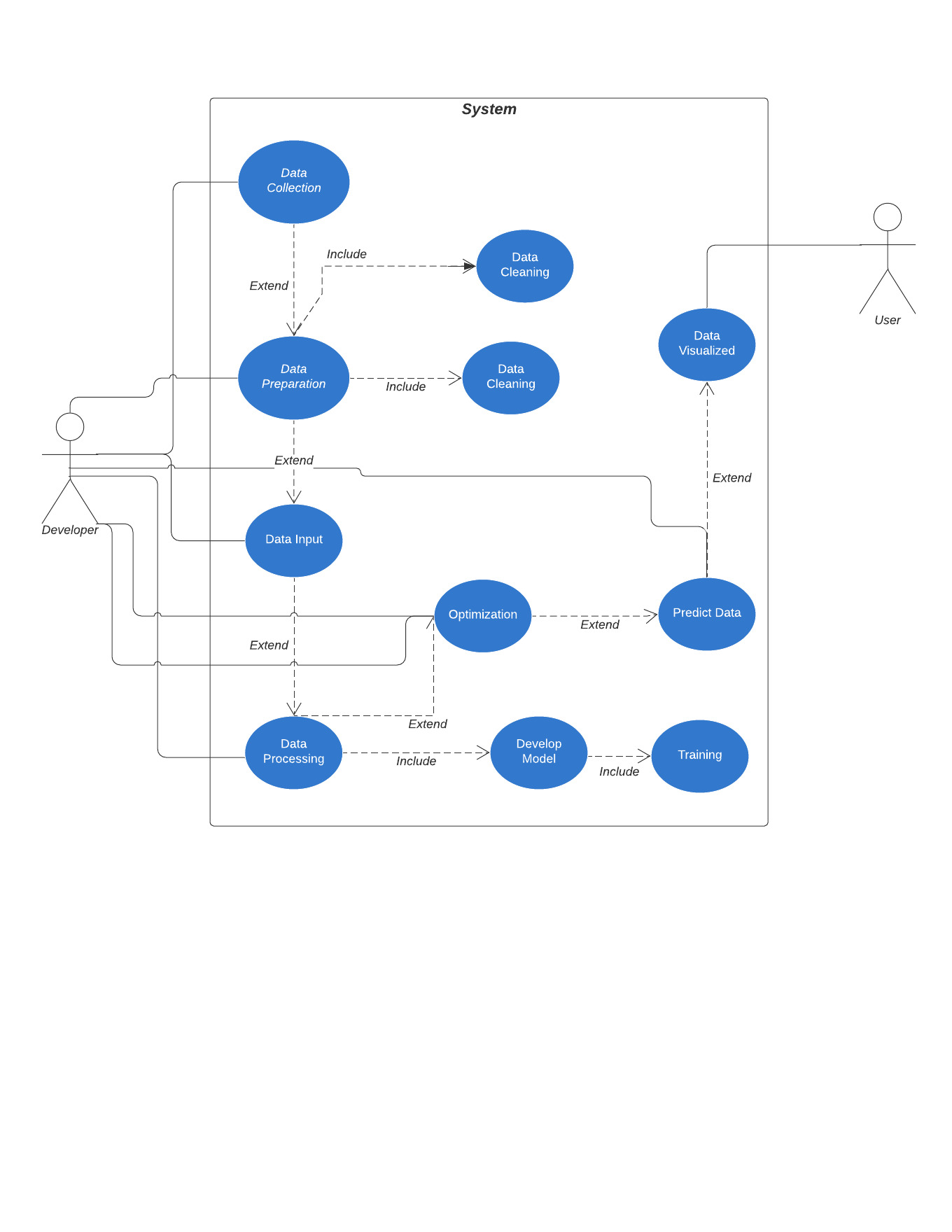
* 1. **Activity Diagram**

****

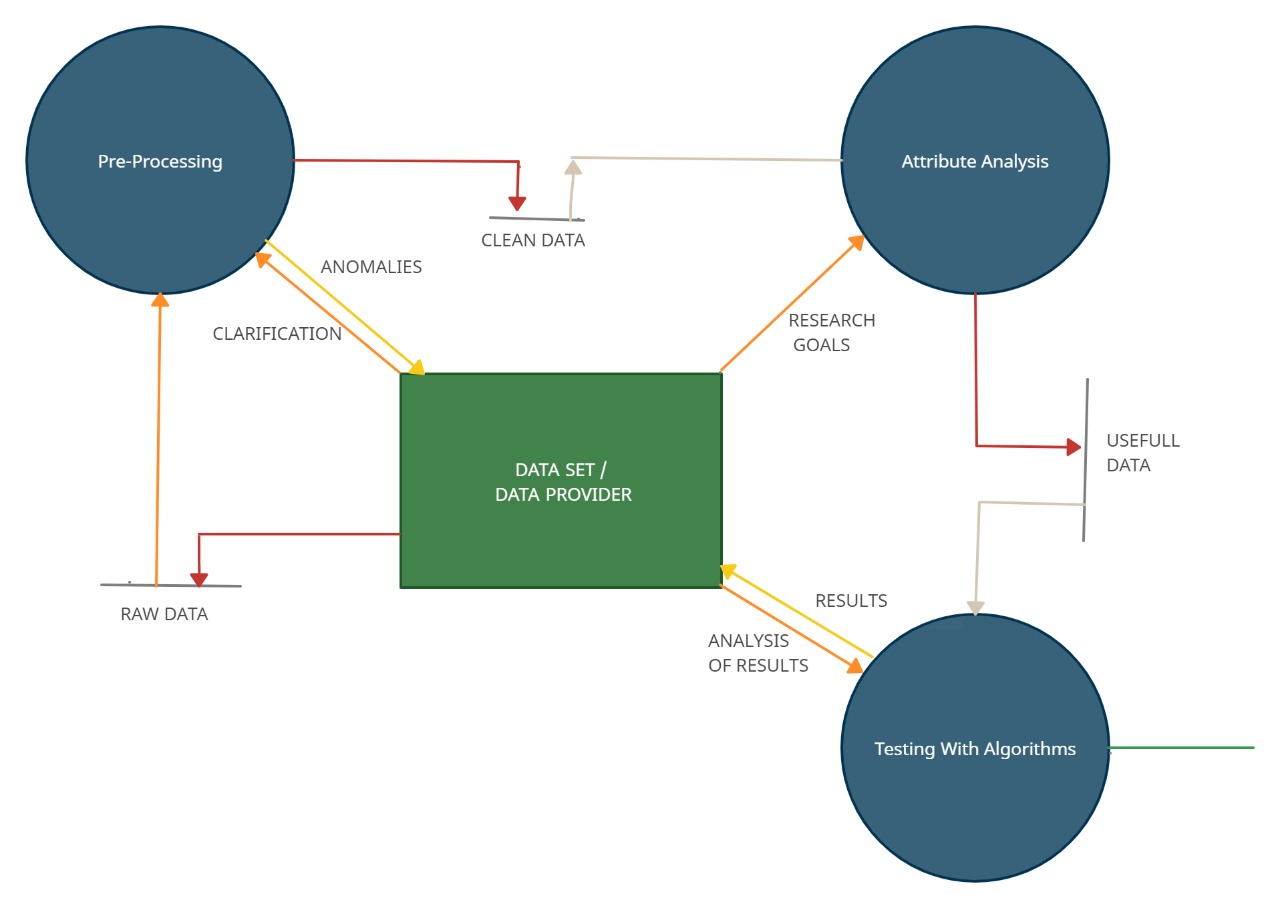
* 1. **Sequence Diagram**



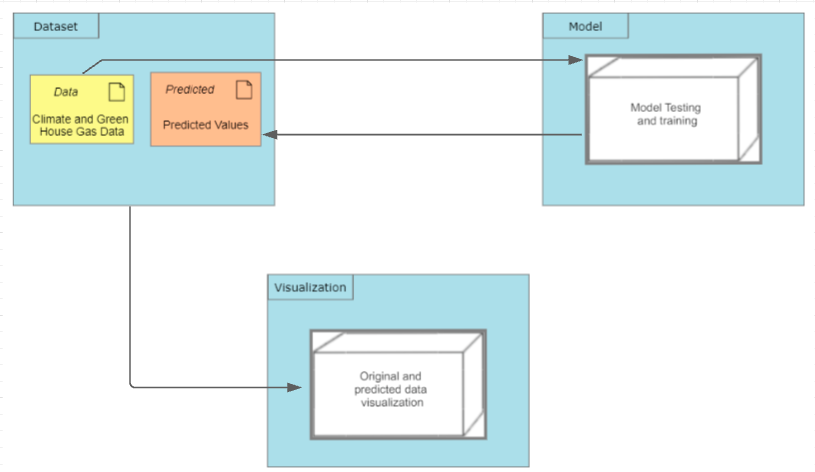
* 1. **Use Case Diagram**

****

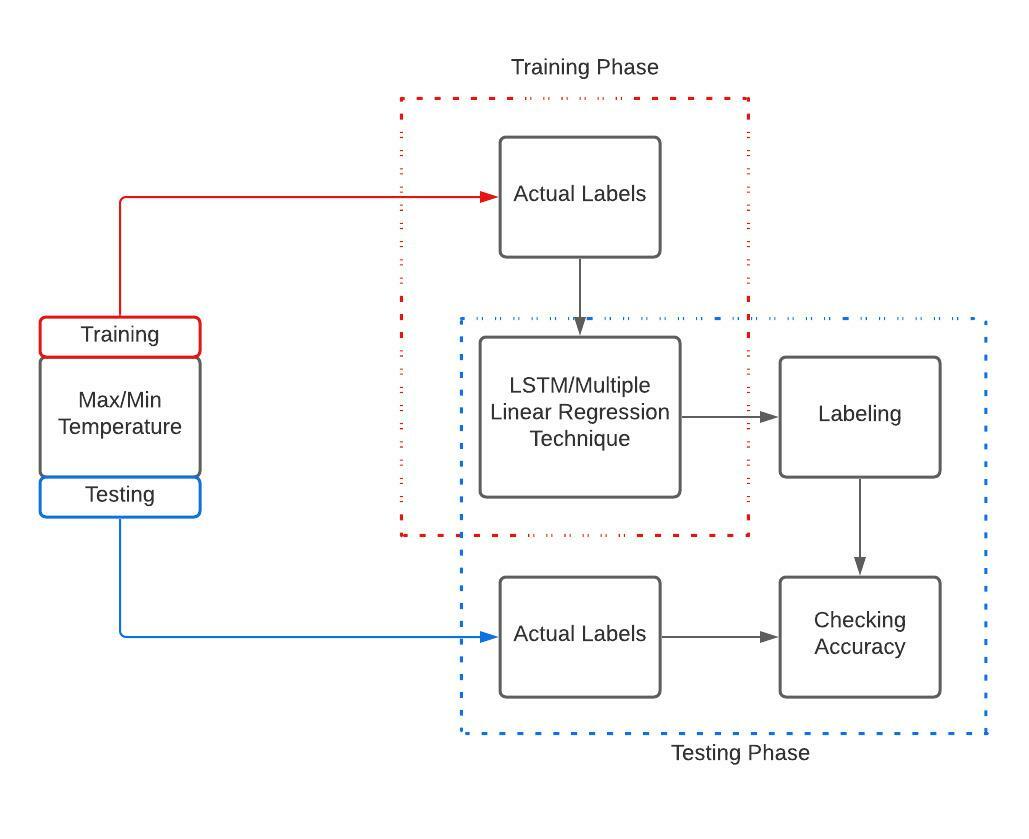
* 1. **Context Diagram**



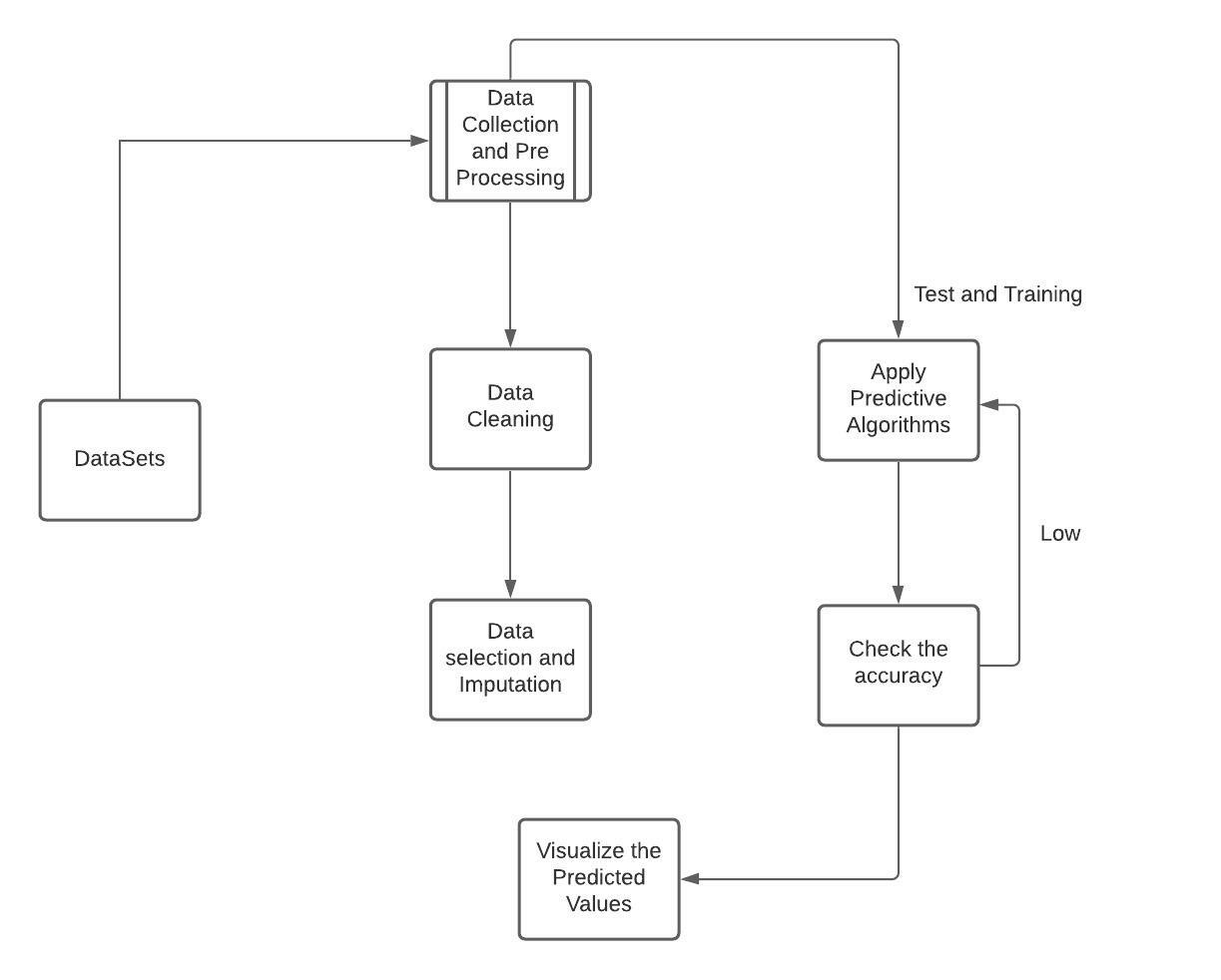
* 1. **Deployment Diagram**

****

* 1. **System Block Diagram**



* 1. **Work Flow Diagram**

****

* 1. **System Architecture Diagram**

