A PROJECT REPORT

ON

Exploratory Analysis of RainFall Data in India for Agriculture

Submitted By:

1.Praveen Raj B

2.Mohan C P

3.Nagha Rathish A

4.Prasanna A E

1. INTRODUCTION

1.1Project Overview

The science of training machines to learn and produce models for future predictions is widely used, and not for nothing. Agriculture plays a critical role in the global economy. With the continuing expansion of the human population, understanding worldwide crop yield is central to addressing food security challenges and reducing the impacts of climate change.

Crop yield prediction is an important agricultural problem. The Agricultural yield primarily depends on weather conditions, rain, temperature, etc. Accurate information about the history of crop yield is an important thing for making decisions related to agricultural risk management and future predictions.

1.2 Purpose

Rainfall has been a major concern these days. Weather conditions have been changing for the time being. Rainfall forecasting is important otherwise, it may lead to many disasters. Irregular heavy rainfall may lead to the destruction of crops, and heavy floods that can cause harm to human life. It is important to exactly determine the rainfall for effective use of water resources, crop productivity, and pre-planning of water structures.The purpose of the project is to build a forecasting machine learning-based model that will be critical to the development of an early warning system that can minimize hazards to people and property while also improving the management of agricultural chores

2.LITERATURE SURVEY

2.1 Existing System

The existing method measures various categories of data by linear regression method in metrics for effective understanding of agriculture in India. A real dataset has been, which consists of past year's rainfall rate according to various seasons. Results of this application help farmers to make correct decisions to harvest a particular crop accordingly to crops seasons.

2.2References

[1]Agriculture water demand management
AUTHOR:Bhatti, A.M., Pongsak, S. and Seigo, N. 2009

[2]Statistical distribution of rainfall in Uttarakhand, India
AUTHOR:Shanu VK and Jahangeer.Applied Water Science 7 (2017): 4765-4776.

[3]Analysis of Changes in Rainfall Magnitude over Haridwar District, India
AUTHOR:Rupali S Ahire and Nikam V World Environmental and Water Resources Congress 2020.May 17-21. | Henderson, Nevada (2020).

[4]Annual rainfall trends and periodicities in Nigeria
AUTHOR:Ayoade JO Nigerian Geographical Journal 16.2 (1973): 167-176.

[5]Changes in extreme climate indices for the northeastern United State.AUTHOR:Brown PJ, Bradley RS, Keimig FT. 2010.
1870–2005.Journal of Climate 23: 6555–6572.

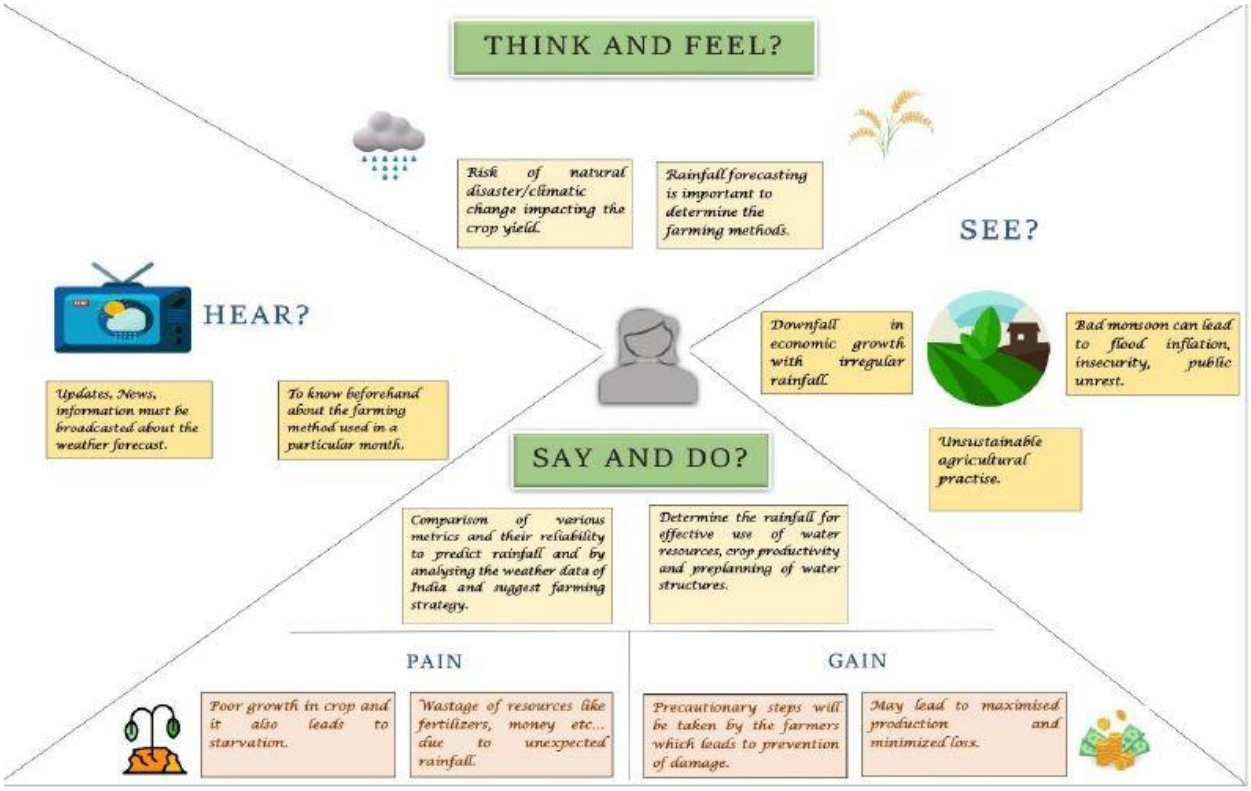
[6]Analysis of rainfall and temperature trends in northeast India. International Journal of climatology.AUTHOR:Jain SK, Kumar V and Saharia M. 2012.

2.3 Problem Statement Definition

Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	Small Farmer	Cultivate land to perform agriculture	I cannot do farming and produce yield	Insufficient of water to crops and irregular rainfall	Bad and depressed which is livelihood for me
PS-2	Large-scale farmer	Do 3 to 4 crop rotations every year	I could only do 2 times a year	Insufficient of water to crops and irregular rainfall	Less satisfied and depressed

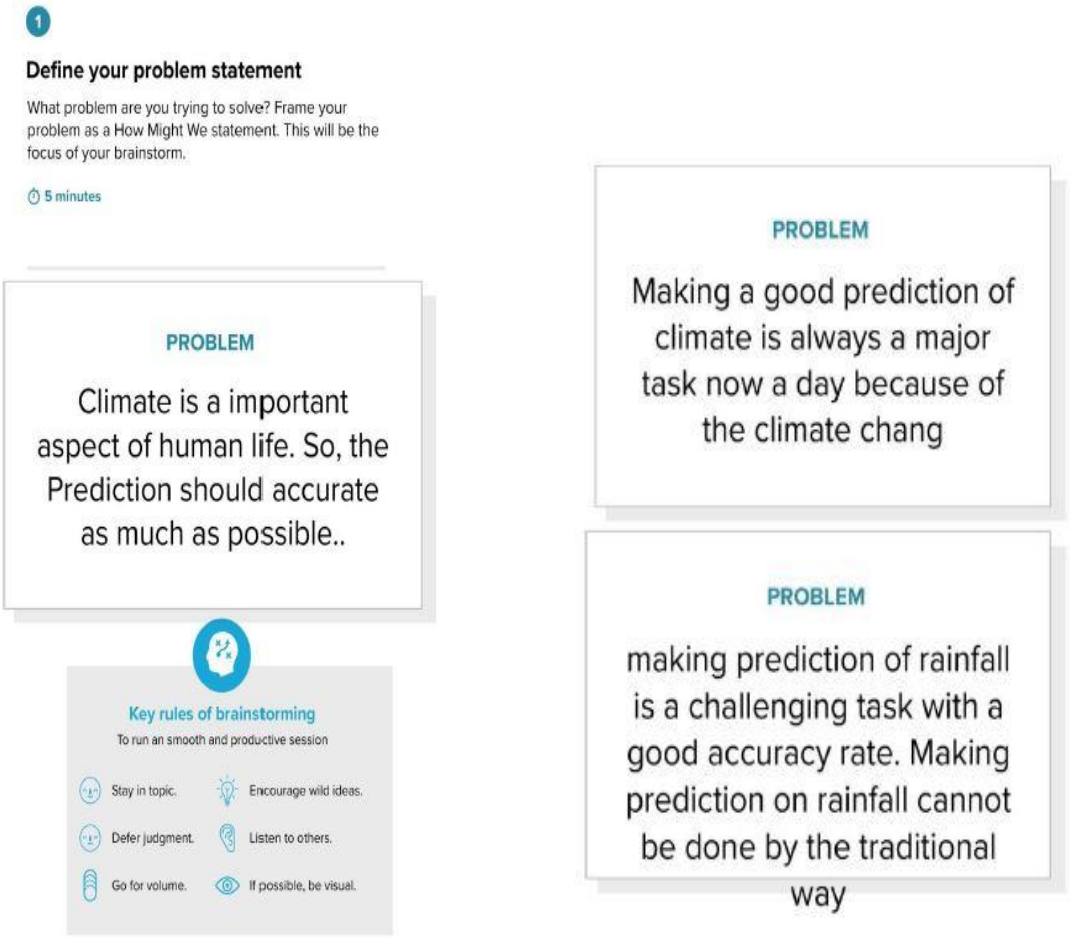
3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

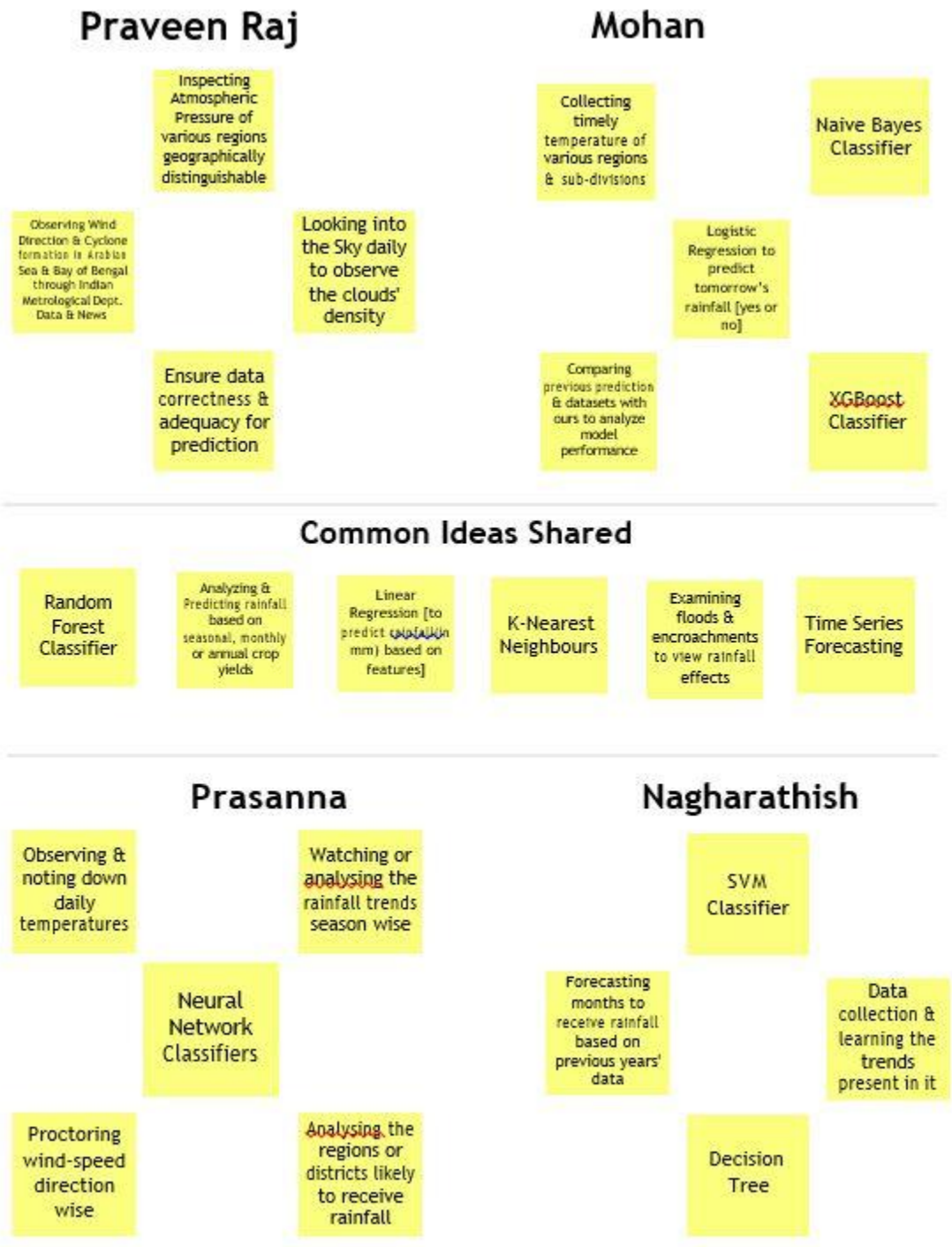


3.2 Ideation & Brainstorming

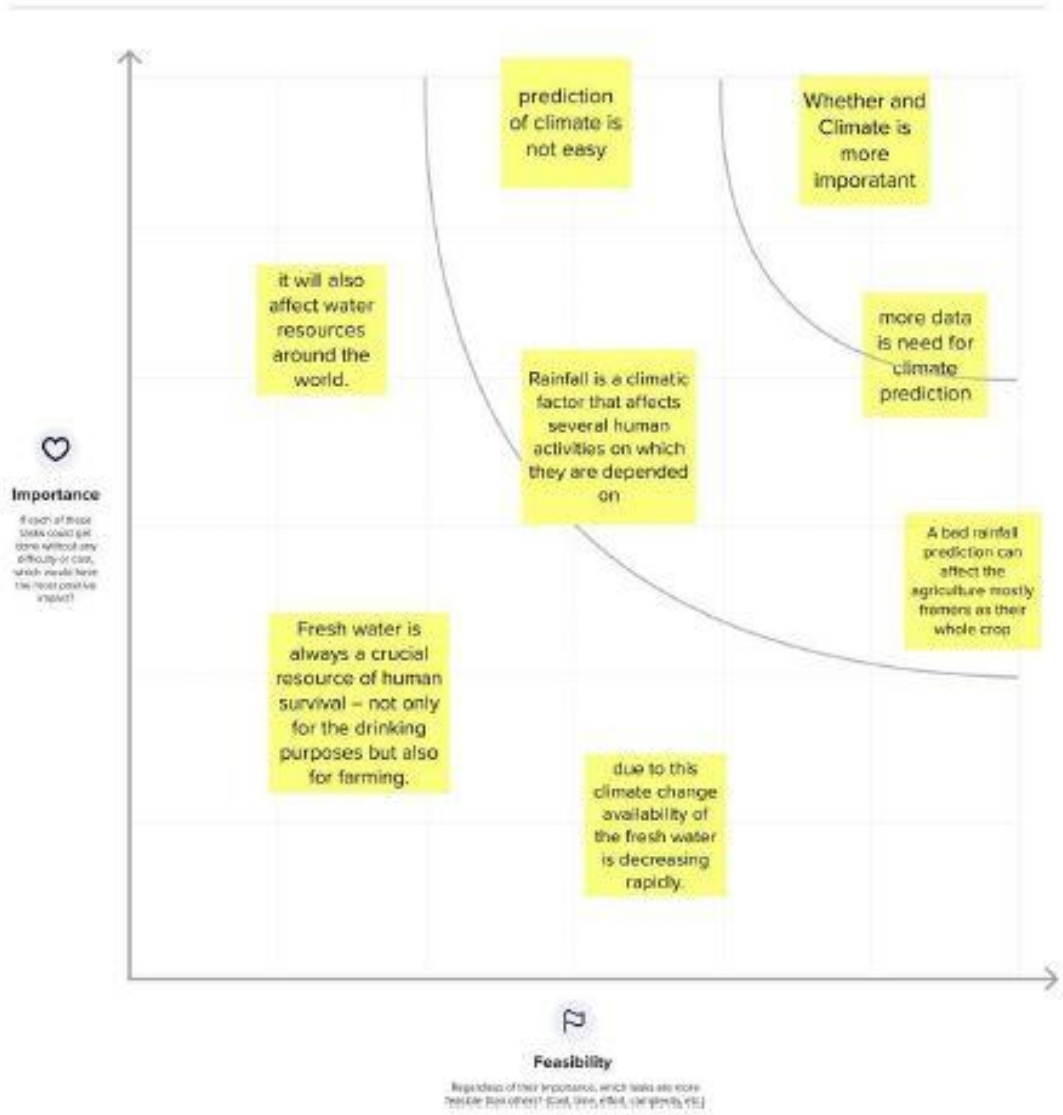
Step-1:Team Gathering,Collaboration and Select the Problem Statement



Step 2: Brainstorm,Idea Listing and Grouping



Step 3: Idea Prioritization

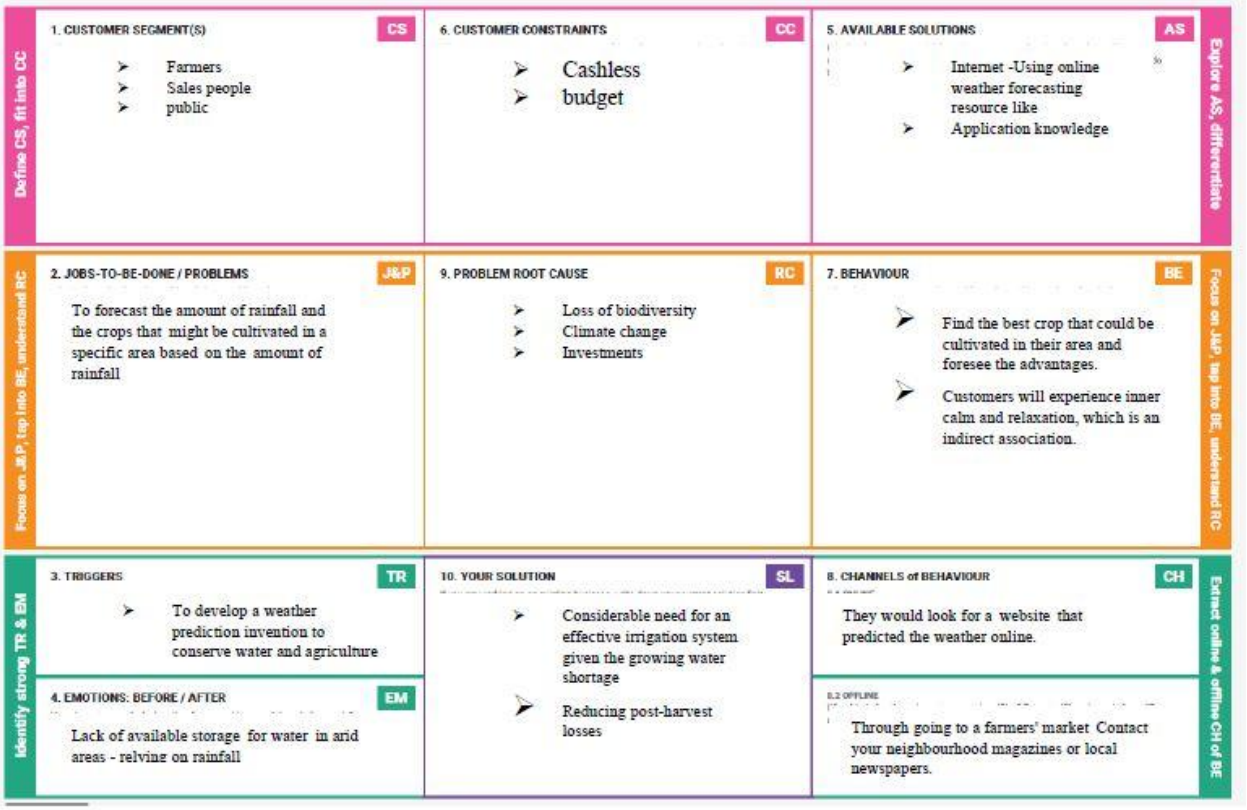


3.3 Proposed Solution

``

4.	Social Impact / Customer Satisfaction	• Different types of crops can be planted for good health. • Helps in producing healthy crops and good fields.
5.	Business Model (Revenue Model)	• This comparative study is conducted concentrating on the following aspects: modeling inputs, Visualizing the data, modeling methods, and pre-processing techniques. The results provide a comparison of various evaluation metrics of these machine learning techniques and their reliability to predict rainfall by analyzing the weather data. We will be using classification algorithms such as Decision tree, Random forest, KNN, and xg boost.
6.	Scalability of the Solution	• When we predict rainfall correctly, ithelps growth of crop and yielding will be better.

3.4 Problem Solution fit



4. **REQUIREMENT ANALYSIS**

4.1 Functional requirement

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Reliability	The prediction will be provided by the system error-free.
FR-4	Performance	The expected output will be produces immediately to the user without much delay.
NFR-5	Availability	The system would be available 24/7
NFR-6	Scalability	The system would be available on web application and any user can login and use it without any disruptions.

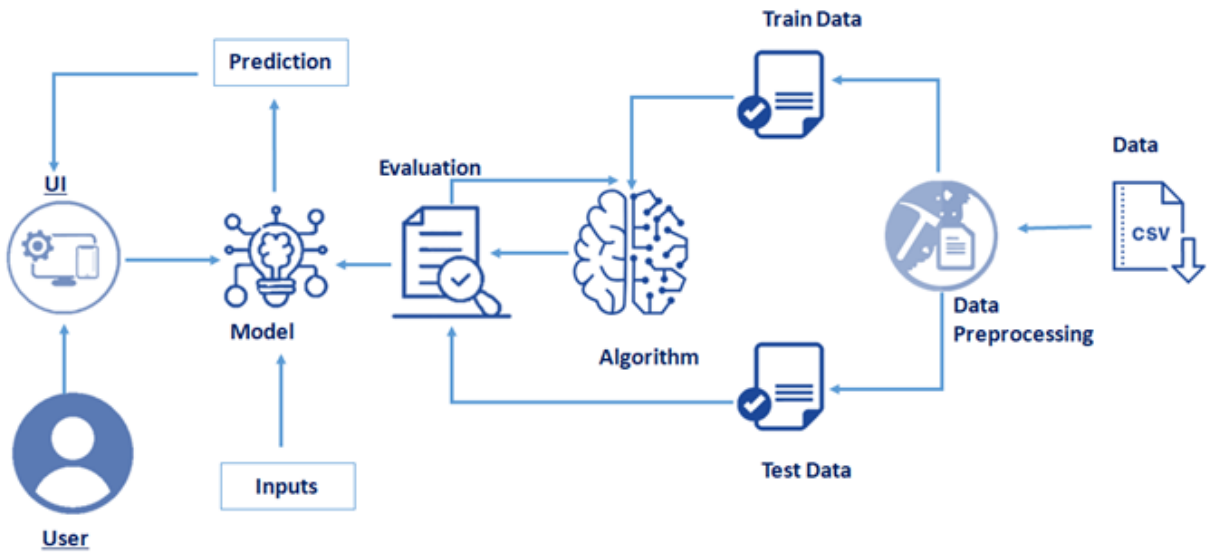
4.2 Non-Functional requirements

Following are the non-functional requirements of the proposed solution.

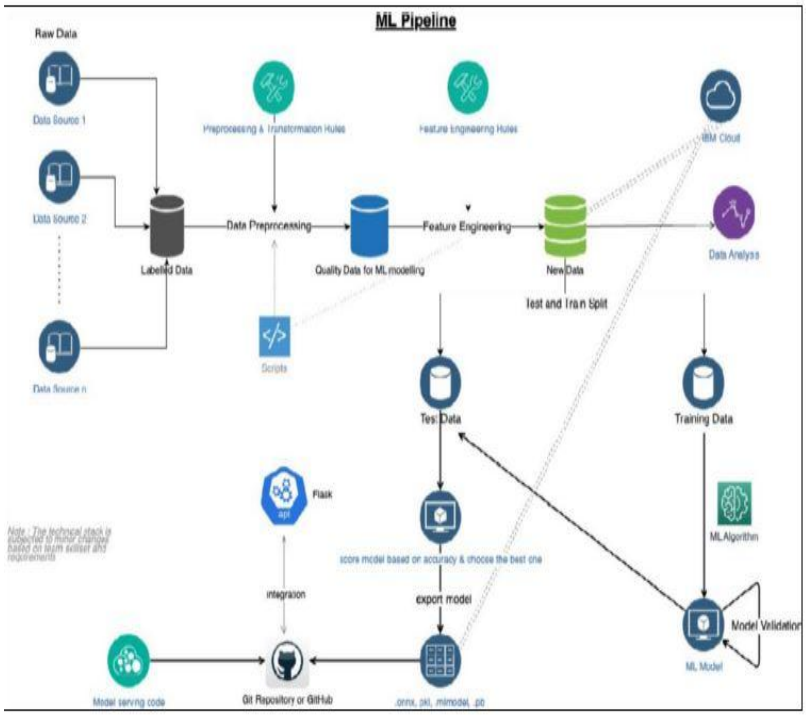
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Can be used anywhere(remote villages to metropolitan cities), anybody (kids to old age)
NFR-2	Security	Security is given over the model, so the user can use this with full trust. However, there are no personal details required to use this.
NFR-3	**Reliability**	Good connectivity and a supporting device can provide good results upto an extent.
NFR-4	**Performance**	This model can give a high accuracy prediction.
NFR-5	**Availability**	Any person can use this and this is an open-source model.

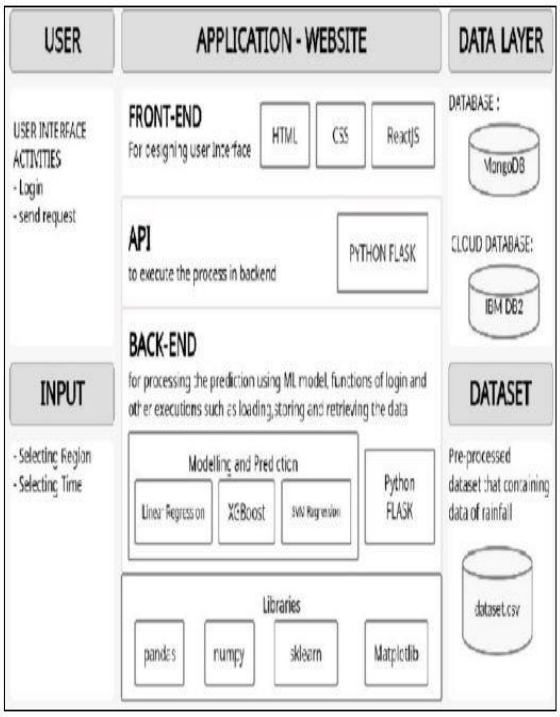
5. PROJECT DESIGN

5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture





5.3 User Stories

User type	Functional Requirement	User Story Number	User story / Task	Acceptance criteria	Priority	Release
Customer (mobile user)	Registration	USN-1	User can register for the application by entering his or her email, password, and confirming the password.	Account specific tasks and actions can be performed	High	Sprint-1
		USN-2	User will receive confirmation email or message once registered for the application	Verify the registered account	High	Sprint-1
		USN-3	Validation of the user can be done directly using email or OTP	Account validated and got access to profile dashboard	Medium	Sprint-1
	Login	USN-4	Enter the username and password to login to the application	Right account credentials should be entered	High	Sprint-1
		USN-5	The existing credentials should be used for login on multiple systems		Medium	Sprint-1
	Dash Board	USN-6	User can search for the region where he/she wants to know the prediction of rainfall	Searching for the region in India will be accepted only	High	Sprint-2

	Prediction	USN-7	User can view the visualization of the rainfall data for a specific region in India or for a specific time period		Medium	Sprint-2
		USN-8	User can change his/her password and can view the account details and search history	Verification will be required and new password should be entered	High	Sprint-2
		USN-9	The prediction or analysis request can be asked for the desired region for future or past events respectively		High	Sprint-2
	News	USN-10	User can give the feedback on the accuracy of the prediction and on the user interface		High	Sprint-3
Customer care Executive	Support	USN-11	Responds to user queries via telephone,email etc	Queries can be raised in situation of doubts	Medium	Sprint-3
		USN-12	The team must analyse all the queries and try to debug and make plans so that such queries wouldn't be raised again		Low	Sprint-3
		USN-13	Organize for a FAQ session where commonly asked doubts can be redressed by the team	The user will get all their doubt clarified	Low	Sprint-3
		USN-14	The team must respond immediately to the queries based on the priority	Queries should get resolved	High	Sprint-3

Core Development Team	Core function	USN-15	Design, develop the application in such a way that the best user interface and maintenance should be taken care of.	Easy and self-understandable user interface	High	Sprint-4
		USN-16	The website is responsive on all the devices and the screen sizes	User experience should be good irrespective of the devices or platforms	Medium	Sprint-4
		USN-17	The updates should be on time with the solutions of the raised queries	The existing functionalities should not affected by the update	High	Sprint-4

6.PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

Milestone & activity

MILESTONE	ACTIVITY
Rainfall Prediction ML Model (Dataset)	Weather Dataset Collection, Data pre-processing, Data Visualization
Registration	As a user, they can register for the application through Gmail. Password is set up.
Login	As a user, they can log into the application by entering email & password
Dashboard	Attractive dashboard forecasting live weather
Rainfall Prediction	User enter the location, temperature, humidity

Testing	Test the Application
Deploy Model	Deploy the model in IBM cloud to make user friendly application

6.2 Sprint Delivery Schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Rainfall Prediction ML Model (Dataset)	USN-1	Weather Dataset Collection, Data pre-processing, Data Visualization.	5	High	Praveen Raj B, Mohan CP
Sprint-1		USN-2	Train Model using Different machine learning Algorithms	5	High	Prasanna AE, Nagha Rathish A
Sprint-1		USN-3	Test the model and give best	10	High	Nagha Rathish A, Mohan CP
Sprint-2	Registration	USN-4	As a user, they can register for the application through Gmail. Password is set up.	5	Medium	Praveen Raj B, Prasanna AE
Sprint-2	Login	USN-5	As a user, they can log into the application by entering email & password	5	Medium	Praveen Raj B, Mohan CP
Sprint-2		USN-6	Credentials should be used for multiple systems and verified	4	Medium	Prasanna AE, Mohan CP
Sprint-2	Dashboard	USN-7	Attractive dashboard forecasting live weather	6	Low	Nagha Rathish A, Mohan CP
Sprint-3	Rainfall Prediction	USN-8	User enter the location, temperature, humidity	10	High	Mohan CP, Prasanna AE
Sprint-3		USN-9	Predict the rainfall and display the result	10	High	Praveen Raj B, Prasanna AE

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-4	Testing	USN-10	Test the application	10	High	Praveen Raj B, Mohan CP
Sprint-4	Deploy Model	USN-11	Deploy the model in IBM cloud to make user friendly application	10	High	Nagha Rathish A, Prasanna AE

7. CODING & SOLUTIONING (Explain the features added in the project along with code)

7.1 Feature 1

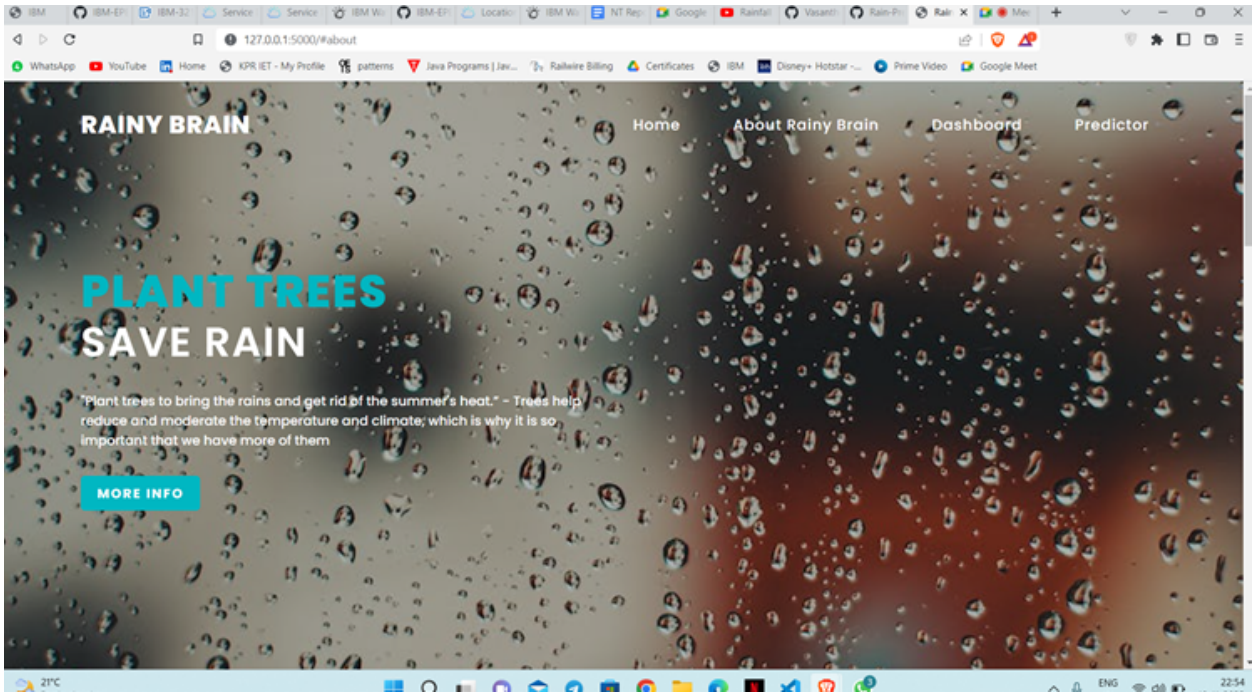
1. IoT Device
2. IBM Watson Platform
3. Node Red
4. Cloudant DB
5. Web UI
6. MIT App Inventor
7. Python code

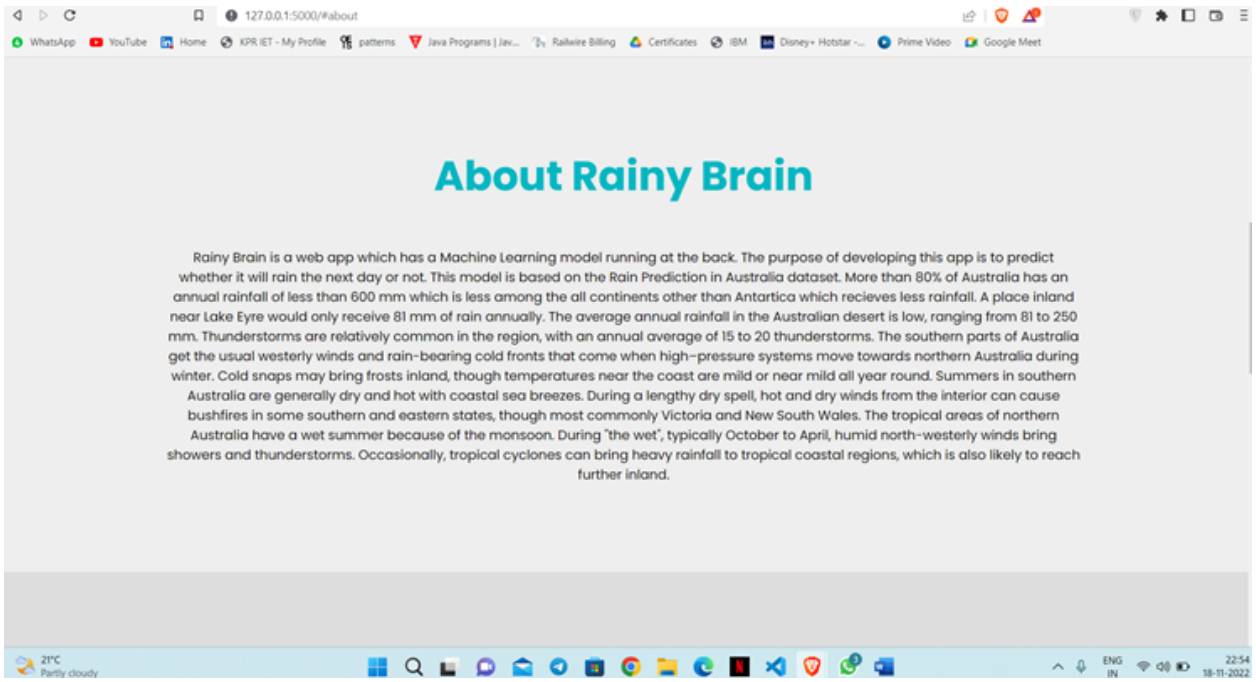
7.2 Feature 2

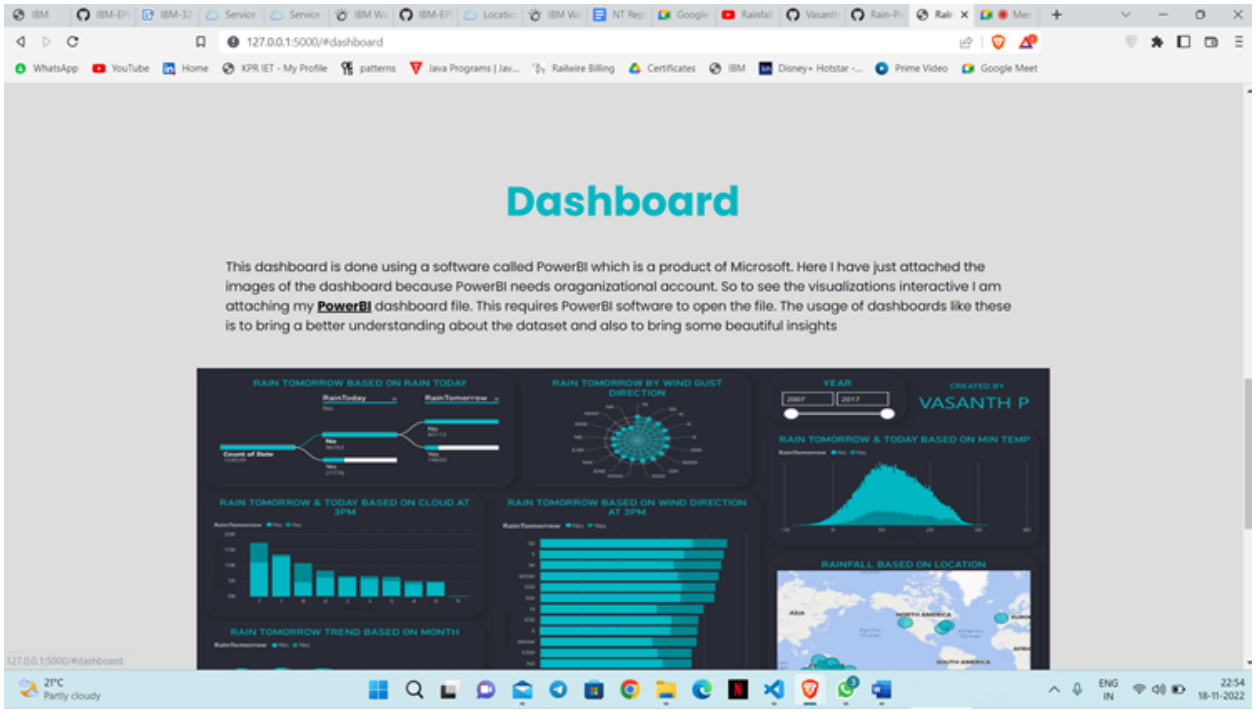
1. Login
2. Wowki

8. TESTING AND RESULTS

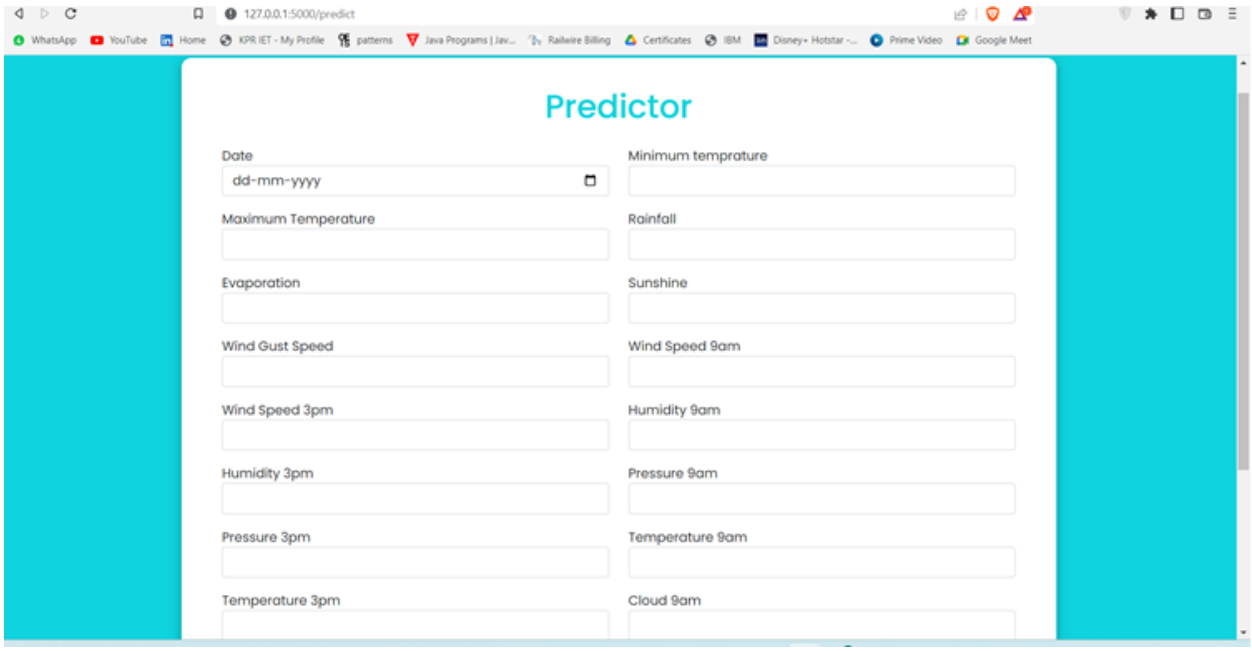
8.1 Test Cases




127.0.0.1:5000/#about
About Rainy Brain
Rainy Brain is a web app which has a Machine Learning model running at the back. The purpose of developing this app is to predict whether it will rain the next day or not. This model is based on the Rain Prediction in Australia dataset. More than 80% of Australia has an annual rainfall of less than 600 mm which is less among the all continents other than Antartica which recieves less rainfall. A place inland near Lake Eyre would only receive 81 mm of rain annually. The average annual rainfall in the Australian desert is low, ranging from 81 to 250 mm. Thunderstorms are relatively common in the region, with an annual average of 15 to 20 thunderstorms. The southern parts of Australia get the usual westerly winds and rain-bearing cold fronts that come when high-pressure systems move towards northern Australia during winter. Cold snaps may bring frosts inland, though temperatures near the coast are mild or near mild all year round. Summers in southern Australia are generally dry and hot with coastal sea breezes. During a lengthy dry spell, hot and dry winds from the interior can cause bushfires in some southern and eastern states, though most commonly Victoria and New South Wales. The tropical areas of northern Australia have a wet summer because of the monsoon. During "the wet", typically October to April, humid north-westerly winds bring showers and thunderstorms. Occasionally, tropical cyclones can bring heavy rainfall to tropical coastal regions, which is also likely to reach further inland.


127.0.0.1:5000/#dashboard
Dashboard
This dashboard is done using a software called PowerBI which is a product of Microsoft. Here I have just attached the images of the dashboard because PowerBI needs oraganizational account. So to see the visualizations interactive I am attaching my PowerBI dashboard file. This requires PowerBI software to open the file. The usage of dashboards like these is to bring a better understanding about the dataset and also to bring some beautiful insights
RAIN TOMORROW BASED ON RAIN TODAY
RAIN TOMORROW BY WIND GUST DIRECTION
YEAR
CREATED BY
VASANTH P
RAIN TOMORROW & TODAY BASED ON MIN TEMP
RAIN TOMORROW & TODAY BASED ON CLOUD AT 3PM
RAIN TOMORROW BASED ON WIND DIRECTION AT 3PM
RAINFALL BASED ON LOCATION
RAIN TOMORROW TREND BASED ON MONTH

8.2 User Acceptance Testing



9.Result:

10.ADVANTAGES & DISADVANTAGES

10.1 Advantages:

There are several reasons for using Exploratory Data Analysis today. Here are a few:

- EDA greatly improves an analyst's core understanding of different variables. They can extract different pieces of information, including averages, mean, minima and maxima, and other relevant points.
- More importantly, EDA can help analysts identify major errors, any anomalies, or missing values in their dataset. This is important before a comprehensive analysis begins, and can help organizations save a great deal of time.
- EDA can also help analysts identify key patterns. They can easily visualize the data through different types of graphs, ranging from box plots to histograms.

10.2 Disadvantages

- Even though it can point you in the right direction towards what is the answer, it is usually inconclusive.
- The main disadvantage of exploratory research is that they provide qualitative data.
- Interpretation of such information can be judgmental and biased.
- Most of the time, exploratory research involves a smaller sample, hence the results cannot be accurately interpreted for a generalized population.
- Many times, if the data is being collected through secondary research, then there is a chance of that data being old and is not updated.

11. CONCLUSION

Thus, we have developed such a type of application which helps the users to understand the behavior of rainfall in India that would help the farmers to plan accordingly.As part of research, we considered adding certain components to the application to make it more useful to the user. Some of the extra Components are like enabling users to register to the application using an existing email or social network account, it will send newsletters and send daily reports of the accuracy while using the application.

12.FUTURE SCOPE

The Future Enhancements of the application can be allowed to support in all the upcoming android versions. History can be set to view all the details in the app even if the particular data is deleted from the database. Statistics could be prepared based on the history of the reports. Sharing of analysis can be allowed so that more of the farms could be benefitted.Daily updates through WhatsApp and SMS could be added. Some of the extra components are like enabling users to register to the application using existing email or social network accounts,and connect with other agricultural people for more technological ideas.

APPENDIX

```
from flask import Flask,render_template,url_for,request,jsonify

from flask_cors import cross_origin

import pandas as pd

import numpy as np
```

```
import datetime
import pickle
app = Flask(__name__, template_folder="template")
model = pickle.load(open("./models/cat.pkl", "rb"))
print("Model Loaded")
@app.route("/",methods=['GET'])
@cross_origin()
def home():
        return render_template("index.html")
@app.route("/predict",methods=['GET', 'POST'])
@cross_origin()
def predict():
        if request.method == "POST":
                # DATE
                date = request.form['date']
                day = float(pd.to_datetime(date, format="%Y-%m-%dT").day)
                month = float(pd.to_datetime(date, format="%Y-%m-%dT").month)
                # MinTemp
                minTemp = float(request.form['mintemp'])
                # MaxTemp
                maxTemp = float(request.form['maxtemp'])
                # Rainfall
                rainfall = float(request.form['rainfall'])
```

```
# Evaporation

evaporation = float(request.form['evaporation'])

# Sunshine

sunshine = float(request.form['sunshine'])

# Wind Gust Speed

windGustSpeed = float(request.form['windgustspeed'])

# Wind Speed 9am

windSpeed9am = float(request.form['windspeed9am'])

# Wind Speed 3pm

windSpeed3pm = float(request.form['windspeed3pm'])

# Humidity 9am

humidity9am = float(request.form['humidity9am'])

# Humidity 3pm

humidity3pm = float(request.form['humidity3pm'])

# Pressure 9am

pressure9am = float(request.form['pressure9am'])

# Pressure 3pm

pressure3pm = float(request.form['pressure3pm'])

# Temperature 9am

temp9am = float(request.form['temp9am'])

# Temperature 3pm

temp3pm = float(request.form['temp3pm'])

# Cloud 9am
```

```
                cloud9am = float(request.form['cloud9am'])

                # Cloud 3pm

                cloud3pm = float(request.form['cloud3pm'])

                # Cloud 3pm

                location = float(request.form['location'])

                # Wind Dir 9am

                winddDir9am = float(request.form['winddir9am'])

                # Wind Dir 3pm

                winddDir3pm = float(request.form['winddir3pm'])

                # Wind Gust Dir

                windGustDir = float(request.form['windgustdir'])

                # Rain Today

                rainToday = float(request.form['raintoday'])

                input_lst = [location , minTemp , maxTemp , rainfall , evaporation , sunshine
,

      windGustDir , windGustSpeed , winddDir9am , winddDir3pm , windSpeed9am ,
windSpeed3pm ,

humidity9am , humidity3pm , pressure9am , pressure3pm , cloud9am , cloud3pm ,
temp9am , temp3pm ,

                                                rainToday , month , day]

                pred = model.predict(input_lst)

                output = pred

                if output == 0:

                  return render_template("after_sunny.html")
```

```
            else:
                return render_template("after_rainy.html")
    return render_template("predictor.html")

if __name__=='__main__':
    app.run(debug=True)
```

Github link: https://github.com/IBM-EPBL/IBM-Project-34599-1660239619