Exploratory Analysis of RainFall Data in India for Agriculture

PROJECT REPORT

Team ID	**PNT2022TMID10440**
Date	**17/11/2022**
Project Name	**Exploratory Analysis of RainFall Data in India for Agriculture**

1. INTRODUCTION

1.1 Project Overview

The main important source of water in any area is rain and it has a dramatic effect on agriculture. Plants get their water supply from natural sources and through irrigation. The yield of crops particularly in rain-fed areas depends on the rainfall pattern, which makes it important to predict the probability of occurrence of rainfall from the past records of hydrological data using statistical analysis. Frequency or probability distribution helps to relate the magnitude of the extreme events like floods, droughts and severe storms with their number of occurrences such that their chance of occurrence with time can be predicted easily. Accurate rainfall estimates during a storm event are invaluable to a forecaster responsible for flood warnings or reservoir operation. Typically concerned with "realtime" forecasting, a forecaster needs to predict actual flows within the next several hours as opposed to simulating a probabilistic design flood. Traditionally, rain gages have been used for measuring precipitation and telemetry for transmitting real-time records from remote gages to a forecast office. A significant drawback associated with gage information is that data is collected at a point and an interpolation scheme is required to produce a rainfall surface and to calculate watershed-average rainfall.

1.2 Purpose

India is an agricultural country and secondary agro based market will be steady with a good monsoon. The economic growth of each year depends on the amount of duration of monsoon rain, bad monsoon can lead to destruction of some crops, which may result in scarcity of some agricultural products which in turn can cause food inflation, insecurity and public unrest. In our analysis we are trying to understand the behavior of rainfall in India over the years, by months and different subdivisions. The total rainfall received in a given period at a location is highly variable from one year to another. The variability depends on the type of climate and the length of the considered period. In general it can be stated that the drier the climate, the higher the variability of rainfall in time. The same hold for the length of the period: the shorter the period the higher the annual variability of rainfall in that period.

2. LITERATURE SURVEY

2.1 Existing problem

Rainfall is a prime input for various engineering design such as hydraulic structures, bridges and culverts, canals, storm water sewer and road drainage system. The detailed statistical analysis of each region is essential to estimate the relevant input value for design and analysis of engineering structures and also for crop planning. This analysis will provide useful information for water resources planner, farmers and urban engineers to assess the availability of water and create the storage accordingly.

2.2 References

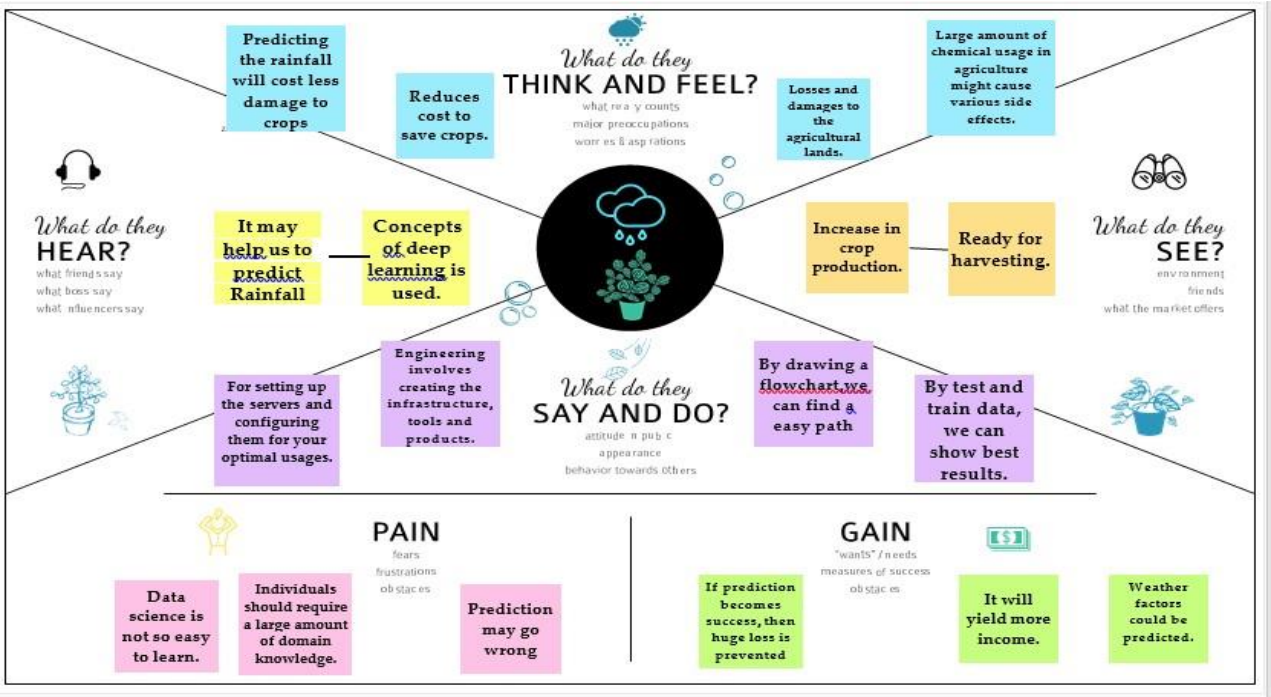
PROJECT TITLE	*AUTHOR*	*OBJECTIVE/OUTCOME*
Spatial analysis of Indian Summer monsoon Rainfall (Mar 26,2014)	Markand Oza C.M.Kishtawal	Understanding the variability in rainfall, analysis of Indian Summer monsoon rainfall using Spatial resolution.
Climate impacts on Indian Agriculture. (16 June,2004)	K.Krishna kumar K.Rupa Kumar R.G.Ashrit N.R.Deshpande J.W.Hansen	Presents about the analysis of Crop-climate relationships for India, using historical predictions.
Exploratory data Analysis of Indian Rainfall Data	Anusha Gajinkar	This Study shows that, India has two monsoon rainfall season one is north west monsoon and second one is south east monsoon.

2.3 Problem Statement Definition

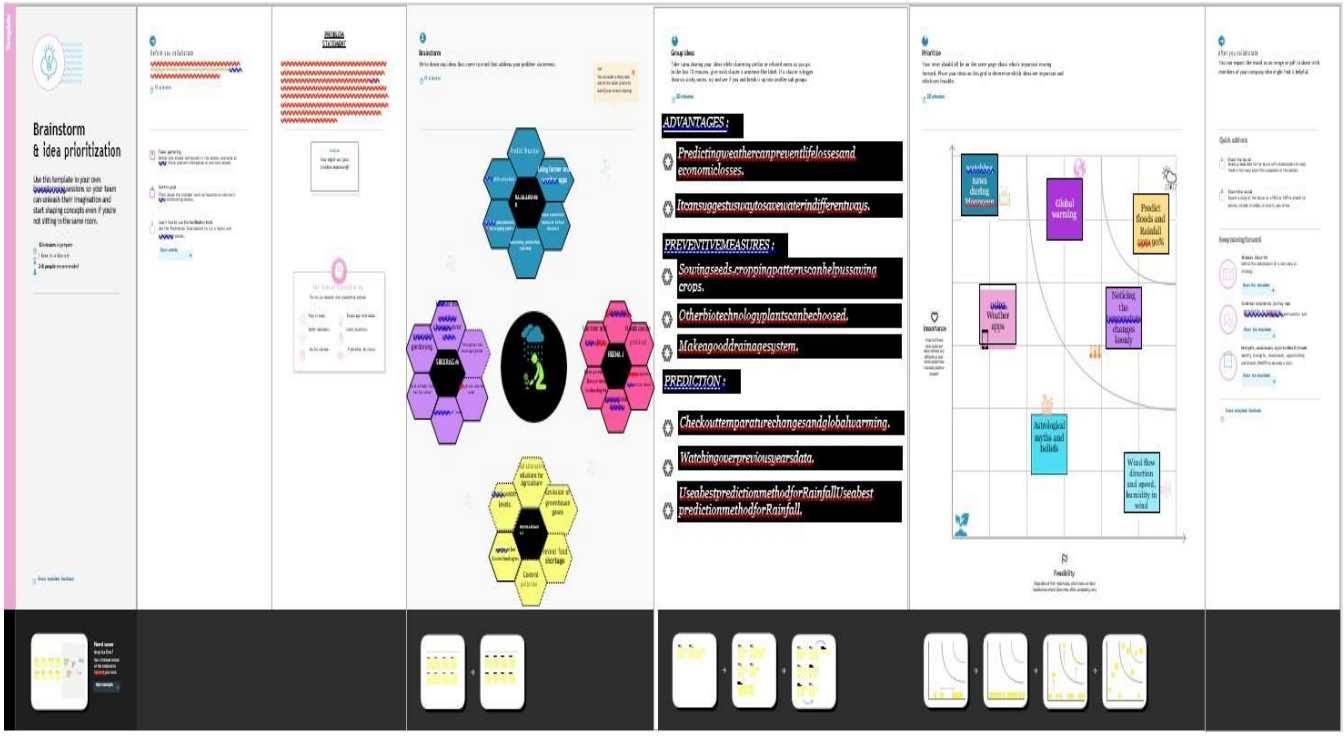
Weather conditions changes then and often. This can lead to Severe threats to all the living beings including human beings. So predicting weather, especially Irregular heavy rainfall can cause huge floods and economic losses. This also decreases crop productivity and may lead into Food shortage. Predicting the Rainfall plays a vital role in our life time. Farmers will get benefit due to this and Our country's GDP will rise. Collection of previous 10 years data may give us an idea about the pattern of Rainfall. Using all these Datas, Appropriate farming activities can be performed. Water is the vital mineral for a life. So, these datas can help us in predicting Rainfall during summer days to save water. Agriculture definitely requires gallons of waters.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



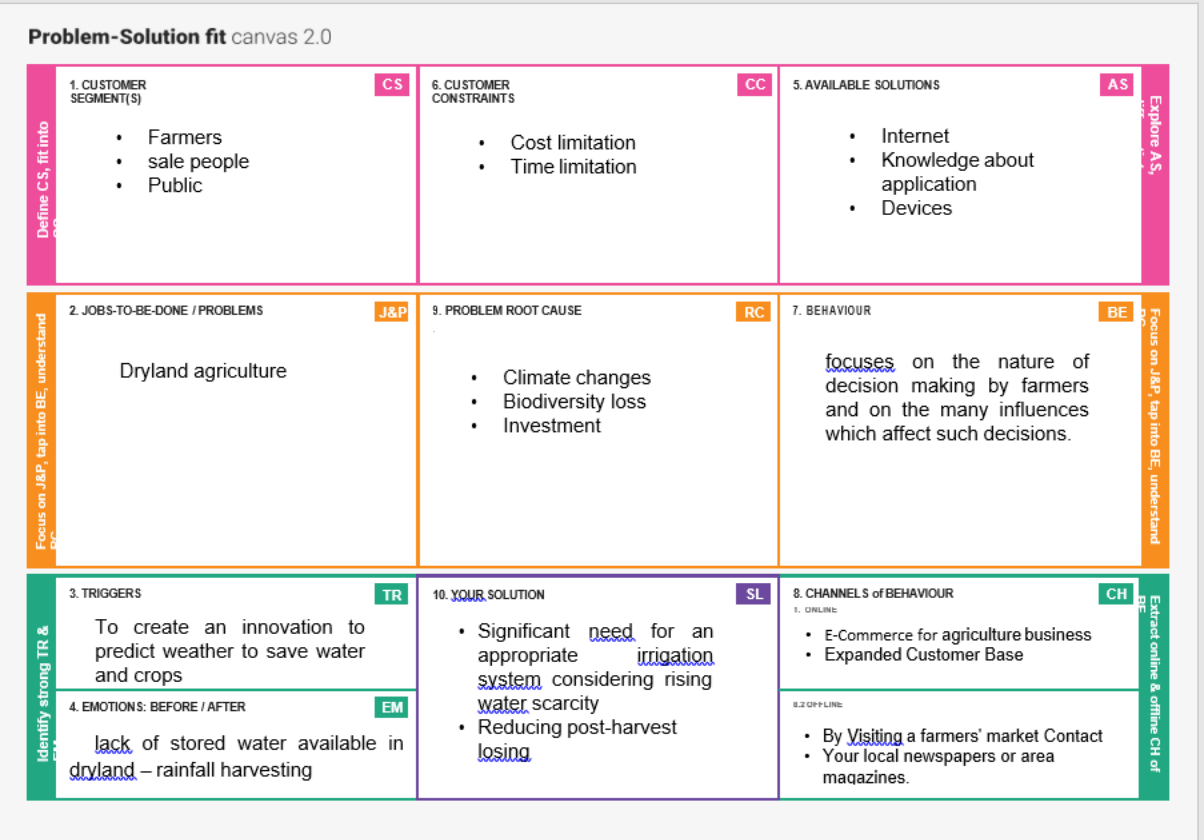
3.2 Ideation & Brainstorming



3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Heavy Rainfall may cause huge threat to all living beings, especially in the field of Agriculture. Droughts could do the same too. It may destroy the crops and cause huge loss to Farmers and dependent field workers. Predicting Rainfall is a major task in both summer and Rainy season.
2.	Idea / Solution description	Analysing the previous 10 years datas can give us a rough idea about Rainfall pattern. Using Data Science, we could solve this and predict the Rainfall upto some good extent.
3.	Novelty / Uniqueness	AI, IOT and so many other fields may require different sensors. We are not going to use any kind of equipment. Time of prediction is very less and easy with affordable cost.
4.	Social Impact / Customer Satisfaction	Farmers (they save crops and money), Vegetable sellers(they knows about vegetable stocks and its emergency)
5.	Business Model (Revenue Model)	This could cost really low as a person should develop knowledge in Data science and probably a gadget to develop this. However, deploying as an App attached with other facilities may cost an extra charge.
6.	Scalability of the Solution	Farmers, Vegetable sellers, Citizens

3.4 Problem Solution fit



4. REQUIREMENT ANALYSIS

4.1 Functional requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Import necessary packages	Importing packages like NumPy, pandas, seaborn, etc
FR-2	Download and load dataset	Download the dataset Load the Appropriate dataset
FR-3	Pre-processing of data	Making data suitable for building a good model
FR-4	Building Machine learningmodel	Choose the best algorithm. Check for the best optimised result.
FR-5	Train the data	Train the model using training data.
FR-6	Test the model	Test the model for the best evaluation and analysing.

4.1 Non Functional requirement

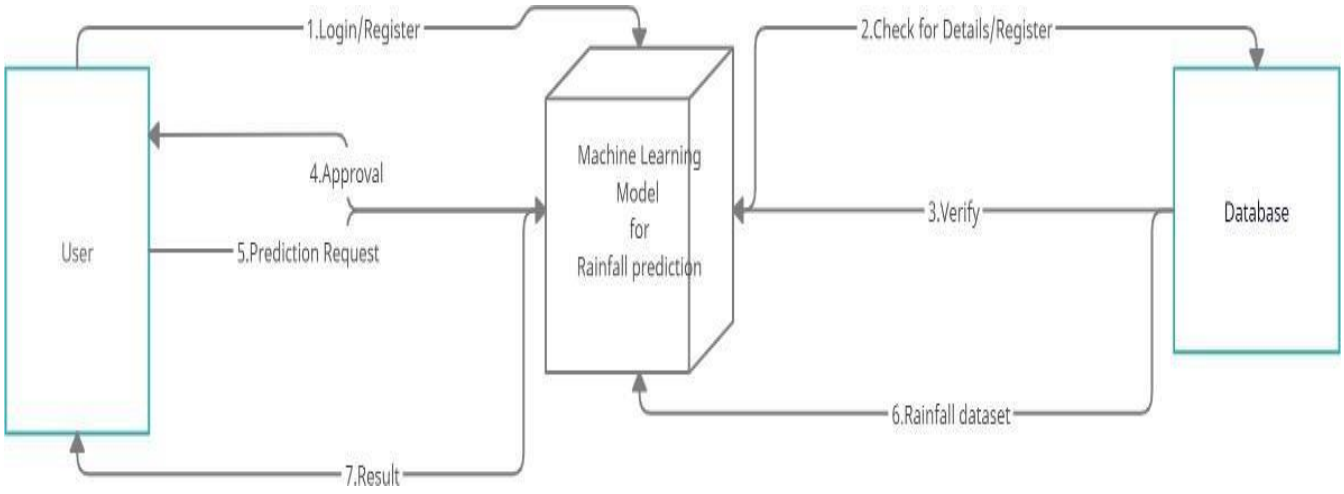
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Can be used anywhere(remote villages to metropolitan cities), anybody (kids to old age)
NFR-2	**Security**	Security is given over the model, so the user can use this with full trust. However, there are no personaldetails required to use 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.
NFR-6	**Scalability**	Farmers, Vegetable sellers, citizens can use this, prediction of data is accurate.

5. PROJECT DESIGN

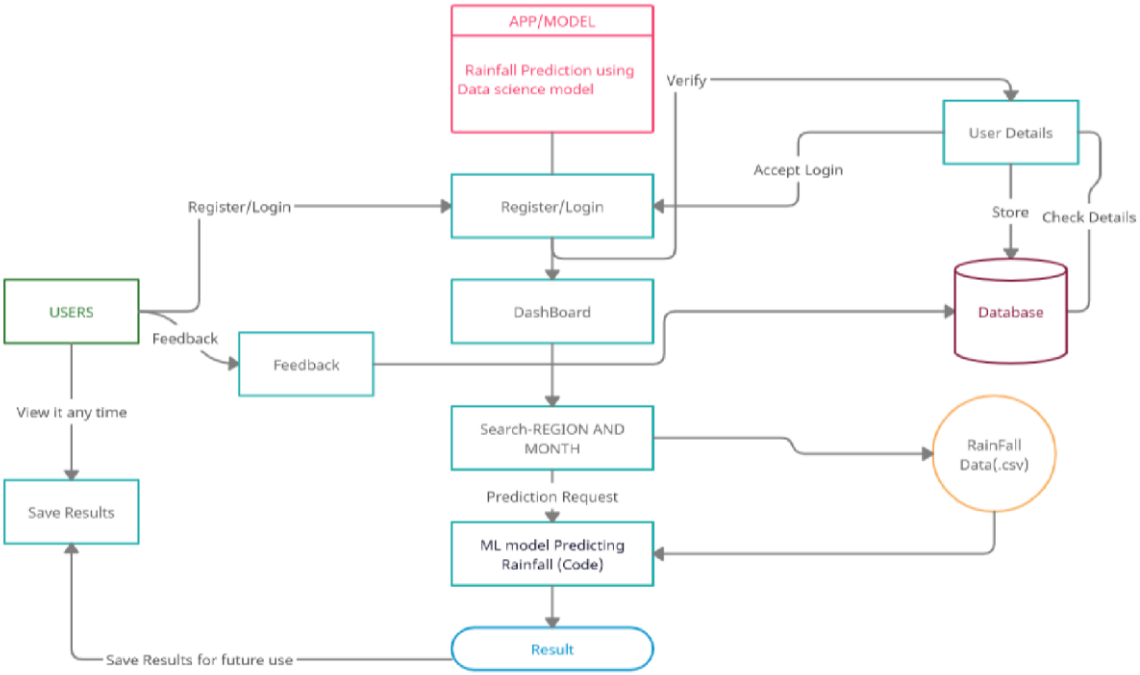
5.1 Data flow diagrams

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the rightamount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

0-LEVEL DATA FLOW DIAGRAM



2-LEVEL DATA FLOW DIAGRAM



5.2 Solution & Technical Architecture

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2

Technology architecture associates application components from application architecture with technology components representing software and hardware components.Its components are generally acquired in the marketplace and can be assembled and configured to constitute the enterprise's technological infrastructure.

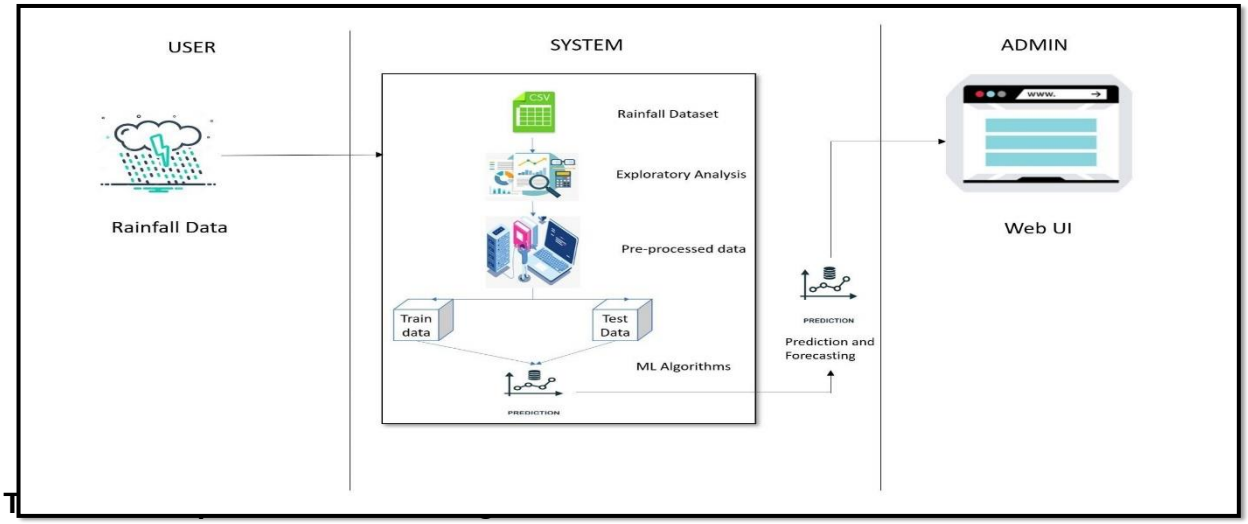


Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g.Web UI, Mobile App, Chatbot etc.	HTML, CSS, JavaScript / Angular Js /React Js etc.
2.	Application Logic-1	Logic for a process in the application	Java / Python
3.	Application Logic-2	Logic for a process in the application	IBM Watson STT service
4.	Application Logic-3	Logic for a process in the application	IBM Watson Assistant
5.	Database	Data Type, Configurations etc.	MySQL, NoSQL, etc.
6.	Cloud Database	Database Service on Cloud	IBM DB2, IBM Cloudant etc.

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Backend framework, CSS styling framework, relational database	PyJWT, Flask, IBM cloud DB
2.	Security Implementations	Request authentication using JWT tokens	HS-256, Encryptions , SSL certs
3.	Scalable Architecture	Support for multiple sample prediction using Excelfile	Pandas , Numpy
4.	Availability	Availability is increased by distributed servers incloud VPS	IBM cloud hosting
5.	Performance	The application is expected to handle multiplepredictions per second	Load balancers , distributed servers

5.3 User Stories

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirmingmy password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation emailonce I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-2

	Login	USN-3	As a user, I can log into the application through my registered email and password.	I can access the dashboard of the system.	High	Sprint-1
		USN-4	User can change their password and can viewtheir search history.	Verification is required and new password should be entered.	HIgh	Sprint-1
		USN-5	The existing credentials should be used forlogin or multiple systems.		Medium	Sprint-1
	Dashboard	USN-6	As a user , I can view the details about the page and navigate through the entire pages.	I can navigate through the pages.	Medium	Sprint-1
	Prediction	USN-7	User can search for the area / place where the user wants to know the prediction of rainfall .	Searching for the region within INDIA only be accepted.	High	Sprint-1
		USN-8	The prediction or analysis for the desired region for the future or past events respective ly.		High	Sprint-1
		USN-9	User can see the visualization of the rainfall data for the specific region in INDIA for a specified time period.		High	Sprint-1
	News	USN-10	User can view the latest news articles relatedto agriculture.	I can view the news article s.	Medium	Sprint-2
Customer care executive	Support	USN-11	User can ask queries about the system.	I can rectify my doubts	High	Sprint-3
		USN-12	The team must analyse all the queries anddebug it in the next update.		High	Sprint-3
		USN-13	Organize for a FAQ session.		Low	Sprint-3

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Delivery

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	Sreerag M,Reena J
Sprint-1		USN-2	Train Model using Different machine learningAlgorithms	5	High	Rajalakshmi R,Sudharsan L
Sprint-1		USN-3	Test the model and give best	10	High	Reena J ,Rajalakshmi R
Sprint-2	Registration	USN-4	As a user, they can register for the application through Gmail. Password is set up.	5	Medium	Rajalakshmi R,Sreerag M
Sprint-2	Login	USN-5	As a user, they can log into the application byentering email & password	5	Medium	Sudharsan L, Reena J
Sprint-2		USN-6	Credentials should be used for multiple systems and verified	4	Medium	Reena J, Sreerag M,Rajalakshmi R
Sprint-2	Dashboard	USN-7	Attractive dashboard forecasting live weather	6	Low	Sreerag M ,Reena J,Sudharsan L
Sprint-3	Rainfall Prediction	USN-8	User enter the location, temperature, humidity	10	High	Rajalakshmi R,Sudharsan L
Sprint-3		USN-9	Predict the rainfall and display the result	10	High	Reena J,Sreerag M
Sprint-4	Testing	USN-10	Test the application	10	High	Sudharsan L,Sreerag M
Sprint-4	Deploy Model	USN-11	Deploy the model in IBM cloud to make user friendly application	10	High	Reena J,Rajalakshmi R

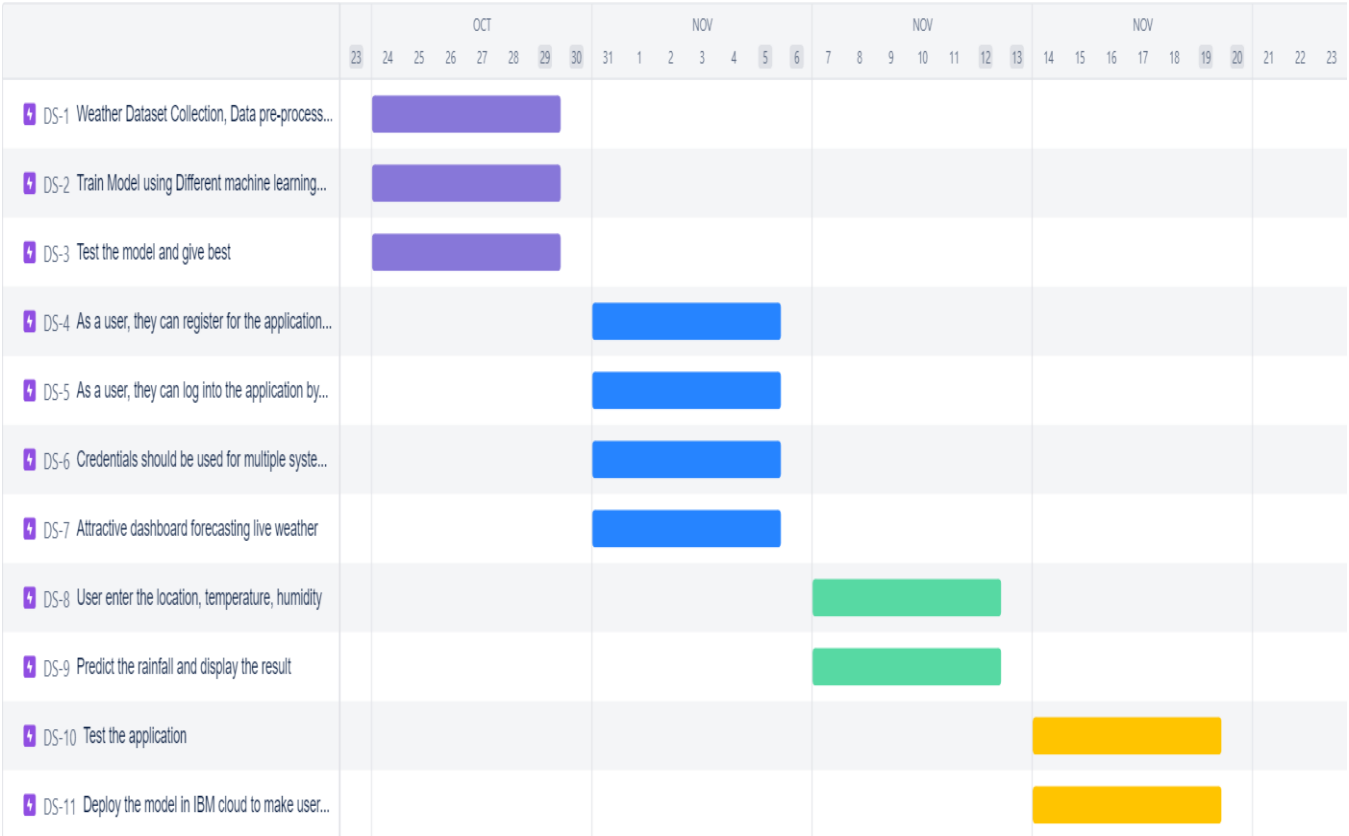
Project Tracker &Velocity

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

6.2 Reports from JIRA

Burndown Chart

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies suchas Scrum. However, burn down charts can be applied to any project containing measurable progress over time



7. CODING & SOLUTIONING

7.1 Feature 1

App.py

```
import numpy as np

import pickle

import joblib
```

```
import matplotlib
import matplotlib.pyplot as plt
import time
import pandas
import os
from sklearn import *
from flask import Flask,request,jsonify,render_template,redirect,url_for
app = Flask(__name__, static_folder='static')
model = pickle.load(open("./rainfall.pkl","rb"))
scale= pickle.load(open("./scale.pkl","rb"))
encoder = pickle.load(open("encoder.pkl","rb"))
@app.route('/')
def home():
  return render_template('index.html')
@app.route('/pred',methods=["POST","GET"])
def pred():
  inp_feature = [x for x in request.form.values()]
  inp_feature=inp_feature[:18]
  print(inp_feature)
  feature_values = [np.array(inp_feature)]
  names = [['Location', 'MinTemp', 'MaxTemp', 'Rainfall', 'WindGustSpeed',
        'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
        'Pressure9am', 'Pressure3pm', 'Temp9am', 'Temp3pm', 'risk', 'RainToday',
        'WindGustDir', 'WindDir9am', 'WindDir3pm']]
  data = pandas.DataFrame(feature_values,columns=names)
  data = scale.fit_transform(data)
  print(data)
  data = pandas.DataFrame(data,columns=names)
  print(data)
  prediction = model.predict(data)
```

```
    if prediction == "Yes":
        return render_template("predict1.html")
    else:
        return render_template("predict2.html")
if __name__ == '__main__':
    app.run(debug= True)
```

Importing Libraries

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import re
import os
import collections
import seaborn as sns
import plotly.express as px
import warnings
warnings.filterwarnings('ignore')
!pip3 install openpyxl
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: openpyxl in /usr/local/lib/python3.7/dist-packages (3.0.10)
Requirement already satisfied: et-xmlfile in /usr/local/lib/python3.7/dist-packages (from openpyxl) (1.1.0)
```

Importing The Dataset

In [3]:
```
df=pd.read_csv('rainfall in india 1901-2015.csv')
df1=pd.read_csv('district wise rainfall normal.csv')
```

Handling Missing Values

In [11]:
```
df.isnull().any()
```

Out[11]:
```
SUBDIVISION    False
YEAR           False
JAN             True
FEB             True
MAR             True
APR             True
MAY             True
JUN             True
JUL             True
AUG             True
SEP             True
OCT             True
NOV             True
DEC             True
ANNUAL          True
Jan-Feb         True
Mar-May         True
Jun-Sep         True
Oct-Dec         True
dtype: bool
```

There is an missing values in dataset of "rainfall in india 1901=2015.csv"

In [12]:
```
df1.isnull().any()
```

Out[12]:
```
STATE_UT_NAME    False
DISTRICT         False
JAN              False
FEB              False
MAR              False
APR              False
MAY              False
JUN              False
JUL              False
AUG              False
SEP              False
OCT              False
NOV              False
DEC              False
ANNUAL           False
Jan-Feb          False
Mar-May          False
Jun-Sep          False
Oct-Dec          False
dtype: bool
```

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
SUBDIVISION     0
YEAR            0
JAN             4
FEB             3
MAR             6
APR             4
MAY             3
JUN             5
JUL             7
AUG             4
SEP             6
OCT             7
NOV            11
DEC            10
ANNUAL         26
Jan-Feb         6
Mar-May         9
Jun-Sep        10
Oct-Dec        13
```

In [10]:
```
df1.isnull().sum
```

Out[10]:
```
.sum of      STATE_UT_NAME  DISTRICT    JAN    FEB    MAR    APR    MAY    JUN    JUL  \
0           False     False  False  False  False  False  False  False  False
1           False     False  False  False  False  False  False  False  False
2           False     False  False  False  False  False  False  False  False
3           False     False  False  False  False  False  False  False  False
4           False     False  False  False  False  False  False  False  False
..            ...       ...    ...    ...    ...    ...    ...    ...    ...
636         False     False  False  False  False  False  False  False  False
637         False     False  False  False  False  False  False  False  False
638         False     False  False  False  False  False  False  False  False
639         False     False  False  False  False  False  False  False  False
640         False     False  False  False  False  False  False  False  False

       AUG    SEP    OCT    NOV    DEC  ANNUAL  Jan-Feb  Mar-May  Jun-Sep  \
0    False  False  False  False  False   False    False    False    False
1    False  False  False  False  False   False    False    False    False
2    False  False  False  False  False   False    False    False    False
3    False  False  False  False  False   False    False    False    False
4    False  False  False  False  False   False    False    False    False
..     ...    ...    ...    ...    ...     ...      ...      ...      ...
636  False  False  False  False  False   False    False    False    False
637  False  False  False  False  False   False    False    False    False
638  False  False  False  False  False   False    False    False    False
639  False  False  False  False  False   False    False    False    False
640  False  False  False  False  False   False    False    False    False

     Oct-Dec
0      False
1      False
2      False
3      False
4      False
..       ...
636    False
637    False
638    False
639    False
640    False

[641 rows x 19 columns]>
```

7.2 Feature 2

In [16]:
```
df.describe()
```

Out[16]:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
count	4090.000000	4090.000000	4090.000000	4090.000000	4090.000000	4090.000000	4090.000000	4090.000000	4090.000000	4090.000000	4090.000000	4090.000000	4090.000000
mean	1958.321271	18.818484	21.644792	27.252494	42.714548	84.868044	228.928020	346.496968	289.897506	197.003056	95.139022	39.548191	18.695575
std	33.148944	33.521719	35.762010	46.829179	67.264863	122.556801	233.535693	269.352685	187.702293	135.266708	99.325638	68.275513	42.185553
min	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.400000	0.000000	0.000000	0.100000	0.000000	0.000000	0.000000
25%	1930.000000	0.600000	0.600000	1.000000	3.000000	8.500000	70.000000	175.225000	155.925000	100.400000	14.600000	0.700000	0.100000
50%	1959.000000	5.900000	6.600000	7.800000	15.500000	36.050000	138.450000	284.300000	259.500000	173.600000	64.650000	9.500000	3.000000
75%	1987.000000	21.950000	26.600000	31.100000	49.375000	94.975000	302.250000	416.700000	377.775000	265.600000	148.100000	45.475000	17.300000
max	2015.000000	583.700000	403.500000	605.600000	595.100000	1168.600000	1609.900000	2362.800000	1664.600000	1222.000000	948.300000	648.900000	617.500000

In [18]:
```
df1.describe()
```

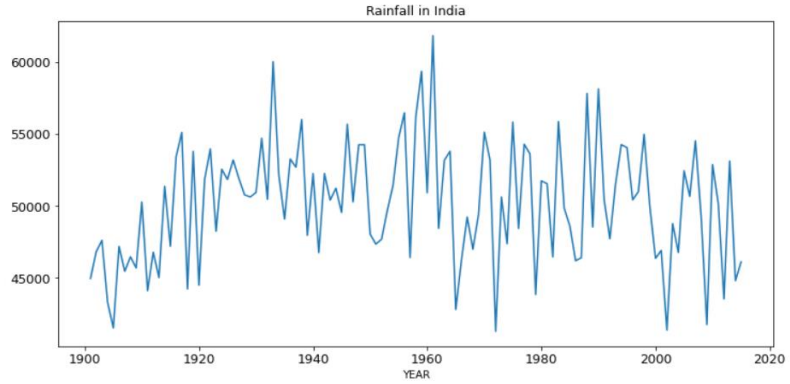
Out[18]:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb
count	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000
mean	18.355070	20.984399	30.034789	45.543214	81.535101	196.007332	326.033697	291.152262	194.609048	90.446334	34.117473	18.150858	1346.969579	39.339470
std	21.082806	27.729596	45.451082	71.556279	111.960390	196.556284	221.364643	152.647325	99.830540	74.990685	59.371274	32.711009	838.878874	47.212773
min	0.000000	0.000000	0.000000	0.000000	0.900000	3.800000	11.600000	14.100000	8.600000	3.100000	1.200000	0.000000	94.600000	0.000000
25%	6.900000	7.000000	7.000000	5.000000	12.100000	68.800000	206.400000	194.600000	128.800000	34.300000	6.600000	5.300000	830.400000	14.700000
50%	13.300000	12.300000	12.700000	15.100000	33.900000	131.900000	293.700000	284.800000	181.300000	62.600000	12.900000	7.900000	1116.200000	27.700000
75%	19.200000	24.100000	33.200000	48.300000	91.900000	226.600000	374.800000	358.100000	234.100000	130.200000	32.300000	14.900000	1530.900000	41.100000
max	144.500000	229.600000	367.900000	554.400000	733.700000	1476.200000	1820.900000	1522.100000	826.300000	517.700000	475.100000	297.700000	7229.300000	335.300000

Data Visualization

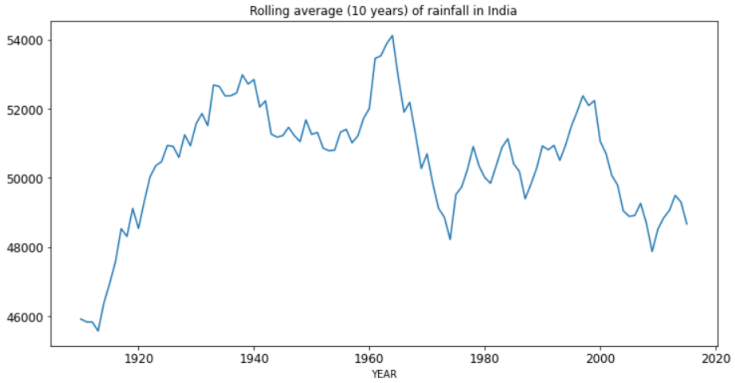
In [33]:
```
(df.groupby(by='YEAR')[['ANNUAL']]
 .sum()
 .plot(figsize=(12, 6), title='Rainfall in India', fontsize=12, legend=False))
```

Out[33]:



In [34]:
```
(df.groupby(by='YEAR')[['ANNUAL']]
 .sum()
 .rolling(10)
 .mean()
 .plot(figsize=(12, 6), title='Rolling average (10 years) of rainfall in India', fontsize=12, legend=False)
)
```

Out[34]:



In [37]:
```
drop_col = ['ANNUAL','Jan-Feb','Mar-May','Jun-Sep','Oct-Dec']

fig, ax = plt.subplots()

(df.groupby(by='YEAR')
 .sum()
 .drop(drop_col, axis=1)
 .T
 .plot(alpha=0.1, figsize=(12, 6), legend=False, fontsize=12, ax=ax)
)

ax.set_xlabel('Months', fontsize=12)
ax.set_ylabel('Rainfall (in mm)', fontsize=12)
```

Out[37]: Text(0, 0.5, 'Rainfall (in mm)')

7.3 DATASET

A data set is a collection of related, discrete items of related data that may be accessed individually or in combination or managed as a whole entity**.** The database itself can be considered a data set, as can bodies of data within it related to a particular type of information, such as sales data for a particular corporate department.

Importing Libraries

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import re
import os
import collections
import seaborn as sns
import plotly.express as px
import warnings
warnings.filterwarnings('ignore')
!pip3 install openpyxl
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: openpyxl in /usr/local/lib/python3.7/dist-packages (3.0.10)
Requirement already satisfied: et-xmlfile in /usr/local/lib/python3.7/dist-packages (from openpyxl) (1.1.0)
```

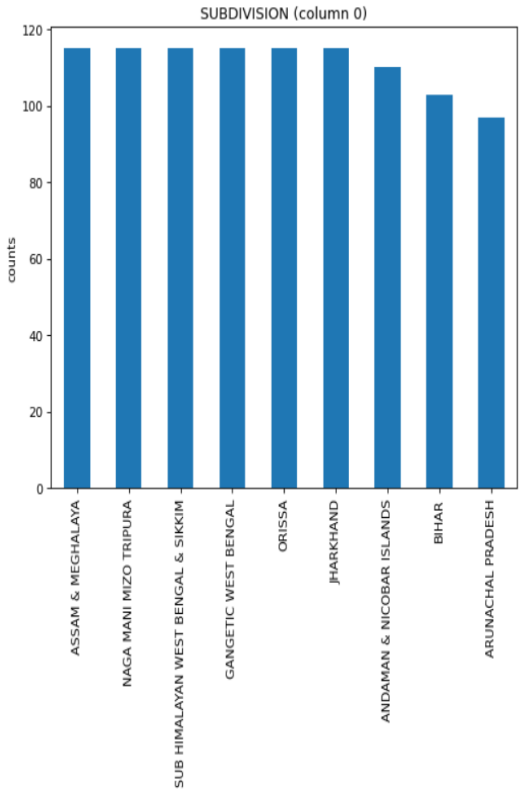
Importing The Dataset

In [3]:
```
df=pd.read_csv('rainfall in india 1901-2015.csv')
df1=pd.read_csv('district wise rainfall normal.csv')
```

In [5]:
```
df.head()
```

Out[5]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.6	3373.2	136.3	560.3	1696.3	980.3
1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2	359.0	160.5	3520.7	159.8	458.3	2185.9	716.7
2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2	284.4	225.0	2957.4	156.7	236.1	1874.0	690.6
3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2	308.7	40.1	3079.6	24.1	506.9	1977.6	571.0
4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	260.7	25.4	344.7	2566.7	1.3	309.7	1624.9	630.8



In [68]:
```
plotCorrelationMatrix(df, 8)
```

8. TESTING

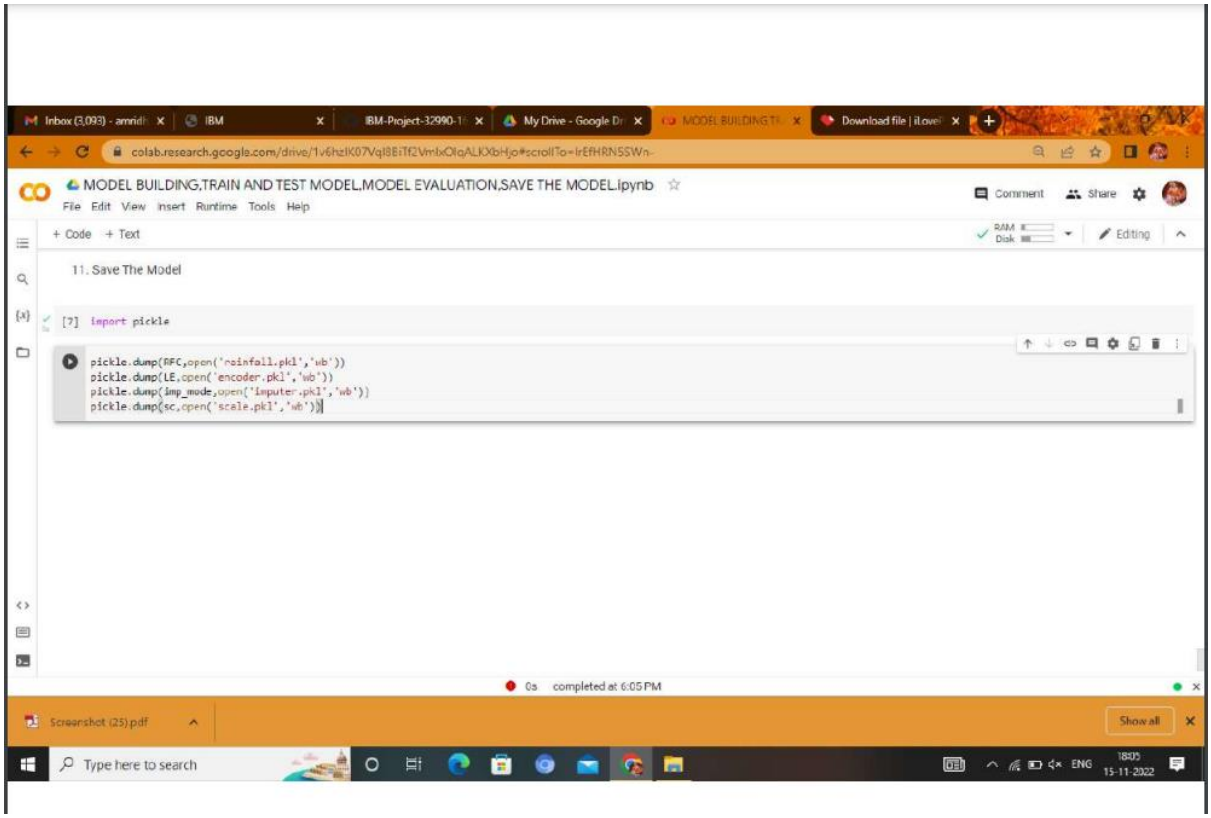
8.1 Test Cases

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	7	0	0	7
Client Application	51	0	0	51
Security	2	0	0	2
Outsource Shipping	3	0	0	3
Exception Reporting	9	0	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

8.2 **User Acceptance Test**

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	2	3	20
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	77

The Rainfall dataset split into a training and test dataset. From 1940-2010 data has given to the training phase, and for testing, from 2011 to 2017 rainfall data used. The prediction model first trained using ANN and Simple RNN, then the model was trained with an Long Short Term Memory. The performance of the built model checked with the test dataset.
In statistics, normality tests are used to determine if a data is well-modelled by a normal distribution and to compute how likely it is for a random variable underlying the data to be normally distributed. Hence assessment of the normality of data is a prerequisite for many statistical tests because normally distributed data is an underlying assumption in all the parametric testing



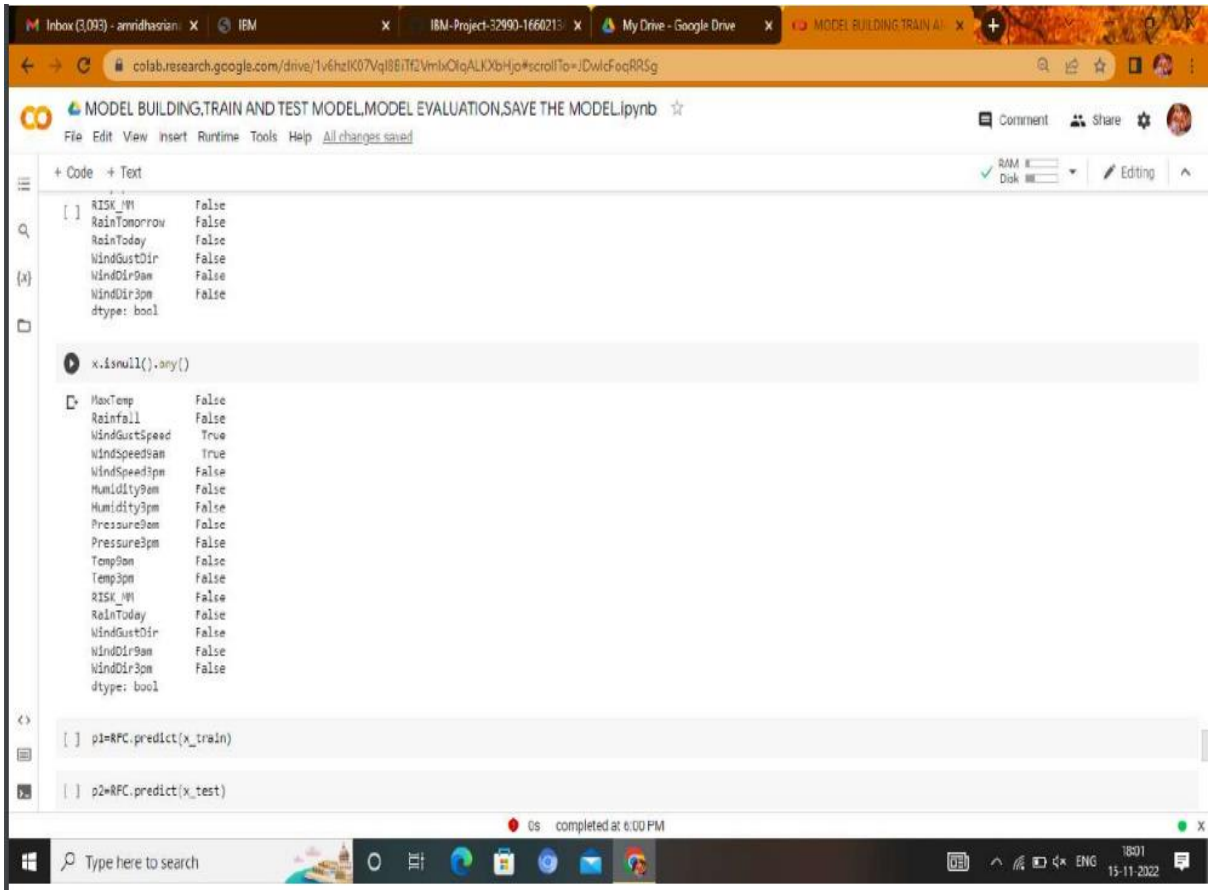
8.2 Performance Testing

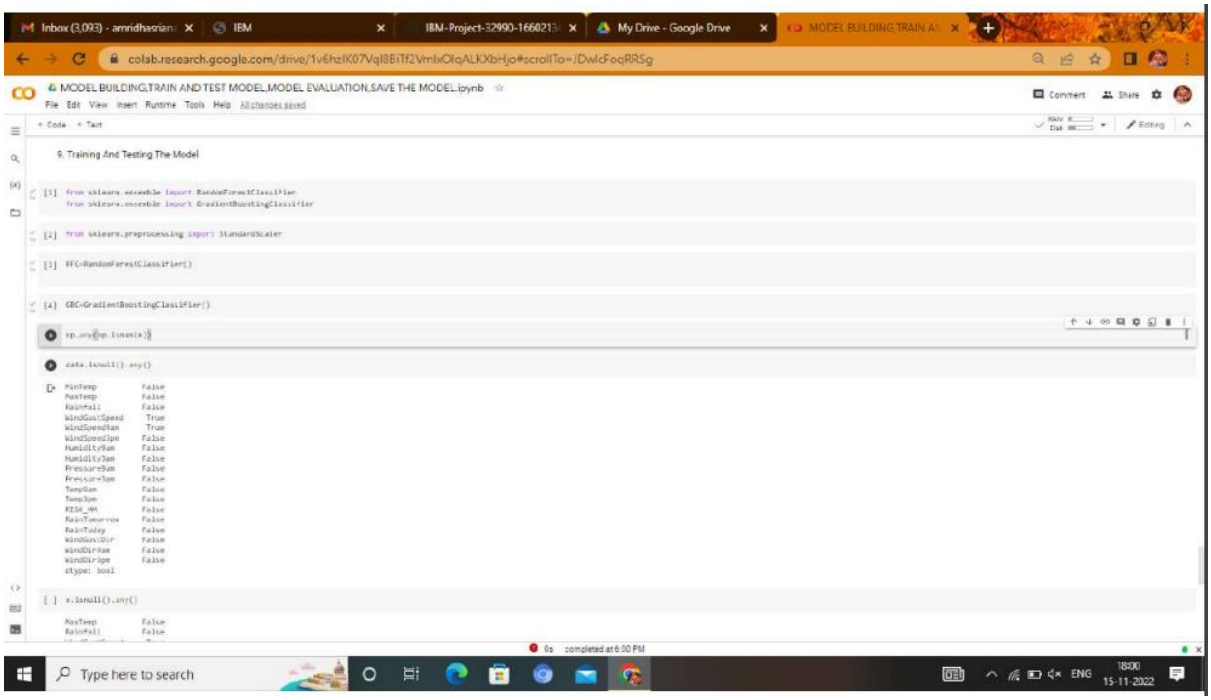
It is a change in the level of data series, usually overtime but sometimes in space. It is a general increase or decrease in the observed values of random variable over a time. In most cases, it is 4 not generally possible to detect trends that are not apparent by inspection, especially for data records of short to moderate length - say 20 years or less. Testing the existence of linear (monotonic) trend (serial correlation) within the whole time series is important in hydrometeorological datas.

Metric Testing

Testing for the existence of linear (monotonic) trend within the whole time series can be done by parametric and nonparametric methods. This test compares the CDF (cumulative distribution function) of sample data with the distribution expected if the data were normal. If the observed difference is adequately large, it will be rejected the null hypothesis of population normality.

Consistency is another desired property for any data. It checks whether or not any data within the data is reasonable. In other words, it checks if there is a surprise data (outlier) compared with the similar family of data. For example, records for rainfall within an area might be increased in three ways: records for additional time periods; records for additional sites with a fixed area; records for extra sites obtained by extending the size of the area.

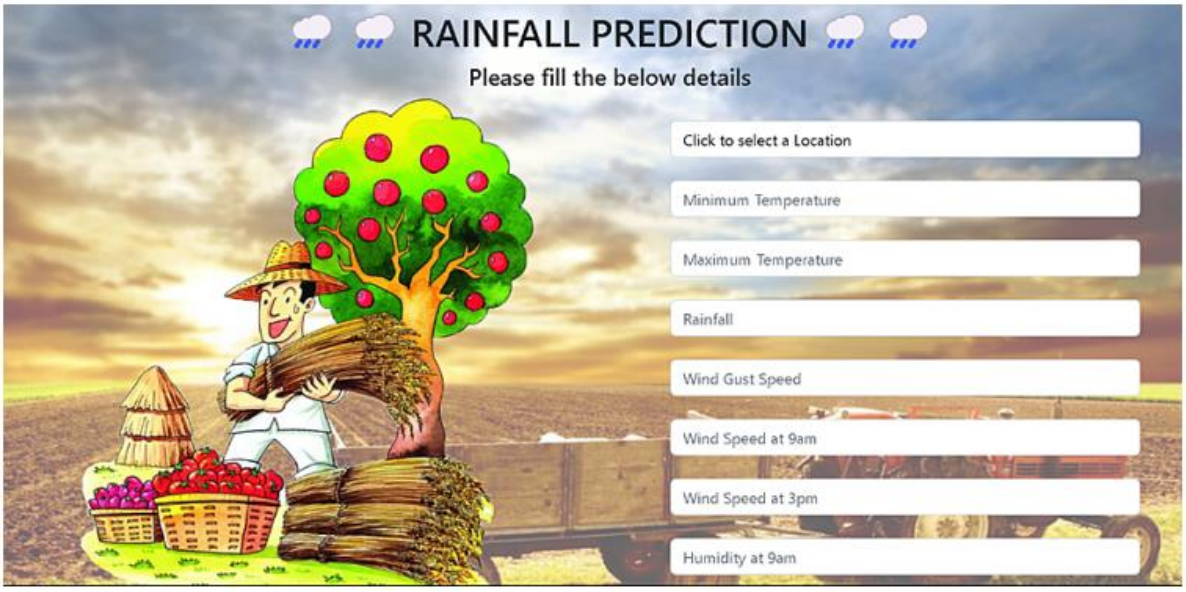




9 PERFORMANCE METRICS

Performance measurement is the process of collecting, analyzing and/or reporting information regarding the performance of an individual, group, organization, system or component . Definitions of performance measurement tend to be predicated upon an assumption about why the performance is being measured.

Classification of Rainfall Intensities	Products	CC	RMSE (mm)	MAE (mm)	NSE
Light rain	3B42V7	0.284	8.75	4.61	−9.86
	3B42RT	0.263	9.99	5.01	−13.16
	Estimates	**0.295**	**6.77**	**4.01**	**−5.45**
Moderate rain	3B42V7	0.161	17.01	13.00	−14.41
	3B42RT	0.124	20.27	14.45	−20.90
	Estimates	**0.163**	**12.63**	**10.28**	**−7.49**
Heavy rain	3B42V7	0.148	24.09	19.79	−11.95
	3B42RT	0.150	27.42	22.23	−15.79
	Estimates	**0.152**	**21.47**	**18.77**	**−9.29**
Rainstorm	3B42V7	0.541	**44.88**	**34.89**	**−0.33**
	3B42RT	0.501	47.05	37.38	−0.46
	Estimates	**0.600**	53.11	43.88	−0.86



create effective performance metrics, you must **start at the end point — with the goals, objectives, or outcomes you want to achieve — and then work backward**. A good performance metric embodies a strategic objective. It is designed to help the organization monitor whether it is on track to achieve its goals.

10.ADVANTAGES AND DISADVANTAGES

Advantages

Improve understanding of variables by extracting averages, mean, minimum, and maximum values, etc. Discover errors, outliers, and missing values in the data. Identify patterns by visualizing data in graphs such as box plots, scatter plots, and histograms. Hence, the main goal is to understand the data better and use tools effectively to gain valuable insights or draw conclusions. Exploratory research offers a great amount of researcher discretion. The lack of structure enables the researcher to direct the progression of the research processes and in that sense, it offers a greater degree of flexibility and freedom. Another pro of exploratory research is the economical way in which the process can be conducted. Exploratory research uses a relatively smaller group of people for defining and understanding the research issue. Analyzing the feasibility and viability of the research issue is another pro of exploratory research. No organization wants to invest time, effort, and resources in an area that is incapable of making value addition to the overall functioning. By carrying out an early study, exploratory research gauges the future importance that the research topic holds and accordingly directs organizational attitude. Exploratory research formulates a greater understanding of a previously unresearched topic and satisfies the researcher uncovers facts and brings new issues to light. In doing so, it helps refine the future research questions. It also helps decide the best approach to reach the objective.

Disadvantages

Exploratory research comes with its own set of cons that can act as roadblocks that impede a seamless data collection experience which lays the groundwork for future probes as well.Exploratory research brings up tentative results and so is inconclusive. The focus of such research is to grasp and formulate a better understanding of the issue at hand. These research insights cannot be relied upon for effective decision-making. Another con of exploratory research is its qualitative data and subsequent analysis. It is difficult to derive accurate insights that can be summarized in an objective manner. The variability in qualitative data itself makes the evaluation of data collected , a difficult and cumbersome process. The small sample used for exploratory research increases the risk of the sample responses being non-representative of the target audience. Smaller groups of people as samples, however useful for a quick study, can hinder a cohesive understanding which not only deteriorates the current quality of research but also adversely impacts the future research carried out along similar lines. Data, when gathered through secondary resources, can supply obsolete information which may not generate any significant contribution to the understanding of an issue in the current scenario. Outdated information is neither actionable nor supportive in offering any sort of clarity under dynamic market conditions.

11. CONCLUSION

The results show that India has two main rainfall season: one is southwest monsoon(advancing monsoon) and other is Northeast monsoon(retreating monsoon). Advancing monsoon contributes almost 80% of the rainfall. Southwest and Northeast part of India receives most of the rainfall during the advancing monsoon. During the retreating monsoon, Andaman & Nicobar Islands, Kerala, Tamil Nadu receive more rainfall as compared to other subdivisions. The trend analysis of Annual rainfall considering India as whole show

decreasing trend however when trend is analysed for all subdivision individually we can see some division showing increasing trend and some showing decreasing trend. It showed that is is import to study subdivision for better forecasting. We considered Tamil Nadu as one of the subdivisions to do further analysis. It receives more rainfall during October and November because of retreating monsoon. Since there are only a few months when the Tamil Nadu gets rains and its location at tropical results in high temperature which in turn results in water scarcity problem. Also because of its geographic location near it is hit sometimes by the cyclones formed in Indian Oceans which results in extreme storms and non normal rainfall.In an interview, Mrutyunjay Mohapatra, the director general of the IMD, explained how climate change is increasing number of days with heavy rainfall. The season started with 33% deficit rainfall but is ending with 10% higher than normal rainfall, with heavy spells of rain resulting in devastating floods in many states.He said the number of heavy rainfall days was increasing because of climate change, which was making predictions more difficult. This type of studies has not been conducted for whole India. Therefore, the present study can be the full package and should be very much helpful to the Indian planners to proposing plans for small and large scale regions.To formulate the management plan for the sustainable development of water resource based sectors and environment, the scientist of others countries can conduct the research like the present study as they need lots of information for developing plan regarding historical, present and future data which can be in any field like hydrology, climatology. However, in the present study, we considered thirty-four meteorological sub-divisions for the research, but to be more accurate, micro level data like district wise data should be incorporated. Then the very high precision micro level management plan will be achieved. Even, the grid wise rainfall study using very advanced microwave remote sensing technology will be very useful for the planners. The ensemble machine learning techniques, deep learning techniques like long-short-term memory (LSTM) network can be used to achieve very high quality forecasting data.

12. FUTURE SCOPE

Apart from predicting weather,algorithms can be used to scan satellite images to automatically derive plant count and production estimates .
Predicting weather accurately doesn’t just help our daily lives but has deeper impact for food security and disaster management. Good news for monsoon-dependent India is that we are getting better at predicting. New technologies, such as Internet of Things (IoT) and Artificial Intelligence are helping meteorological experts to give better information to predict agricultural output and natural disasters. As for future scope we can’t able to use linear regression when it comes to huge amount of data set and as its doesn’t give accurate result. So, for predicting huge volume of dataset we can develop a neural network system for more better results and accurate prediction of the weather forecasting. Also we connect analysing process to IOT technology. Because without data we can not perform analysis and prediction because IOT is major source of data. So IOT will generate data from devices which helps to take initiative to improve decision making

13. APPENDIX

SOURCE CODE

Login HTML

```
<!Doctype Html>
<Html>
<Head>
<Title>
EDA of Rainfall LOGIN!!
</Title>

<style type=text/css>
body
{
height: 125vh;
margin-top: 20px;
padding: 30px;
font-family: sans-serif;
}
</style>
</Head>
<Body>
<h1 style="color:white;">
<center> EXPLORATORY ANALYSIS OF RAIN FALL DATA IN INDIA FOR AGRICULTURE</h1> </center>
<h2 style="color:white;">
<center> <marquee> A Single Gentle Rain Makes the grass Many Shades Greener </marquee></h2>
<Title>
LOGIN PAGE
</Title>
<center><style type=text/css>
```

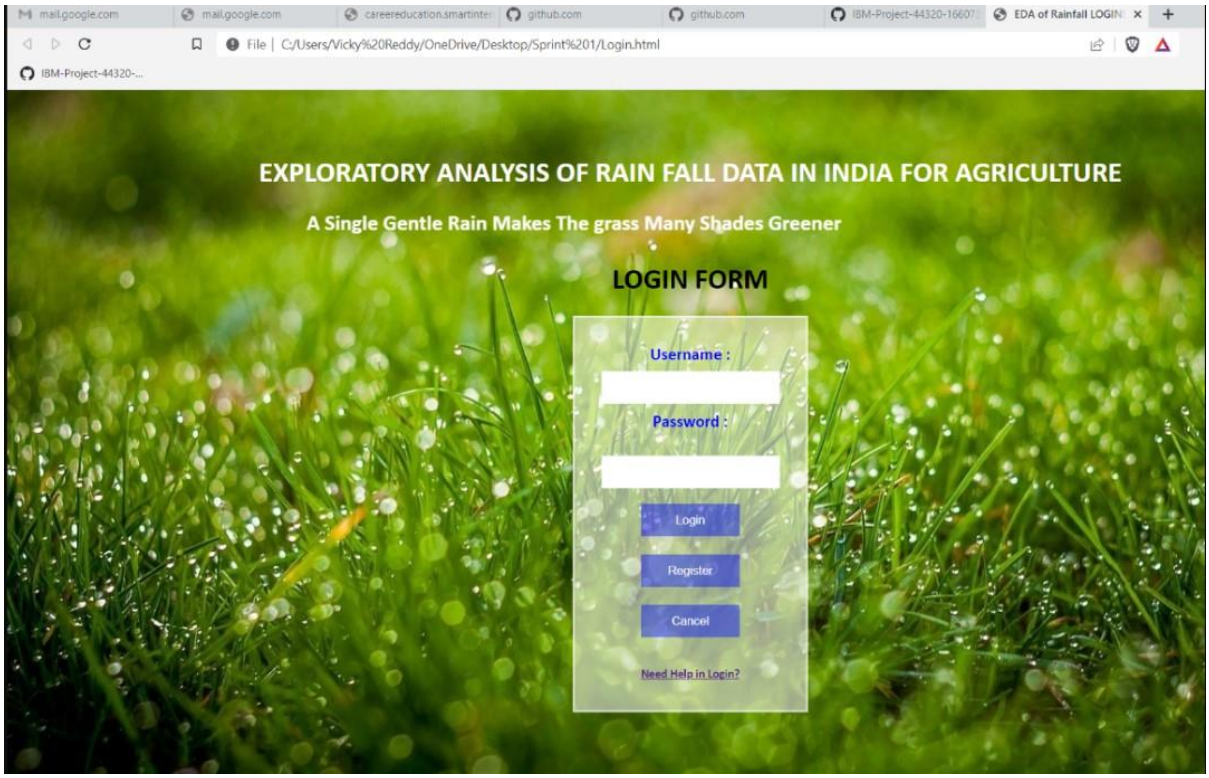
```
Body {
  font-family: Calibri, Helvetica, sans-serif;
font-size: 190,90;
background-image: url("nature-green-water_drops-leaves-grass-field.jpg");
background-position: center;
background-repeat: no-repeat;
background-attachment: fixed;
background-size: cover;
}
<style>
Body {
  font-family: Calibri, Helvetica, sans-serif;
  background-color: white;
}
button {
        background-color: rgba(0, 13, 255, 0.446);
       width: 100%;
        color: rgb(255, 255, 255);
        padding: 15px;
        margin: 10px 18px;
        border: blue;
        cursor: pointer;
         }
 form {
        border: 3px solid #ffffff8a;
        background-color: #ffffff8a;
padding: 10px 18px;
       width:50%;
    margin-left:25%;
    margin-right:25%;
```

```
    color: blue;
    }
 input[type=text], input[type=password] {
        width: auto;
        margin: 8px 0;
        padding: 10px 18px;
        display: inline-block;
        border: 2px blue;
        box-sizing: border-box;
    }
 button:hover {

padding: 10px 18px;
       width:50%;
    margin-left:25%;
    margin-right:25%;
    }
.subbtn
{
       padding: 10px 18px;
       width:50%;
    margin-left:25%;
    margin-right:25%;
}
  .cancelbtn {
       padding: 10px 18px;
       width:50%;
    margin-left:25%;
    margin-right:25%;
}
```

```
.regbtn {
        padding: 10px 18px;
       width:50%;
    margin-left:25%;
    margin-right:25%;
}
    }
.container {
        padding: 25px;
        background-image: url("rain7.jpg");
background-position: center;
background-repeat: no-repeat;
background-attachment: fixed;
background-size: cover;
    }
</style>
</head>    <center><body background="rain7.jpeg"></center>
<center><style type=text/css>
Body {
  font-family: Calibri, Helvetica, sans-serif;
font-size: 1000,1000;
}
}
<style>
</style>
</head>
<body>
    <center> <h1>  LOGIN FORM </h1> </center>
    <form style="margin: auto; width: 220px;">
        <div class="container">
```

```
        <h3> <label>Username : </label>
         <input type="text"  name="username" required><br>
         <label>Password : </label>   <h3>
         <input type="password"  name="password" required>  <br>


<button type="button" class="subbtn"id="login">Login</button>
<a href="ibmregister.html">
<a href="./ibmregister.html"><button type="button"
class="regbtn"id="register">Register</button></a>
<button type="button" class="cancelbtn"> Cancel</button>
<br>
         <h5 style="color:blue;">
         <a href="#"> Need Help in Login? </a>
      </div>
   </form>
</body>
</html>
</p>
</Body>
</Html>
```

OUTPUT



Registration HTML

```
<!DOCTYPE html>
<html>
<head>
<meta name="viewport" content="width=device-width, initial-scale=1">
<title> REGISTER HERE!!</title>
<center><body background="rain7.jpg"></center>
<center><style type=text/css>
Body {
  font-family: Calibri, Helvetica, sans-serif;
font-size: 30,90;

background-image: url("nature.jpg");
```

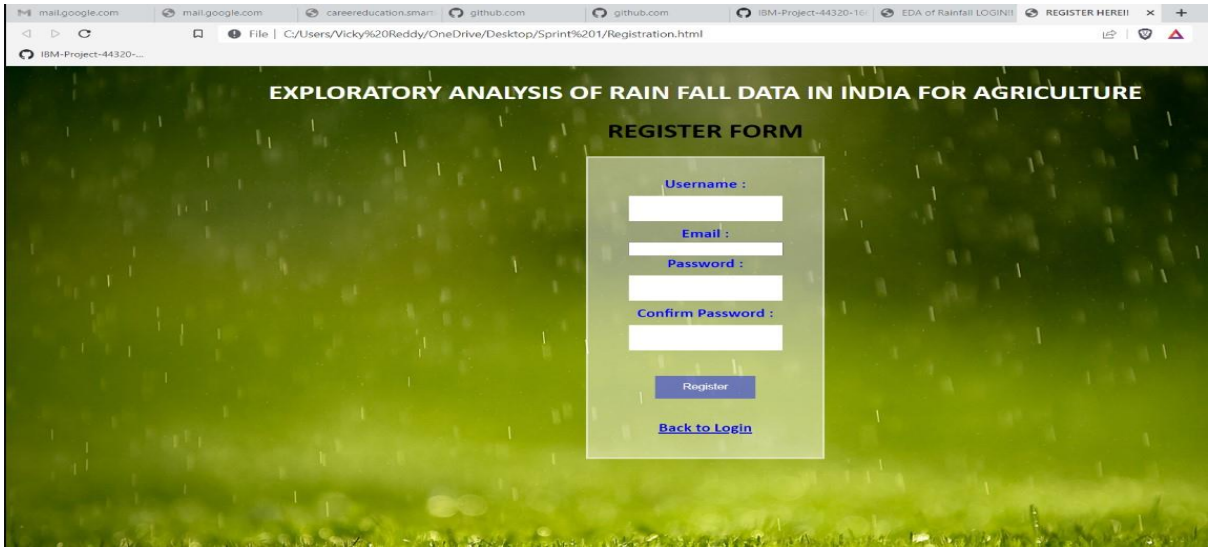
```
background-position: center;
background-repeat: no-repeat;
background-attachment: fixed;
background-size: cover;
}
}
<style>

Body {
  font-family: Calibri, Helvetica, sans-serif;
  background-color: mediumturquoise;
}
button {
        background-color: rgba(0, 13, 255, 0.446);
       width: 100%;
        color: white;
        padding: 15px;
        margin: 10px 10px;
        border: none;
        cursor: pointer;
         }
 form {
        border: 3px solid #ffffff78;
        background-color: #ffffff78;
padding: 10px 18px;
       width:50%;
    margin-left:25%;
    margin-right:25%;
    color: blue;
    }
```

```
 input[type=text], input[type=password] {
        width: auto;
        margin: 8px 0;
        padding: 12px 20px;
        display: inline-block;
        border: 2px black;
        box-sizing: border-box;
    }
 button:hover {

padding: 10px 18px;
       width:50%;
    margin-left:25%;
    margin-right:25%;
    }
.regbtn
{

        padding: 10px 18px;
       width:50%;
    margin-left:25%;
    margin-right:25%;
}

  .canczelbtn {

        padding: 10px 18px;
       width:50%;
    margin-left:25%;
    margin-right:25%;
```

```
}
   }


 .container {
       padding: 25px;

background-image: url("rain7.jpg");
background-position: center;
background-repeat: no-repeat;
background-attachment: fixed;
background-size: cover;
   }
</style>
</head>

<body>
  <h1 style="color:white;">
   <center> EXPLORATORY ANALYSIS OF RAIN FALL DATA IN INDIA FOR AGRICULTURE</center></h1>
 <div class="container">
       <h1>REGISTER FORM</h1>
      <form style="margin: auto; width: 220px;">
          <h3> <label>Username : </label>
          <input type="text" name="username" required size="15"><br>
           <label>Email : </label>
<input type="email" name="Email" required size="20">  <br>
<label>

<label>Password : </label>
          <input type="password"  name="password" required size="15">  <br>
```

```
        <label>Confirm Password : </label>
    <input type="password"  name="password" required size="15">  <br> <h3>
<a href="#"><button type="button" class="regbtn"id="reg">Register</button></a>
<h4><a href="ibmlogin.html"> Back to Login </a>


     </form>
    </div>
</body>
</html>
```

OUTPUT



WEATHER DASHBOARD

```
<!DOCTYPE html>
<html lang="en">

<head>
```

```
    <meta charset="utf-8">
    <title>Weather Dashboard!!</title>
    <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.4.1/css/bootstrap.min.css"
      integrity="sha384-
Vkoo8x4CGsO3+Hhxv8T/Q5PaXtkKtu6ug5TOeNV6gBiFeWPGFN9MuhOf23Q9Ifjh"
crossorigin="anonymous">
    <script src="https://kit.fontawesome.com/958828ca48.js" crossorigin="anonymous"></script>
    <link rel="stylesheet" href="style.css">
</head>

<body>
    <!--Nav Bar-->
    <nav class="navbar navbar-dark justify-content-center myNav">
      <span class="navbar-brand mb-0 h1 navSpan">WEATHER DASHBOARD</span>
    </nav>
 <!--Main container of page-->
    <main class="container col-12">
      <section class="row">

        <!--List div-->
        <div class="col-lg-3 col-md-12   listDiv">
          <div id="1Day row searchLabel">
            <label for="one-day-input ">Search for a City</label>
          </div>
          <form class="row">
            <input type="text" id="city-input" class="col-7 offset-1 "><br>
            <button id="add-city" type="submit" value="" class="col-2 searchButton"><i
                class="fas fa-search"></i></button>
          </form>
          <div id="OneDayWeather"></div>
          <ul class="list" style="list-style-type:none;"> </ul>
        </div>
```

```
            <!--Weather div-->
          <div class="col-lg-9 col-md-12   weatherDiv">

             <!--1 Day forecast div-->
            <div class="row" id="dayForecast"></div>

             <!--Heading div-->
            <div class='row'>
               <h4 class="forecast">5 Day Forecast:</h4>
            </div>

             <!--5 Day weather row-->
            <div class="row">
               <div class="col-lg-2 ml-4 fiveDay  col-md-10  offset-md-2" id="nextDay"></div>
               <div class="col-lg-2 ml-4 fiveDay  col-md-10  offset-md-2" id="dayTwo"></div>
               <div class="col-lg-2 ml-4 fiveDay  col-md-10  offset-md-2" id="dayThree"></div>
               <div class="col-lg-2 ml-4 fiveDay  col-md-10  offset-md-2" id="dayFour"></div>
               <div class="col-lg-2 ml-4 fiveDay  col-md-10  offset-md-2" id="dayFive"></div>
            </div>
          </div>
       </section>
    </main>

     <!--Scripts-->
    <script src="https://code.jquery.com/jquery-3.4.1.min.js"></script>
    <script src="https://cdnjs.cloudflare.com/ajax/libs/moment.js/2.24.0/moment.min.js"></script>
    <script src="javascript.js"></script>
</body>
</html>
```

CSS

```
.myNav {
    background-color: #343A41;
}

.navSpan {
    font-size: 35px;
}

.weatherDiv {
    background-color: #FFFFFF
}

.fiveDay {
    border: solid;
    border-color: #0060ff;
    border-width: 2px;
    margin: 10px 5px;
    padding: 4px;
    background-color: #007AFA;
    color: white;
    border-radius: 5px;
}

main {
    background-color: #F8F9FA;
}

#dayForecast {
    border-color: #F4F4F4;
    border: solid;
```

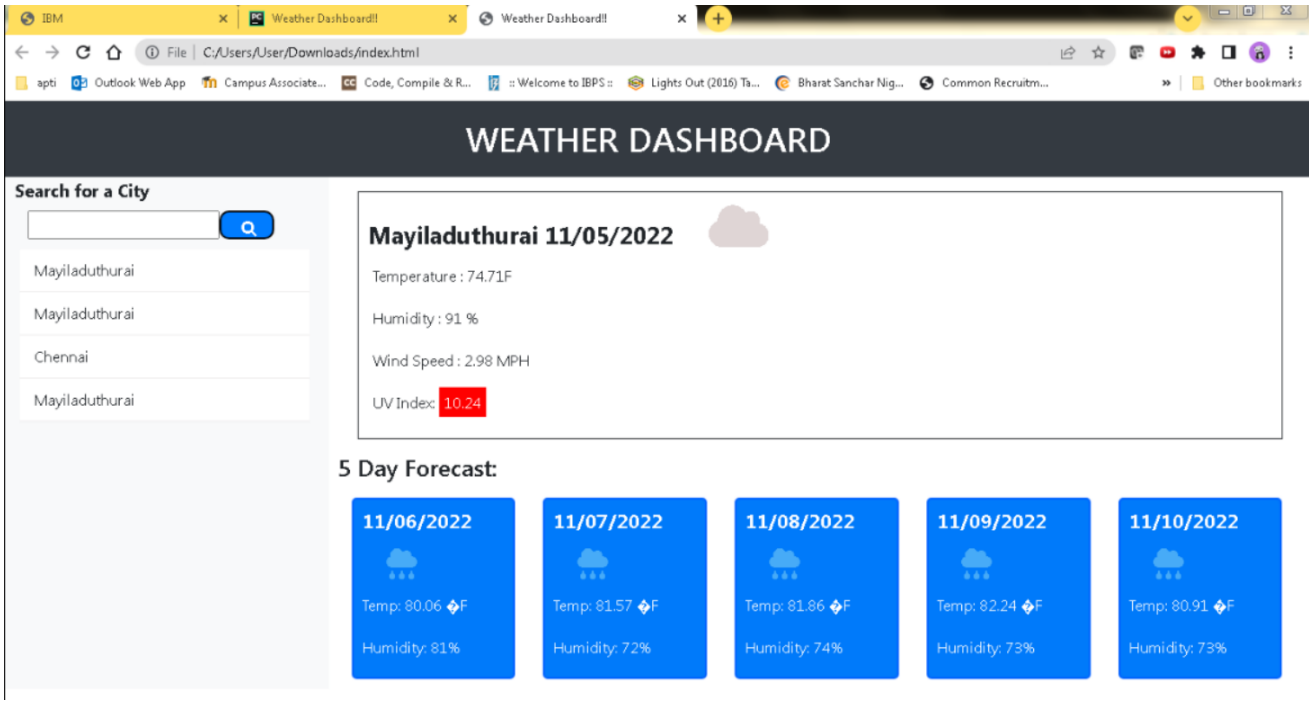
```
    border-width: 1px;
    background-color: white;
    padding: 10px;
    margin: 15px;
}

.bigger {
    font-size: 25px;
    font-weight: bold;
}

.head {
    font-weight: bold;
    font-size: 20px;
    padding: 2px 2px 5px 5px;
    margin-bottom: 10px;
}

.fa-sun {
    color: #FFAB4D !important;
}

.fa-cloud-rain {
    color: #47abf7 !important;
}

.fa-cloud {
    color: rgb(223, 213, 213);
}
```

```
.fa-smog {
    color: lightgrey;
}

.fas, .far {
    padding-left: 30px;
    padding-bottom: 10px;
}

.forecast {
    margin-left: 10px;
    ;
}

.UvIndex, .windSpeed, .humidity, .tempClass {
    margin-bottom: 10px;
    padding: 5px;
    font-size: 16px;
}

.searchButton {
    background-color: #007CFF;
    border-radius: 10px;
}

.searchButton :hover :active {
    border: solid;
    border-width: 3px;
    border-color: black;
}
```

```
label {
    font-size: 18px;
    font-weight: bold;
}

.list {
    background-color: #FFFFFF;
    margin: 10px 5px;
}

li {
    border-bottom: solid;
    border-color: rgb(240, 238, 238);
    border-width: .5px;
    margin-left: -40px;
    padding: 10px 15px;
}

li:hover {
    border: solid;
    border-color: black;
    border-width: 2px;
}

.fa-search {
    color: white;
    text-align: center;
    width: 100%;
    position: relative;
```

```
}

.fa-search::before {
    position: absolute;
    left: 20%;
}

.low {
    color: white;
    background-color: green;
    padding: 7.5px;
    border-radius: 3px;
}

.high {
    color: white;
    background-color: red;
    padding: 5px;
}

.medium {
    color: white;
    background-color: #fbc02d;
    padding: 5px;
    border-radius: 1.5px;
}

.icon {
    background-color: white;
}
```

OUTPUT



GITHUB LINK

IBM-EPBL/**IBM-Project-34328-1660234262**

PROJECT DEMO LINK

https://drive.google.com/file/d/1vNcH9pLlXlhCGlwrj-qlg9q_P7p479Ix/view?usp=share_link

TEAM MEMBERS

1) **SUDHARSAN L**
2) **SREERAG M**
3) **REENA J**
4) **RAJALAKSHMI R**