

IBM-NALAIYA THIRAN

Exploratory Analysis of Rainfall Data in India for Agriculture *Machine Learning*

Submitted by

1. Roselin Janice NR
2. Nithishree N
3. Prasanna Venkatesan S
4. Karunakaran T

TEAM ID - PNT2022TMID08012

Department of Computer Science And Engineering, Adhiyamaan College Of
Engineering - 635 109.

Institution Mentor: Gopika Krishnan G S

Industrial Mentor: Lalitha Gayathri

TABLE OF CONTENTS

1.INTRODUCTION

1.1 Project Overview

1.2 Purpose

2. LITERATURE SURVEY

2.1 Existing problem

2.2 References

2.3 Problem Statement Definition

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

3.2 Ideation & Brainstorming

3.3 Proposed Solution

3.4 Problem Solution fit

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

4.2 Non-Functional requirements

5. PROJECT DESIGN

5.1 Data Flow Diagrams

5.2 Solution & Technical Architecture

5.3 User Stories

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

6.2 Sprint Delivery Schedule

6.3 Reports from JIRA

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

7.1 Feature 1

7.2 Feature 2

7.3 Database Schema (if Applicable)

8. TESTING

8.1 Test Cases

8.2 User Acceptance Testing

9. RESULTS

9.1 Performance Metrics

10. ADVANTAGES & DISADVANTAGES

11. CONCLUSION

12. FUTURE SCOPE

13. APPENDIX

Source Code

GitHub & Project Demo Link

1.INTRODUCTION

1.1 Project Overview

Rainfall prediction remains a serious concern and has attracted the attention of governments, industries, risk management entities, as well as the scientific community. Rainfall is a climatic factor that affects many human activities like agricultural production, construction, power generation, forestry and tourism, among others. To this extent, rainfall prediction is essential since this variable is the one with the highest correlation with adverse natural events such as landslides, flooding, mass movements and avalanches. These incidents have affected society for years. Rainfall also determines how fast a crop will grow from seed, including when it will be ready for harvesting. A good balance of rain and proper irrigation can lead to faster-growing plants, which can cut down on germination time and the length between seeding and harvest. Therefore, having an appropriate approach for rainfall prediction makes it possible to do agricultural processes in a calculated way. To solve this uncertainty, we used various machine learning techniques and models to make accurate and timely predictions. This paper aims to provide an end to end machine learning life cycle right from Data preprocessing to implementing models to evaluating them. Data Preprocessing steps include imputing missing values, feature transformation, encoding categorical features, feature scaling and feature selection. We implemented models such as Logistic Regression, Decision Tree, K Nearest Neighbour and Random Forest Classifier. For evaluation purposes, we used Accuracy, Precision, Recall, F-Score and Area Under Curve as evaluation metrics. For our experiments, we train our classifiers using Australian weather data gathered from various weather stations in Australia.

1.2 Purpose

- The main purpose of this project work is to find the best prediction model i.e.the best machine learning technique which will accurately predict the rainfall which will be useful in the agricultural process.
- Determining the rainfall for effective use of water resources, crop productivity, and pre-planning of water structures.
- Comparing with other algorithms and building the prediction model with the most efficient algorithm among them.

2. LITERATURE SURVEY

2.1 Existing problem

The rainfall has been predicted with the help of deep learning algorithms. Two deep learning techniques which were used are Multilayer Perceptron and Linear Regression. Although some of the approaches showed good performances in predicting rainfall, most approaches do not provide any transparent reasons behind predicted outcomes.

2.2 References

- V. Brahmananda Rao ,K. Hada 1994: An experiment with linear regression in forecasting of spring rainfall over south Brazil
- K. Hrona_, P. Filzmoserb and K. Thompsonc 2009 : Linear regression with compositional explanatory variables.
- A. Bardossy and E. J. Plate. Space-time model for daily rainfall using atmospheric circulation patterns. Water Resources Research, 28(5):1247–1259,
- S. P. Charles, B. C. Bates, I. N. Smith, and J. P. Hughes. Space-time model for daily rainfall using atmospheric circulation patterns. Hydrological Processes, 18:1373–1394, 2004.

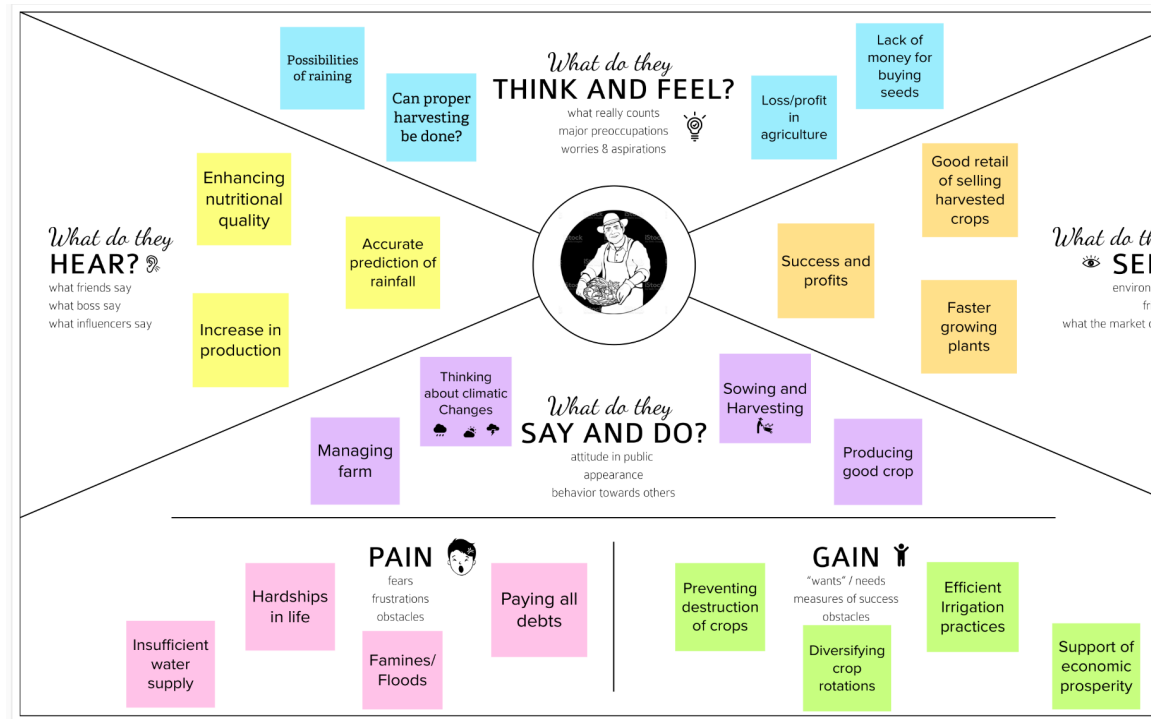
2.3 Problem Statement Definition

Rainfall is one of the most complex and difficult elements of the hydrology cycle to understand and to model due to the complexity of the atmospheric processes that generate rainfall and the tremendous range of variation over a wide range of scales both in space and time. Heavy rainfall prediction is a major problem for meteorological department as it is closely associated with the economy and life of human. It is a cause for natural disasters like flood and drought which are encountered by people across the globe every year. Accuracy of rainfall forecasting has great importance for countries like India whose economy is largely dependent on agriculture. Due to dynamic nature of atmosphere, Statistical techniques fail to provide good accuracy for rainfall forecasting. Thus, accurate rainfall prediction is one of the greatest challenges in operational hydrology. On a worldwide scale, large numbers of attempts have been made by different researchers to predict rainfall accurately using various techniques. But due to the nonlinear nature of rainfall, prediction accuracy obtained by these techniques is still below the satisfactory level.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviours and attitudes. It is a useful tool to help teams better understand their users. Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges.



3.2 Ideation & Brainstorming

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

Step-1: Team Gathering, Collaboration and Select the Problem Statement In this step team members gather and provide their ideas and collaborate those ideas and select their problem statement. The ideas should be relevant to their problem statement.

1

Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

🕒 5 minutes

PROBLEM

The main aim of our application is to predict rainfall which will help farmers to strategize farming.



Key rules of brainstorming

To run a smooth and productive session



Stay in topic.



Encourage wild ideas.



Defer judgment.



Listen to others.

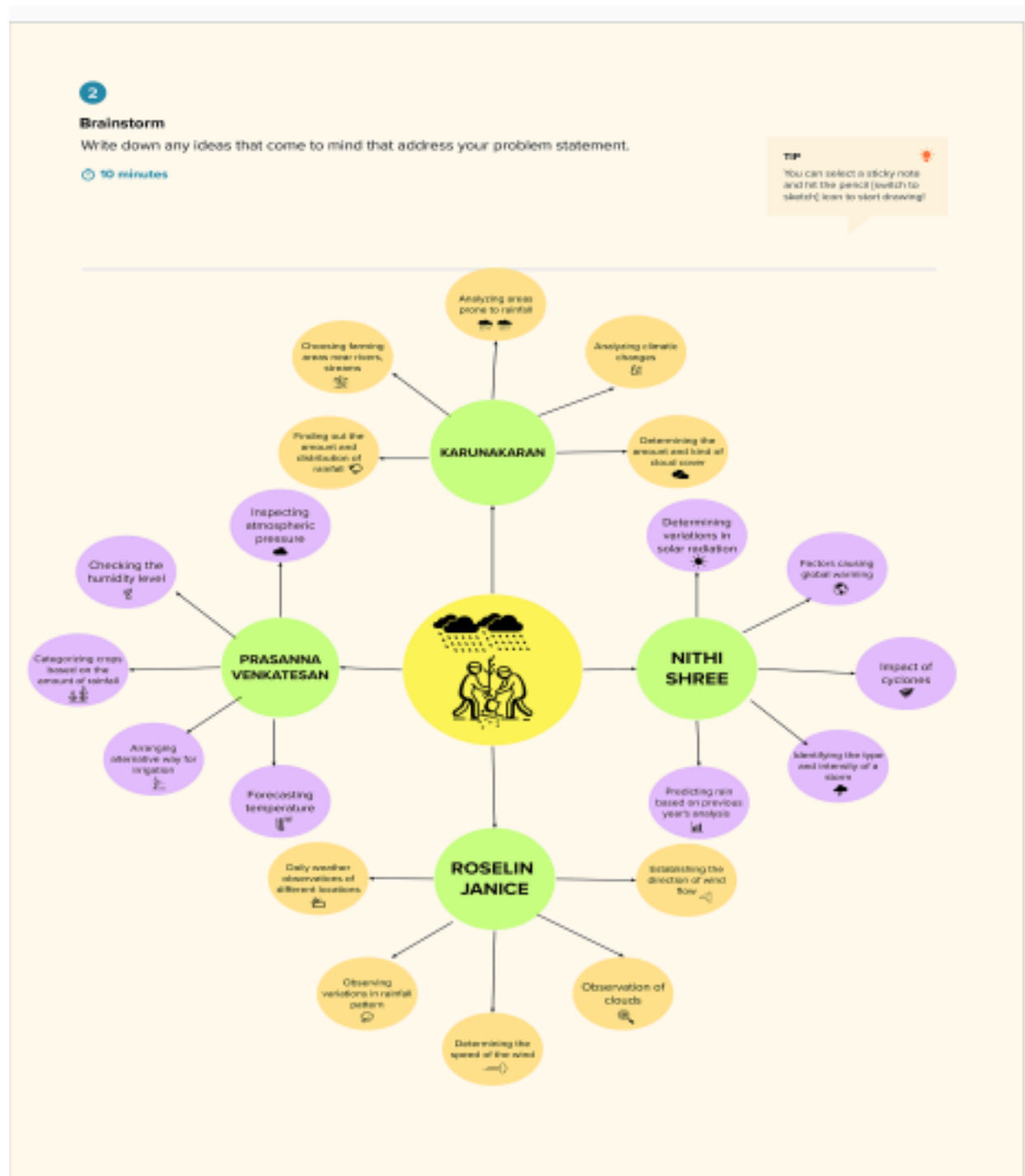


Go for volume.



If possible, be visual.

Step-2: Brainstorm, Idea Listing and Grouping. In this step team members gather and provide their ideas and collaborate those ideas and select their problem statement. The ideas should be relevant to their problem statement.



3

Group ideas

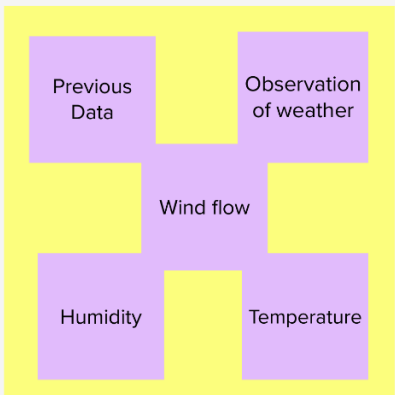
Take turns sharing your ideas while clustering similar or related notes as you go. In the last 10 minutes, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you can break it up into smaller sub-groups.

🕒 20 minutes

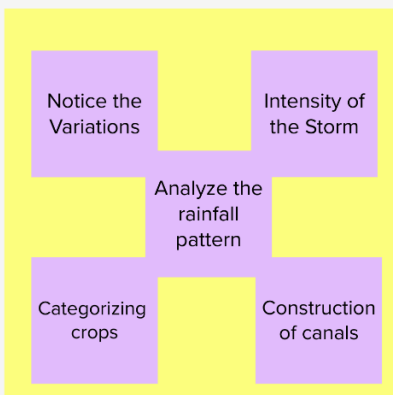
TIP

Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.

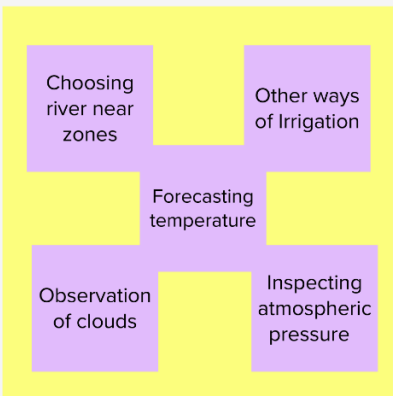
Prediction



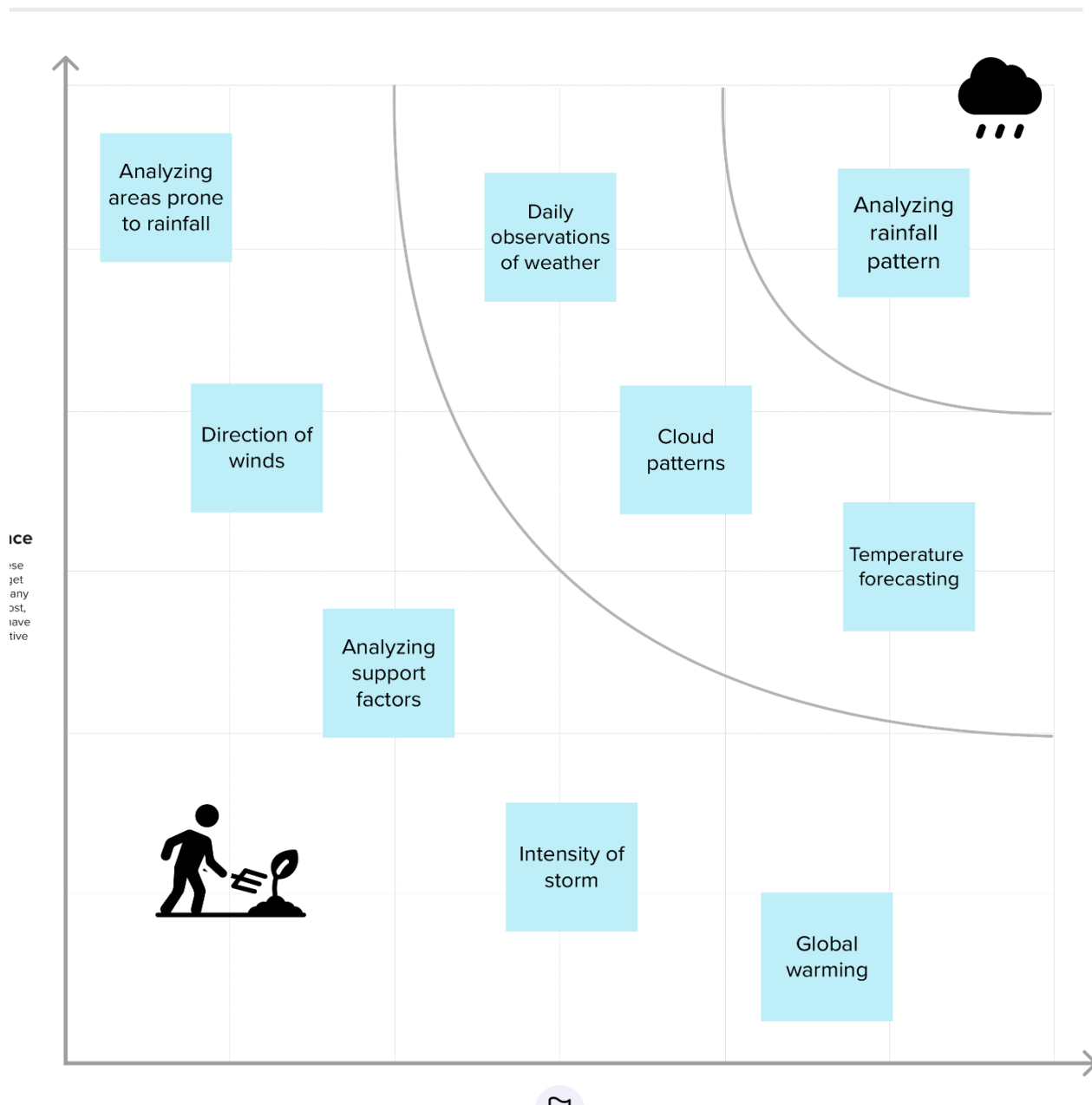
Prevention



Alternative Measures



Step-3: Idea Prioritization As mentioned, idea prioritization is just a part of the idea management process. Having a structured idea management process and a systematic way of gathering, evaluating and prioritizing new ideas takes time. To make it work, the entire idea management process should be integrated into everyday ways



3.3 PROPOSED SOLUTION.

The project aims at predicting the rainfall that may help in effective use of water resources, crop productivity and pre-planning of water structures. Time series data is collected to analyze the pattern in rainfall.

	Parameter	Description
1	Problem statement (problem to be solved)	Agriculture is highly dependent on weather conditions. Due to unpredictable rainfall there is a great impact in crop productivity.
2	Idea/Solution description	By analyzing the pattern using the factors causing rainfall and determining the amount of rainfall in different locations. Analyzing the inner relationships between the agriculture parameters
3	Novelty/ Uniqueness	Suggesting the type of crop to yield based on predicting the amount of rainfall. To analyse large amounts of dataset in consideration of collecting weather data from multiple locations.
4	Social Impact/ Customer Satisfaction	This helps the farmers ineffective usage of water resources ,preventing destruction of crops and do agriculture more smartly in a much better and calculated way.
5	Business Model (financial Benefit)	As the demand for weather forecasting is growing faster over the years, this project will help in achieving the accuracy in rainfall prediction which will have a great impact in the Indian economy
6	Scalability of Solution	The future work of the project would be the improvement of architecture if needed and other weather scenarios.

3.4 PROBLEM SOLUTION FIT

Project Name: India for Agriculture			Project Design Phase 1: Problem Fit Template			Customer: FARMERS WHO DO NOT HAVE ACCESS TO WEATHER FORECASTING		
Define CS, fit into CC	<div>1. CUSTOMER SEGMENT(S)<div>Who is your customer? I.e. working parents of 0-5 y.o. kids</div></div> <div>The customer of this product will be the farmers who involve in agricultural process.</div>	<div>6. CUSTOMER CONSTRAINTS<div>What constraints prevent your customers from taking action or limit their choices of solutions? I.e. spending power, budget, no cash, network connection, available devices.</div><div>Some of the limitations may be that this may not be much reachable to all the farmers.</div><div>Uneducated farmers may find it difficult to use.</div><div>Farmers may be unaware of this product.</div></div>	<div>5. AVAILABLE SOLUTIONS<div>Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? I.e. pen and paper is an alternative to digital notetaking</div><div>Predicting rainfall pattern is already an available solution</div><div>Pros: This helps in predicting the rainfall conditions</div><div>Cons: Does not give accurate prediction May not be helpful in agricultural process</div></div>	Explore AS, differentiate				
	<div>2. JOBS-TO-BE-DONE / PROBLEMS<div>Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides.</div><div>Unpredictable rainfall may lead to destructions of crops which induces a great loss.</div><div>Farmers find difficult in scheduling their irrigation process</div><div>This has a great impact in crop productivity</div></div>	<div>9. PROBLEM ROOT CAUSE<div>What is the real reason that this problem exists? What is the back story behind the need to do this job? I.e. customers have to do it because of the change in regulations.</div><div>The real reason behind unpredictable rainfall may be due to factors causing global warming and changes in weather conditions periodically</div><div>The rainfall prediction has become inevitable as agriculture plays a huge part in the Indian economy</div></div>	<div>7. BEHAVIOUR<div>What does your customer do to address the problem and get the job done? I.e. directly related: find the right solar panel installer, calculate usage and benefits; Indirectly associated: customers spend free time on volunteering work (I.e. Greenpeace)</div><div>Farmers usually do traditional practices such as checking the present state of atmospheric features such as humidity, wind direction etc...</div><div>They may also check online resources for weather forecasting</div></div>	Focus on J&P, tap into BE, understand RC				
Identify strong TR & EM	<div>3. TRIGGERS<div>What triggers customers to act? I.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news.</div><div>By using this they can do agriculture in a better and smart way which may trigger other farmers to use this.</div></div>	<div>10. YOUR SOLUTION<div>If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality. If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.</div><div>By analysing the time series data of rainfall in different location, we can predict the rainfall</div><div>This product is unique in such a way that it can give an accurate prediction of rainfall, gives weather alerts to farmers and also gives suggestions on type of crop to be planted.</div></div>	<div>8. CHANNELS of BEHAVIOUR<div>8.1 ONLINE What kind of actions do customers take online? Extract online channels from #7</div><div>8.2 OFFLINE What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.</div><div>In online the customer can enter the data related to factors causing rainfall and see the result whether it will rain or not.</div><div>They may also get the historical data of rainfall in various locations.</div></div>	Identify strong TR & EM				
	<div>4. EMOTIONS: BEFORE / AFTER<div>How do customers feel when they face a problem or a job and afterwards? I.e. lost, insecure > confident, in control - use it in your communication strategy & design.</div><div>BEFORE: Hardships, hopelessness, depression</div><div>AFTER: confidence, smart work, success,</div></div>							

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENTS

FR-1

Functional Requirement-User Registration

Sub Requirement-User can register and signup the account through Gmail

FR-2

Functional Requirement-User Confirmation

Sub Requirement-The system sends an email confirmation when a new user account is created.

FR-3

Functional Requirement-Authentication

Sub Requirement-Ensures user's authentication by securing with a password

FR-4

Functional Requirement-Analysis and Display

Sub Requirement-The system analyses the rainfall pattern and displays the output to the use.

4.2 Non-Functional requirements

NFR-1

Usability

- In this web application users can easily navigate between tabs.
- This application can be accessible by the users easily.

NFR-2

Security

- This application allows users to create an account to access the rainfall data. ●
- This application will not grant access until the user creates a strong password.
- After certain number of login attempts, this application will lock the account to protect from unauthorized users.

NFR-3

Reliability

- This application will maintain accuracy of information.
- This provides well defined interfaces and readability features.

NFR-4

Performance

- This application will provide less response time in a Chrome desktop browser.
- All the rainfall data are updated dynamically and displayed in the interface

NFR-5

Availability

- The users can view the data related to rainfall whenever required.
- In the case of unplanned system downtime, all features will be available again after some time.

NFR-6

Scalability

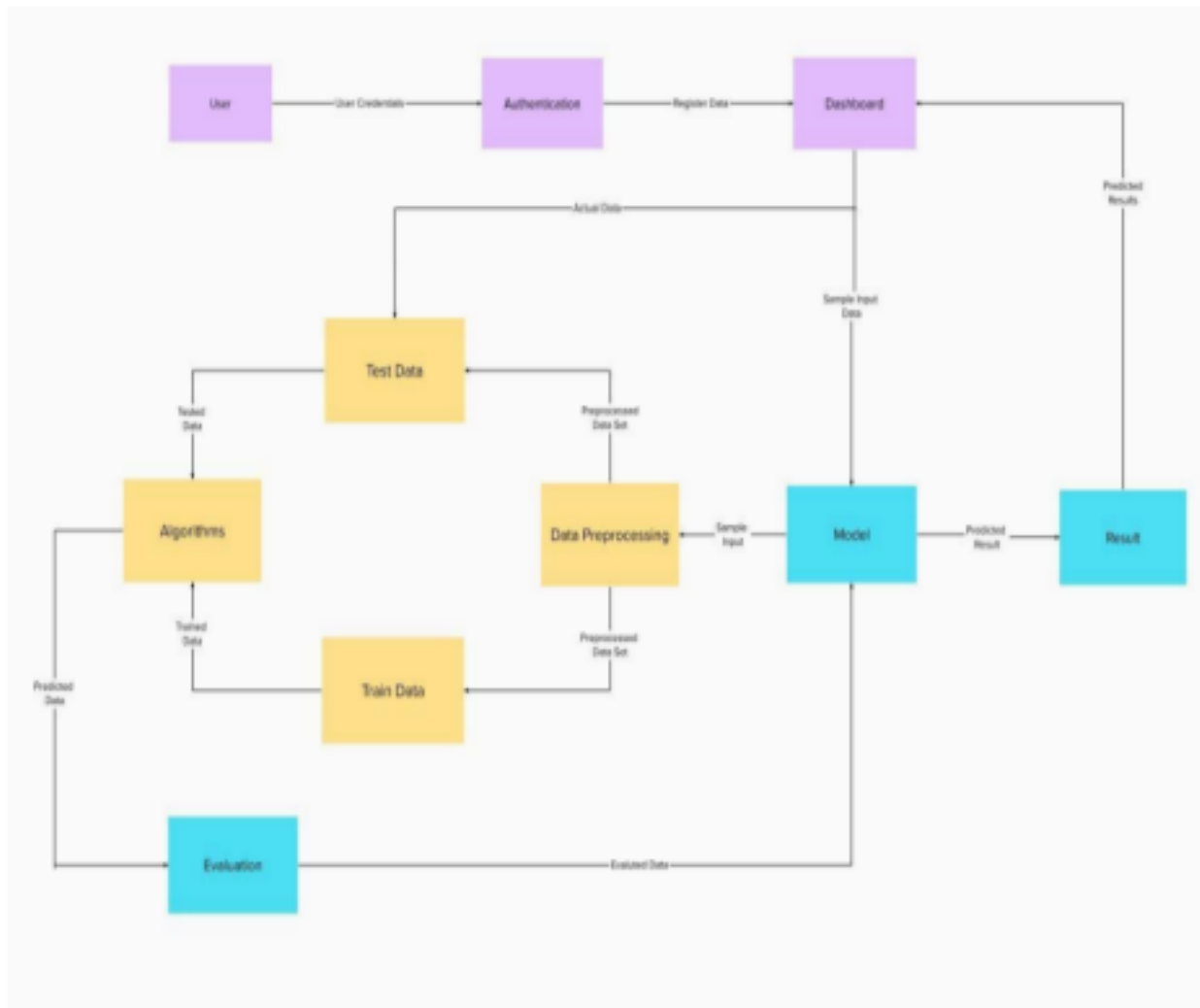
- This application will predict the rainfall data at a faster time.
- This application will be scalable enough to support good number of visitors at the same time while maintaining optimal performance.

5. PROJECT DESIGN

5.1 Data Flow Diagrams

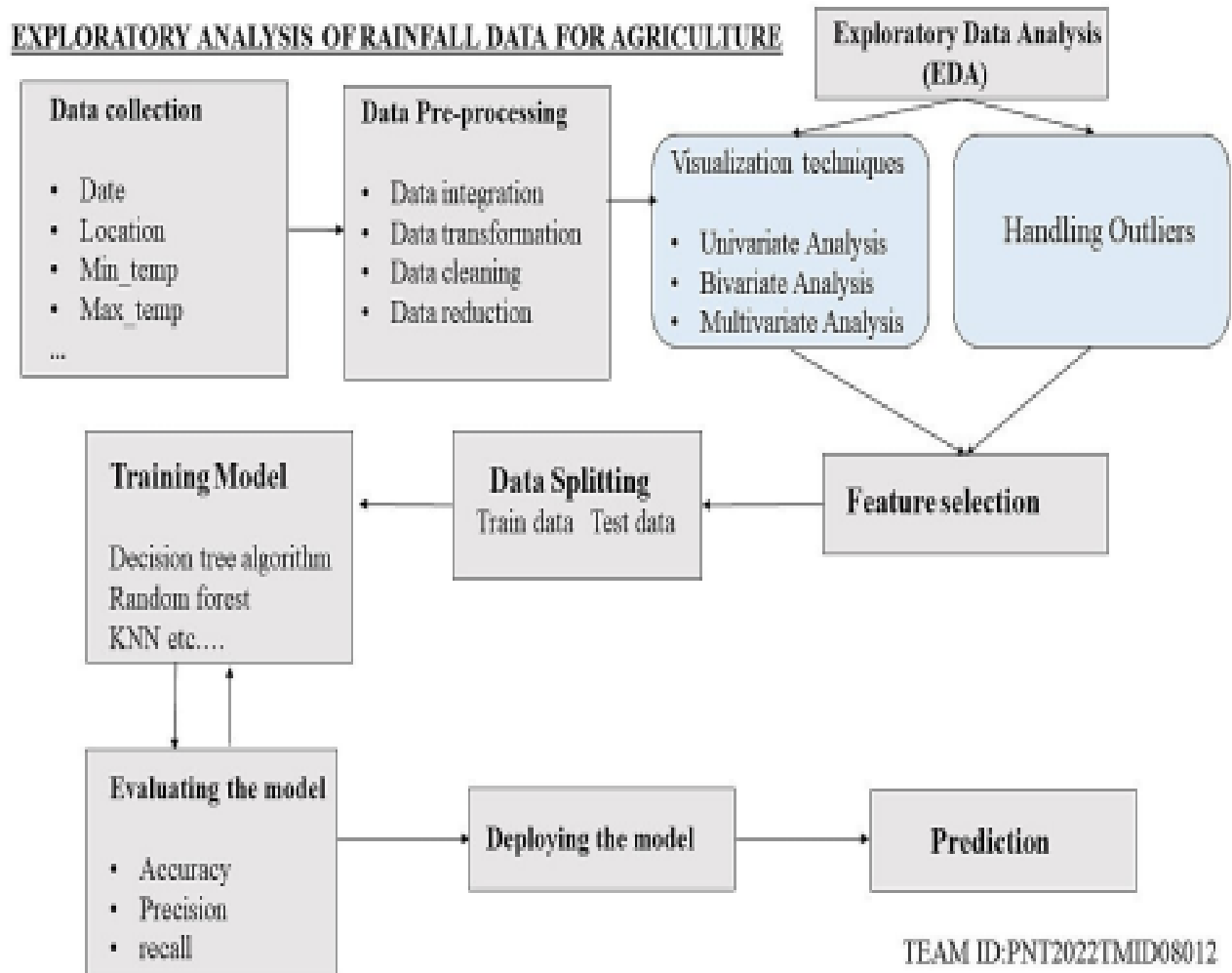
A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored. A data flow diagram shows the way information flows through a process or system. It includes data inputs and outputs, data stores, and the various sub processes the data moves through. DFDs are built using standardized symbols and notation to describe various entities and their relationships. A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. Data flowcharts can range from simple, even hand-drawn process overviews, to in-depth, multi

level DFDs that dig progressively deeper into how the data is handled. They can be used to analyze an existing system or model a new one. Like all the best diagrams and charts, a DFD can often visually “say” things that would be hard to explain in words, and they work for both technical and nontechnical audiences, from developer to CEO.

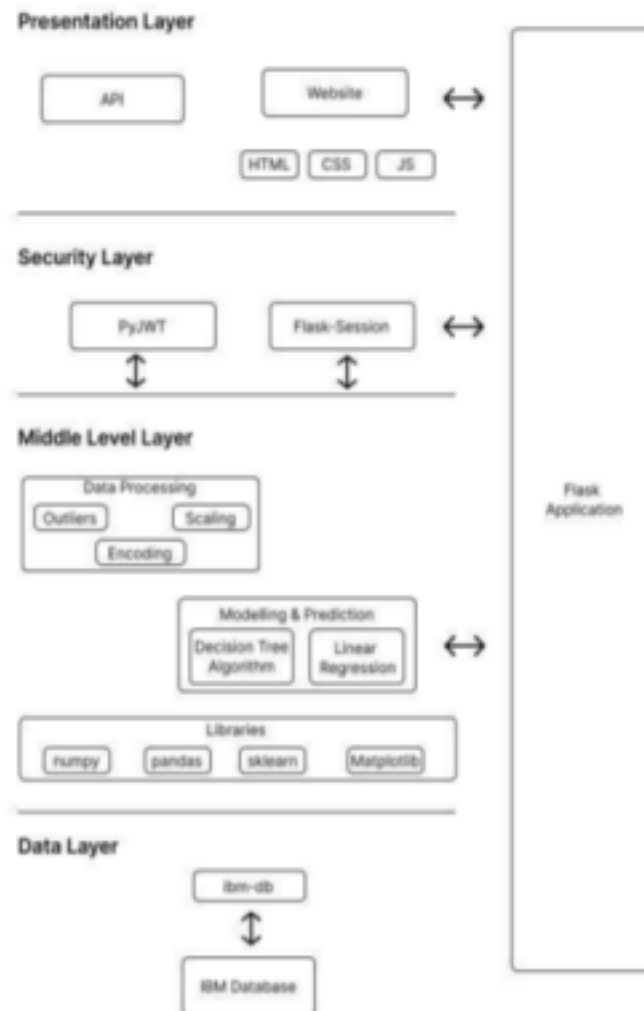


5.2 SOLUTION AND TECHNICAL ARCHITECTURE

Solution architecture:



Technical architecture:



5.3 User Stories

Sprint-01:

USN-1 :As a user, I can register or login to create a dashboard for my processing

USN-2:Once I enter the dashboard I can input values for a single sample prediction

USN-4:As a user I can get the predicted results

Sprint-02:

USN-3:Once I enter the dashboard I can input values for multiple sample prediction

Sprint-03:

USN-5:As a user I can view the detailed report of my prediction

Sprint-04:

USN-6:As a developer I can access dashboard's settings and view the API token

6. PROJECT PLANNING & SCHEDULING**6.1 Sprint Planning & Estimation**

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Members
Sprint-1	Upload data	USN-1	As a user, I can upload the needed data for the prediction of the rainfall	2	High	1.Roselin 2.Prasanna
Sprint-4	Test Vital Page	USN-2	As a user, I will get the prediction result and accuracy on the test vital page.	3	High	1.Nithishree 2.Karunakaran
Sprint-2	Dashboard	USN-3	Dashboard displays the prediction results of the rainfall data.	2	Low	1.Prasanna 2.Karunakaran
Sprint-1	Data Pre Processing	USN-5	As an Administrator, I should clean my data and prepare it	2	High	1.Roselin Janice 2.Karunakaran

			for model building			
--	--	--	--------------------	--	--	--

			by doing pre processing activities such as resizing, visualizing the dataset and			
Sprint-2	Model Building	USN-6	As an Administrator, I need to build the model using Random Forest Classifier for rainfall prediction	3	High	1.Prasanna 2.Karunakaran 3.Nithishree
Sprint-3	Deployment of Model	USN-7	As an Administrator, I need to deploy the Machine Learning model that was built.	2	Medium	1.Roselin Janice
Sprint-3	Building Frontend of the application	USN-8	As an Administrator, I need to build the website for the application using HTML, CSS etc.	2	High	1.Nithishree 2.karunakaran

Sprint-4	Connecting the ML model, Frontend and Backend	USN-9	As an Administrator, I can integrate the deployed model and web application using python flask server.	3	High	1.Roselin Janice 2.Prasanna
----------	---	-------	--	---	------	--------------------------------

Project Tracker, Velocity & Burndown Chart

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	30	6 Days	24 Oct 2022	29 Oct 2022	30	30 Oct 2022
Sprint-2	30	6 Days	31 Oct 2022	05 Nov 2022	30	06 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	13 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	20 Nov 2022

Velocity:

Sprint	Average Velocity
Sprint-1	5
Sprint-2	5
Sprint-3	3.33

Sprint-4	3.33
-----------------	-------------

6.2 Sprint Delivery Schedule

Sprint-1

- Total Story Points - 20
- Duration - 7 Days.
- Sprint Start Date - 24 Oct 2022.
- Sprint End Date (Planned) - 29 Oct 2022.
- Story Points Completed (as on Planned End Date) - 20
- Sprint Release Date (Actual) - 30 Oct 2022.

Sprint-2

- Total Story Points - 20
- Duration - 7 Days.
- Sprint Start Date - 31 Oct 2022.
- Sprint End Date (Planned) - 05 Nov 2022.
- Story Points Completed (as on Planned End Date) - 20
- Sprint Release Date (Actual) - 06 Nov 2022.

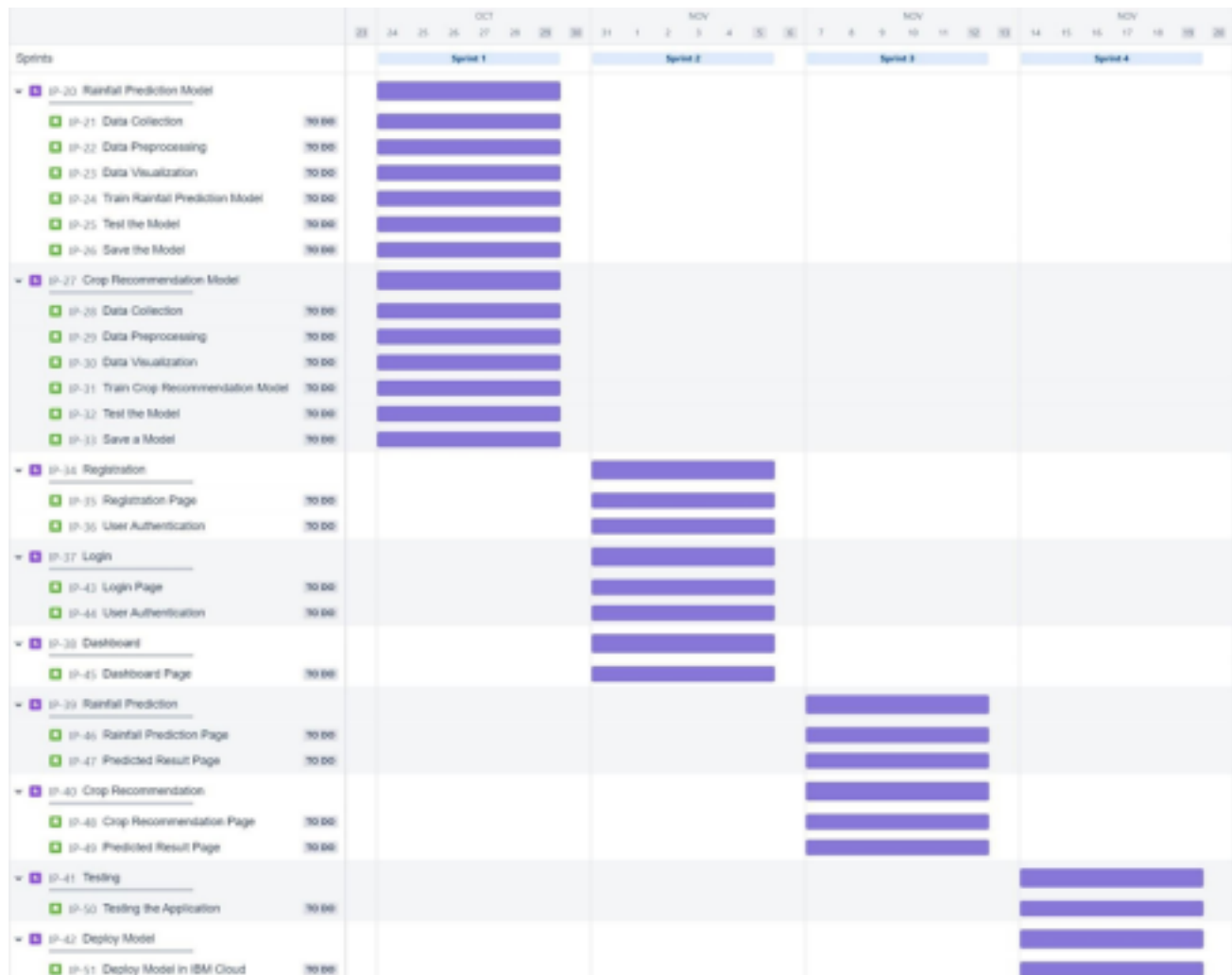
Sprint-3

- Total Story Points -20
- Duration - 7 Days.
- Sprint Start Date - 07 Nov 2022.
- Sprint End Date (Planned) - 12 Nov 2022.
- Story Points Completed (as on Planned End Date) - 20
- Sprint Release Date (Actual) - 13 Nov 2022

. Sprint-4

- Total Story Points - 20
- Duration - 7 Days.
- Sprint Start Date - 14 Nov 2022.
- Sprint End Date (Planned) - 19 Nov 2022.
- Story Points Completed (as on Planned End Date) - 20
- Sprint Release Date (Actual) - 20 Nov 2022.

6.3 REPORTS FROM JIRA



7. CODING & SOLUTIONING

7.1 Feature 1

The classification algorithms Random Forest, KNN, decision tree, and logistic regression were feed with specific features and all classifier parameters gives the best classification performance, and the results from all methods were positive. Thus the results show the Random Forest algorithm provides an accuracy of 98% that is higher than that of the other

three algorithms. It is highly believed that the proposed system can reduce the risk of chronic diseases by predicting them earlier and also reduces the cost for diagnosis, treatment, and doctor consultation. Thus we increased the four standard metrics - accuracy, precision, f1-score and recall using the Random Forest Algorithm and our proposed model as been achieved.

```
[33]: from sklearn.ensemble import RandomForestClassifier
      rf = RandomForestClassifier()
      rf.fit(x_train,y_train)
      predictions = rf.predict(x_test)
      print(confusion_matrix(y_test, predictions))
      print(classification_report(y_test, predictions))
      print(accuracy_score(y_test, predictions))
```

```
[[16667  824]
 [ 2490 2604]]
```

	precision	recall	f1-score	support
0	0.87	0.95	0.91	17491
1	0.76	0.51	0.61	5094
accuracy			0.85	22585
macro avg	0.81	0.73	0.76	22585
weighted avg	0.85	0.85	0.84	22585

```
0.8532654416648218
```

8. TESTING

8.1. Test Cases

A test case template is a document containing an organized list of test cases for different test scenarios that check whether or not the software has the intended functionality. A test case is a set of steps carried out to test a specific feature of an application. This report shows the number of test cases that have passed, failed, and untested.

A set of test inputs, execution conditions, and expected results developed for a particular objective, such as to exercise a particular program path or to verify compliance with a specific requirement. To write the test case, we must have the requirements to derive the inputs, and the test scenarios must be written so that we do not miss out on any features for testing.

3. Test Case Analysis

This report shows the number of test cases that have passed, failed, and untested

Section	Total Cases	Not Tested	Fail	Pass
Handling missing values	6	0	0	6
Client Application	49	0	0	49
testing the model	3	0	0	3
Turning the model	4	0	0	4
Exception Reporting	8	0	0	8
Final Report Output	2	0	0	2
Control version	2	0	0	2

8.2 User Acceptance Testing

In UAT design phase, test engineers are preparing UAT testcases as per the business requirements. AT test coverage should be with Alpha and beta testing. After having complete idea about business requirements and have a discussion with BA or Product Owner one can proceed with UAT Test case design/mapping to UAT test suite.

Approach: Alpha Testing and Beta Testing

- UAT Test scenarios & Testcases prepared based on business needs in both functional and non-functional aspects
- UAT Testcases can be set of existing testcases and maintained as a separate UAT Test suite.
- UAT Test scenarios & cases once designed should be reviewed by BA or Product Owner.
- UAT Testcases target is customer environment based in-terms of Test data and Servers.

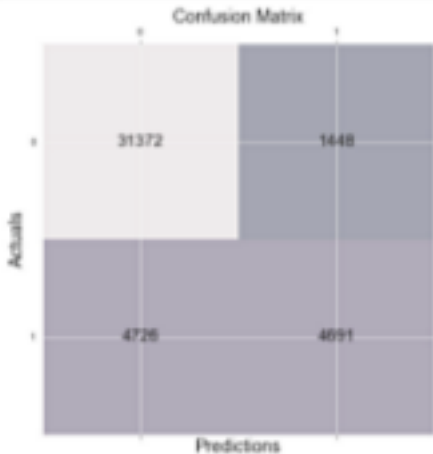
Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved.

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	8	2	1	2	13r
Duplicate	1	0	3	0	4
External	2	2	0	2	6
Fixed	8	1	4	16	35
Not Reproduced	0	0	1	0	1
Skipped	0	1	0	1	3
Won't Fix	0	4	1	1	6
Totals	19	10	10	22	68

9.RESULTS

9.1. PERFORMANCE METRICS

S.N o.	Parameter	Values	Screenshot
1.	Metrics 1	<p>Classification Model: Random Forest</p> <p>Confusion Matrix – [[31372 1448] [4726 4691]]</p> <p>Accuracy Score 0.8538248455145963</p> <p>Classification Report – Accuracy: 0.8538248455145963 Precision: 0.7641309659553673 Recall: 0.49814165870234683 F1-score: 0.6031113396760092</p>	<p>Random forest Confusion matrix</p> <pre>conf_matrix = metrics.confusion_matrix(y_test,t1)</pre> <pre>fig,ax = plt.subplots(figsize=(7.5,7.5)) ax.imshow(conf_matrix,alpha=0.3) for i in range(conf_matrix.shape[0]): for j in range(conf_matrix.shape[1]): ax.text(x=j, y=i, s=conf_matrix[i,j], va='center', ha='center',size='xx-large') plt.xlabel('Predictions',fontsize=18) plt.ylabel('Actuals',fontsize=18) plt.title('Confusion Matrix',fontsize=18) plt.show()</pre>  <pre>t1 = Rand_forest.predict(X_test_scaled)</pre> <pre>print("Rand_forest:",metrics.accuracy_score(y_test,t1))</pre> <p>Rand_forest: 0.8538248455145963</p>

10. ADVANTAGES & DISADVANTAGES

10.1. Advantages:

- As Weather conditions have been changing for the time being this helps people to know about the rainfall prediction
 - To avoid unnecessary floods by opening dams with the help of rainfall prediction
 - Farmers and fisherman will get the most advantage of these rainfall details so that we they can plan accordingly
- During the monsoon days it helps the government to find the evacuation areas to avoid loss of human life and costly things

10.2. DisAdvantages:

- As the data was collected from limited places so it helps only for the people who located in those areas.
- In case the data was collected being wrong the algorithm will produce the wrong prediction
- As of now have collecting only a limited number of data set, In feature, we will make the algorithm to work worldwide

11. CONCLUSION

For predicting rainfall, here this project uses machine learning techniques. This project uses almost 1.45 lakh data of Australia in excel format. After the training dataset have been trained, they have tested it by predicting some unseen day's temperature and found accurate results. Finally Random Forest Classifier have performed well when compared to other algorithms. It is more suited to predict the tomorrow's rainfall.

12. FUTURE SCOPE:

With the change in climatic conditions and rainfall patterns this can lead to flash floods causing catastrophic damage to the environment. The system can be further enhanced with a flood prediction system along with rainfall prediction. Evacuation areas can be included along with the flood prediction system in such a way that the system recommends the user as well as to the community if there might be an occurrence of flood. A recommendation system integrated with the prediction system shall sound good for society

13. APPENDIX

Source Code

Importing the dataset

```
import pandas as pd
df = pd.read_csv('weather.csv')
Df
```

Handling missing values

```
df = df.dropna(axis = 0, how ='any')
df.shape
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['WindGustDir'] = le.fit_transform(df['WindGustDir'])
df['WindDir9am'] = le.fit_transform(df['WindDir9am'])
df['WindDir3pm'] = le.fit_transform(df['WindDir3pm'])
df['RainToday'] = le.fit_transform(df['RainToday'])
df['RainTomorrow'] = le.fit_transform(df['RainTomorrow'])
```

Splitting The Data Into Train And Test

```
from sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test = train_test_split(x,y, test_size = 0.2)
```

```
print(' x_train.shape : ',x_train.shape)
print(' y_train.shape : ',y_train.shape)
print(' x_test.shape : ',x_test.shape)
print(' y_test.shape : ',y_test.shape)
x_train.shape : (90340, 16)
y_train.shape : (90340,)
x_test.shape : (22585, 16)
y_test.shape : (22585,)
```

Model Building

```
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
```

```
#Logistic Regression
```

```
from sklearn.linear_model import LogisticRegression
```

```
lr = LogisticRegression()
```

```
lr.fit(x_train,y_train)
```

```
predictions = lr.predict(x_test)
```

```
print(confusion_matrix(y_test, predictions))
```

```
print(classification_report(y_test, predictions))
```

```
print(accuracy_score(y_test, predictions))
```

```
[[16562  929]
```

```
 [ 2568 2526]]
```

	precision	recall	f1-score	support
0	0.87	0.95	0.90	17491
1	0.73	0.50	0.59	5094

accuracy			0.85	22585
macro avg	0.80	0.72	0.75	22585
weighted avg	0.84	0.85	0.83	22585

```
0.8451627186185522
```

```
from sklearn.tree import DecisionTreeClassifier
```

```
dt = DecisionTreeClassifier()
```

```
dt.fit(x_train,y_train)
```

```
predictions = dt.predict(x_test)
```

```
print(confusion_matrix(y_test, predictions))
```

```
print(classification_report(y_test, predictions))
```

```
print(accuracy_score(y_test, predictions))
```

```
[[14935 2556]
```

```
 [ 2385 2709]]
```

	precision	recall	f1-score	support
0	0.86	0.85	0.86	17491
1	0.51	0.53	0.52	5094

accuracy			0.78	22585
macro avg	0.69	0.69	0.69	22585

weighted avg 0.78 0.78 0.78 22585

0.7812264777507195

```
from sklearn.ensemble import RandomForestClassifier
```

```
rf = RandomForestClassifier()
```

```
rf.fit(x_train,y_train)
```

```
predictions = rf.predict(x_test)
```

```
print(confusion_matrix(y_test, predictions))
```

```
print(classification_report(y_test, predictions))
```

```
print(accuracy_score(y_test, predictions))
```

```
[[16667  824]
```

```
 [2490 2604]]
```

```
      precision    recall  f1-score   support
```

```
0       0.87       0.95       0.91     17491
```

```
1       0.76       0.51       0.61      5094
```

```
accuracy                0.85    22585
```

```
macro avg       0.81    0.73    0.76    22585
```

```
weighted avg     0.85    0.85    0.84    22585
```

0.8532654416648218

```
from sklearn import metrics
```

```
from sklearn.neighbors import KNeighborsClassifier
```

```
k_nearest_neighbour_model = KNeighborsClassifier()
```

```
k_nearest_neighbour_model=k_nearest_neighbour_model.fit(x_train,y_train)
```

```
y_pre= k_nearest_neighbour_model.predict(x_test)
```

```
#display KNN classification Metrics for SVM
```

```
a=round(metrics.precision_score(y_test,y_pre)*100,2)
```

```
b=round(metrics.accuracy_score(y_test,y_pre)*100,2)
```

```
c=round(metrics.recall_score(y_test,y_pre)*100,2)
```

```
d=round(metrics.f1_score(y_test,y_pre)*100,2)
```

```
print('\nPrecision: ',str(a))
```

```
print('Accuracy: ',str(b))
```

```
print('Recall: ',str(c))
```

```
print('F1-score: ',str(d))
```

```
#display classification report for KNN
print('\nClassification Report:\n',metrics.classification_report(y_test,y_pre))
```

```
#display confusion matrix
print('\nConfusion Matrix: \n',metrics.confusion_matrix(y_test,y_pre))
#plot confusion matrix
sns.heatmap(metrics.confusion_matrix(y_test,y_pre), annot = True)
plt.show()
```

Precision: 68.38

Accuracy: 83.43

Recall: 49.37

F1-score: 57.34

Classification Report:

	precision	recall	f1-score	support
0	0.86	0.93	0.90	17491
1	0.68	0.49	0.57	5094
accuracy			0.83	22585
macro avg	0.77	0.71	0.74	22585
weighted avg	0.82	0.83	0.82	22585

Confusion Matrix:

```
[[16328 1163]
 [ 2579 2515]]
```

```
import pickle
pickle.dump(rf,open('rainfall.pkl', 'wb')) # model
pickle.dump(le,open('encoder.pkl','wb'))# encoder saving
pickle.dump(sc, open ('scale.pkl', 'wb'))# scaling the dat
a
```

app.py

```
from pyexpat import features, model
import numpy as np
import pickle
#import joblib
```

```
#import matplotlib
#import matplotlib.pyplot as plt
#import time
import pandas
#import os
from flask import Flask, request, jsonify, render_template, redirect, url_for
```

```
# Declare a Flask app
app = Flask(__name__,template_folder='template')
```

```
model = pickle.load(open("rainfall.pkl",'rb'))
scale = pickle.load(open("scale.pkl",'rb'))
```

```
@app.route('/')
def home():
    return render_template("home.html")
```

```
@app.route('/chance/',methods=['GET', 'POST'])
def chance():
    return render_template("chance.html")
```

```
@app.route('/nochance/',methods=['GET', 'POST'])
def nochance():
    return render_template("noChance.html")
```

```
@app.route('/help/')
def help():
    return render_template("help.html")
@app.route('/contact/')
def contact():
    return render_template("contact.html")
```

```
@app.route('/about/')
def about():
    return render_template("about.html")
```

```
@app.route('/predict',methods=["POST","GET"])
```



```

def predict():
    res = " "
    # If a form is submitted
    if request.method == "POST":
        input_feature=[x for x in request.form.values() ]
        features_values=[np.array(input_feature)]
        names = [['Location','MinTemp','MaxTemp','Rainfall','WindGustSpeed',
        'WindSpeed9am','WindSpeed3pm','Humidity9am','Humadity3pm',
        'Pressure9pm','Pressure3am','Temp9pm','Temp3pm','RainyToday',
        'WindGustDir','WindDir9pm','WindDir3pm']]
        data = pandas.DataFrame(features_values,columns=names)
        data = scale.fit_transform(data)
        data = pandas.DataFrame(data,columns=names)

        #Get prediction
        prediction = model.predict(data)

    else:
        prediction = ""

    if prediction == 1:
        return redirect(url_for('chance'))

    elif prediction == 0:
        return redirect(url_for('nochance'))

    return render_template("index.html", output = res)
#Running the app

if __name__ == "__main__":
    app.run(debug = True,host='0.0.0.0',port=80)

```

chance.html

```

<!DOCTYPE html>
<html>
<head>
    <title>Rainfall Prediction</title>
<style>
body {

```

```
background: linear-gradient(45deg, rgb(169, 250, 247), rgb(216, 77, 255)) fixed;
}
div.gallery {
border: 1px solid rgb(0, 0, 0);
}

div.gallery:hover {
border: 1px solid #777;
}

div.gallery img {
width: 366px;
height: 280px;
}

div.desc {
padding: 15px;
text-align: center;
}

* {
box-sizing: border-box;
}

.responsive {
margin-top: 15vh;
padding: 0 6px;
float: left;
width: 24.99999%;
}

@media only screen and (max-width: 700px) {
.responsive {
width: 49.99999%;
margin: 6px 0;
}
}

@media only screen and (max-width: 500px) {
.responsive {
```

```
width: 100%;
}
}
```

```
.clearfix:after {
  content: "";
  display: table;
  clear: both;
}
```

```
.icon img {
  float: left;
  width: 100px;
  height: 100px;
  background: #555;
  margin-left: 5%;
}
```

```
</style>
```

```
</head>
```

```
<body>
```

```
<div class="icon">
```

```

```

```
</div>
```

```
<center><h1>Chances of rain today</h1>
```

```
<br>
```

```
<h2>You can follow any of the following process for effective agriculture!</h2></center>
```

```
<div class="responsive">
```

```
<div class="gallery">
```

```
<a target="_blank"
```

```
href="https://affordabletreeservice.com.au/shared/content/uploads/Land-clearing-hero-1-1024x550.jpg">
```

```

```

```
</a>
```

```
<div class="desc">Land Clearing</div>
```

```
</div>
```

```
</div>
```

```
<div class="responsive">
  <div class="gallery">
    <a target="_blank" href="https://www.britannica.com/explore/savingearth/wp
content/uploads/sites/4/2019/04/0000156670-1024x681.jpg ">
      
    </a>
    <div class="desc">Ensure proper drainage</div>
  </div>
</div>
```

```
<div class="responsive">
  <div class="gallery">
    <a target="_blank" href="https://4.imimg.com/data4/VQ/KS/MY-3597886/img_0112-
500x500.jpg">
      
    </a>
    <div class="desc">Set up a rain cover</div>
  </div>
</div>
```

```
<div class="responsive">
  <div class="gallery">
    <a target="_blank" href="https://soilsmatter.files.wordpress.com/2015/02/sensor-based
nitrogen-application-to-corn-maize-in-the-midwest-photo-raun.jpg?w=1200">
      
    </a>
    <div class="desc">Pre-plant Fertilizer</div>
  </div>
</div>
</body>
</html>
```

nochance.html:

```
<!DOCTYPE html>
<html>
<head>
  <title>Rainfall Prediction</title>
</style>
  body {
    background: linear-gradient(45deg, rgb(250, 250, 169), rgb(255, 202, 77)) fixed;

  }
  div.gallery {
    border: 1px solid rgb(0, 0, 0);
  }
  div.gallery:hover {
    border: 1px solid #777;
  }

  div.gallery img {
    width: 366px;
    height: 280px;
  }

  div.desc {
    padding: 15px;
    text-align: center;
  }

  * {
    box-sizing: border-box;
  }

  .responsive {
    margin-top: 15vh;
    padding: 0 6px;
    float: left;
    width: 24.99999%;
  }

  @media only screen and (max-width: 700px) {
    .responsive {
```

```
width: 49.99999%;
margin: 6px 0;
}
}
```

```
@media only screen and (max-width: 500px) {
  .responsive {
    width: 100%;
  }
}
```

```
.clearfix:after {
  content: "";
  display: table;
  clear: both;
}
```

```
.icon img {
  float: left;
  width: 100px;
  height: 100px;

  margin-left: 5%;
}
```

</style>

</head>

<body>

<div class="icon">

</div>

<center><h1>No chances of rain today</h1>

<h2>You can follow any of the following process for effective agriculture!</h2></center>

<div class="responsive">

<div class="gallery">

<div class="desc">Calculate the water needed for Irrigation</div>

</div>

</div>

<div class="responsive">

<div class="gallery">

<a target="_blank" href="https://media.istockphoto.com/id/1126541751/photo/hands
planting-the-seeds-into-the

dirt.jpg?s=612x612&w=0&k=20&c=aVUr7F_H4ZSJX89Nmiw59F8WvneKeg

YsBoOiDQw0SA=">

<div class="desc">Sowing</div>

</div>

</div>

<div class="responsive">

<div class="gallery">

<a target="_blank" href="https://media.istockphoto.com/id/1151784210/photo/ripe-rice
field-and-sky-background-at
sunset.jpg?s=612x612&w=0&k=20&c=DZz4wxlbPXnMhmoTsEV06uYKup9MEZTtRFe2XkDb
0mY=">

<div class="desc">Sundrying / Threshing of crops</div>

</div>

</div>

<div class="responsive">

<div class="gallery">

<a target="_blank"

href="https://static.vecteezy.com/system/resources/previews/010/508/297/large_2x/old
farmers-spray-fertilizer-or-chemical-pesticides-in-the-rice-fields-chemical-fertilizers-free
photo.jpg">

<div class="desc">Apply Agricultural chemicals</div>
</div>
</div>
</body>
</html>

```

## home.html

```

<!DOCTYPE html>
<html>
 <head>
 <meta charset="utf-8">
 <title>Rainfall Prediction Webpage</title>

 <link rel="stylesheet" href="{{ url_for('static', filename='css/home.css') }}">

 <link rel="preconnect" href="https://fonts.googleapis.com">
 <link rel="preconnect" href="https://fonts.gstatic.com" crossorigin>
 <link
href="https://fonts.googleapis.com/css2?family=Raleway&family=Roboto:wght@100;500&
display=swap" rel="stylesheet">
 <style type="text/css"></style>
 </head>

 <body>
 <header>
 <div class="header1">

 </div>
 <div class="navbar">

 <div class="nav">HOME</div>
 <div class="nav"><a href="{{ url_for('predict')
}}">PREDICTOR</div>
 <!-- <div class="nav">ABOUT</div> -->
 <!-- <div class="nav"><a href="{{ url_for('help')

```



```

}}">HELP</div> -->
 <!-- <div class="nav"><a href="{{ url_for('contact')
}}">CONTACT</div> -->

</div>
 </header>
 <div>
 <div class="head1">
 Forecast Rainfall
 </div>
 <div class="body1">
 We serve as an early warning system to exactly determine the rainfall
for effective use of water resources,
 crop productivity, and pre-planning of water structures.
 </div>
 </div>
</body>
<footer><p>IBM - Nalaiya Thiran</p></footer>
</html>

```

index.html

```

<!DOCTYPE html>
<html>
 <head>
 <meta charset="UTF-8">
 <title>Rainfall prediction Webpage</title>
 <link rel="stylesheet" href="{{ url_for('static', filename='css/index.css') }}">

 <style>
 .head1{

 color: rgb(10, 73, 112);
 font-family: 'Century Gothic';
 font-size: 60px;
 width: 40%;
 margin-left: 30%;
 text-align: center;
 margin-top: 2vh;

```

```
padding: 5px;
}
```

```
</style>
</head>
```

```
<body>
 <div class="login">
 <div>
 <div class="head1">
 RAINFALL PREDICTION
 </div>

 <form action="{{ url_for('predict')}}" method="post">

 <div class="body1">
 <label for="Location">Location:</label>
 <select id="Location" name="Location">
 <option value=2>Albury</option>
 <option value=4>Badgeryscreek</option>
 <option value=10>Cobar</option>
 <option value=11>CoffsHarbour</option>
 <option value=21>Moree</option>
 <option value=24>Newcastle</option>
 <option value=26>NorahHead</option>
 <option value=27>Norfolksland</option>
 <option value=30>Penrith</option>
 <option value=34>Richmond</option>
 <option value=37>Sydney</option>
 <option value=38>SydneyAirport</option>
 <option value=42>Waggawagga</option>
 <option value=45>Williamtown</option>
 <option value=47>Wollongong</option>
 <option value=9>Canberra</option>
 <option value=40>MountGinini</option>
 <option value=5>Ballarat</option>
 <option value=6>Bandigo</option>
```

```
<option value=35>Sale</option>
<option value=19>MelborneAirport</option>
<option value=18>Melborne</option>
<option value=20>Mildura</option>
<option value=25>Nhill</option>
<option value=33>Portland</option>
<option value=44>Watsonia</option>
</select>
```

```
<label for="MinTemp">MinTemp:</label>
<input type = "number" step="any" id="MinTemp" name="MinTemp" value="MinTemp">
```

```
<label for="MaxTemp">MaxTemp:</label>
<input type = "number" step="any" id="MaxTemp" name="MaxTemp"
value="MaxTemp">
```

```
<label for="Rainfall">Rainfall:</label>
<input type = "number" step="any" id="Rainfall" name="Rainfall" value="Rainfall">
```

```
<label for="WindGustSpeed">WindGustSpeed:</label>
<input type = "number" step="any" id="WindGustSpeed" name="WindGustSpeed"
value="WindGustSpeed">
```

<br>

<br>

```
<label for="WindSpeed9am">WindSpeed9am:</label>
<input type = "number" step="any" id="WindSpeed9am" name="WindSpeed9am"
val="WindSpeed9am">
```

```
<label for="WindSpeed3pm">WindSpeed3pm:</label>
<input type = "number" step="any" id="WindSpeed3pm" name="WindSpeed3pm"
val="WindSpeed3pm">
```

```
<label for="Humidity9am">Humidity9am:</label>
<input type = "number" step="any" id="Humidity9am" name="Humidity9am"
val="Humidity9am">
```

```
<label for="Humidity3pm">Humidity3pm:</label>
<input type = "number" step="any" id="Humidity3pm" name="Humidity3pm"
val="Humidity3pm">
```

<br>

```


 <label for="Pressure9am">Pressure9am:</label>
 <input type = "number" step="any" id="Pressure9am" name="Pressure9am"
val="Pressure9am">

 <label for="Pressure3pm">Pressure3pm:</label>
 <input type = "number" step="any" id="Pressure3pm" name="Pressure3pm"
val="Pressure3pm">

 <label for="Temp9am">Temp9am:</label>
 <input type = "number" step="any" id="Temp9am" name="Temp9am" val="Temp9am">

 <label for="Temp3pm">Temp3pm:</label>
 <input type = "number" step="any" id="Temp3pm" name="Temp3pm" val="Temp3pm">

<label for="RainToday">RainToday</label>
<select id="RainToday" name="RainToday">
 <option value="1">Yes</option>
 <option value="2">No</option>
</select>

<label for="WindGustDir">WindGustDir</label>
<select id="WindGustDir" name="WindGustDir">
 <option value=14>W</option>
 <option value=15>WNW</option>
 <option value=0>WSW</option>
 <option value=7>NE</option>
 <option value=13>NNW</option>
 <option value=10>N</option>
 <option value=2>NNE</option>
 <option value=1>SW</option>
 <option value=6>ENE</option>
 <option value=11>SSE</option>
 <option value=12>S</option>
 <option value=9>NW</option>
 <option value=3>SE</option>
 <option value=8>ESE</option>
 <option value=5>E</option>
```





```
<style>
body {
 font-family: Roboto, sans-serif;
 margin: 0;
}

html {
 box-sizing: border-box;
}

.about-section {
 padding: 255px;
 text-align: center;
 background-image: url("https://c4.wallpaperflare.com/wallpaper/722/662/336/samsung-galaxy-polyart-pastel-wallpaper-thumb.jpg");
 color: white;
 background-repeat: no-repeat;
 background-size: cover;
}

.title {
 color: grey;
}

</style>
</head>
<body>

<div class="about-section">
 <h1>CONTACT US</h1>
 <p><h3>Ask any query to us, our team is wearing the thinking caps. Feel free to pen down your thoughts to the below given mail.</h3></p>
 <p><h2>ibmnt@gmail.com</h2></p>
</div>

</body>
</html>
```

## help.html

```
<!DOCTYPE html>
```

```
<html>
```

```
<head>
```

```
<meta charset="utf-8">
```

```
<title>Rainfall Prediction Webpage</title>
```

```
<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/twitter
bootstrap/5.1.3/css/bootstrap.min.css">
```

```
<script type="text/javascript"
src="https://cdnjs.cloudflare.com/ajax/libs/jquery/3.6.0/jquery.min.js"></script>
```

```
<style type="text/css">
```

```
body{
```

```
background: #eee;
```

```
padding-top: 20px;
```

```
font-family: system-ui;
```

```
}
```

```
.header{
```

```
border-radius: 20px 20px 0px 0px;
```

```
padding: 10px 0px;
```

```
background: purple;
```

```
color: #fff;
```

```
width: 100%;
```

```
display: flex;
```

```
align-content: center;
```

```
justify-content: center;
```

```
}
```

```
.faq-item{
```

```
margin-bottom: 40px;
```

```
margin-top: 40px;
```

```
}
```

```
.faq-body{
```

```
display: none;
```

```
margin-top: 30px;
```

```
}
```

```
.faq-wrapper{
```



```

width: 75%;
margin: 0 auto;
}
.faq-inner{
padding: 30px;
background: aliceblue;
}
.faq-plus{
float: right;
font-size: 1.4em;
line-height: 1em;
cursor: pointer;
}
hr{
background-color: #9b9b9b;
}
</style>
</head>

```

```
<body>
```

```

<div class="container">
<div class="row">
<div class="faq-wrapper">
<div class="header">
<h1>FAQs</h1>
</div>
<div class="faq-inner">
<div class="faq-item">
<h3>

```

What are the steps to be followed inorder to use the website for Rainfall Prediction ?

```

+
</h3>

```

```
<div class="faq-body">
```

It's very simple. Just follow the steps given below.<br><br>

<b>Step 1 :</b> Open the website. On the top, you 'll be provided with various tabs, namely Home, Predictor, About, FAQ and Contact.<br>

<b>Step 2 :</b> Select Predict Page.You 'll be redirected to the Predictor Page.<br>

**Step 3 :** Then specify the values for all the attributes in the space provided.<br>

**Step 4 :** Once this is done, we are good to go with the prediction.<br>

**Step 5 :** Click the PREDICT button given below.<br>

</div>

</div>

<hr>

<div class="faq-item">

<h3>

What are the different attributes in Predictor Page?

<span class="faq-plus">&plus;</span>

</h3>

<div class="faq-body">

Predictor Page has **17** attributes. They are Location, Minimum Temperature (MinTemp), Maximum Temperature (MaxTemp), Rainfall, Wind Gustation Speed (WindGustSpeed), WindSpeed 9am, WindSpeed 3pm, Humidity 9am, Humidity 3pm, Pressure 9am, Pressure 3pm, Temperature 9am(Temp 9am), Temperature 3pm(Temp 3pm), RainToday, Wind Gust Direction (WindGustDir), Wind Direction 9am (WindDir 9am), Wind Direction 3pm (WindDir 3pm).

</div>

</div>

<hr>

<div class="faq-item">

<h3>

What activities can be done if there is a chance of rain ?

<span class="faq-plus">&plus;</span>

</h3>

<div class="faq-body">

Given below are the various processes to be done on a rainy day.<br>

It includes <br>

1. Land Clearing<br>
2. Ensure Proper Drainage<br>
3. Set up a Rain Cover<br>
4. Pre Plant Fertilizer<br>

</div>

</div>

<hr>

<div class="faq-item">

<h3>

What activities can be done if there is no chance of rain ?

<span class="faq-plus">&plus;</span>

</h3>

<div class="faq-body">

Given below are the various processes to be done if there is no chance of rain.<br>

It includes <br>

1. Calculate the water needed for Agriculture<br>

2. Sowing<br>

3. Sun Drying / Threshing<br>

4. Apply Agricultural Chemicals<br>

</div>

</div>

</div>

</div>

</div>

</div>

<script type="text/javascript">

\$(".faq-plus").on('click',function(){

\$(this).parent().parent().find('.faq-body').slideToggle();

});

</script>

</body>

</html>