

# **Exploratory Analysis of Rainfall Data in India for Agriculture**

## **Machine Learning**

### **Submitted by**

**SANDHIYA S  
BHUVANESHWARI P  
ELAKKIYA A  
MALINIPRIYA B**

**TEAM ID - PNT2022TMID42699**

**Department of Computer Science and Engineering,**

**JAI SHRIRAM ENGINEERING COLLEGE**

**Faculty Mentor:** Meenalochini M

**Industrial Mentor:** Lalitha Gayathri

#### **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:**

With the changes in the climatic conditions and irregular pattern of weather conditions, predicting their occurrence for preventing lifeloss to humankind and environment is an utmost societal needed problem of the society. Drastic changes in climate have occurred over the past years and with change in revolution proper preventive measures are needed. Heavy rainfall can lead to floods. Flash floods are catastrophic. Climate change is increasing the frequency, intensity and magnitude of disasters, leading to a higher number of deaths and injuries, as well as increased property and economic losses. In the past 20 years, 90% of major disasters have been caused by

weather-related events such as heatwaves, storms, floods and droughts, according to the UN Office for Disaster Risk Reduction (UNISDR). Natural disasters are increasing in strength and frequency. Shifting weather patterns make predictions and emergency planning difficult. Hence, we focus on the effective prediction of the probability of the flood occurring in a particular region and recommending an evacuation area nearby by performing an exploratory analysis of the data collected.

## **1.2. Purpose**

To design a disaster management system by forecasting a flood event to control flood risk by recommending an evacuation area from flood hazard areas which ultimately helps to manage the environment and water resource system. This also serves a purpose of the Early warning system by training a model and selecting the best prediction algorithm among the classifiers. The occurrence of flash floods can cause catastrophic damage to the society. They first mainly affect the people living near to the riverbeds. Evacuating them from the hazard areas and providing them the shelter they needed. With the irregular change in climate patterns, it's been difficult to predict the occurrence of floods using traditional methods leading to massive destruction. Thus to cope with flash floods and to handle critical situations new methodologies are invented to overcome such difficulties. Technology has to be more aware to reduce the loss that a flash flood would make. In the modernizing era, it's made even easier to predict the occurrence of floods and recommend nearby evacuation areas. Hazard areas that are prone to destruction and devastating loss are monitored regularly and the rainfall readings are collected, integrated from multiple resources, curated, mined, analyzed and prediction is done over patterns. With the prediction, recommendation areas are listed for the society. Early warning systems are climate change adaptation measures that use integrated communication systems to help communities prepare for dangerous climate-related events. An early warning system's success saves lives and jobs, land and infrastructure, and supports long-term sustainability.

## **2. LITERATURE SURVEY:**

### **2.1. Existing Solutions**

This paper deals with the idea of predicting floods using the algorithm Artificial neural networks (ANN) and with the support of the Internet of Things. This system looks after the humidity, temperature, pressure, rainfall, and river water level periodically to the temporal correlative information for flood prediction analysis. Flood data is dynamic and non-linear in

nature. The sensors read the data and inform the system. With those values, the prediction is done and the decision is taken on the occurrence of a flood.[1]

Precipitation in any form such as snow, rain or hail can affect the routine of the society. Therefore predicting the occurrence of rainfall beforehand and warning the society about the day's condition can be helpful in a lot more ways. Providing accurate results for forecasting rainfall has been a major issue with the drastic change in climatic conditions. Using a fusion of machine learning techniques can help in providing much more accurate results about the occurrence of rainfall. Four supervised learning algorithms have been used to get out the accurate results for prediction. The four effective algorithms that result in accurate prediction are decision tree, Naïve Bayes, K-nearest neighbors, and support vector machines. The effectiveness of the algorithm is checked by incorporating the technology known as fuzzy logic. A twelve year historical weather data of city Lahore is considered for training, validating and for testing. In such a way that this fusion model outperformed other existing models.[2]

The drastic change in climatic conditions has caused severe impact on the society and environment. A country's economic and financial condition is mainly dependent on the country's agriculture. Farming and agriculture are considered to be India's backbone of economic conditions. In such a way any climate change affects the agricultural development which in directly affects the economic and financial conditions of the country. Therefore predicting the occurrence of rainfall is one of the most important aspect for the safety of the society as well as the country with its economic conditions. Loss in agriculture could lead to famine and create a huge economic crisis. Prediction made should be to the point. The traditional methods of predicting rainfall have gone out of control with the drastic change in climatic conditions and development of the country. With the rise in global warming conditions, rough humidity and change in the oceans predicting rainfall with any modest technologies that results in the precise results is an utmost need of the society. Applying machine learning classification algorithms to predict the accurate results of rainfall has been implemented. UCI repository dataset has been considered for training, validating and testing.[3]

With nature being unpredictable the intensity of the rainfall varies according to the climatic conditions and the pressure of the wind. Under such conditions, urban floods can be a great disaster

---

for society. This paper deals with a classification-based real-time flood prediction model with the support of a numerical analysis model based on hydraulic theory and the required machine learning models. The Flood database has been created beforehand with the help of the Environmental Protection Agency-Storm Water Management model and from a two-dimensional inundation model. Using the Latin hypercube sampling and probabilistic neural network are used for categorizing the flood depth data into five categories. This machine learning model is constructed to identify the respective cumulative volume if the observed rainfall data is entered. Therefore a system that's capable of generating a real-time flood map by cumulative volume of each grid to the cumulative volume using linear regression and nonlinear regression. The developed system can predict the rainfall-induced flooding potential in such a way that reduces the risk due to disaster and minimizes damage to health and properties. Therefore a useful disaster management system has been developed for preventing huge losses due to disasters.[4]

On a high note, research has been continuously carried out on achieving efficient and accurate prediction technology or systems. With the help of machine learning techniques and algorithms, prediction can be made easy to obtain accurate and earlier results such as making the required arrangements and evacuating people from the hazard areas. Over the two decades, neural networks have shown an extraordinary outcome in predicting the occurrence of floods with the given rainfall data providing better results and cost-effective solutions. This paper is novel in the way of analyzing databases by Multi-layer perceptron classifier to read data such as dynamic identification, deficit treatment, data validation, and data cleaning to be carried across the database. Advancements in every note can provide better results based on the preprocessing of data.[5]

## **2.2. References**

- [1] Swapnil Bande, Virendra V. Shete, "Smart flood disaster prediction system using IoT& neural networks", 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)
  
  - [2] Atta-ur Rahman, Sagheer Abbas, Mohammed Gollapalli, Rashad Ahmed, Shabib Aftab, Munir Ahmad, Muhammad Adnan Khan, Amir Mosav, "Rainfall Prediction System Using Machine Learning Fusion for Smart Cities", 2022 May National Library of Medicine
-

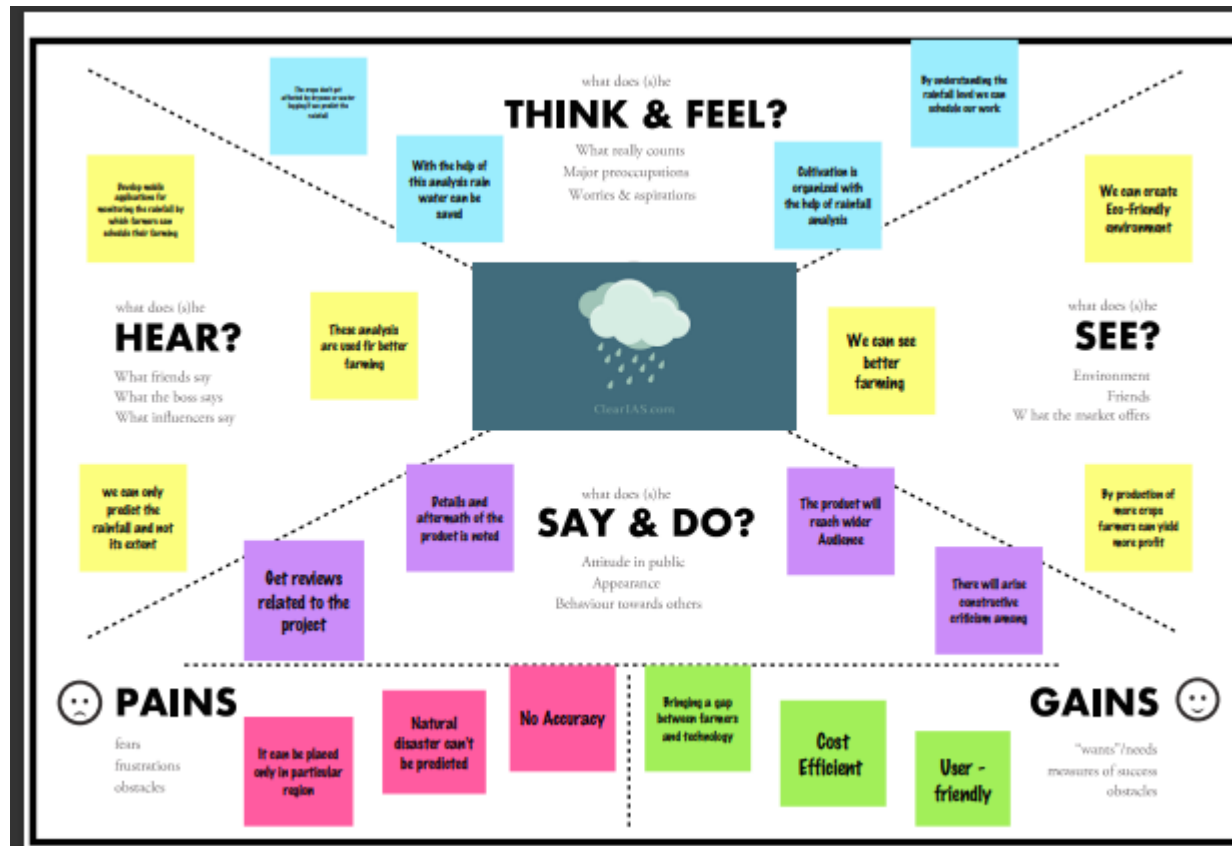
- [3] Vikas Kumar, Vishal Kumar Yadav, Er. Sandeep Dubey, “Rainfall Prediction using Machine Learning”, Ijrasnet Journal For Research in Applied Science and Engineering Technology, 2022.
- [4] Ho Jun Keum, Kun Yeun Han & Hyun Il Kim, “Real-Time Flood Disaster PredictionSystem by Applying Machine Learning Technique”, KSCE Journal of Civil Engineering 24, 2835-2848(2020)
- [5] Thegeshwar Sivamoorthy, Asif Mohammed Ansari, Dr. B. Sivakumar, V. Nallarasan, “Flood Prediction Using ML Classification Methods on Rainfall Data”, IJRASET Journal For Research in Applied Science and Engineering Technology

### 2.3. Problem Statement Definition

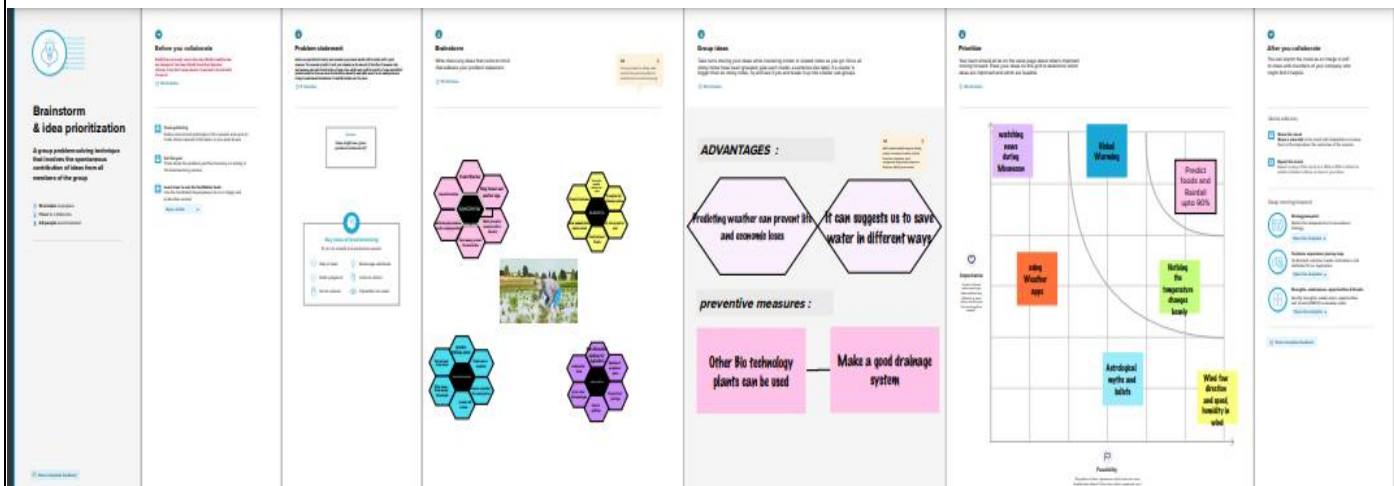
Problem Statement (PS)	I am (Customer)	I’m trying to	But	Because	Which makes me feel
PS-1	Farmer	Predict the heavy rainfall to take precautionary measures to protect the crops from destruction	Can’t predict the heavy rainfall	The climate changes made difficult to make the season as usual	Sad and anxious
PS-2	Department agencies	Make a continuous and good supply of the crops	The crops are destructed and are in shortage.	The heavy rainfall can’t be predicted beforehand	Frustrated

### 3. IDEATION AND PROPOSED SOLUTION:

#### 3.1. Empathy Map Canvas:



#### 3.2. Ideation and Brainstorming:



### 3.3. Proposed Solution:

S.No.	Parameter	Description
1.	Problem Statement	To analyze the rainfall data in India in order to help the growth of agriculture by predicting the heavy rainfall to prevent flooding and damage in crops.
2.	Idea / Solution description	In our analysis we are trying to understand the behavior of rainfall in India over the years, by months and different subdivisions. The trend analysis of Annual rainfall considering India as whole, show that it is important to study subdivision for better forecasting.
3.	Novelty / Uniqueness	Rainfall prediction is important as heavy rainfall can lead to many disasters. The prediction helps people to take preventive measures and moreover the prediction should be accurate. So, it is important to analyse rainfall in order to prevent damage in crops.



4.	Social Impact / Customer Satisfaction	India is an agricultural country and secondary agro based market will be steady with a good monsoon. Thus, the prediction of amount of rainfall can play a major role in agricultural field.
5.	Business Model (Revenue Model)	<ul style="list-style-type: none"> <li>● To create customer value by satisfying needs of a farmer (i.e.) predicting when the rainfall is high and providing early warning.</li> <li>● Marketing the product among farmers and farmers associations.</li> </ul>
6.	Scalability of the Solution	The Scalability of the project is ensuring that model's high-performance computing can support agriculture by delivering more accurate predictions, using higher resolution and more complex modelling, greater use of ensembles and vastly increased volumes of data of all forms.

### 3.4. Solution Fit

Project Title: Exploratory Analysis of Rainfall data in India for Agriculture

Project Design Phase-I - Solution Fit

Team ID: PNT2022TMID21508

Define CS, fit into CC	<b>1. CUSTOMER SEGMENT(S)</b> <span>CS</span> <p>(1) Farmers and farming industries</p> <p>(2) Government departments and agencies</p> <p>(3) Scientific journals.</p>	<b>6. CUSTOMER CONSTRAINTS</b> <span>CC</span> <ul style="list-style-type: none"> <li>Budget</li> <li>Network connection in rural areas</li> <li>Basic statistical knowledge</li> </ul>	<b>5. AVAILABLE SOLUTIONS</b> <span>AS</span> <p>An Exploratory Study on Occurrence and Impact of Climate Change on Agriculture in Tamil Nadu, India - examine the occurrence of climate change in Tamil- -Nadu, and its impact on rainfall pattern which is a primary constraint for agricultural production</p> <p>Flood forecasting using Internet of things and artificial neural networks - India is one of the worst flood-affected countries in the world based on the annual rainfall. They use number of IOT and based techniques but the challenge is that no one has attempted the possibility of occurrence of flood rainfall intensity.</p>	Explore AS, differentiate
	<b>2. JOBS-TO-BE-DONE / PROBLEMS</b> <span>J&amp;P</span> <ul style="list-style-type: none"> <li>To create customer value by satisfying needs of a farmer (i.e.) predicting when the rainfall is high and providing early warning.</li> <li>Marketing the product among farmers and farmers associations.</li> </ul>	<b>9. PROBLEM ROOT CAUSE</b> <span>RC</span> <p>An important aspect to be understood regarding the relationship between rainfall and agriculture is that rainfall is the major factor in the growth and production of food crops both at the germination and fruit development stage. But with a change in the world's climate, temperatures will rise and rainfall will increase in some places. In other places, rainfall will decrease. As a result of global warming, the world's climate is changing and its effect is being felt the world over. And one of the most important parameters of climate is rainfall. So in order to find an effective solution for finding the right time for the cultivation of crops, an algorithm is needed to predict the rainfall rate and derive an useful model out of it.</p>	<b>7. BEHAVIOUR</b> <span>BE</span> <p>The model's high-performance computing can support agriculture by delivering more accurate predictions, using higher resolution and more complex modelling, greater use of ensembles and vastly increased volumes of data of all forms.</p>	

<b>3. TRIGGERS</b> <span>TR</span> <p>By monitoring the data and metrics mentioned above, farmers find a wealth of benefits, including higher production quality and quantity. Other benefits include: Save costs: smart farming leads to lower costs on labor, water, and nutrients for crops. Save water resources: knowing the exact rainfall for each crop can help optimize watering, thus preventing overwatering, which can impact not only crop health, but the environment.</p>	<b>10. YOUR SOLUTION</b> <span>SI</span> <p>In our analysis we are trying to understand the behavior of rainfall in India over the years, by months and different subdivisions. The trend analysis of Annual rainfall considering India as whole, shows that it is important to study subdivision for better forecasting. We can also extend the scope of the project by predicting the probability of flood.</p>	<b>8. CHANNELS of BEHAVIOUR</b> <span>CH</span> <p><b>8.1 ONLINE</b> The prediction is done online through a server-client model.</p> <p><b>8.2 OFFLINE</b> The predicted result can be downloaded and made offline.</p>
<b>4. EMOTIONS: BEFORE / AFTER</b> <span>EM</span> <p>Anxious &gt; Clear and calm (i.e) the farmers can be pre-prepared to face the heavy rainfall and can reduce the destruction of crops.</p>		

8.	Machine Learning Model	Predictive modeling is a statistical technique using machine learning and data mining to predict and forecast likely future outcomes with the aid of historical and existing data	Predictive modeling	
9.	Infrastructure (Server)	Application Deployment on Local System Local Server Configuration: built-in flask web server	Flask web server	
Non Functional Requirements				

4.2.

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Flask	Micro web framework written in Python
2.	Security Implementations	Basic HTTP authentication, Session based authentication, User Login Tracking	Flask Security

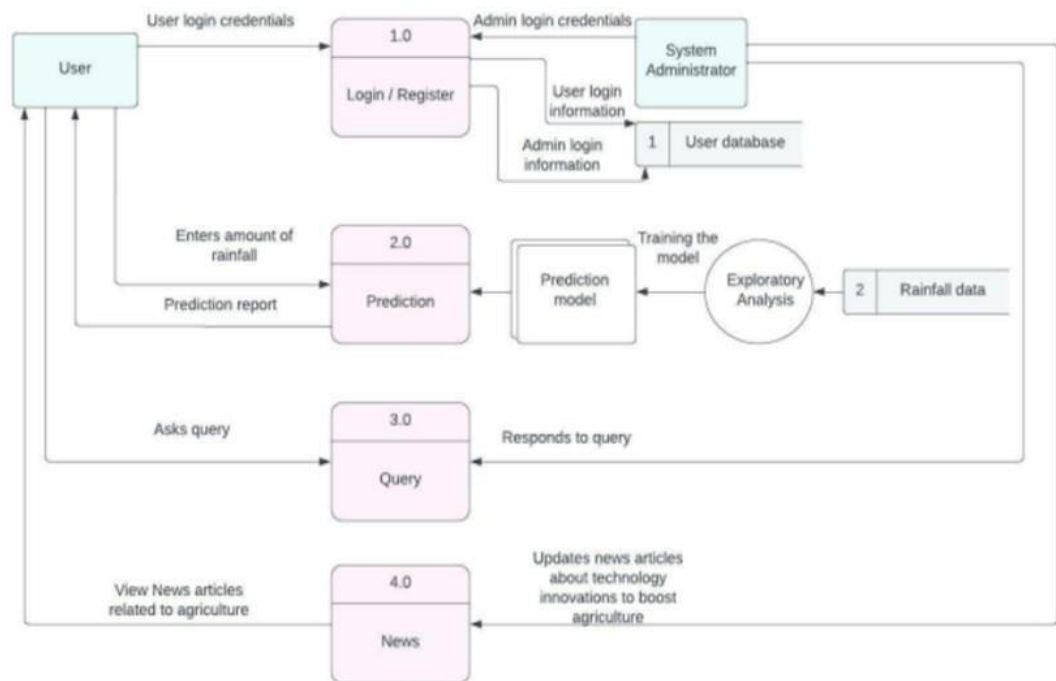
--

3.	Scalable Architecture	Size is everything, and Flask’s status as a microframework means that you can use it to grow a tech project such as a web app incredibly quickly. Its simplicity of use and few dependencies enable it to run smoothly even as it scales up and up.	Flask
4.	Availability	Higher compatibility with latest technologies and allows customization	Flask

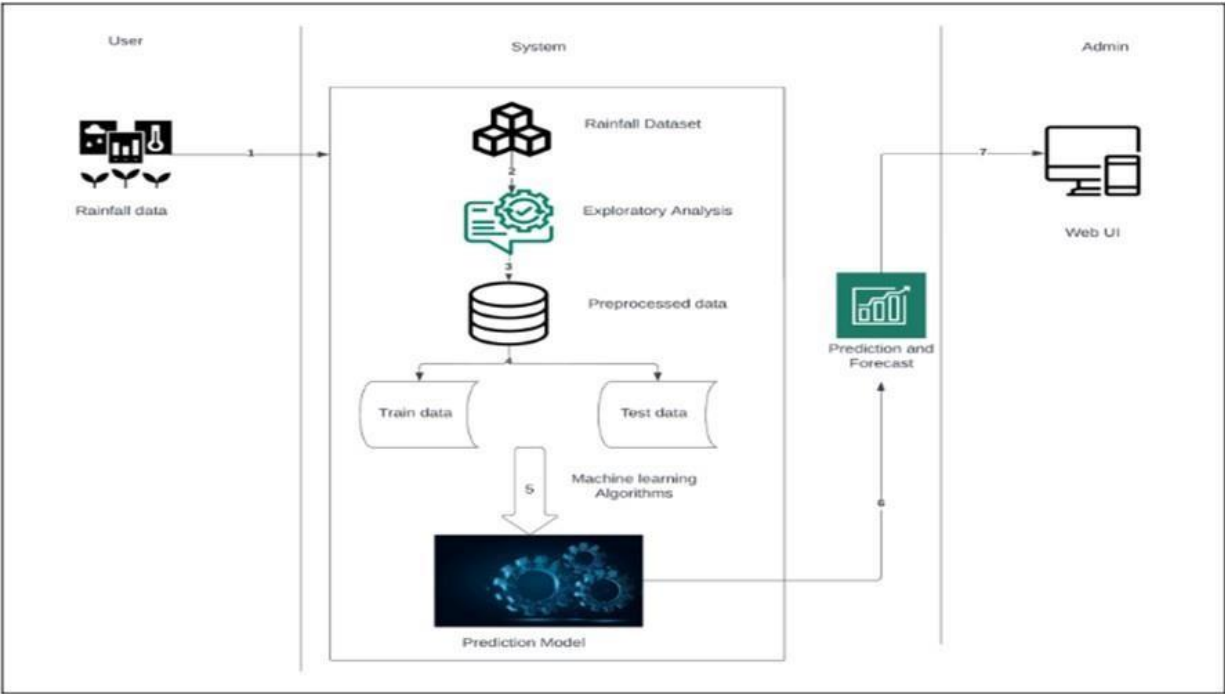
5.	Performance	<ul style="list-style-type: none"> <li>• Integrated support for unit testing.</li> <li>• RESTful request dispatching.</li> <li>• Uses Jinja templating.</li> <li>• Support for secure cookies (client side sessions)</li> <li>• 100% WSGI 1.0 compliant.</li> </ul>	Flask
----	-------------	---	-------

## 5. PROJECT DESIGN

### 5.1. Data Flow Diagram:



5.2.        **Solution And Technical  
Architecture:**







	Registration	USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	Medium	Sprint-2
	Login	USN-3	As a user, I can log into the application by entering email & password	I can access the system	High	Sprint-1
	Dashboard	USN-4	As a user, I can view the details about the system and can navigate	I can navigate through pages	High	Sprint-3

**Project Tracker, Velocity & Burndown Chart: (4 Marks)**

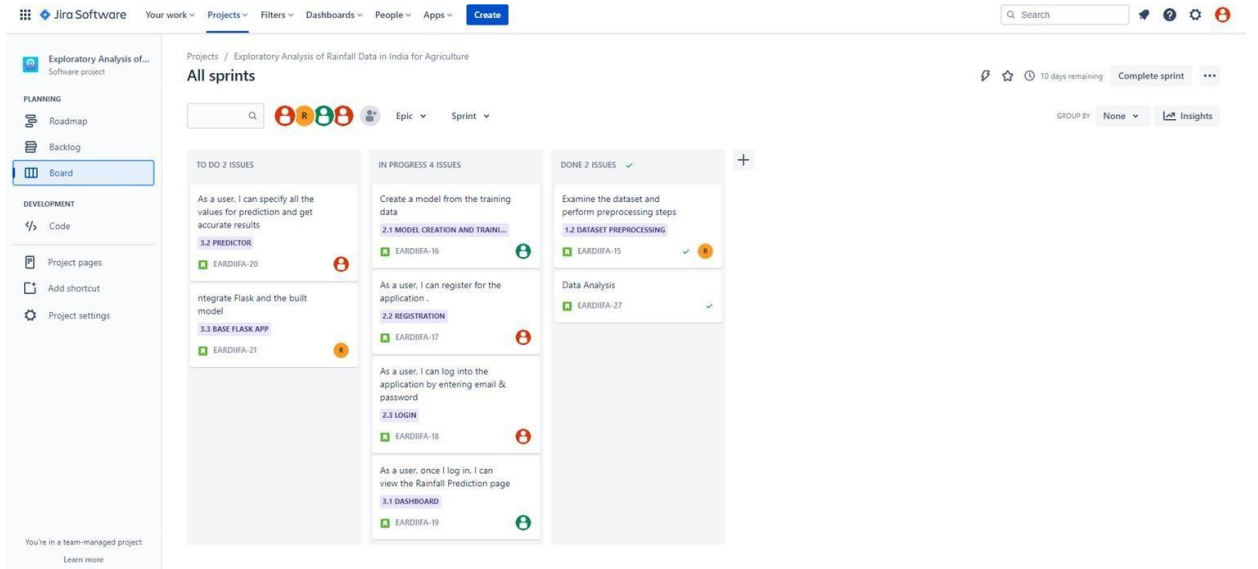
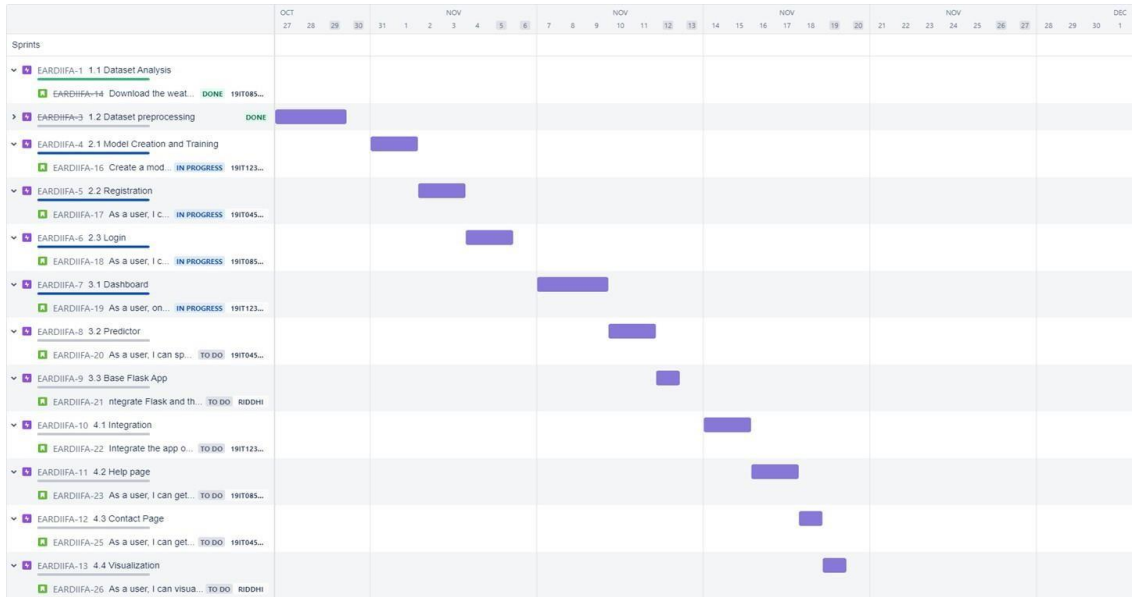
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	3 Days	10 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-2	20	5 Days	13 Nov 2022	17 Nov 2022	20	17 Nov 2022
Sprint-3	20	6 Days	18 Nov 2022	29 Nov 2022	20	23 Nov 2022
Sprint-4	20	6 Days	24 Nov 2022	29 Nov 2022	20	29 Nov 2022

**Product Backlog, Sprint Schedule, and Estimation (4 Marks)**

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Rainfall Prediction ML Model (Dataset)	USN-1	Weather Dataset Collection, Data preprocessing, Data Visualization.	5	High	Sandhiya S , Malinipriya B
Sprint-1		USN-2	Train Model using Different machine learning Algorithms	5	High	Sandhiya S , Malinipriya B
Sprint-1		USN-3	Test the model and give best	10	High	Sandhiya S , Malinipriya B
Sprint-2	Registration	USN-4	As a user, they can register for the application through Gmail. Password is set up.	5	Medium	Bhuvaneshwari P , Elakkiya A
Sprint-2	Login	USN-5	As a user, they can log into the application by entering email & password	5	Medium	Bhuvaneshwari P , Elakkiya A
Sprint-2		USN-6	Credentials should be used for multiple systems and verified	4	Medium	Bhuvaneshwari P , Elakkiya A
Sprint-2	Dashboard	USN-7	Attractive dashboard forecasting live weather	6	Low	Sandhiya S , Malinipriya B
Sprint-3	Rainfall Prediction	USN-8	User enter the location, temperature, humidity	10	High	Bhuvaneshwari P , Elakkiya A
Sprint-3		USN-9	Predict the rainfall and display the result	10	High	Malinipriya B, Bhuvaneshwari P
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-4	Testing	USN-10	Test the application	10	High	Elakkiya A, Malinipriya B
Sprint-4	Deploy Model	USN-11	Deploy the model in IBM cloud to make user friendly application	10	High	Sandhiya S, Bhuvaneshwari P

## 6.3. Reports From Jira





## 7.2. Feature 2: To get predictions when the user enters the values and connecting to the deployed ML model using scoring end point

```
def predict():
    res = " "

    # If a form is submitted

    if request.method == "POST":

        Location = request.form.get('Location')

        MinTemp = request.form['MinTemp']

        MaxTemp = request.form['MaxTemp']

        Rainfall = request.form['Rainfall']

        WindGustSpeed = request.form['WindGustSpeed']

        WindSpeed9am = request.form['WindSpeed9am']

        WindSpeed3pm = request.form['WindSpeed3pm']

        Humidity9am = request.form['Humidity9am']

        Humidity3pm = request.form['Humidity3pm']

        Pressure9am = request.form['Pressure9am']

        Pressure3pm = request.form['Pressure3pm']

        Temp9am = request.form['Temp9am']

        Temp3pm = request.form['Temp3pm']

        RainToday = request.form.get('RainToday')

        WindGustDir = request.form.get('WindGustDir')

        WindDir9am = request.form.get('WindDir9am')

        WindDir3pm = request.form.get('WindDir3pm')

        new_row =
        {'Location':Location,'MinTemp':MinTemp,'MaxTemp':MaxTemp,'Rainfall':Rainfall,'WindGustSpeed':WindGustSpeed,'WindSpeed9am':WindSpeed9am,'WindSpeed3pm':WindSpeed3pm,'Humidity9am':Humidity9am,'Humidity3pm':Humidity3pm,'Pressure9am':Pressure9am,'Pressure3pm':Pressure3pm,'Temp9am':Temp9am,'Temp3pm':Temp3pm,'RainToday':RainToday,'WindGustDir':WindGustDir,'WindDir9am':WindDir9am,'WindDir3pm':WindDir3pm} print(new_row)

        new_df =
        pd.DataFrame(columns=['Location','MinTemp','MaxTemp','Rainfall','WindGustSpeed','WindSpeed9am','WindSpeed3pm','Humidity9am','Humidity3pm','Pressure9am','Pressure3pm','Temp9am','Temp3pm','RainToday','WindGustDir','WindDir9am','WindDir3pm'])
```

```
ed9am','WindSpeed3pm','Humidity9am','Humidity3pm','Pressure9am','Pressure3pm','Temp
9a m','Temp3pm','RainToday','WindGustDir','WindDir9am','WindDir3pm']) new_df =
new_df.append(new_row,ignore_index=True)
```

```
labeled =
new_df[['Location','MinTemp','MaxTemp','Rainfall','WindGustSpeed','WindSpeed9am','Wind
Sp
eed3pm','Humidity9am','Humidity3pm','Pressure9am','Pressure3pm','Temp9am','Temp3pm'
,'R ainToday','WindGustDir','WindDir9am','WindDir3pm']]
```

```
X = labeled.values print(X)
```

```
payload_scoring = {"input_data": [{"field":
[['Location','MinTemp','MaxTemp','Rainfall','WindGustSpeed','WindSpeed9am','WindSpeed3
pm
','Humidity9am','Humidity3pm','Pressure9am','Pressure3pm','Temp9am','Temp3pm','RainyT
od ay','WindGustDir','WindDir9am','WindDir3pm']], "values": X.tolist()}]}
```

```
response_scoring =
requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/73230b85-51ea-45d
b-baa7-e86b5d528fbe/predictions?version=2022-11-14',
json=payload_scoring,headers={'Authorization': 'Bearer ' + mltoken})
```

```
print("Scoring response") predictions =
```

```
response_scoring.json() print(predictions) output =
```

```
predictions['predictions'][0]['values'][0][0]
```

```
print(output)
```

```
else:
```

```
output = "" if output == 1:
```

```
return redirect(url_for('chance'))
```

```
elif output == 0:
```

### 7.3.

```
return redirect(url_for('nochance'))  
return render_template("index.html", output = res)
```

#### **Feature 3 : To navigate between pages**

```
<div class="navbar">  
  <ul>  
    <div class="nav"><a href="">HOME</a></div>  
    <div class="nav"><a href="{{ url_for('predict') }}">PREDICTOR</a></div>  
    <div class="nav"><a href="{{ url_for('help') }}">HELP</a></div>  
    <div class="nav"><a href="{{ url_for('contact') }}">CONTACT</a></div>  
  </ul>  
</div>
```

## 8.2. User Acceptance Testing

### 1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the Project Exploratory Analysis of Rainfall data in India for Agriculture at the time of the release to User Acceptance Testing (UAT).

## 2. 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	0	0	0	0	0
Duplicate	0	0	0	0	0
External	0	0	0	0	0
Fixed	0	0	0	0	0
Not Reproduced	0	0	0	0	0
Skipped	0	0	0	0	0
Won't Fix	0	0	0	0	0
Totals	0	0	0	0	0

## 3. Test Case Analysis

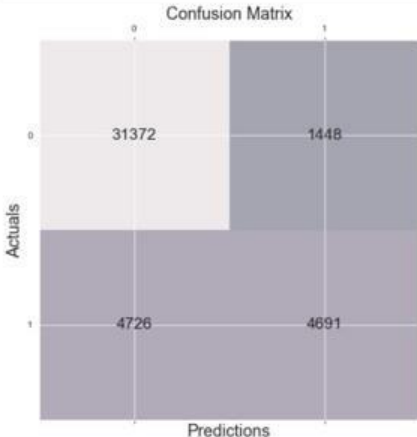
This report shows the number of test cases that have passed, failed, and untested

Section	Total Cases	Not Tested	Fail	Pass
Home Page	2	0	0	2
Predict Page	4	0	0	4



9. RESULTS:

9.1. Performance Metrics

S.N	Parameter	Values	Screenshot																														
1.	Metrics	<p><b>Classification Model:</b> <b>Random Forest</b></p> <p><b>Confusion Matrix –</b> [[31372 1448] [ 4726 4691]]</p> <p><b>Accuracy Score-</b> 0.8538248455145963</p> <p><b>Classification Report –</b> Accuracy: 0.8538248455145963 Precision: 0.7641309659553673 Recall: 0.49814165870234683 F1-score: 0.6031113396760092</p>	<p><b>Random forest Confusion matrix</b></p> <pre>conf_matrix = metrics.confusion_matrix(y_test,t1)</pre> <pre>fig,ax = plt.subplots(figsize=(7.5,7.5)) ax.matshow(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>																														
			<pre>print("***10, "Classification Report", "***10)</pre> <pre>print("-"*30)</pre> <pre>print(classification_report(y_test, t1))</pre> <pre>print("-"*30)</pre> <pre>***** Classification Report *****</pre> <table><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr><tr><td>0</td><td>0.87</td><td>0.96</td><td>0.91</td><td>32820</td></tr><tr><td>1</td><td>0.76</td><td>0.50</td><td>0.60</td><td>9417</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.85</td><td>42237</td></tr><tr><td>macro avg</td><td>0.82</td><td>0.73</td><td>0.76</td><td>42237</td></tr><tr><td>weighted avg</td><td>0.85</td><td>0.85</td><td>0.84</td><td>42237</td></tr></table> <pre>-----</pre>		precision	recall	f1-score	support	0	0.87	0.96	0.91	32820	1	0.76	0.50	0.60	9417	accuracy			0.85	42237	macro avg	0.82	0.73	0.76	42237	weighted avg	0.85	0.85	0.84	42237
	precision	recall	f1-score	support																													
0	0.87	0.96	0.91	32820																													
1	0.76	0.50	0.60	9417																													
accuracy			0.85	42237																													
macro avg	0.82	0.73	0.76	42237																													
weighted avg	0.85	0.85	0.84	42237																													

2.	Tune the Model	Hyperparameter Tuning & Validation Method - <u>RandomizedSearchCV</u>	<h2>Hyperparameter Tuning</h2> <pre> from sklearn.ensemble import RandomForestRegressor rf = RandomForestRegressor(random_state = 42) from pprint import pprint # Look at parameters used by our current forest print('Parameters currently in use:\n') pprint(rf.get_params()) </pre> <p>Parameters currently in use:</p> <pre> {'bootstrap': True,  'ccp_alpha': 0.0,  'criterion': 'mse',  'max_depth': None,  'max_features': 'auto',  'max_leaf_nodes': None,  'max_samples': None,  'min_impurity_decrease': 0.0,  'min_impurity_split': None,  'min_samples_leaf': 1,  'min_samples_split': 2,  'min_weight_fraction_leaf': 0.0,  'n_estimators': 100,  'n_jobs': None,  'oob_score': False,  'random_state': 42,  'verbose': 0,  'warm_start': False} </pre>
----	----------------	---	--

			<pre> n_estimators = [10,20,30,50] max_features = ['auto', 'sqrt'] max_depth = [int(x) for x in np.linspace(10, 50, num = 8)] min_samples_split = [4, 8, 10] min_samples_leaf = [2, 4, 6] bootstrap = [True, False] # Create the random grid random_grid = {'n_estimators': n_estimators,                'max_features': max_features,                'max_depth': max_depth,                'min_samples_split': min_samples_split,                'min_samples_leaf': min_samples_leaf,                'bootstrap': bootstrap}  from sklearn.model_selection import RandomizedSearchCV rf = RandomForestRegressor() rf_random = RandomizedSearchCV(estimator = rf,param_distributions = random_grid, cv=5)  rf_random.fit(X_train_scaled, y_train)  Fitting 5 folds for each of 100 candidates, totalling 500 fits  RandomizedSearchCV(cv=5, estimator=RandomForestRegressor(), n_iter=100,                   n_jobs=-1,                   param_distributions={'bootstrap': [True, False],                                      'max_depth': [10, 15, 21, 27, 32, 38,                                      44, 50],                                      'max_features': ['auto', 'sqrt'],                                      'min_samples_leaf': [2, 4, 6],                                      'min_samples_split': [4, 8, 10],                                      'n_estimators': [10, 20, 30, 50]},                   random_state=35, verbose=2)  best_params = rf_random.best_params_  print ('Best Parameters is', best_params)  Best Parameters is {'n_estimators': 50, 'min_samples_split': 10, 'min_samples_ leaf': 6, 'max_features': 'sqrt', 'max_depth': 21, 'bootstrap': False}  print(f'Accuracy =: {round(rf_random.score(X_train_scaled, y_train) * 100, 2)}%') Accuracy = 75.87% </pre>
--	--	--	---

## **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:**

Floods are the most common natural disasters and have widespread effect flood forecasting is hence an important research area and various possible solutions have been presented in literature to this end the input data were selected based on a correlation and uncertainty analysis of the rainfall and flood data and a classification based real-time flood prediction model was developed heavy rainfall that may occur in urban areas was analyzed in advance and the expected range of an urban flood was predicted in real time using the proposed model

## **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.