EXPLORATORY ANALYSIS OF RAINFALL DATA FOR AGRICULTURE IN INDIA

TEAM ID - PNT2022TMID52964

Department of Electronics and Communication

SSN College of Engineering

Submitted by:

Pranav Narayan. B

Prasanna Venkatesh. P

Sabari. E

Somrithran. S

Faculty Mentor: Ramani. B

Industrial Mentor: Lalitha Gayathri

1. INTRODUCTION:

1.1. Project Overview:

There is always an ambiguity when it comes to changes in the climate and pattern of weather conditions.Rainfall has been a major concern these days. Predicting their occurrence for preventing lifeloss to humankind and environment is an utmost societal needed problem of the society.Irregular heavy rainfall may lead to the destruction of crops, heavy floods that can cause harm to human life. Many disasters have occurred over the past years and with change in revolution proper preventive measures are needed. 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. 90% of major disasters in the last 2 decades have been caused by weather-related events such as heatwaves, storms, floods and droughts, according to the UN Office for Disaster Risk Reduction (UNISDR). It is important to accurately determine the rainfall for effective use of water resources, crop productivity, and pre-planning of water structures.Hence, we focus on predicting the occurrence of rainfall in a particular region and recommending an evacuation area nearby by performing an exploratory analysis of the rainfall data collected.

1.2. Purpose:

Weather forecasting can help with a farmer's business decisions. Forecasts can help them plan for the many day-to-day decisions. These decisions include crop irrigation, time to fertilize, and what days are suitable for working in the field. The decisions that farmers make will result in a profitable crop or failure.To produce a successful crop, a farmer needs to be aware of the moisture, light, and temperature. Detailed weather information, which includes past records, present weather and future forecasts are required. Many decisions decide the cost of the crop. As an example, a farmer can save water and money by not irrigating when rain is in the forecast.It's important for a farmer to know the correct time to apply fertilizer, along with the application rate and type of fertilizer to use. Bad weather at the time of application can wash away the field's profits. It must be dry enough so that the fertilizer doesn't wash away yet the field must be moist enough so that the fertilizer gets down into the soil.By forecasting rainfall in a given area, a farmer is able to refine an irrigation schedule. Rain forecasting at a particular time allows a farmer to plant and use the available water in the farm to water the crop until rains come. When the rains come, a farmer needs to harvest the rain water to ensure that they have sufficient water for the crops. Such an irrigation schedule is designed to safeguard that no water is wasted.

2. LITERATURE SURVEY:

S.NO	PAPER	AUTHOR	YEAR	INFERENCE
1.	Impacts of irrigation and rainfall on agricultural production under climate change	GN Gurjar and Sanjay Swami	2019	The hydrological cycle is predicted to be more intense, with higher annual average rainfall as well increased drought. Thus surface water availability showed a general increase over all 3 basins (though future populations projections would need to be considered to project per capita water availability.
2.	Analysis of rainfall prediction using machine learning data mining and satellite techniques	Nikhilkumar B. Shardoor, Mandapati Venkateswar Rao	2018	Effective rainfall is computed considering the different attributes, such as Max Temp, Min Temp, Wind Speed, Humidity, evaporation, Cloud form, Radiation, Sunshine and Rain FallHourly, Monthly, yearly and etc. There few approaches used to predict the rainfall are, Empirical Method, Physical Method and Statistical Method.
3.	Analysis and prediction of rainfall using machine learning techniques	Anurag Kumar, Lalsingh Chouhan	2022	This proposes a method based on the Multiple Linear Regression (MLR) and Support Vector Regression(SVR). Multiple regression is used to predict the values with the help of descriptive variables and is a statistical method.SVR works on the principle of structural risk minimization from statistical learning theory and establishes a hyperplane that can predict the distribution of data. Comparing the different performance matrices, we can conclude that SVR accuracy is better than MLR

4.	Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting	Ari Yair Barrera-Animas, Lukumon Oyedele, Muhamma Bilal, Taofeek Dolapo Akinosho Juan , Manuel Davila Delgado, Lukman Adewale Akanbi	2022	Originally adopted for natural language processing and time series modelling, RNNs are now being explored for meteorological time-series To overcome the limitation of long term forecast of RNN, , a variation of the ANNs called Long short-term memory (LSTM) Networks has been developed with the inclusion of memory cells that regulate the passage of information in and out of its cells

5.	Prediction of Rainfall using Artificial Neural Networks	A.Kala, Dr. S. Ganesh Vaidyanathan	2018	Artificial Neural Network(ANN) such as Feed Forward Neural Network(FFNN) model is used to predict rainfall. ANN is based on self-adaptive mechanism in which the model learns from historical data and captures functional relationship between them. The accuracy is measured using confusion matrix and RMSE. The results show that prediction using ANN produces acceptable accuracy.

2.1. References

[1] Sanjay Swamy, "Impacts of irrigation and rainfall on agricultural production under climate change", 2019 International Journal of Chemical Studies

[2] Nikhilkumar B. Shardoor, Mandapati Venkateswar Rao, "Analysis of rainfall prediction using machine learning data mining and satellite techniques", 2018 International Journal of Engineering and Technology

[3] Anurag Kumar, Lalsingh Chouhan, "Analysis and prediction of rainfall using machine learning techniques",2022 International Research Journal of Engineering and Technology

[4] Ari Yair Barrera- Animas, Lukumon Oyedele, Muhamma Bilal, Taofeek Dolapo Akinosho Juan , Manuel Davila Delgado, Lukman Adewale Akanbi, "Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting", 2022 Machine Learning with Applications 7 100204

[5] A.Kala, Dr. S. Ganesh Vaidyanathan,"Prediction of Rainfall using Artificial Neural Networks", 2018 International Conference on Inventive Research in Computing Applications (ICIRCA)

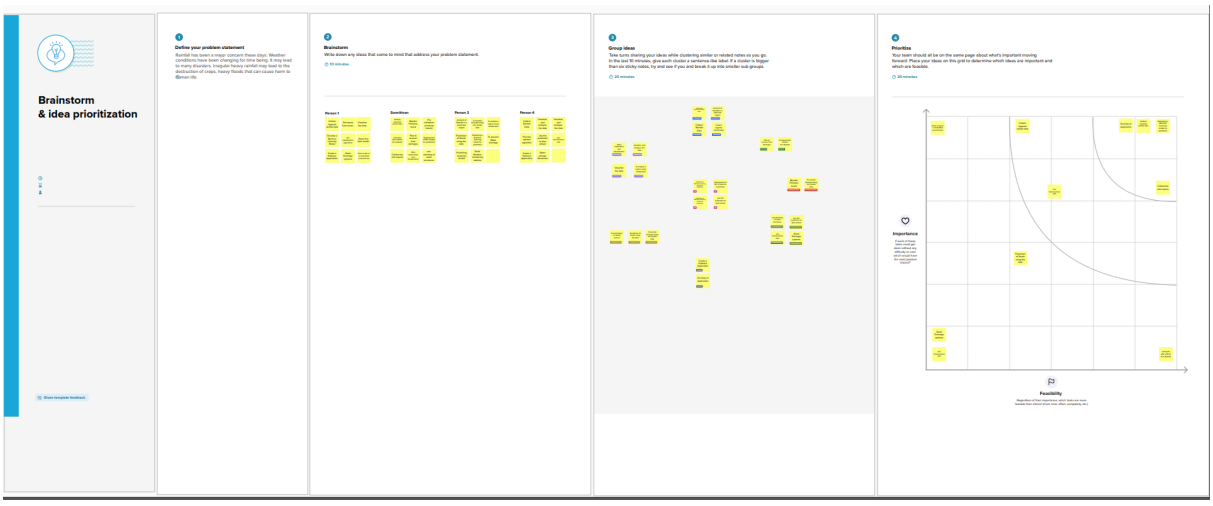
2.2. Problem Statement Definition

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

3. IDEATION AND PROPOSED SOLUTION:

3.1. Empathy Map Canvas:

3.2. Ideation and Brainstorming:

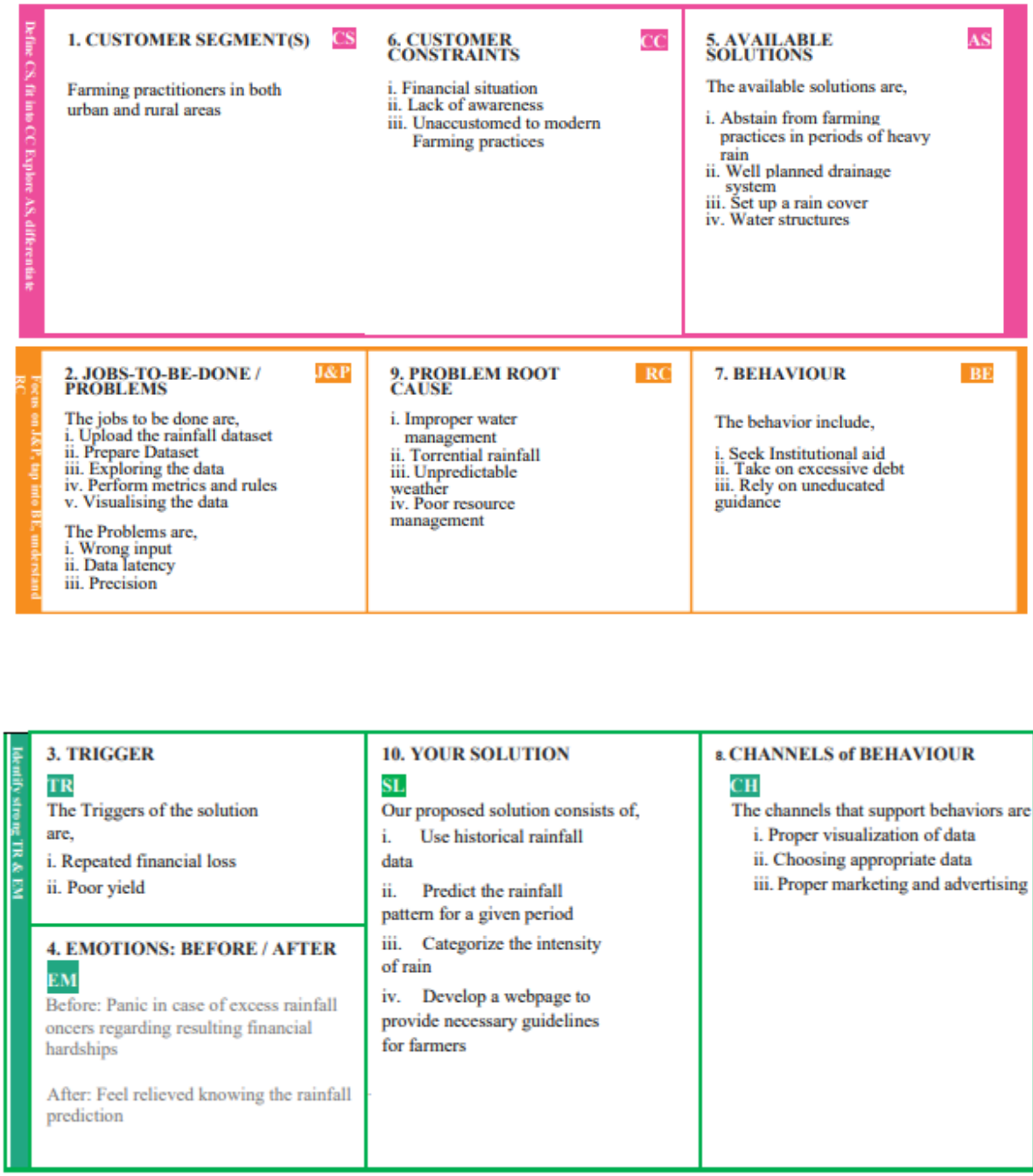


3.3. Proposed Solution:

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Rainfall is a highly influential factor with respect to agriculture. Irregular heavy rainfall often results in heavy floods, destruction of crops and other socio-economic hazards. It is thereby crucial to exactly determine the rainfall for effective use of water resources, crop productivity, and pre-planning of water structures.
2.	Idea / Solution description	Smart prediction of rainfall using exploratory analysis of rain fall data in India to enable efficient allocation of water resources for agricultural purposes.
3.	Novelty / Uniqueness	The solution is reliant on the rainfall data in India for agriculture. The solution is enabled by the use of data analytics and data visualisation tools.
4.	Social Impact / Customer Satisfaction	Efficient management of water resources leading to minimized cost of operation. Farmers enjoy increased yield and thereby increased revenue. This enables them to settle existing debts if any, and an overall improved quality of life.
5.	Business Model (Revenue Model)	Increased revenue for agricultural practitioners as a result of increased yield.
6.	Scalability of the Solution	Highly Scalable. Proposed solution will help improve the workflow of the overall agricultural process, thereby making life easier for farmers.

3.4. Solution Fit

Project Title: Exploratory Analysis of Rain Fall Data in India for Agriculture **Team ID:** PNT2022TMID52964



4. REQUIREMENT ANALYSIS:

4.1. Functional Requirements

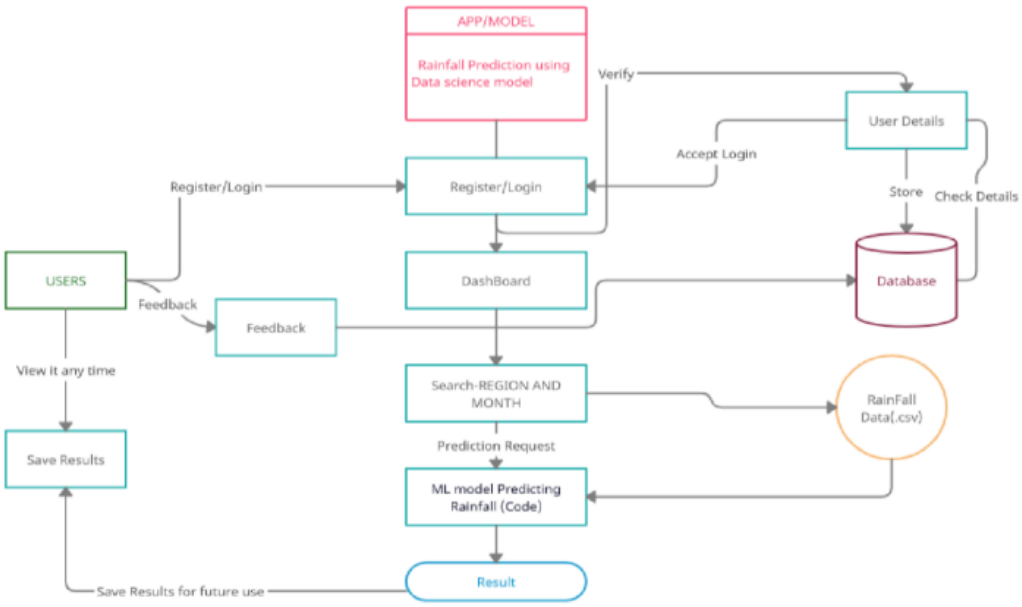
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form
FR-2	User Confirmation	Confirmation via OTP
FR-3	Dataset details	Download and load the appropriate dataset followed by data pre-processing
FR-4	Build a Machine Learning Model	Train the model using training data and find the best algorithm
FR-5	Prediction Model	Forecast weather conditions for a region and generate alerts for flood or drought
FR-6	Testing	Test the model using various performance metrics
FR-7	Application	Develop a website to display and visualize the weather forecast for the region of concerned user

4.2. Non Functional Requirements

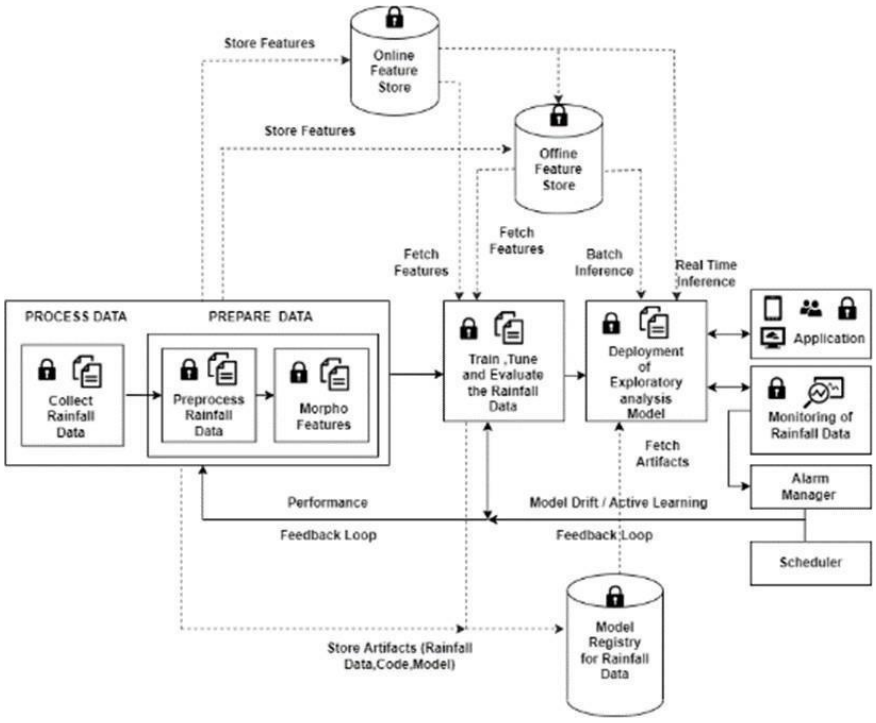
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The usability is countrywide deployment (Villages to Cities) and made available in all regional languages so everybody can be benefited
NFR-2	**Security**	Security will be provided over the website and doesn't require personal details
NFR-3	**Reliability**	Can be accessed anywhere with good connectivity
NFR-4	**Performance**	It can provide accurate data or alerts to all users without delay and provide 24/7 availability
NFR-5	**Availability**	It will be an open-source model and hosted on internet so that users can browse it anytime
NFR-6	**Scalability**	The project can be scaled to domains other than agriculture with improved accuracy

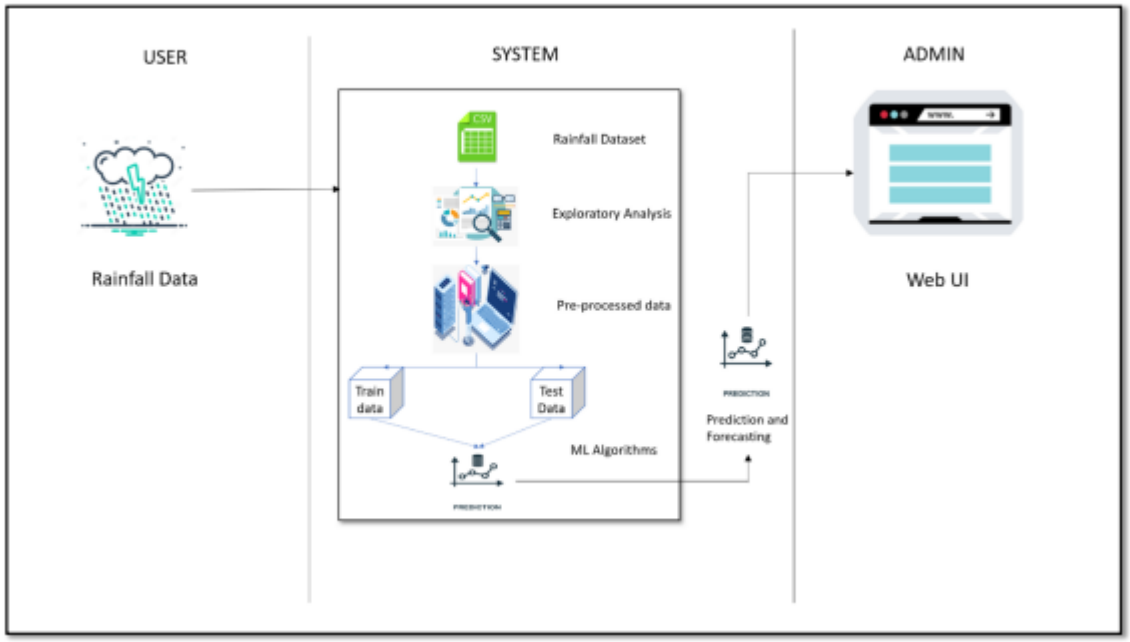
5. PROJECT DESIGN

5.1. Data Flow Diagram:



5.2. Solution And Technical Architecture:





5.3.User Stories:

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Web Page)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password	I can access the system	High	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-2
	Prediction	USN-7	User can search for the area/place where the user wants to know the prediction of rainfall	Only places in India will be accepted	High	Sprint-2
		USN-8	User can see the visualization of the rainfall for a specific region over a duration and also alerts for adverse weather condition		High	Sprint-2
		USN-9	User can give suggestions and feedback on the accuracy of prediction and user interface		High	Sprint-3
Customer Care Executive	Contact	USN-9	User can ask queries about the system	I can rectify my doubts	Medium	Sprint-3
		USN-10	The team must analyse queries and make plans to rectify the queries based on priorities	Queries must be responded or solved	High	Sprint-3
	Chat Bot	USN-11	Implement chat bot to respond to queries	Instant response	Low	Sprint-3
Administrator	Login	USN-12	I can register for the application by entering my email and password	I can view and update the system	High	Sprint-3
	User Experience	USN-13	The website is responsive on all the devices and the screen sizes	User experience should be good	High	Sprint-3

6. PROJECT PLANNING & SCHEDULING:

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

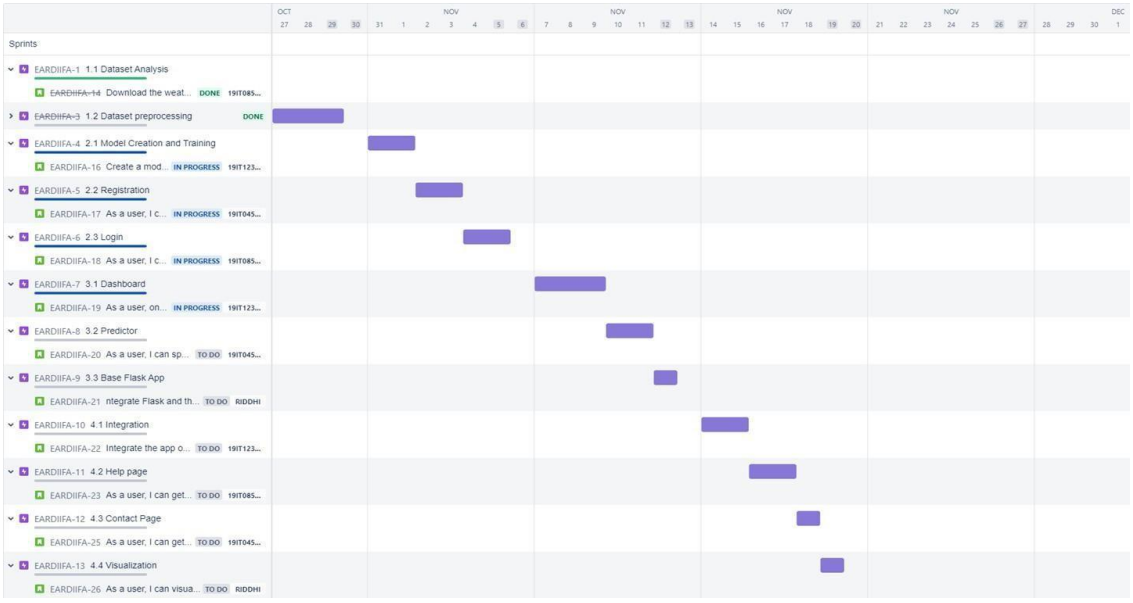
Use the below template to create product backlog and sprint schedule

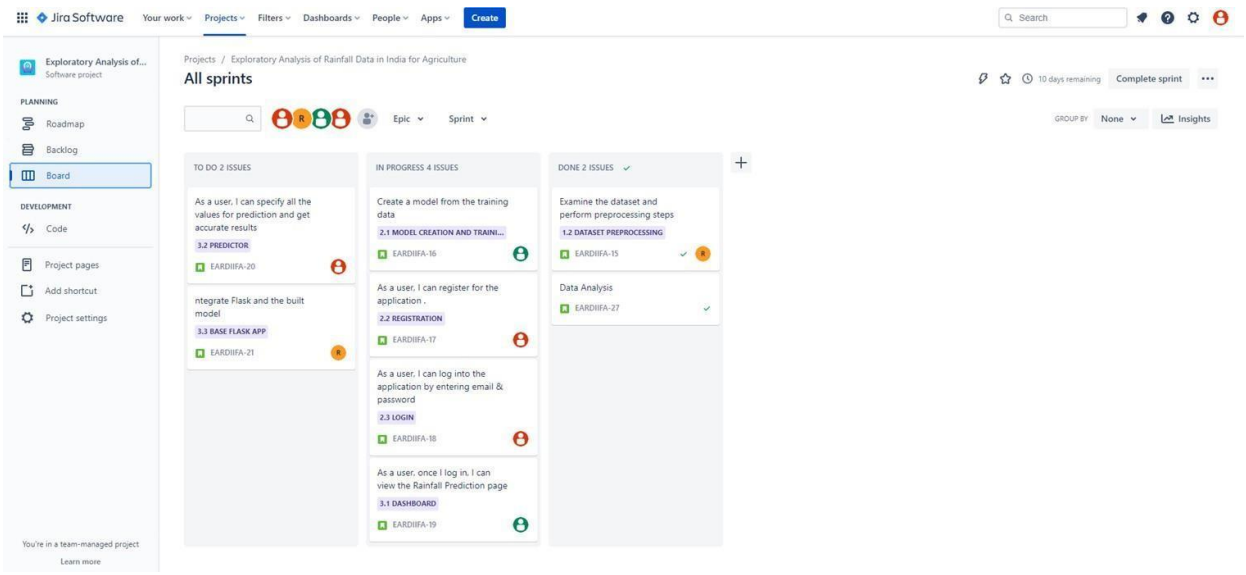
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Rainfall Prediction ML Model (Dataset)	USN-1	Weather Dataset Collection, Data preprocessing, Data Visualization.	5	High	Somrithran S, Pranav Narayana B
Sprint-2		USN-2	Train Model using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

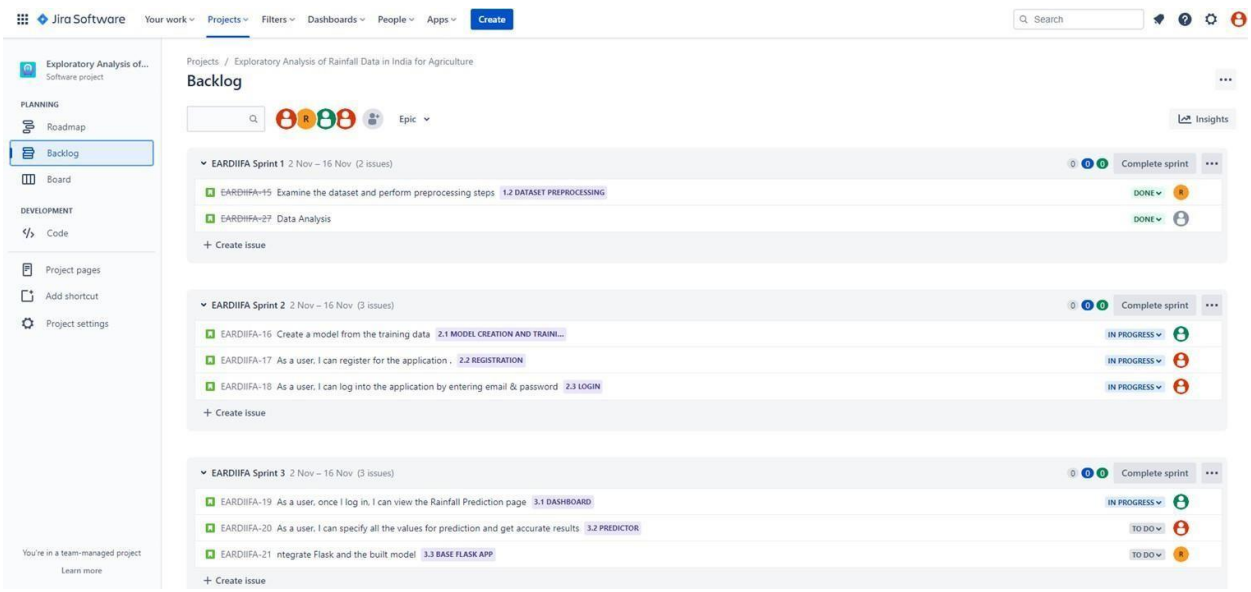
Project Tracker, Velocity & Burndown Chart: (4 Marks)

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	6 Days	31Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-2	20	6 Days	05 Nov 2022	10 Nov 2022	20	10 Nov 2022
Sprint-3	20	6 Days	10 Nov 2022	15 Nov 2022	20	15 Nov 2022
Sprint-4	20	6 Days	15 Nov 2022	21 Nov 2022	20	21 Nov 2022

Reports From Jira







7. CODE AND SOLUTIONING:

7.1. Feature 1: To retrieve information from IBM cloud account using API Key

```
import requests

import json

API_KEY = "PQBr9MBF7mFuSh2VVLfOE-liIA04VH-h5VEk8EfjFIuw"

token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":
API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})

mltoken =

token_response.json()["access_token"]

print("ML Token",mltoken)

header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}
```

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,'WindGustsp
eed':WindGustSpeed,'WindSpeed9am':WindSpeed9am,'WindSpeed3pm':WindSpeed3pm,'Hu
midity9am':Humidity9am,'Humidity3pm':Humidity3pm,'Pressure9am':Pressure9am,'Pressure
```

```
    3pm':Pressure3pm,'Temp9am':Temp9am,'Temp3pm':Temp3pm,'RainToday':RainToday,'WindG
    ustDir':WindGustDir,'WindDir9am':WindDir9am,'WindDir3pm':WindDir3pm}
    print(new_row)
     new_df =
    pd.DataFrame(columns=['Location','MinTemp','MaxTemp','Rainfall','WindGustSpeed','WindSpe
    ed9am','WindSpeed3pm','Humidity9am','Humidity3pm','Pressure9am','Pressure3pm','Temp9a
    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','WindSp
    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','WindSpeed3pm
    ','Humidity9am','Humidity3pm','Pressure9am','Pressure3pm','Temp9am','Temp3pm','RainyTod
    ay','WindGustDir','WindDir9am','WindDir3pm']], "values": X.tolist()}]}

     response_scoring =
    requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/73230b85-51ea-45
    d 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:
```

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

7.3. 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. TESTING:

8.1. Test Cases

Test case ID	Feature Type	Component	Test Scenario	Prerequisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Comments	TC for Automation(Y/N)	BUG ID
HomePage_TC_001	UI	Home Page	Verify all the UI elements in Home page rendered properly	HTML	1. Enter URL and click go 2. Verify all the UI elements displayed or not		All the UI elements rendered properly	Working as expected	Pass		N	
HomePage_TC_002	Functional	Home page	Verify the Data Entry page can be reachable.	HTML, CSS	1. click the predict tab in navigation bar. 2. Verify all the UI elements displayed or not.		User should navigate to Predictor page	Working as expected	Pass		N	

Predict_Page_TC_003	UI	Predict Page	Verify all the UI elements in Predict page rendered properly	HTML,CSS	1. Enter URL and click go 2. Verify all the UI elements displayed or not		All the UI elements rendered properly	Working as expected	Pass		N	
PredictPage_TC_004	Functional	Predict Page	Enter all the values and verify the prediction	Flask	1. Enter URL and click go 2. Enter the values for 17 attributes 3. Click Predict	NewCastle 13.4 22.6 0.6 44 21 24 70 78 1007.7 1007.1 34 32 Yes WSW NNW ESE	Redirect to corresponding html page (chance/no chance)	Working as expected	Pass		N	
OutputPage_TC_005	Functional	Chance Page	Verify whether it is redirected to chance page		1. Enter URL and click go 2. Enter the values and click predict button	Prediction = 1	Redirect to chance page	Working as expected	Pass		N	
					3. If prediction equals one, chance page is displayed.							
OutputPage_TC_006	Functional	No chance Page	Verify whether it is redirected to no chance page		1. Enter URL and click go 2. Enter the values and click predict button 3. If prediction equals zero, no chance page is displayed.	Prediction = 0	Redirect to no chance page	Working as expected	Pass		N	

S.NO	Test Scenerios
1	Verify all the UI elements in Home page rendered properly.
2	Verify the Data Entry page can be reachable.
3	Verify all the UI elements in Predict page rendered properly
4	Enter all the values and verify the prediction
5	Verify whether it is redirected to chance page
6	Verify whether it is redirected to no chance page

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).

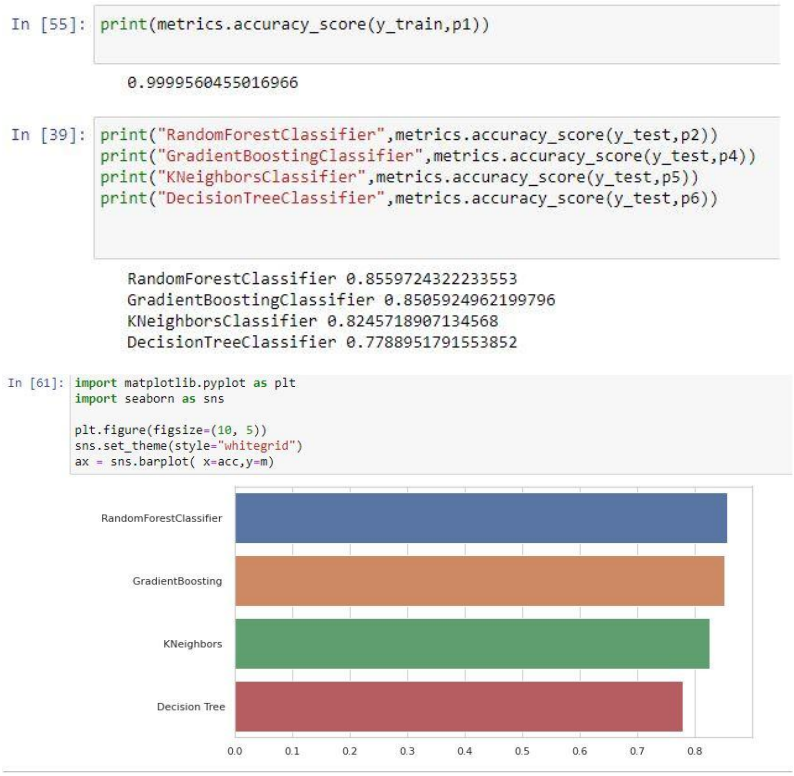
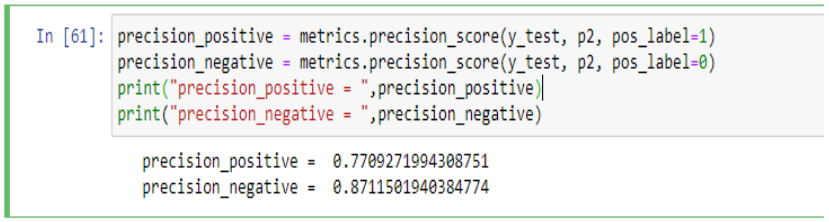
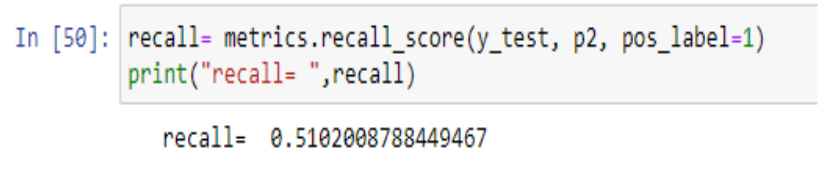
2. Defect Analysis

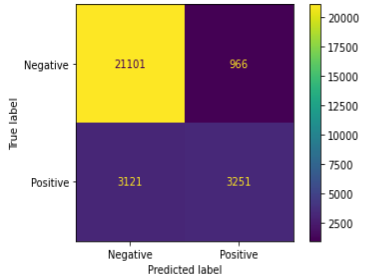
This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

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

9. RESULTS:

9.1. Performance Metrics

S.NO.	PARAMETER	SCREENSHOT
1.	**Accuracy of various model** RandomForestClassifier = 0.855972432223 GradientBoostingClassifier =0.85059249621 KNeighborsClassifier = 0.824571890713 DecisionTreeClassifier = 0.778895179155	
2.	**Precision** **precision_positive = 0.7709271994308751** **precision_negative = 0.8711501940384774**	
3.	**Recall** **recall= 0.5102008788449467**	

4.	**Confusion Matrix**	In [64]:
		`metrics.plot_confusion_matrix(RFC, x_test,y_test , display_labels=['Negative', 'Positive'])`
		`/opt/conda/envs/Python-3.9/lib/python3.9/site-packages/sklearn/utils/deprecation.py:87: Fut`
		`unction `plot_confusion_matrix` is deprecated in 1.0 and will be removed in 1.2. Use one of`
		`or ConfusionMatrixDisplay.from_estimator.`
		`warnings.warn(msg, category=FutureWarning)`
		Out[64]: `<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fa92c752eb0>`
		 
		In [63]:
		`confusion = metrics.confusion_matrix(y_test, p2)`
		`confusion.ravel()`
		Out[63]: `array([21101, 966, 3121, 3251])`

10. ADVANTAGES & DISADVANTAGES:

10.1. Advantages:

- Provides more information regarding rainfall, which is especially valuable in uncertain weather conditions
- Avoid unnecessary floods by opening dams aided by effective rainfall prediction
- Biggest beneficiaries include farmers and fishermen, who can plan their livelihood according to the predictions
- Crucial during monsoon in helping the government find evacuation areas to avoid loss of human life.

10.2. DisAdvantages:

- Geographically limited: Since the data was collected from limited places , predictions are useful only for the people located in those areas.
- Inaccurate data collection leads to poor performance of the prediction model
- Limited data set: Dataset should be enhanced for worldwide operability

11. CONCLUSION:

Rainfall has been a major concern these days. Predicting their occurrence for preventing lifeloss to humankind and environment is an utmost societal needed problem of the society.Irregular heavy rainfall may lead to the destruction of crops, heavy floods that can cause harm to human life. A model was developed to predict rainfall based on historical data in India, thereby helping farmers and fishermen better plan their livelihood around rainfall.

12. FUTURE SCOPE:

Enhancing the dataset helps in overcoming the geographical limitations of the current model based in India. A bigger and enhanced dataset provides worldwide operability to the rain forecasting model. Rain forecasting at a particular time allows a farmer to plant and use the available water in the farm to water the crop until rains come. Such an irrigation schedule is designed to safeguard that no water is wasted. A recommendation system integrated with the prediction system shall ultimately has a positive impact on society.