A PROJECT REPORT

ON

EXPLORATORY ANALYSIS OF RAINFALL DATA IN INDIA FOR AGRICULTURE

Submitted By:

1.Nithya U
2.Manikandan S
3.Preethika M
4.Poorna S

1. INTRODUCTION

1. Project Overview
2. Purpose

2. LITERATURE SURVEY

1. Existing problem
2. References
3. Problem Statement Definition

3. IDEATION & PROPOSED SOLUTION

1. Empathy Map Canvas
2. Ideation & Brainstorming
3. Proposed Solution
4. Problem Solution fit

4. REQUIREMENT ANALYSIS

1. Functional requirement
2. Non-Functional requirements

5. PROJECT DESIGN

1. Data Flow Diagrams
2. Solution & Technical Architecture
3. User Stories

6. PROJECT PLANNING & SCHEDULING

1. Sprint Planning & Sprint Delivery Schedule

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

1. Feature 1
2. Feature 2

8. TESTING

9. RESULTS

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

India is an agricultural nation, thus a healthy monsoon will keep the secondary agro-based economy steady. The length of the monsoon season determines how quickly the economy grows each year. A severe monsoon may destroy some crops, which might lead to a shortage of certain agricultural items, which could then lead to food inflation, insecurity, and discontent among the populace. In our investigation, we want to comprehend how India's rainfall has changed over time, by month and other sub divisions. Predicting crop yields is a significant agricultural issue. Weather factors including temperature, rainfall, and other factors heavily influence agricultural productivity. For the purpose of managing agricultural risk and generating projections for the future, accurate knowledge of crop production history is crucial.

Rainfall has been a major concern these days. Weather conditions have been changing for time being. Rainfall forecasting is important otherwise, it may lead to many disasters. Irregular heavy rainfall may lead to the destruction of crops, heavy floods that can cause harm to human life. It is important to exactly determine the rainfall for effective use of water resources, crop productivity, and pre-planning of water structures. This comparative study is conducted concentrating on the following aspects: modeling inputs, Visualizing the data, modeling methods, and pre-processing techniques. The results provide a comparison of various evaluation metrics of these machine learning techniques and their reliability to predict rainfall by analysing the weather data.

1.2 Purpose

Rainfall has been a major source of concern recently. For the time being, weather patterns are shifting. Rainfall forecasting is critical since it may lead to a variety of calamities. Irregular severe rainfall can destroy crops and produce flooding, which can endanger human life. It is critical to precisely estimate rainfall in order to make the best use of water resources, increase crop output, and prepare ahead of time for water

infrastructure. The project's goal is to construct a forecasting machine learning-based model that will be important in the development of an early warning system that can minimise threats to people and property while also enhancing agricultural work management.

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.A bad rainfall prediction can affect the agriculture mostly framers as their whole crop is depend on the rainfall and agriculture is always an important part of every economy. So, making an accurate prediction of the rainfall somewhat good.Now climate change is the biggest issue all over the world. Peoples are working on to detect the patterns in climate change as it affects the economy in production to infrastructure. So as in rainfall also making prediction of rainfall is a challenging task with a good accuracy rate. Making prediction on rainfall cannot be done by the traditional way, so scientist is using machine learning and deep learning to find out the pattern for rainfall prediction.

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.

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 helps 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.


What are all the operation done in this
How the visualization of data done
What's new in this
How the data set are generated
How this is used in Agriculture
Says
Takes much time to complete process
May be this isn't the best model
Is the data sets are accurate
Is that prediction becomes true
How they handle data set
Thinks
USER
Does
Feels
Check the website
Exploratory data analysis on dataset
analyze the weather forecasting data.
Showcase rainfall rate of a particular state,city,area
predicting rainfall rate of a region
Decision Making
Doubting
Excitement
Fear
Happy

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.

Nithya

the software is free to use

use the weather report for rain production

Give details of safty measure in rainfall

create attractive user interface design

by rainfall data help in the indian agriculture

create user manual for user help

Manikandan

we can use web application in all the device like android,jos,windows

using this data we can identify the furthure weather conditional also

collect the datas from previous year and months it give accurate reports

we can analyse rainfall rate and it give daily update about the climate

This application is based on location based on it was able to find the weather conditional about any place

This application occupied less storage so we can instant in low end device

preethika

The application user friendly

Preduct flood due to rainfall

predict the threat for the crops

Prediction history is stored

give safty precautions on heavy rainfall

the application is open source

Poorna

Analyzing the features of the data according to the crops need

The data includes the rainfall monthly,seasonally,and annually

Exploiting data mining technique for rainfall prediction

Creating a well integrated system which can predict the rainfall and updation

Train,test and then analyze the results of applied mode is under consideration

can effectively predict the rainfall through previous data training

3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem tobesolved)	● Climate is a important aspect of human life. So, the Prediction should accurate as much as possible. In this paper we try to dealwith the prediction of the rainfall which is also a major aspect of human life and which provide the major resource of humanlife which is Fresh Water. ● Now climate change is the biggest issue all over the world. Peoples are working onto detectthe patterns in climatechange as it affects the economy in production to infrastructure
2.	Idea / Solution description	● In rainfall also making predictionof rainfall is a challenging task with a good accuracy rate. Making prediction on rainfall cannot be doneby the traditional way, so scientist is using machine learning and deep learning to findout thepattern for rainfall prediction. ● Provides extrasupport to maintain

		theagriculture.
3.	Novelty / Uniqueness	● This application is useful forthebeginners in agriculture. ● Seed maturity selection features are available.
4.	Social Impact/ Customer Satisfaction	● Different types of crops can be plantedfor goodhealth. ● Helps in producing healthy crops andgood fields.
5.	Business Model(Revenue Model)	● This comparative study is conducted concentrating on the following aspects: modeling inputs, Visualizingthe data,modeling methods, and pre- processing techniques. The results providea comparison of various evaluation metrics of these machine learning techniques and theirreliability to predictrainfall by analyzing the weather data. Wewill be using classification algorithms such as Decisiontree, Random forest,KNN, and xgboost.

Problem Solutionfit

CUSTOMER SEGMENT(S) /CS

- Farmers
- Sale Peoples
- public

JOBS-TO-BE-DONE / J&P PROBLEMS

. To forecast the amount of rainfall and the crops that might be cultivated in a specific area based onthe amount of rainfall.

TRIGGERS/ TR

To develop a weatherprediction invention to conserve waterand agriculture.

EMOTIONS:

EM Lack of available storagefor water in arid areas - relyingon rainfall

AVAILABLE SOLUTIONS/ AS

Internet -Using onlineweather forecasting resources like Ø Application Knowledge

CUSTOMER CONSTRAINTS /CC

a. Cashless
b. Budget

BEHAVIOUR/ BE

- Find the best crop that could be cultivated in their area andforesee the advantages.
- Customers will experience inner calm and relaxation, which is direct association.

- **PROBLEM ROOT CAUSE/ RC**
    a. Loss of biodiversity
    b. climate change
    c. Investments

YOUR SOLUTIONSL

a. Considerable need for an effective irrigation system given the growing water shortage
b. Reducing post-harvest losses

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONA

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement(Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration throughLinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Reliability	Theprediction will be provided by the systemerror-free.
FR-4	Performance	The expected output will be produces immediatelyto the user withoutmuch delay.
NFR-5	Availability	The system wouldbe available 24/7

NFR-6	Scalability	Thesystem would be available on web application and any user can login and use it without anydisruptions.

Non-Functional requirements

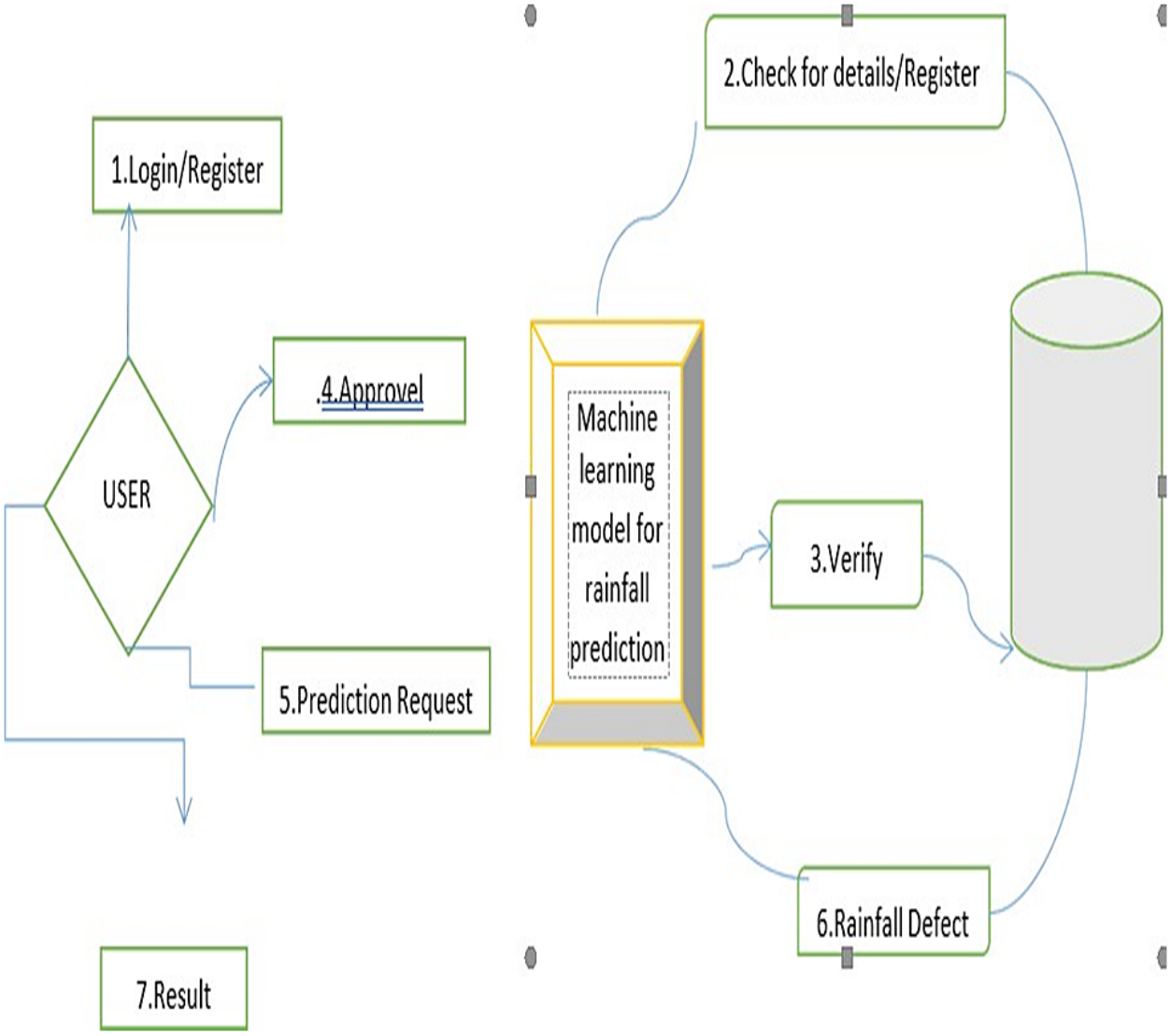
Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Can be used anywhere(remote villages to metropolitan cities), anybody (kids to old age)
NFR-2	Security	Security is given over the model, so the user can usethis with full trust. However, there are no personaldetails required to use this.
NFR-3	**Reliability**	Goodconnectivity and a supporting devicecanprovide goodresults upto an extent.
NFR-4	**Performance**	Thismodel can give a high accuracy prediction.
NFR-5	**Availability**	Anyperson can use this and this is an open-sourcemodel.

DATA FLOW DIAGRAM

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 amountof the systemrequirement graphically. It shows how data entersand leaves the system, what changes the information, and where data is stored.

LEVEL DATA FLOW DIAGRAM



Solution Architecture:

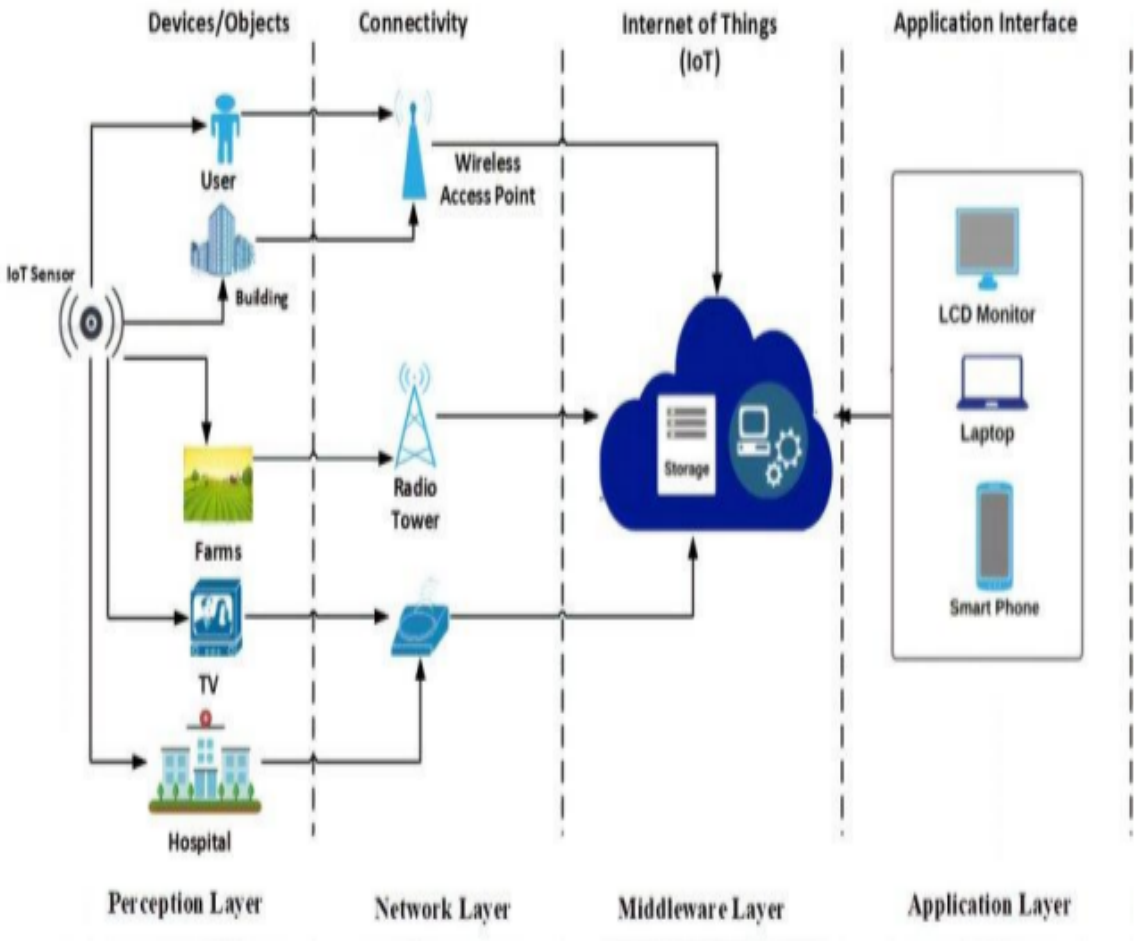
Solution architecture is a complex process – with many sub-processes – that bridges the gapbetween business problems and technology solutions. Its goals are to:

1. Find the best tech solution to solve existing businessproblems.
2. Describe the structure, characteristics, behavior, and other aspectsof the software to projectstakeholders.
3. Define features, development phases, and solutionrequirements.
4. Provide specifications according to which the solution is defined, managed,and delivered

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2Technology architecture associates application components from application architecturewith technology components representing software and hardware components. Its components are generally acquired in the marketplace and can be assembled and configured to constitute the enterprise's technological infrastructure.

Solution & technical architecture



User Stories

User type	Functional Requirement	Userstory number	Userstory / Task	Acceptancecriteria	Priority	Release
Customer (mobileuser)	Registration	USN-1	I can sign up for the application as a user by providing my email address, a password, and a password confirmation.	I can accessmydashboardor account.	High	Sprint-1
		USA-2	When I register for the application as a user, I will get aconfirmation email.	can click theconfirm buttonin a confirmation email.	High	Sprint-2
	Login	USA-3	I can access the application as a user by using my registered email and password.	can get to the system's dashboard.	High	Sprint-1
		USA-4	The user has the option to check their search history and change their password.	New passwordmust be supplied and verification is necessary.	High	Sprint-1

		USA-5	For numeroussystems, the already-existing login information should be used.		Medium	Sprint-1

	Dash Board	USA-6	I can view the page's specifics and navigate through the entire page as a user.	is able to traverse thepages.	Medium	Sprint-1
	Prediction	USA-7	The user canconduct a searchfor the location or area where they want to receive rainfall forecasts.	Only searchesfor regions in India are acceptable.	High	Sprint-1
		USA-8	the analysis or forecast for the selected area for either upcoming or historical events.		High	Sprint-1

		USA-9	The visualisation of the rain is available to users.information for a given time period for a particular region of INDIA.		High	Sprint-1
	News	USA-10	Users can view the most recent agricultural-related news articles.	I am able tosee the news stories.	Medium	Sprint-2
Customercare Executive	Support	USA-11	User inquiries about the system are welcome.	I can clear upmy doubts.	High	Sprint-3
		USA-12	The group must examine each query and fix it inthe upcomingupdate.		High	Sprint-3
		USA-13	Prepare a FAQsession.		Low	Sprint-3

Core DevelopmentTeam	Core Function	USA-14	The programmeshould be designed and developed in such a way that the optimal user interface and		High	Sprint-1

			maintenance are considered.			
		USA-15	On all devices and screen sizes,the websiteis	The user experience should be positive regardless of the platforms or devices.	High	Sprint-1
		USA-16	The updatesmust contain the answers to the asked questions in a timely manner.	The updatesshouldn't havean impact on the already available functionalities.	High	Sprint-1

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 othe r 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

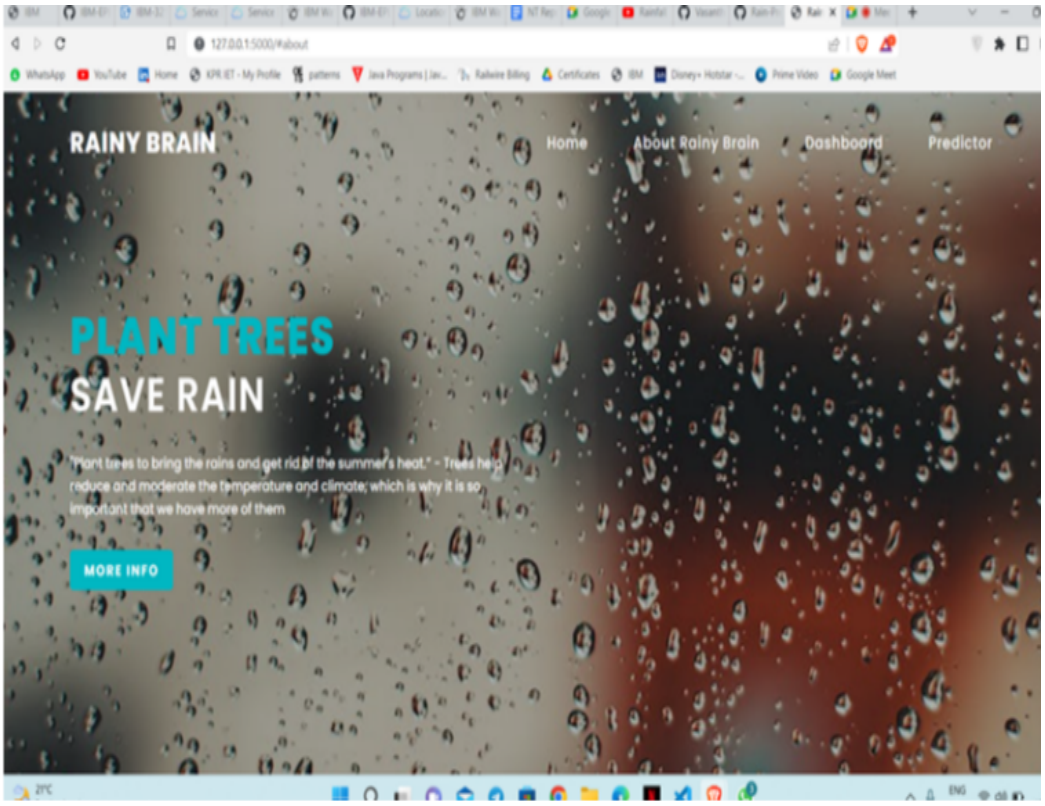
standard metrics - accuracy,precision,f1-score and recall using the Random Forest Algorithm and our proposed model as been achieved.

```
[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 AND RESULTS

8.1 Test Cases



Dashboard

This dashboard is done using a software called PowerBI which is a product of Microsoft. Here I have just attached the images of the dashboard because PowerBI needs oraganizational account. So to see the visualizations interactive I am attaching my **PowerBI** dashboard file. This requires PowerBI software to open the file. The usage of dashboards like these is to bring a better understanding about the dataset and also to bring some beautiful insights

9. RESULT

Advantages

Precipitation and other Earth-observing datasets from NASA are used for forecasting tropical cyclones

Monitoring soil moisture conditions and freshwater availability

Monitoring flood and drought conditions

It facilitates policy decisions regarding the cropping pattern, sowing date, construction of roads and providing drinking water to urban and rural areas.

Disadvantages

*Offering inconclusive results

*Lack of standardized analysis

*small sample population

*Outdated information that can adversely affect the authenticity of information.

10. conclusion:

As a conclusion, we created a particular kind of application that enables users to comprehend how India's rainfall behaves, enabling farmers to make informed plans. As part of our research, we thought about incorporating a few extra features to the application to improve user utility. Some of the added features include the ability for users to sign up for the application using an existing email address or social media account, the ability to get newsletters and daily reports on the application's accuracy.

11.FUTURE SCOPE

The Future Enhancements of the application can be allowed to support in all the upcoming android versions. History can be set to view all the details in the app even if the particular data is deleted from the database. Statistics could be prepared based on the history of the reports. Sharing of analysis can be allowed so that more of the farms could be benefitted.Daily updates through WhatsApp and SMS could be added. Some of

the extra components are like enabling users to register to the application using existing email or social network accounts,and connect with other agricultural people for more technological ideas.

12.APPENDIX

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

```
 rainfall = float(request.form['rainfall'])
# Evaporation
evaporation = float(request.form['evaporation'])
 # Sunshine
sunshine = float(request.form['sunshine'])
# Wind Gust Speed
windGustSpeed = float(request.form['windgustspeed'])
# Wind Speed 9am
windSpeed9am = float(request.form['windspeed9am'])
 # Wind Speed 3pm
 windSpeed3pm = float(request.form['windspeed3pm'])
 # Humidity 9am
 humidity9am = float(request.form['humidity9am'])
# Humidity 3pm
 humidity3pm = float(request.form['humidity3pm'])
 # Pressure 9am
pressure9am = float(request.form['pressure9am'])
 # Pressure 3pm
pressure3pm = float(request.form['pressure3pm'])
# Temperature 9am
temp9am = float(request.form['temp9am'])
# Temperature 3pm temp3pm = float(request.form['temp3pm'])
# Cloud 9am
cloud9am = float(request.form['cloud9am'])
# Cloud 3pm
 cloud3pm = float(request.form['cloud3pm'])
 # Cloud 3pm
 location = float(request.form['location'])
 # Wind Dir 9am
```

```
winddDir9am = float(request.form['winddir9am'])
# Wind Dir 3pm
winddDir3pm = float(request.form['winddir3pm'])
# Wind Gust Dir
windGustDir = float(request.form['windgustdir'])
# Rain Today
rainToday = float(request.form['raintoday']) input_lst = [location , minTemp ,
maxTemp , rainfall , evaporation , sunshine ,

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

humidity9am , humidity3pm , pressure9am , pressure3pm , cloud9am , cloud3pm
, temp9am , temp3pm ,
rainToday , month , day] pred = model.predict(input_lst) output = pred
If output == 0:
    return render_template("after_sunny.html")
else: return render_template("after_rainy.html")
    return render_template("predictor.html")

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

13.Github link:

https://github.com/IBM-EPBL/IBM-Project-33552-1660222589