

EXPLORATORY ANALYSIS OF RAINFALL DATA IN INDIA FOR AGRICULTURE

TEAM ID: PNT2022TMID10697

1.INTRODUCTION

1.1 PROJECT OVERVIEW:

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. We know that Agriculture is the primary source of the Indian economy. During the last 10 years, there have been vast improvements in technology and this has increased the rate of global warming, pollution of air, water, noise, dust, etc. This resulted in drastic changes in climate and weather conditions. Rainfall is a key part of the hydro logical cycle and alteration of its patterns directly affects the water resources. Changes in the pattern have become a major issue for harvesting crops. Hence, the research on changes in rainfall occurrences is the most sustainable water resource management.

Technology is much more advanced now. Machine Learning has become trending for predictions. It contains various algorithms that can help us in predicting our required value. One major concern is the selection of algorithms. We have to select the algorithms based on our problem statement. Supervised Algorithms are classified as classification and regression algorithms. Regression algorithms are perfect for predicting when a dataset has a single variable (dependent variable) or more variables (independent variables).

Regression analysis is an important tool for modeling and analyzing information. It is used for predictive analysis, for instance, forecasting rainfall or weather, predicting trends in business, finance, and marketing. It can also be used for correcting errors and also provide quantitative support.

1.1.1 REGRESSION ANALYSIS:

Regression analysis deals with the dependence of one variable (called as dependent variable) on one or more other variables, (called as independent variables) which is useful for estimating and/ or predicting the mean or average value of the former in terms of known or fixed values of the latter. For example, the salary of a person is based on his/her experience here, the experience attribute is independent variable salary is dependent variable. Simple linear regression defines the relationship between a single dependent variable and a single independent variable. The below equation is the general form of regression. $y = \beta_0 + \beta_1 x + \varepsilon$ where β_0 and β_1 are parameters, and ε is a probabilistic error term. Regression analysis is a vital tool for modeling and analyzing information. It is used for predictive analysis that is forecasting of rainfall or weather, predicting trends in business, finance, and marketing. It can also be used for correcting errors and also provide quantitative support.

1.1.2 MACHINE LEARNING TECHNIQUES:

Random forest is an ensemble learning method and supervised algorithm of machine learning technique. The ensemble learning method is a learning process in which different algorithms or the same algorithm is repeatedly combined to provide the most accurate prediction technique. It combines several similar algorithms, namely the various decision trees, to produce trees like a forest. Thus, the term “Random Forest” applies to both classification and regression functions. The basic procedures required in carrying out the random forest algorithm are first to choose K records at random from the data collection. And then create a decision tree using these K records. Decide how many trees you want to use in your algorithm and replicate the previous steps. Each forest tree estimates the division under which the new record belongs when it comes under classification. Eventually, the new form is granted to the class that receives the most votes. Since several trees and every tree are trained on a sample of data, the random forest algorithm is not biased.

1.1.3 LIBRARIES AND PLATFORMS:

NUMPY is a numerical Python application that provides fast mathematical calculation functions for computation. It can be used to read data in arrays and for computing procedures. PANDAS can read and write various files and directories. Furthermore, data processing of information frameworks makes the data source file highly performed and easy to use. SEABORN is based on matplotlib, which is used for data visualization in python data that provides a high-level platform for displaying appealing and detailed statistical graphics. MATPLOTLIB is a Python twodimensional plotting library that produces high-quality reports in various hardcopy and graphical formats. Matplotlib is used in Python scripts, Jupyter notebooks, IPython shells, Web frameworks, and four graphical user interface toolkits. Matplotlib strives

to make things easy, fast and also make complex things possible. Using a few code lines, you can generate graphs, histograms, bar charts, scatter plots, etc. Anaconda Navigator is a user interface software application that allows you to quickly launch applications and access conda packages, configurations, and channels without using command-line functions. It is compatible with Linux, Windows, and OS. Jupyter notebook is available in anaconda navigator. It is an open-source computational notebook that allows researchers to combine source code, computational performance, descriptive language, and multimedia tools into a single document.

1.2 PURPOSE:

The purpose of the study is the prediction of the rainfall using historical monthly data based on artificial intelligence methodologies such as support vector machine and artificial neural network. The extraction procedures/algorithms will produce the output by classification of the data according to the categories using SVM and ANN. The similar data will be grouped for the accurate and precise information that will predict rainfall more correctly and with perfect figures. The accurate and exact predictions will help in developing the more appropriate strategies for agriculture and water reserves and will also be informed about the flood to implement precautionary measures. Amongst all weather happenings, rainfall plays the most imperative part in human life.

Human civilization to a great extent depends upon its frequency and amount to various scales. Several stochastic models have been attempted to forecast the occurrence of rainfall, to investigate its seasonal variability, to forecast yearly/monthly rainfall over some geographical area. This is the monthly data with all parameters of rainfall including wind speed, direction, air pressure, humidity and temperature. The aim of the proposed study is too effective and efficient in predicting the rainfall with accuracy and precision.

Rainfall prediction is significant not only on the micro but also on the macro level. The study is of significance with respect to its vital contribution in the field of agriculture, water reserve management, flood prediction and management with an intention to ease the people by keeping them updated with the weather and rainfall prediction. It is also important to be utilized by the agricultural industries for keeping their crops safe and ensure the production of seasonal fruits and vegetables by updated rainfall prediction. The rainfall prediction will impressively help in dealing with the increasing issue of water resource management; as water is a scarce resource and it needs to get saved for the benefit of human beings themselves. Also, it will help the people to manage and plan their social activities accordingly.

2.LITERATURE SURVEY

2.1 EXISTING PROBLEM:

Over the previous decade, academic and commercialized databases have been extending at exceptional rates. Capture advanced perception from such databases is hard, expansive and timeconsuming if done manually. It is hopeless when data exceeds definite limits of size and complexity. For this reason, during the previous years the automated analysis and visualization of huge multidimensional datasets has been the center of attention on scientific research. The fundamental aim is to observe rules and relationships in the data, thereby gaining attain to invisible and potentially valuable knowledge. Artificial Neural Networks are a hopeful part of this broad field. Motivated by advances in biomedical research, they shape a class of algorithms that goal to reproduce the neural structures of the brain. The reason is that ANN (Artificial Neural Network) model is based on 'prediction' by smartly 'analyzing' the trend from an already existing voluminous historical set of data.

Apart from ANN, the other models are either mathematical or statistical. These models have been found to be very accurate in calculation, but not in prediction as they cannot adapt to the irregularly varying patterns of data which can neither be written in form of a function, nor deduced from a formula. These real-life situations have been found to be better interpreted by 'artificial neurons' which can learn from experience, i.e by back propagation of errors in next guess and so on. This may lead to a compromise in accuracy but give us a better advantage in 'understanding the problem', duplicating it or deriving conclusions from it. Amongst all weather happenings, rainfall plays the most imperative part in human life. Human civilization to a great extent depends upon its frequency and amount to various scales. Several stochastic models have been attempted to forecast the occurrence of rainfall, to investigate its seasonal variability, to forecast yearly/monthly rainfall over some geographical area.

2.2 REFERENCES:

- [1] Thirumalai, Chandrasekhar, et al. "Heuristic prediction of rainfall using machine learning techniques." 2017 International Conference on Trends in Electronics and Informatics (ICEI). IEEE, 2017.
- [2] Geetha, A., and G. M. Nasira. "Data mining for meteorological applications: Decision trees for modeling rainfall prediction." 2014 IEEE International Conference on Computational Intelligence and Computing Research. IEEE, 2014

[3] Parmar, Aakash, Kinjal Mistree, and Mithila Sompura. "Machine learning techniques for rainfall prediction: A review." 2017 International Conference on Innovations in information Embedded and Communication Systems. 2017.

[4] Dash, Yajnaseni, Saroj K. Mishra, and Bijaya K. Panigrahi. "Rainfall prediction for the Kerala state of India using artificial intelligence approaches." Computers & Electrical Engineering 70 (2018): 66-73.

[5] Dr.C K Gomathy, Article: An Effective Innovation Technology In Enhancing Teaching And Learning Of Knowledge Using Ict Methods, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrct) E-Assn: 2395-5325 Volume3, Issue 4,P.No-10-13, April '2017

2.3 PROBLEM STATEMENT DEFINITION:

Rainfall forecasting is very important because heavy and irregular rainfall can have many impacts like the destruction of crops and farms, damage of property so a better forecasting model is required for an early warning that can reduce the risks to life and property and also helps to manage the agricultural farms in a better way. Heavy rainfall is a cause for natural disasters like floods and drought that square measure encountered by individuals across the world each year. Many models are developed to evaluate the rainfall and for predicting the likeliness of rain. These models are based on both supervised and unsupervised machine learning algorithms. Taking into consideration of overall rainfall will not help us to know if it rains in specific conditions. Accuracy is the major concern in machine learning. We are going to understand the data and then train the model accordingly to predict whether if it rains under given conditions or not.

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. 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 AND PROPOSED SOLUTIONS

3.1 EMPATHY MAP CANVAS:

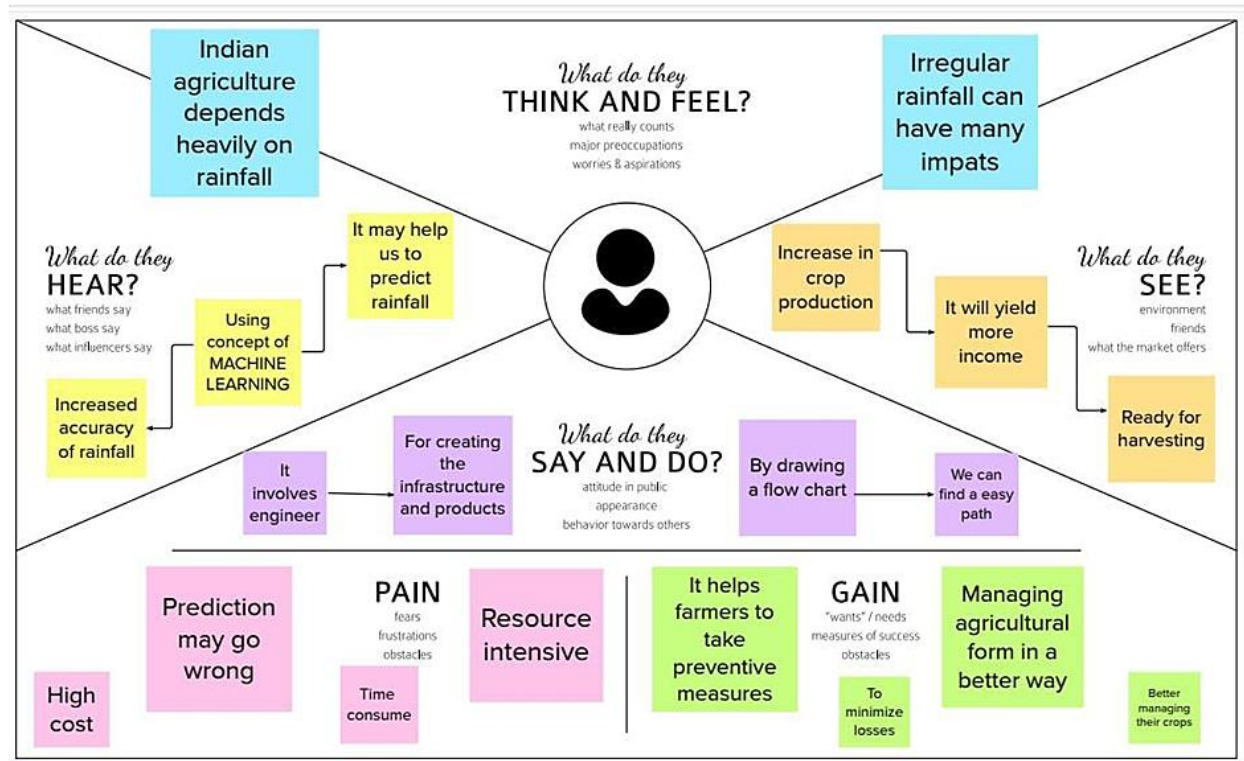


figure 3.1 empathy map

An empathy map **helps to map what a design team knows about the potential audience**. This tool helps to understand the reason behind some actions a user takes deeply. This tool helps build Empathy towards users and helps design teams shift focus from the product to the users who are going to use the product.

3.2 IDEATION & BRAINSTORMING:

Ideation is often closely related to the practice of brainstorming, a specific technique that is utilized to generate new ideas. A principal difference between ideation and brainstorming is that **ideation is commonly more thought of as being an individual pursuit, while brainstorming is almost always a group activity**.

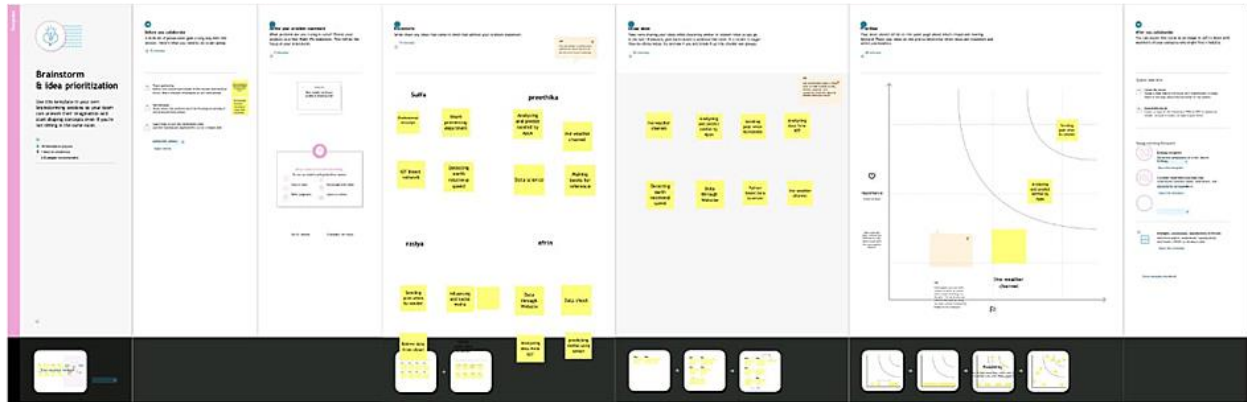


figure 3.2 ideation & brainstorming

3.3 PROPOSED SOLUTION:

S.NO	PARAMETER	DESCRIPTION
1	Problem Statement (Problem to be solved)	<p>Rainfall has been major concern these days.</p> <p>Rainfall is a key part of the hydro logical cycle and alternation of its pattern directly affects the water resources.</p> <p>Changes in the pattern have become a major issues for harvesting crops.</p>
2	Idea/ Solution description	<p>Technology is much more advanced now. Machine learning has become trending for predictions.</p> <p>It can contains various algorithms that can help us in predicting our required value.</p>
3	Novelty / Uniqueness	<p>The use of machine learning techniques has increased the accuracy of rainfall prediction systems by exploring the hidden pattern of historical weather data. This application is useful for beginners in agriculture.</p>

4	Social impacts / Customer satisfaction	<p>Helps in producing fields and healthy crops.</p> <p>Any type of malfunction in the weather sensor can also compromise the accuracy of the proposed rainfall prediction system.</p>
5	Business model(revenue model)	<p>This comparative study is conducted concentrating on the following aspects: modeling inputs, visualizing the data, modeling methods, and preprocessing techniques.</p>
6	Scalability of the solution	<p>Machine learning used for accurate prediction in which the given dataset is cleaned and normalized before the classification process begin.</p>

table 3.3 proposed solution

3.4 PROBLEM SOLUTION FIT:

Problem-solution fit is a term used to describe the point validating that the base problem resulting in a business idea really exists and the proposed solution actually solves that problem. Validate that the problem exists: When you validate your problem hypothesis using real-world data and feedback.

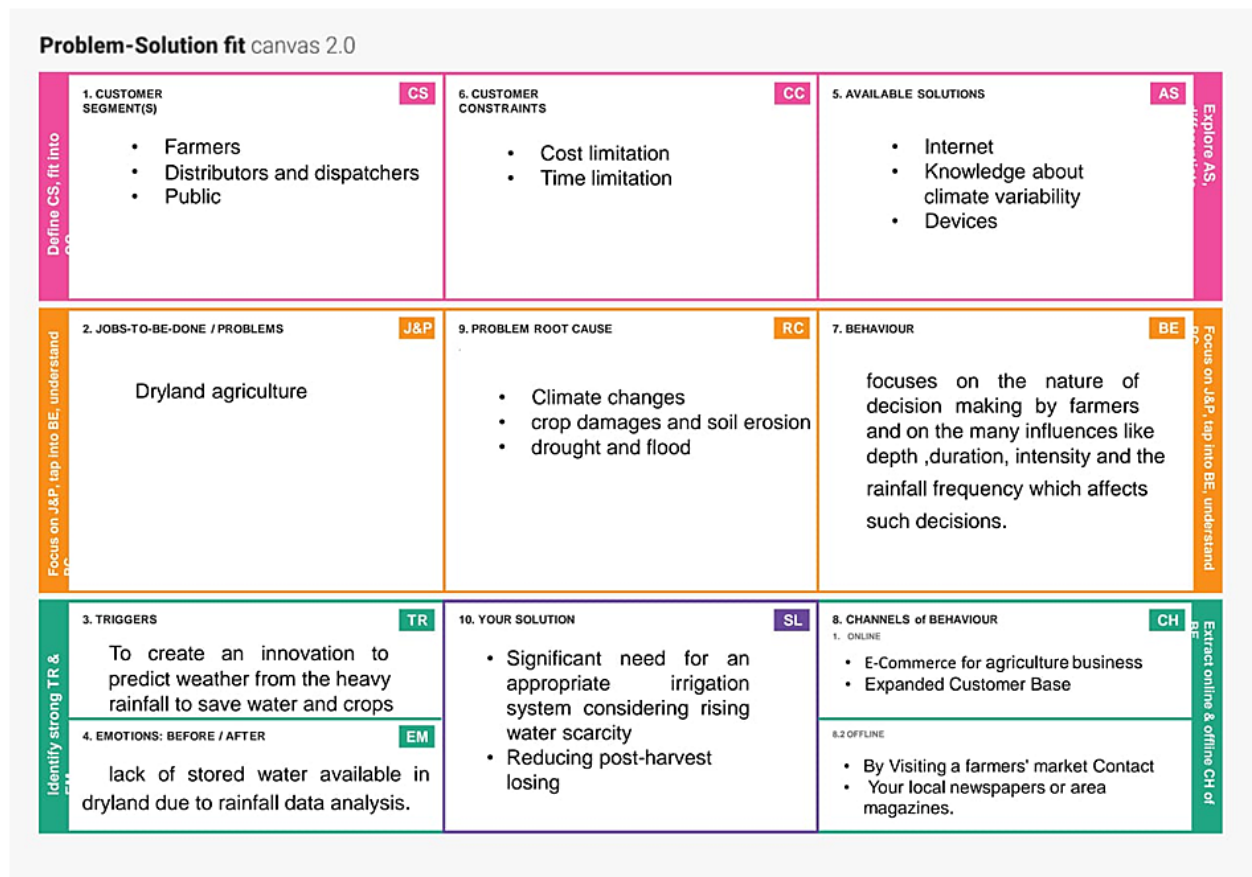


figure 3.4 problem solution fit

4.REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT:

Functional requirements are **the desired operations of a program, or system as defined in software development and systems engineering**. The systems in systems engineering can be either software electronic hardware or combination software-driven electroni

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	<ul style="list-style-type: none">• Registration through Form• Registration through Gmail• Registration through LinkedIn
FR-2	User Confirmation	<ul style="list-style-type: none">• Confirmation via Email• Confirmation via OTP
FR-3	App installation	<ul style="list-style-type: none">• Installation through link• Installation through play store
FR-4	Settings geofence	<ul style="list-style-type: none">• Setting by user to analysis rainfall
FR-5	User Interface	<ul style="list-style-type: none">• User Login Form.• Admin Login Form.
FR-6	Analysis Rainfall Data	<ul style="list-style-type: none">• Analyzing data via app• Analyzing data via SMS
FR-7	Dataset	<ul style="list-style-type: none">• Data collection for Rainfall prediction is taken from both the weather prediction website and contains several atmospheric parameters.• That values be held under such limits for good data analysis performance.
FR-8	Preprocessing of dataset	<ul style="list-style-type: none">• Ensures the consistency of extraction performance.• The collection of data used in this system includes rainfall data from many regions within India.
FR-9	APC	<ul style="list-style-type: none">• APC used for compensate the rainfall attenuation.

table 4.1 functional requirement

4.2 NON FUNCTIONAL REQUIREMENT:

Nonfunctional Requirements (NFRs) **define system attributes such as security, reliability, performance, maintainability, scalability, and usability.** They serve as constraints or restrictions on the design of the system across the different backlogs.

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	<ul style="list-style-type: none">• The prediction helps people to take preventive measures and moreover the prediction should be accurate.
NFR-2	Security	<ul style="list-style-type: none">• Model is essential for an early warning that can minimize risks to life and property and also managing the agricultural farms.• It reduce the impacts like destruction of crops and farms, damage of property.
NFR-3	Reliability	<ul style="list-style-type: none">• Portable• Easy to access• Flexibility• Scalability
NFR-4	Dynamicity	<ul style="list-style-type: none">• Every ML application is a case of Dynamical Machine Learning.
NFR-5	Availability	<ul style="list-style-type: none">• Exactly determine the rainfall for effective use of water resources, crop productivity and pre-planning of water structures.• Get rainfall details at anytime.• Know the current weather.
NFR-6	Scalability	<ul style="list-style-type: none">• Farmers need not worry about their crops.
NFR-7	Valuability	<ul style="list-style-type: none">• The system should be able to delivery promptly to the financing authority.
NFR-8	Desirability	<ul style="list-style-type: none">• Navigation should be made easy.• The user should be able to search and find the information he needs without much hassle.

table 4.2 non functional requirement

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

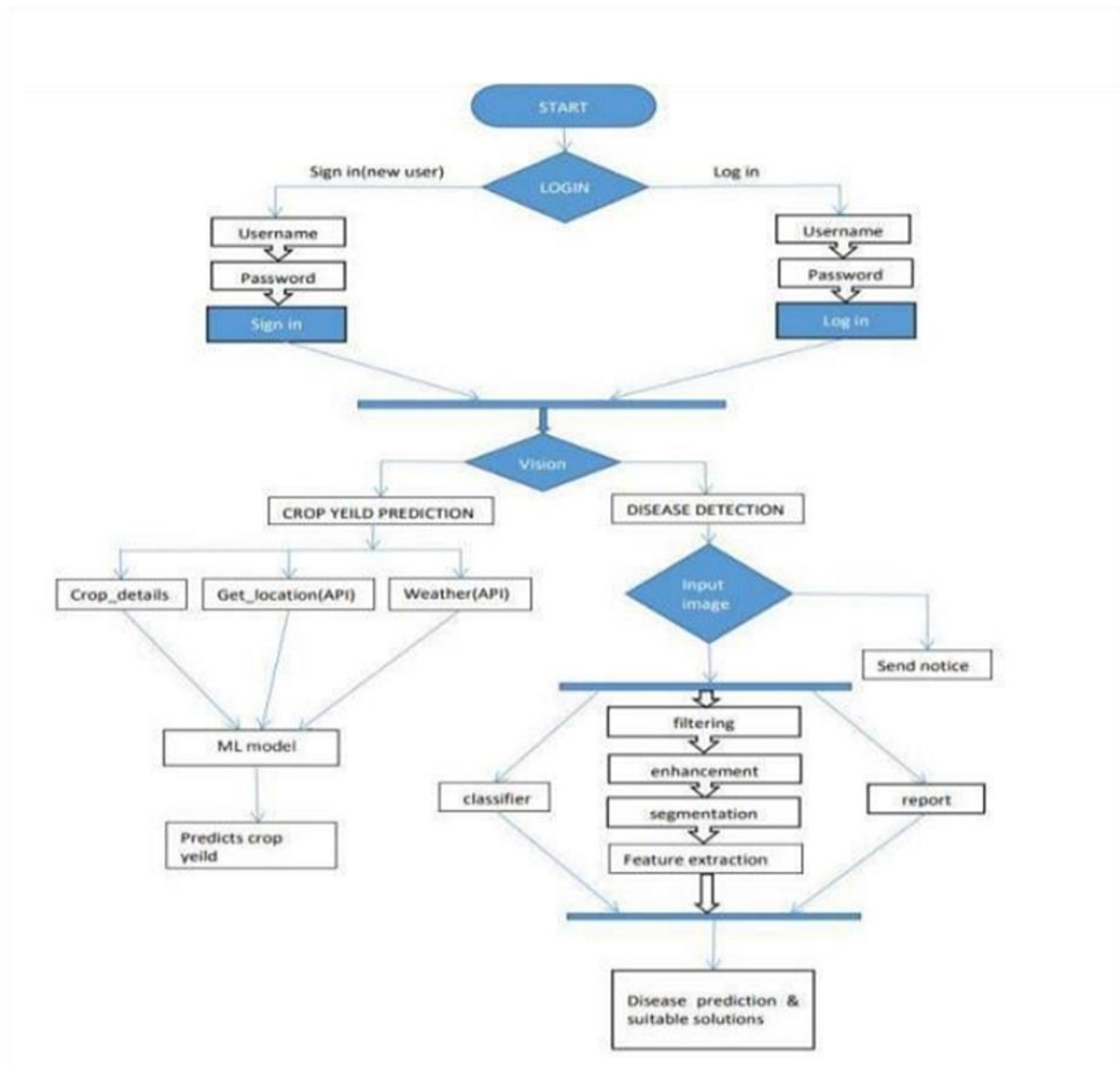


figure 5.1 data flow diagram

5.2 SOLUTION & TECHNICAL ARCHITECTURE:

Solution architecture (SA) is **an architectural description of a specific solution**. SAs combine guidance from different enterprise architecture viewpoints (business, information and technical), as well as from the enterprise solution architecture (ESA).

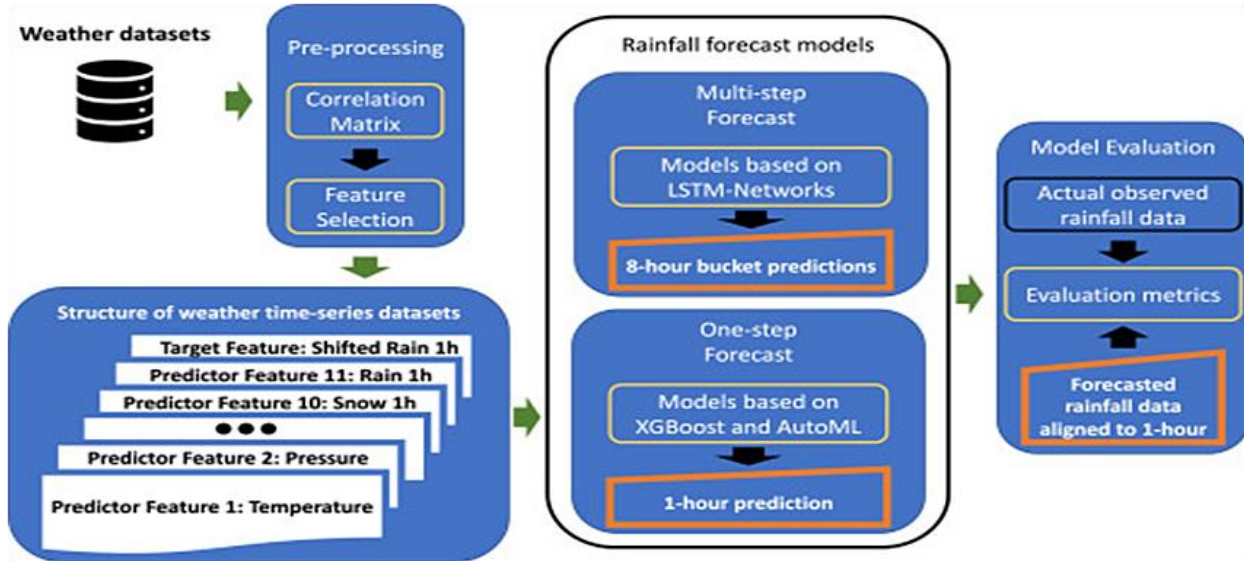


figure 5.2 technical architecture

s.no	Components	Description	Technology
1	User interface	Web GUI	HTML, CSS, java script/Angular Js / React Js.
2	Application logic-1	HYDRO-LOGICAL MODELLING	Python toolkit
3	Application logic-2	RUN OFF MODELLING	ArcGIS analyst tool
4	Database	Gauge dataset , merge data set, satellite only data sets.	MySQL, NoSQL.
5	Cloud database	A database service on cloud provides data security by design.	IBM DB2, NoSQL
6	File storage	Cloud stores the meteorological function	AWS, CWS

7	External API-1	Weather stack	REST API
8	External API-2	Visual crossing API	IBM cloud API
9	Machine learning model	MLP is the most popular neural network for predict rainfall.	Object recognition model
10	Infrastructure	Application deployment on local system / cloud local server configuration:4GB RAM , I5 core Cloud server configuration: memory, space, speed	Local, cloud foundry, kubernetes.

table 5.2 components and technology

S.no	Characteristics	Description	Technology
1	Open source frameworks	Software tools are used to processing the hydro logical data.	Standard Microsoft windows, openmeteo.org frameworks.
2	Security implementations	Security protocols and schemes for ensuring the availability, integrity and the confidentiality of the system and data.	SHA-256, Encryption, OWSAP.
3	Scalable architecture	The micro services that not only handle a large no of tasks or requests at the time.	REST, micro services framework, python
4	Availability	Agribale pocket rain gauge	Digital database forecasting, PWS.

5	Performance	Provides the categorical statics , descriptive statics and decomposition on the basis of rainfall intensity classification.	TRMM, GPM, IMERG.
---	-------------	---	-------------------

table 5.2 application characteristics

5.3 USER STORIES:

A user story is **an informal, general explanation of a software feature written from the perspective of the end user**. Its purpose is to articulate how a software feature will provide value to the customer. It's tempting to think that user stories are, simply put, software system requirements.

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	3	High	Sulfa
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	3	High	Seyed Rasiyammal
Sprint-1		USN-3	As a user, I can register for the application through LinkedIn	1	Low	Preethika
Sprint-1		USN-4	As a user, I can register for the application through Form	2	Medium	Afrin Shahnaj Rasiyammal
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	2	High	Sulfa
Sprint-1	Dashboard	USN-6	As a user , I can view the details about the page and navigate through the entire pages	1	Medium	Seyed Rasiyammal
Sprint-1	Prediction	USN-7	User can search for the area / place where the user wants to know the prediction of rainfall .	2	High	Preethika
		USN-8	The prediction or analysis for the desired region for the future or past events respectively	2	Medium	Afrin Shahnaj

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
		USN-9	User can see the visualization of the rainfall data for the specific region in INDIA for a specified time period.	2	Medium	Sulfa
		USN-10	Remote Sensing data is a powerful tool for Estimating crop yield	1	Low	Afrin Shahnaj
Sprint-2	News	USN-11	User can view the latest news articles related to agriculture	7	High	Preethika
Sprint-2		USN-12	User can change his/her password and can view the account details and search history	7	High	Sayed Rasiyammal
Sprint-3	Support	USN-13	User can ask queries about the system.	5	High	Sulfa
Sprint-3	Support(Customer care executive)	USN-14	The team must analyse all the queries and debug it in the next update.	5	High	Sayed Rasiyammal
		USN-15	User can give the feedback on the accuracy of the prediction and on the user interface	4	High	Preethika
Sprint-4	Core Function	USN-16	Design and develop the application in such a way that the best user interface and maintenance should be taken care of	5	High	Afrin Shahnaj
		USN-17	The updates should be on time with the solutions of the raised queries	4	High	Sayed Rasiyammal
		USN-18	The website is responsive on all the devices and the screen sizes. User experience should be good irrespective of the devices or platforms.	5	High	Sulfa

table 5.3 user stories

6.PROJECT PLANNING & DESIGNING

6.1 SPRINT PLANNING AND ESTIMATION:

Sprint planning requires some level of **estimation**. The team needs to define what can or cannot be done in the sprint: **estimated** effort vs capacity.

Sprint	Total story points	Duration	Sprint start date	Sprint end date	Story points completed	Sprint release date
Sprint-1	19	6 days	24 Oct 2022	31 Oct 2022	-	31 Oct 2022
Sprint-2	14	6days	31 Oct 2022	05 nov 2022	-	05 nov 2022
Sprint-3	14	6 days	07 nov 2022	12 nov 2022	-	12 nov 2022
Sprint-4	14	6 days	14 nov 2022	19 nov 2022	-	19 nov 2022

table 6.1 sprint planning

6.2 SPRINT DELIVERY SCHEDULE:

A sprint schedule is a **document that outlines sprint planning from end to end**. It's one of the first steps in the agile sprint planning process—and something that requires adequate research, planning, and communication.

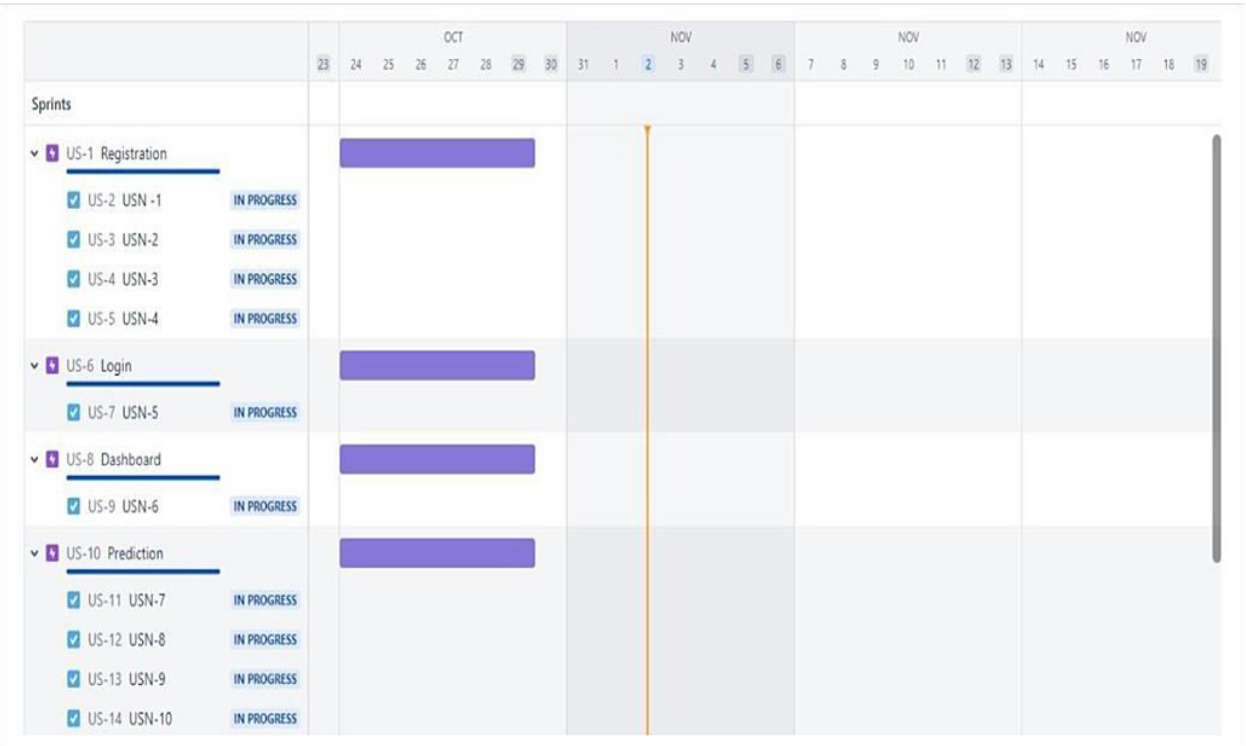


figure 6.2.1 sprint delivery plan 1

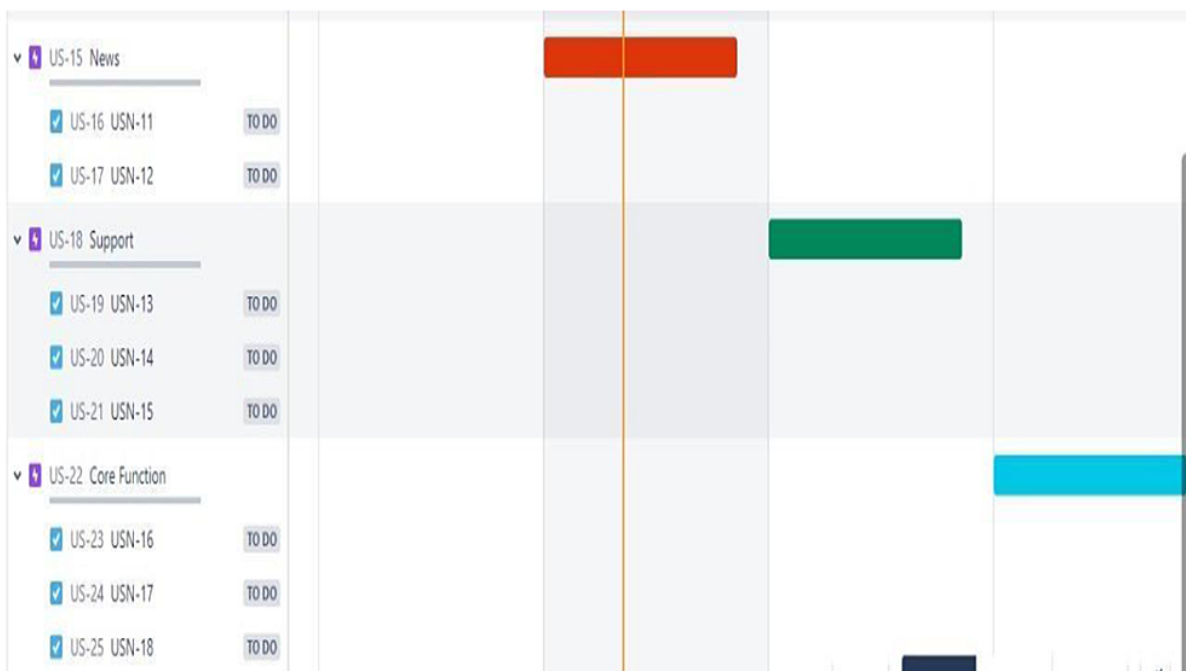


figure 6.2.2 sprint delivery plan 2

6.3 REPORTS FROM JIRA:

Reporting **helps you track and analyze your team's work throughout a project**. Jira Software has a range of reports that you can use to show information about your project, versions, epics, sprints, and issues

The Burnup report **tracks your team's progress as they complete a sprint**. It shows the amount of work that has been completed and compares it with the work that was planned. This helps you identify problems such as scope creep or a deviation from the planned project path.

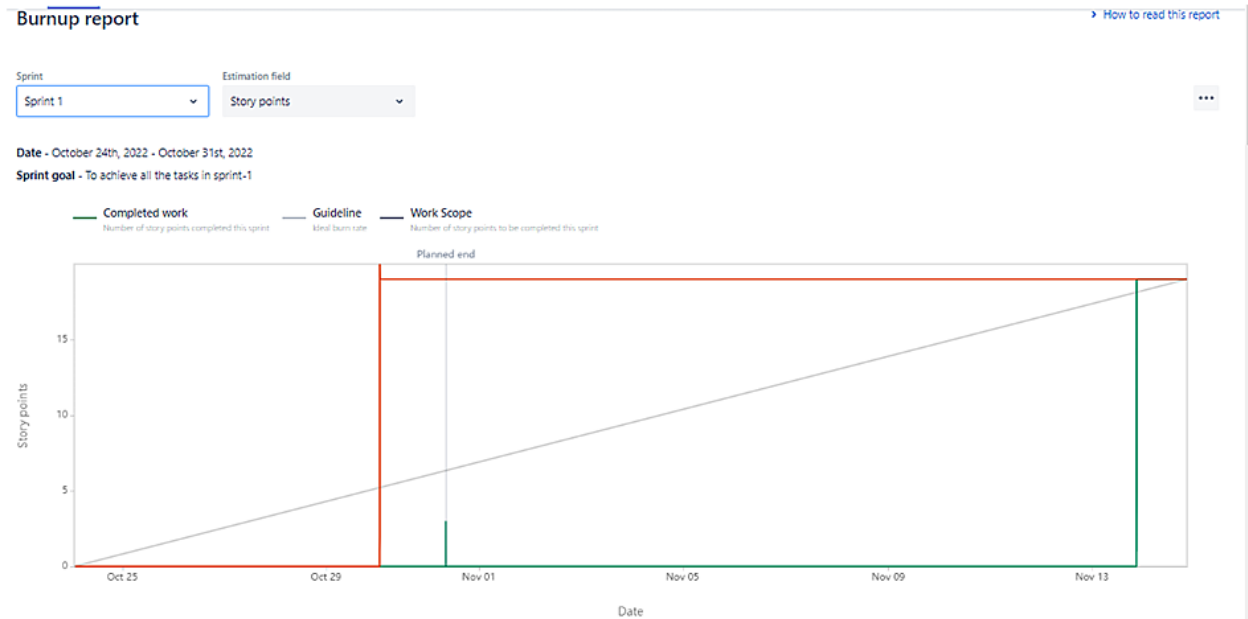


figure 6.3.1 burn up chart - sprint 1

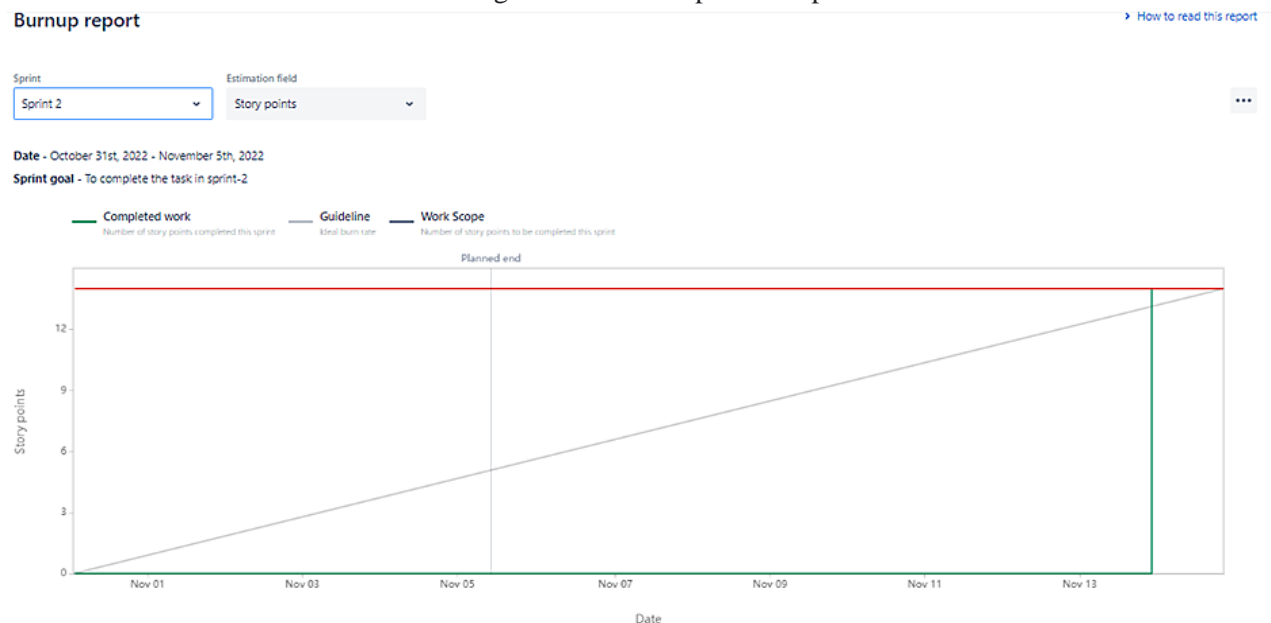


figure 6.3.2 burn up chart - sprint 2

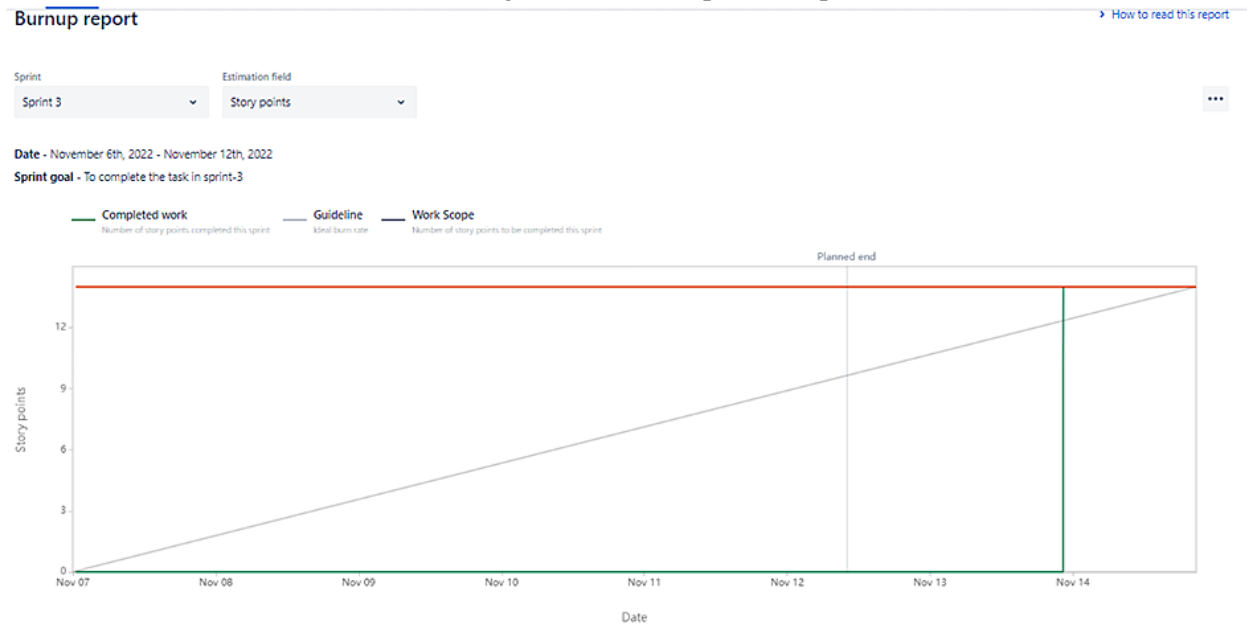


figure 6.3.3 burn up report - sprint 3

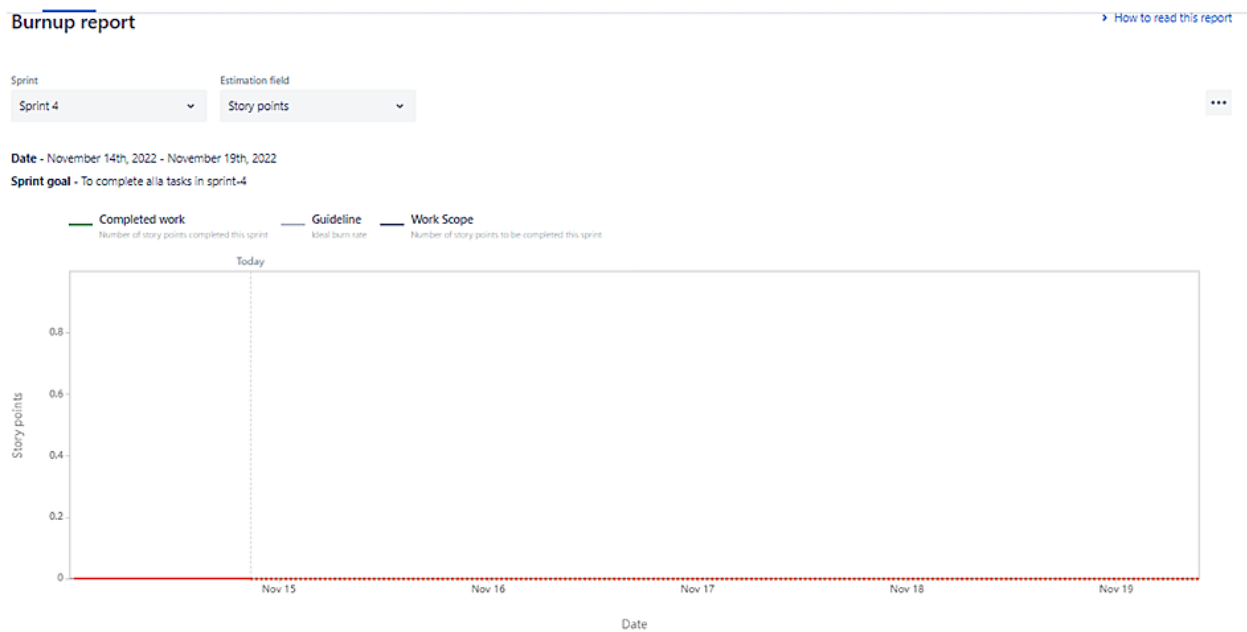


figure 6.3.4 burn up report - sprint 4

8.TESTING

8.1 TEST CASES:

Test cases **define what must be done to test a system**, including the steps executed in the system, the input data values that are entered into the system and the results that are expected throughout test case execution.

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Comments	TC for Automation(Y/N)	BUG ID	Executed By
LoginPage_TC_001	UI	Home Page	Verify user is login by entering email,password,and confirming password.		1.Enter URL and click go 2.Enter the email id, password and confirm password.	https://rainfalldata.w3spaces.com	Login/ registering for the application	Working as expected	Pass				Sulfa
LoginPage_TC_002	UI	Home Page	Verify the can access the dashboard with the LinkedIn login.		1. Enter the URL and click enter 2. enter the valid mail id in the Email text box. 3. enter the valid password in the password text box. 4. click on the join now button in linked in.	https://rainfalldata.w3spaces.com/	Application should show below UI elements: a. email text box b. password text box c. join now button d. shows the dashboard page	Working as expected	pass				Afrin Shahnaj
LoginPage_TC_003	Functional	Home page	Verify user is able to log into application with Valid credentials and get the confirmation mail.		1.Enter URL and click go 2.Click on My Account dropdown button 3.Enter Valid username/email in Email text box 4.Enter valid password in password text box	Username: lbmmsec@gmail.com password: Testing123	Application should send the confirmation mail	Working as expected	Pass				Preethika

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Comments	TC for Automation(Y/N)	BUG ID	Executed By
LoginPage_TC_004	Functional	Login page	Verify user is able to log into application with Valid credentials		1. Enter URL(https://shopenzer.com/) and click go 2. Click on My Account dropdown button 3. Enter Valid username/email in Email text box 4. Enter valid password in password text box	Username: lbmmsec@gmail.com password: Testing123	User should navigate to the home page.	Working as expected	Pass				Sayed Rasiyammal
LoginPage_TC_005	Functional	Login page	Verify user is able to log into application with Invalid credentials		1. Enter URL(https://shopenzer.com/) and click go 2. Click on My Account dropdown button 3. Enter Valid username/email in Email text box 4. Enter Invalid password in password text box	Username: chalam@gmail.com password: Testing123678686786876876	Application should show "Incorrect email or password" validation message.	Working as expected	pass				Sulfa

figure 8.1 test case report

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Comments	TC for Automation(Y/N)	BUG ID	Executed By
LoginPage_TC_005	Functional	Login page	Verify user is able to log into application with Invalid credentials		1. Enter URL(https://shopenzer.com/) and click go 2. Click on My Account dropdown button 3. Enter Valid username/email in Email text box 4. Enter Invalid password in password text box	Username: chalam@gmail.com password: Testing123678686786876876	Application should show "Incorrect email or password" validation message.	Working as expected	pass				Sulfa
LoginPage_TC_006	Functional	Login page	Verify user is able to log into application with Invalid credentials		1. Enter URL(https://shopenzer.com/) and click go 2. Click on My Account dropdown button 3. Enter Invalid username/email in Email text box 4. Enter Invalid password in password text box	Username: lbmmsec@gmail.com password: Testing654	Application should show "Incorrect email or password" validation message.	Working as expected	pass				Afrin Shahnaj

figure 8.1 test case report

8.2 USER ACCEPTANCE TESTING:

The purpose of this document is to briefly explain the test coverage and open issues of the exploratory analysis of rainfall data in India for agriculture project at the time of the release to User Acceptance Testing(UAT).

8.2.1 DEFECT ANALYSIS:

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	11	3	2	3	19
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	12	2	4	19	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	4	2	1	7
Totals	26	12	13	25	76

table 8.2.1 defect analysis

8.2.2 TEST CASE ANALYSIS:

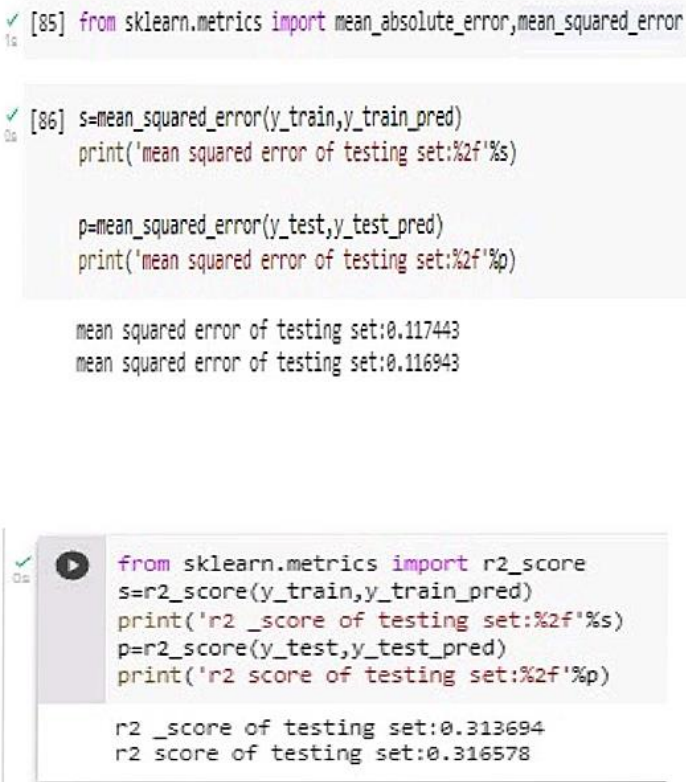
	Total Cases	Not Tested	Fail	Pass
Print Engine	4	0	0	4
Client Application	25	0	0	25
Security	1	0	0	1
Outsource Shipping	3	0	0	3
Exception Reporting	5	0	0	5
Final ReportOutput	4	0	0	4
Version Control	2	0	0	2

table 8.2.2 test case analysis

9. RESULTS

9.1 PERFORMANCE METRICS:

Performance testing is a **non-functional software testing technique** that determines how the **stability, speed, scalability, and responsiveness** of an application holds up under a given workload.

S.N o.	Parameter	Values	Screenshot
1	Metrics	Regression Model: 1.MAE-Linear Regression MSE-Linear Regression 2.R2 score-Linear Regression	 <pre>[85] from sklearn.metrics import mean_absolute_error,mean_squared_error [86] s=mean_squared_error(y_train,y_train_pred) print('mean squared error of testing set:%2f'%s) p=mean_squared_error(y_test,y_test_pred) print('mean squared error of testing set:%2f'%p) mean squared error of testing set:0.117443 mean squared error of testing set:0.116943 from sklearn.metrics import r2_score s=r2_score(y_train,y_train_pred) print('r2 _score of testing set:%2f'%s) p=r2_score(y_test,y_test_pred) print('r2 score of testing set:%2f'%p) r2 _score of testing set:0.313694 r2 score of testing set:0.316578</pre>

		Classification Model: 1. Confusion Matrix –Logistic Regression 2. Accuracy Score- Logistic Regression & 3. Classification Report – Logistic Regression	<pre>[] logreg = LogisticRegression() logreg.fit(X_train_res, y_train_res) LogisticRegression()</pre> <pre>[] y_pred2 = logreg.predict(X_test) print(confusion_matrix(y_test,y_pred2)) print(accuracy_score(y_test,y_pred2)) print(classification_report(y_test,y_pred2))</pre> <pre>[[17601 5116] [1515 4860]] 0.772067922452908</pre> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.92</td><td>0.77</td><td>0.84</td><td>22717</td></tr><tr><td>1</td><td>0.49</td><td>0.76</td><td>0.59</td><td>6375</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.77</td><td>29092</td></tr><tr><td>macro avg</td><td>0.70</td><td>0.77</td><td>0.72</td><td>29092</td></tr><tr><td>weighted avg</td><td>0.83</td><td>0.77</td><td>0.79</td><td>29092</td></tr></tbody></table>		precision	recall	f1-score	support	0	0.92	0.77	0.84	22717	1	0.49	0.76	0.59	6375	accuracy			0.77	29092	macro avg	0.70	0.77	0.72	29092	weighted avg	0.83	0.77	0.79	29092
	precision	recall	f1-score	support																													
0	0.92	0.77	0.84	22717																													
1	0.49	0.76	0.59	6375																													
accuracy			0.77	29092																													
macro avg	0.70	0.77	0.72	29092																													
weighted avg	0.83	0.77	0.79	29092																													
2	Tune the Model	1. Hyperparameter Tuning – Logistic Regression(Grid search) 2. Validation method-Hold-out-cross validation	<pre>from sklearn.model_selection import GridSearchCV model=LogisticRegression() grid_vals={'penalty':['l1','l2'],'Rainfall':[0,1,0.2]} grid_lr=GridSearchCV(estimator=model,param_grid=grid_vals, scoring='accuracy',cv=6,refit=True,return_train_score=True)</pre> <pre>[109] grid_lr.fit(X_train,y_train) preds=grid_lr.best_estimator_.predict(X_test)</pre> <pre>[113] from sklearn.model_selection import train_test_split</pre> <pre>X,y=np.arange(10).reshape((5,2)),range(5) X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=11)</pre>																														

Model	Accuracy	Precision	Recall	F1 Score
Model A	0.85	0.82	0.88	0.85
Model B	0.78	0.75	0.80	0.77
Model C	0.92	0.90	0.94	0.92
Model D	0.88	0.85	0.90	0.88

10. ADVANTAGES & DISADVANTAGES

10.1 ADVANTAGES:

- High prediction accuracy
- Hold perfectly good for large scale datasets with large number of variables
- Integral variable selection based on importance and variable interaction
- Deals efficiency with data having missing values
- Computation of relation between variables and classification
- Proximity calculation between cases
- Can be used for unsupervised learning and outlier detection
- Internal unbiased estimation of the generation error

10.2 DISADVANTAGES:

However, it has some minor disadvantages include necessity of product separation after precipitation and generation of the large volume of salt containing solutions. There is also difficulty in maintaining a constant product quality throughout the whole precipitation process if the precipitation is carried out discontinuously.

11.CONCLUSION

The overall aim is to define various ML techniques that are useful in predicting rainfall. The goal of this research is to design accurate and efficient model by applying lesser number of attributes and tests. Firstly, the data is pre processed and then it is used in the model. K-nearest neighbor with the 87% and random forest classifier with approximately 88% are the most efficient classification algorithms. However, decision tree classifier gives the least accuracy with 73% . we can further expand this research covering other ML techniques such as time series, clustering and association rules and other ensemble techniques. Taking into consideration the limitations of this study , there is a needed to build to more complex and combination of models to get higher accuracy for rainfall prediction system. Study can also formulated using greater articulate monitoring for particular area and create this kind of model for enormous dataset so that calculation rate can be increased with better precision and with more accuracy.

12.FUTURE SCOPE

Predicting the rainfall of a specific geographic location would be a challenge. Improvising the prediction model to predict the weather conditions and even predicting the loses of rainfall. Coping with the changing parameter values and making the code compatible for the changes in the parameter values. Improvising the ANN algorithm to further reduce the mean squared error.

