Nalaiya Thiran

Professional Readiness for Innovation, Employability,Enterpreneurship.

Estimate the Crop Yield using Data Analytics

R.Gokila	952019104006
B.Jeevadharshini	952019104009
S.Bhavani	952019104002
R.Sakthi Priya	952019104021

In partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

P.S.R.R COLLEGE OF ENGINEERING
SIVAKASI- 626140

ANNA UNIVERSITY::CHENNAI 600025

NOV/DEC-2022

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Project Overview

Agriculture forms the basis for food security and hence it is important. In India, majority of the population i.e., above 55% is dependent on agriculture as per the recent information. Agriculture is the field that enables the farmers to grow ideal crops in accordance with the environmental balance. Farmers also grow non-food items like rubber, cotton, jute etc. More than 70% of the household in the rural area depend on agriculture. This domain provides employment to more than 60% of the total population and has a contribution to GDP also (about 17%). In the farm output, India ranks second considering the world wide scenario. This is the widest economic sector and has an important role regarding the framework of socio-economic fabric of India. Farming depends on various factors like climate and economic factors like temperature, irrigation, cultivation, soil, rain fall, pesticide and fertilizers. Historical information regarding crop yield provides major input for companies engaged in this domain. These companies make use of agriculture products as raw materials, animal feed, paper production and so on. The estimation of production of crop helps these companies in planning supply chain decision like production scheduling. The industries such as fertilizers, seed, agrochemicals and agricultural machinery plan production and activities like marketing based on the estimates of crop yield. Farmers experience was the only way for prediction of crop yield in the past days. Technology penetration into agriculture field has led to automation of the activities like yield estimation, crop health monitoring etc.

Achieving maximum crop yield at minimum cost is one of the goals of agricultural production. Early detection and management of problems associated

with crop yield indicators can help increases yield and subsequent profit. Additionally, these predictions could be used to maximize crop prediction when potential exists for favorable growing conditions.

Crop yield prediction helps the farmers in various ways by providing the record of previous crop yield. Knowing what crops has been grown, and how much area of it had been shown historically, combined with the prices .

This project works on achieving the more quality of the crop that will help the farmers to gain more money. In this project we have collected the datasets of all the factors that are depends of the crops of several years. Using this data the prediction is obtained to show that the harvest of the crop that is growth in that region.

1.2 Purpose

It is been observed that farmers are facing the problem at the time of the yield of the crop because of the rapid changes in the weather where it effect the yield of the crop. Decrease the quality of the crop and which in turn provide less income to the farmers. This project works on achieving the more quality of the crop that will help the farmers to gain more money. In this project we have collected the datasets of all the factors that are depends of the crops of several years. Using this data the prediction is obtained to show that the harvest of the crop that is growth in that region.

2. LITERATURE SURVEY

2.1 Existing Problem

A research group conducted a work with an objective of accurate prediction of crop yield through big data analytics to assess various crop yield influencing

factors such as Area under Cultivation (AUC) interims of hectors, Annual Rainfall (AR) rates and Food Price Index (FPI) and to develop relationship among these parameters. Regression Analysis (RA) methodology was applied to examine the selected factors and their impact on crop prediction and final yield. RA methodology is a multivariable investigation practice which can categorize the factors in to groups such as explanatory and response variables and helps to assess their interaction to obtain a resolution. All the selected factors of the present study design known as AR, AUC and FPI were measured for a period of 10 years between the years 1990-2000. A novel method called Linear Regression (LR) is applied to analyze the relationship between explanatory variables (AR, AUC, FPI) and the crop yield considered as response variable. Study reported that the R2 value for the studied factors clearly indicate that crop yield is principally depends on AR. Study also reported that the other two factors (AUC and FPI) screened were also found to have significant impact after the AR. Study shall be continued to analyze the impact of for other substantial factors like Minimum Support Price (MSP), Cost Price Index (CPI), Wholesale Price Index (WPI) etc. and their relationship on the yields of different crops.

2.2 References

[1] Dakshayini Patil, M.S. Shirdhonkar.,(2017) "Rice Crop Yield Prediction using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering.

[2] David B. Lobell(2013), "The use of satellite data for crop yield gap analysis", Field Crops Research.

[3] Dhivya B H, Manula R, Siva Bharathi S. Madhumathi R.(2017)," A Survey on Crup Yield Prediction based on Agricultural Data", International Journal of Innovative Research in Science, Engineering and Technology.

[4] Jharna Majumdar, Sneha Naraseeyappa, Shilpa Ankalaki(2017), "Analysis of agriculture data using datamining techniques: application of big data". Journal of Big data.

[5] Majumdar J, Ankalaki S(2016), "Comparison of clustering algorithms using quality metrics with invariant features extracted from plant leaves", International Conference on Computational Science and Engineering.

2.3 Problem Statement Definition

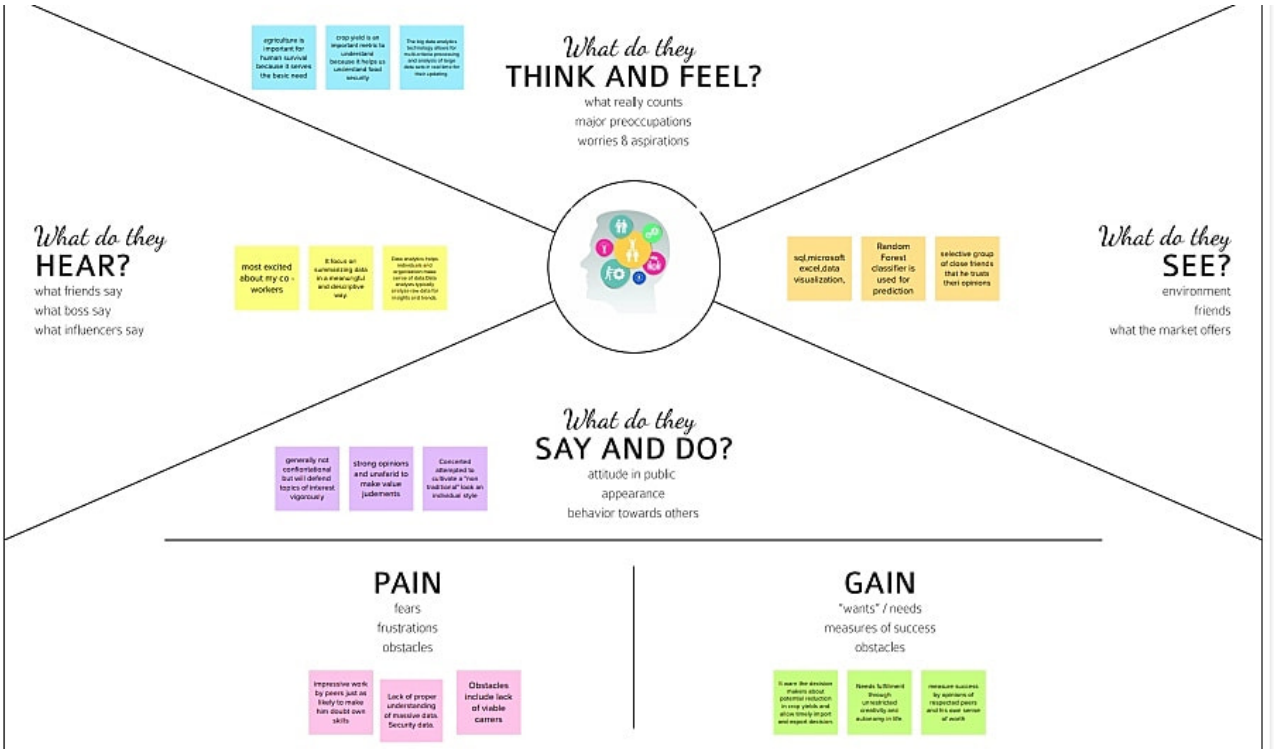
India is an agricultural country. Yield of each crop depends on its dependent factors. It is very important to predict the yield of a crop to help farmers. Crop Yield Prediction is predicting the yield of a crop in future based on the dependent factors. Crop yield is dependent on factors like rainfall, pressure, temperature and area orthe geographical location. This is achieved by

(a) Designing a system to estimate crop yield.

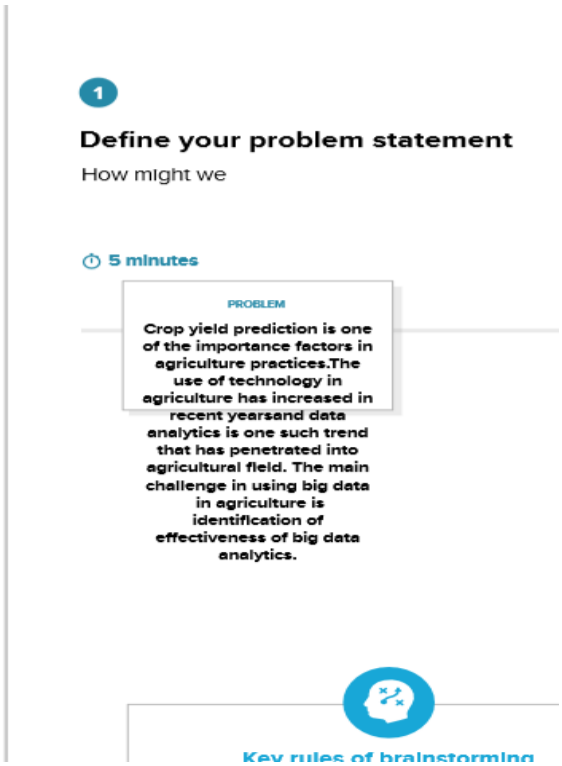
(b) Providing graphical user interface to view estimation result and historical datasets.

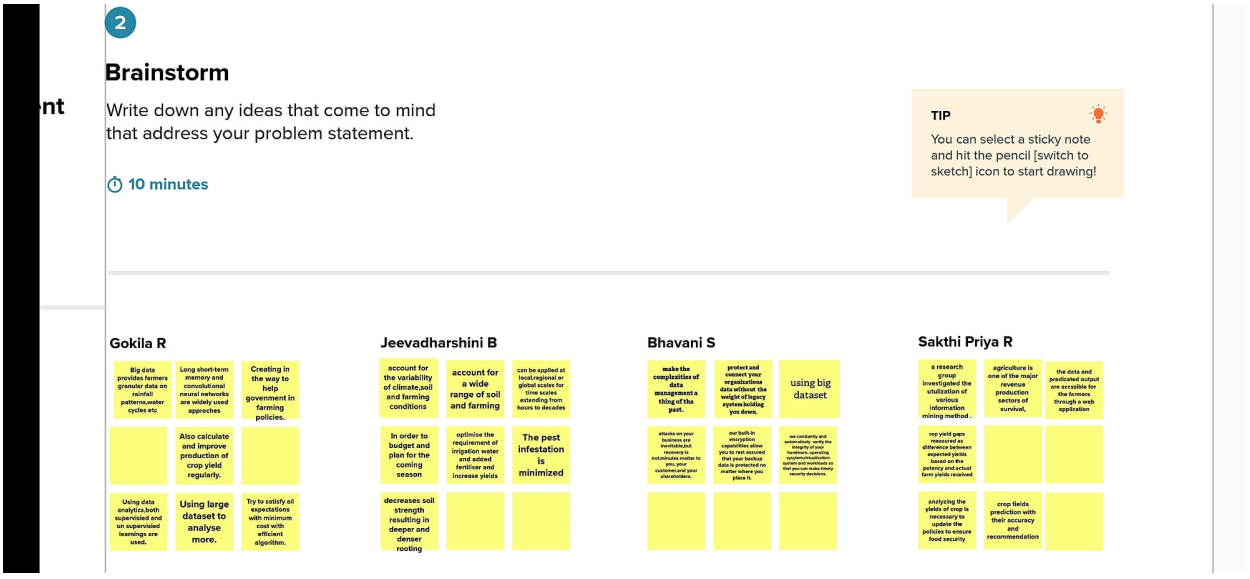
3. IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation and Brainstorming



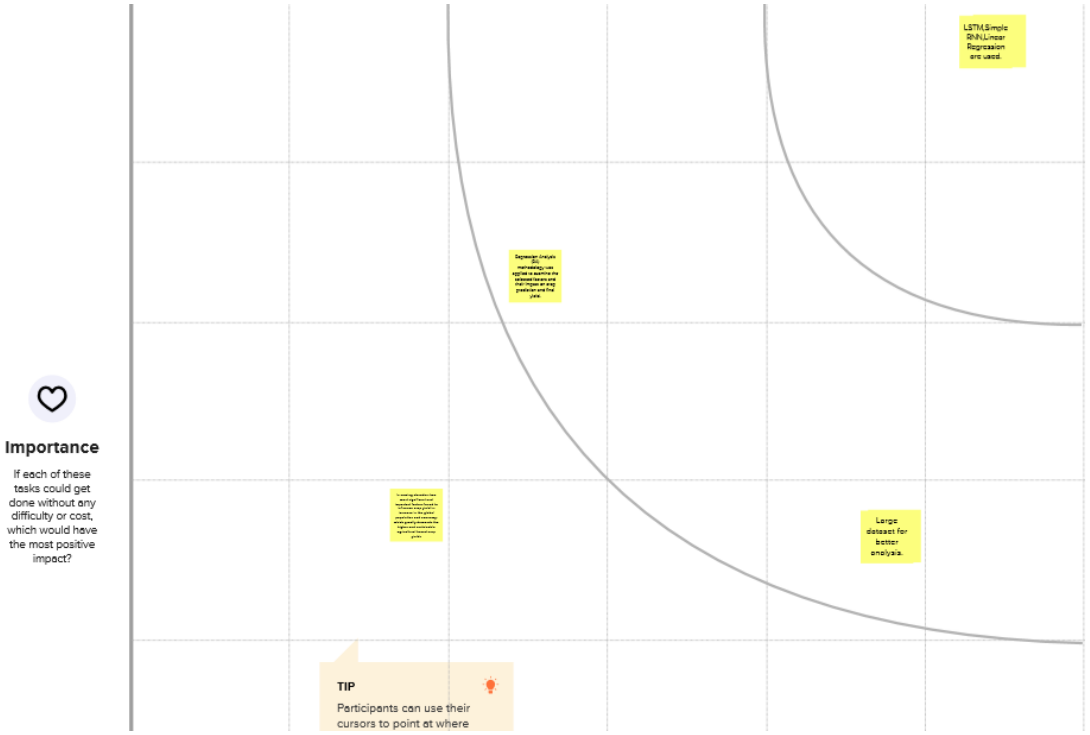


3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

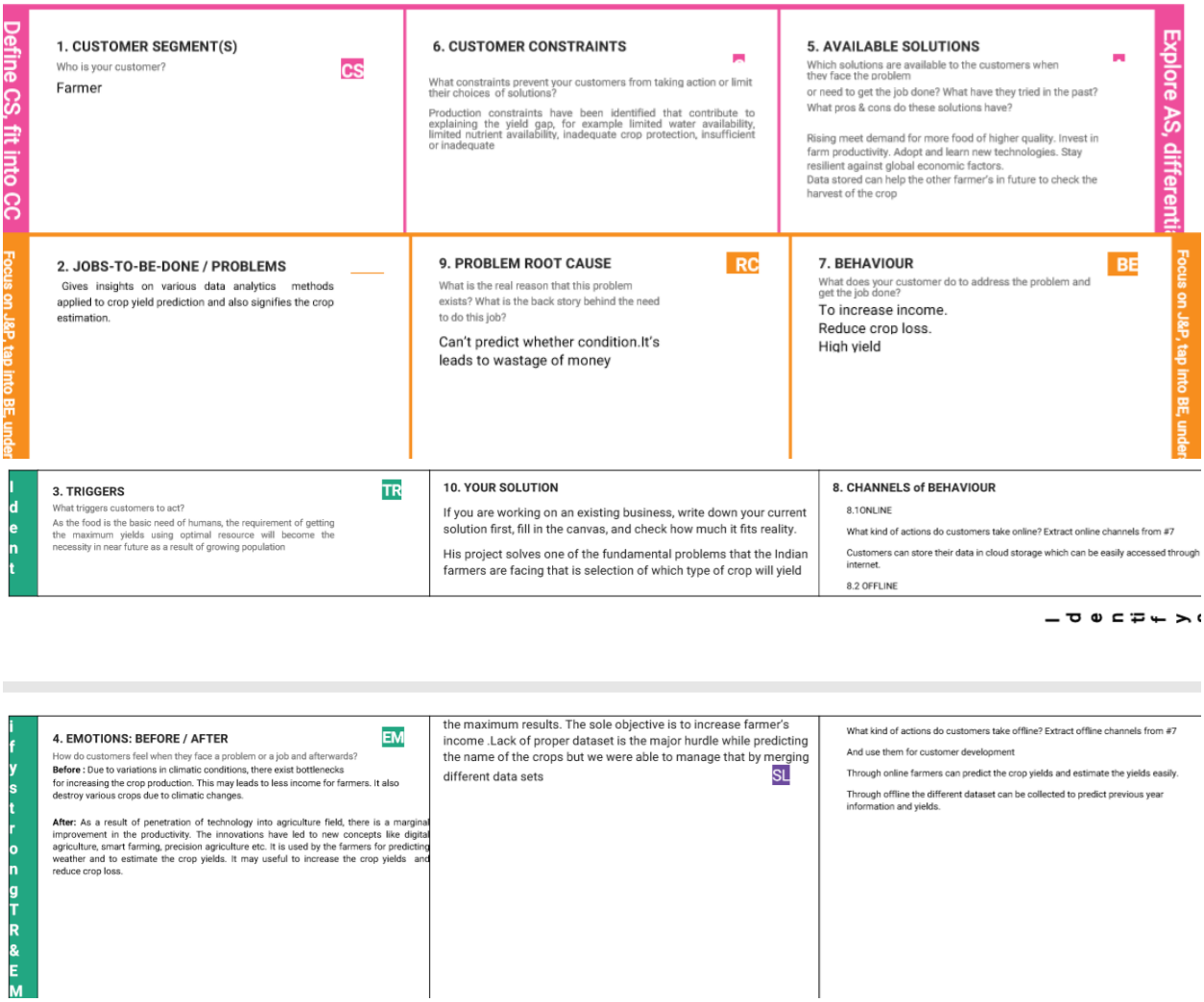


3.3 Proposed Solution

S.No	Parameter	Description
1	Problem Statement (Problem to be solved)	Crop yield prediction is predicting the yield of a crop in future based on the dependent factors. Crop yield is dependent on factors like weather. This is achieved by (a) Designing a system to predict yield. (b) Providing graphical user interface to view(estimate) predict results.
2	Idea / Solution description	To predict the yield, one of the machine learning algorithms called Multiple Linear Regression algorithm is used. The result of prediction is plotted via graph.
3	Novelty / Uniqueness	By using of data analytics in crop yield estimation helps in analysing some important visualization, creating dashboard and by going through these we will get most of the insights of crop production in India.
4	Social Impact / Customer Satisfaction	Customer(farmer) can satisfied by increased income, reduced crop loss and high yield.
5	Business Model (Revenue Model)	Initially an raw data set was collected and subjected to analysis. From the dataset, it is subjected to

		feature selection for make a predictive modelling. Final representation represents the graphical result which is helpful for analysis or estimation of crops in specified rainfall.
6	Scalability of the Solution	Effectively analysis large dataset. Easy to predict by using previous data.

3.4 Problem Solution Fit



4 REQUIREMENT ANALYSIS

4.1 Functional Requirement

FR No	Functional Requirement (Epic)	Sub Requirement(Story/SubTask)
FR-1	User login	Login through internet or app
FR-2	Login through internet or app	User can update their profile with name, mobile number and password
FR-3	Analyse the dataset	Analyse the dataset and process data preprocessing to avoid noise data.
FR-4	Choose the crop	Through which the user can choose particular crop for their convenience.
FR-5	Predict result	The result will be predicted based on the previous year data in the way of production per hectare for particular rainfall measure in that area.
FR-6	Estimation of the result	The graphical representation shown the estimation analysis of the crop to increase more yield.

4.2 Non-Functional Requirements

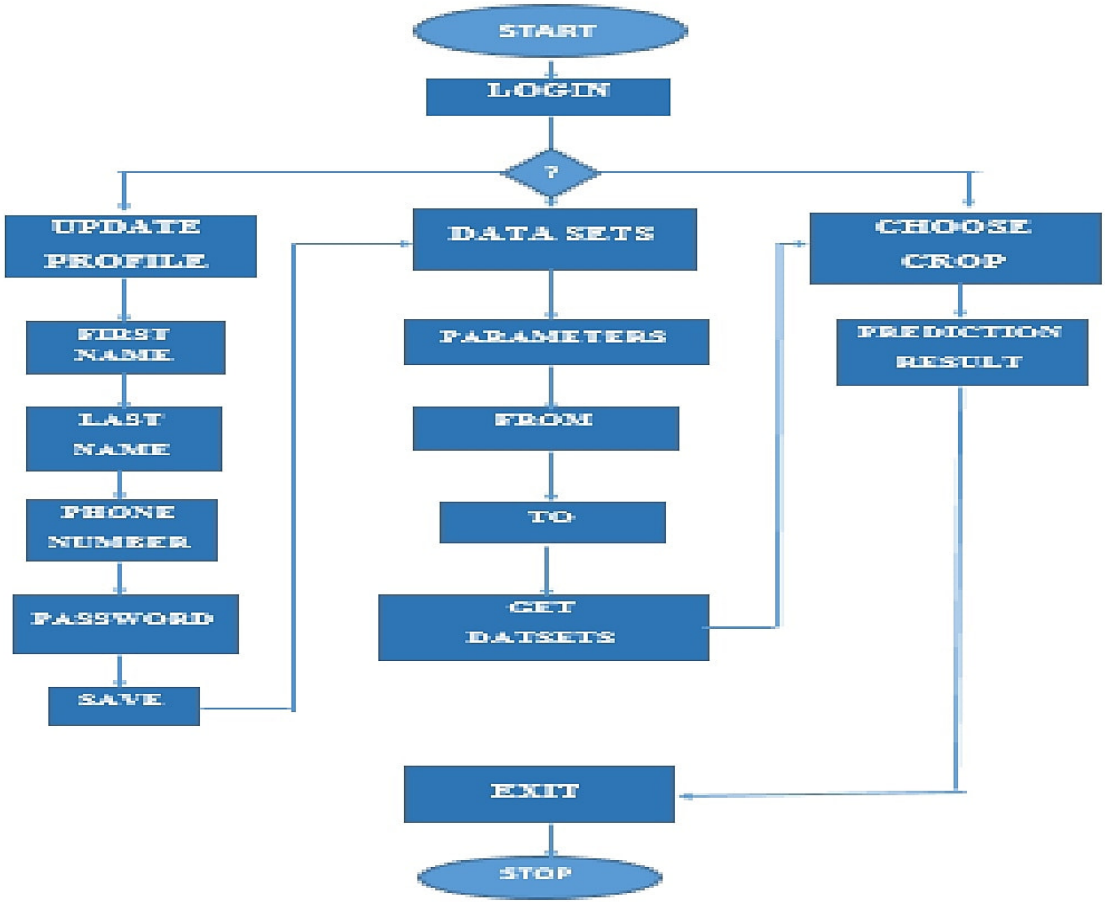
NFR No	Non-FunctionalRequirement	Description
NFR-1	Usability	Contains easy user interface in order to use by uneducated people also
NFR-2	Security	Information of the personal data are kept secure. By using data analytics, no loss or corruption of the data in dataset. The structure of the system is kept feasible enough so that there should not be any problem from the users' point of view
NFR-3	Reliability	The best technique for rainfall is Simple RNN with a mean absolute error of 22.14 mm. After applying various techniques we found out that in Crop Yield and Crop Name . Random Forest yields the best result with minimum mean absolute error
NFR-4	Performance	Performance analysis is done to find out whether the proposed system is time efficient and accurate. It is essential that the process of performance analysis and definition must be conducted in

		parallel. The application's load time should not be more than one second for users.
NFR-5	Availability	User can predict and estimate the crop yield throughout the period at any time. Platforms & tools used in this project are widely used. So the skilled manpower is readily available in the industry.
NFR-6	Scalability	It can also work with an large dataset without performance degradation. All changes should be in positive direction, there will be increased level of efficiency and better customer service

5. PROJECT DESIGN

5.1 Data Flow Diagram

A data flow diagram is a way of representing a flow of data through a process or a system. The data flow diagram also provides information about the outputs and inputs of each entity and the process itself.



(1) User (farmer) can login into project. User can also update their profile using First name, last name, phone number, password.

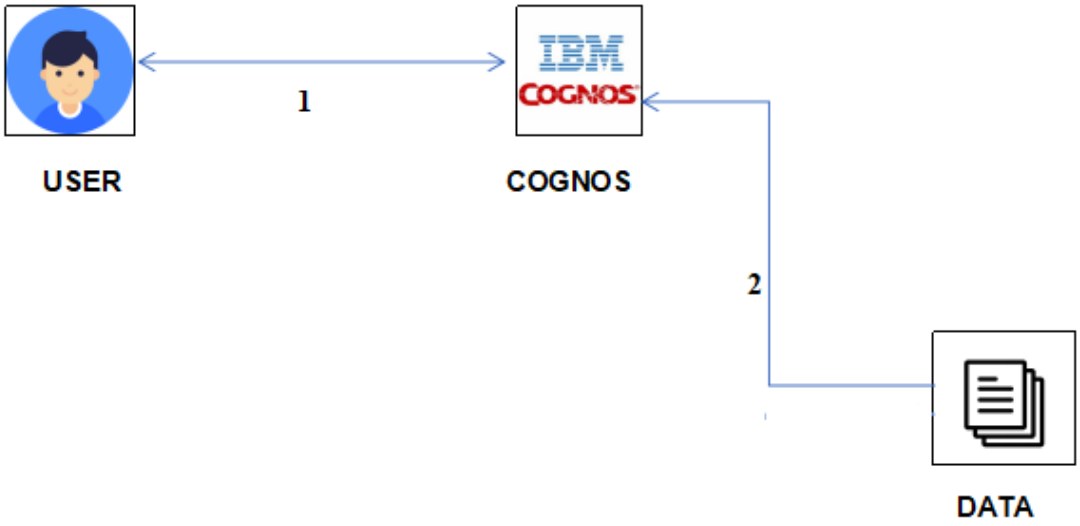
(2) User can enter the available dataset(rainfall, area, production etc).

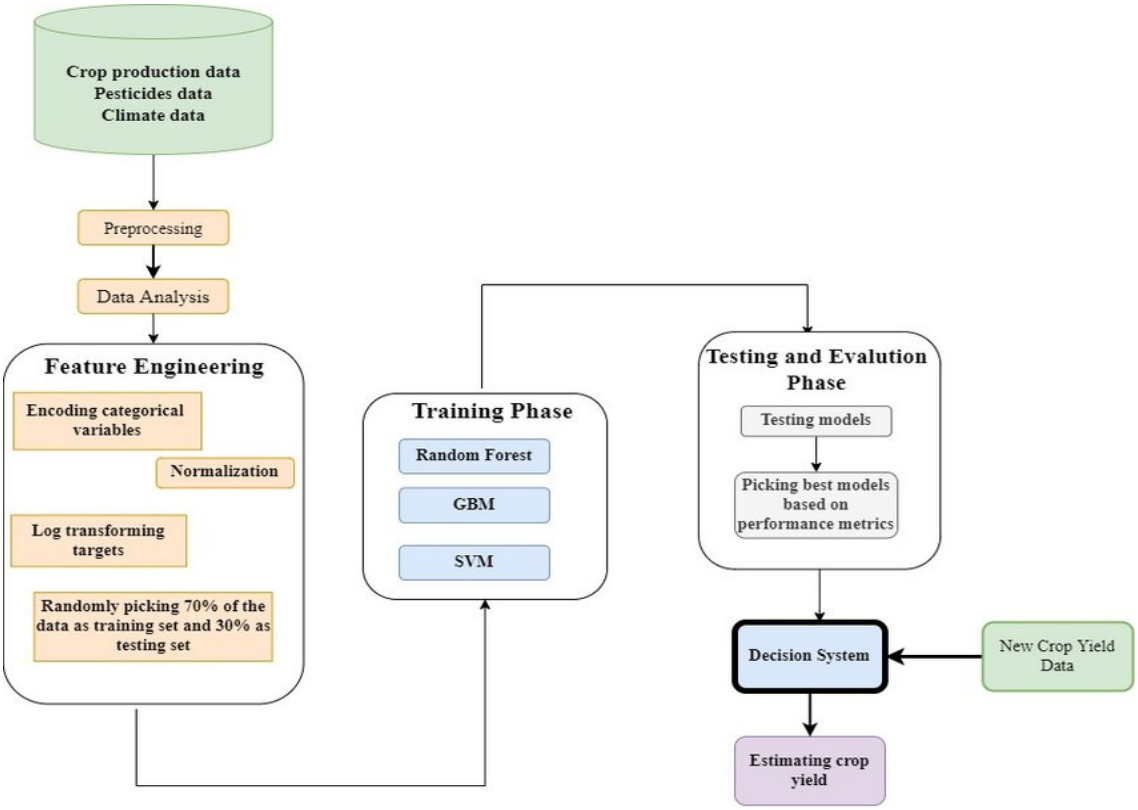
(3) The project analyse the dataset and show the prediction in graphical Format.

(4) Then the user can use these prediction to increase the crop yield and reduce the crop loss.

(5) Then user can exit.

5.2 Solution & Technical Architecture





5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	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 will receive confirmation email once I have registered for the application	I can register & access the dashboard with Facebook Login	Low	sprint 1
		USN-4	As a user, I can register for the application through Gmail		Medium	sprint 1
	Login	USN-5	As a user, I can		High	sprint 1

			log into the application by entering email & password			
	Dashboard	USN-6	Creating an interactive dashboard from the datasets			sprint 2
Customer (Web user)						
Customer Care Executive						
			In data pre-processing module data is cleaned and only necessary attributes are taken for further analysis		Medium	sprint 2
			Prediction is result of Apriori and Naïve bayes which predicts the crop yield in quintals.		High	sprint 3
			Final representation		High	sprint 4

			represents the graphical result of K-means and Naïve bayes which is helpful for analysis of crops in specified rainfall.			

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

S.no	Milestone	Activities	Start Date	End Date
1	Solution Requirement	Creating the IBM Cognos for creating dashboard and data visualization charts.	22-Aug-2022	24-Aug-2022
2	Project Objectives	Prepare the project objectives.	22-Aug-2022	24-Aug-2022
3	Project Flow	Prepare the project flow.	22-Aug-2022	24-Aug-2022
4	IBM Cloud Account	Creating IBM cloud account.	22-Aug-2022	24-Aug-2022
5	IBM Cognos Analytics	Creating IBM cognos account.	22-Aug-2022	24-Aug-2022
6	Working with the Dataset	Understanding The Dataset Loading The Dataset.	24-oct-2022	19-Nov-2022
7	Data Visulization Charts	➢ Seasons With Average Productions ➢ With Years Usage of Area And Production ➢ Top 10 States with Most	24-oct-2022	19-Nov-2022

		Area ➢ State With Crop Production ➢ States With the Crop Production Along with Season		
8	Creating the Dashboard	Creating The Dashboard	24-oct-2022	19-Nov-2022
9	Export the Analytics	Export The Analytics	24-oct-2022	19-Nov-2022
10	Ideation Phase	➢ Literature Survey On The Selected Project & Information Gathering Prepare ➢ Empathy Map ➢ Ideation	22-Aug-2022	17-Sept-2022
11	Project Design Phase - I	➢ Proposed Solution ➢ Problem Solution Fit ➢ Solution Architecture	22-Aug-2022	17-Sept-2022
12	Project Design Phase - II	➢ Customer Journey ➢ Functional Requirement ➢ Data Flow Diagrams ➢ Technology Architecture	22-Aug-2022	01-Oct-2022
13	Project Planning Phase	➢ Prepare Milestone & Activity List ➢ Sprint Delivery Plan	17-Oct-2022	22-Oct-2022
14	Project Development Phase	➢ Project Development - Delivery of Sprint-1 ➢ Project Development - Delivery of Sprint-2 ➢ Project Development -	24-Aug-2022	19-Nov-2022

		Delivery of Sprint-3 ➢ Project Development - Delivery of Sprint-4		

6.2 Sprint Delivery Schedule

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 by entering my Agri - id card and request..	2	High	Gokila R
		USN-3	As a user, I can register for the application through Gmail	2	Medium	jeevadharshini B
	Login	USN-4	As a user, I can Call and request or Approach for datase	2	High	Bhavani S
	Working with the Dataset	USN-5	To work on the given dataset, Understand the Dataset.	2	High	Sakthi Priya R
		USN-6	Load the dataset to Cloud platform	10	High	Gokila R

			then Build the required Visualizations.			
Sprint-2	Data Visualization Chart	USN-7	Using the Crop production in Indian dataset, create various graphs and charts to highlight the insights and visualizations.			
			*Build a Visualization to showcase Average Crop Production by Seasons	4	Medium	Sakthi Priya R
			*Showcase the Yearly usage of Area in Crop Production	4	Medium	Jeevadharshini B
			*Build a visualization to show case top 10 States in Crop Yield Production by Area.	4	Medium	Bhavani S
			*Build the required	4	Medium	Sakthi Priya R

			Visualization to showcase the Crop Production by State. *Build Visual analytics to represent the Sates with Seasonal Crop Production using a Text representation	4	Medium	Gokila R
Sprint-3	Creating The dashboard	USN-8	Create the Dashboard by using the created visualizations	20	High	Gokila R Jeevadharshini B Bhavani S Sakthi Priya R
Sprint-4	Export The Analytics	USN-9	Export the created Dashboard	20	High	Gokila R Jeevadharshini B Bhavani S Sakthi Priya R

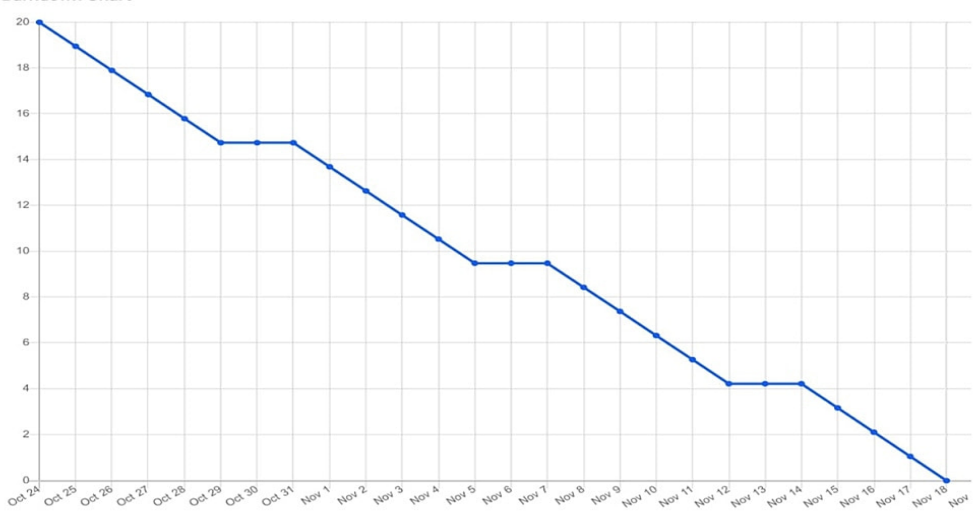
6.3 Reports from JIRA

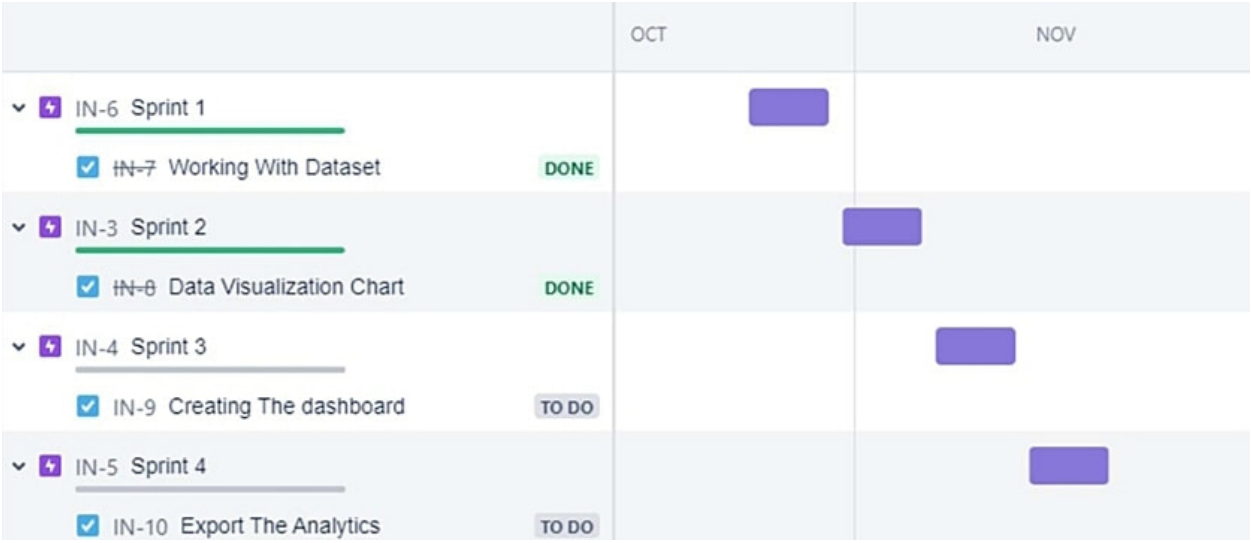
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

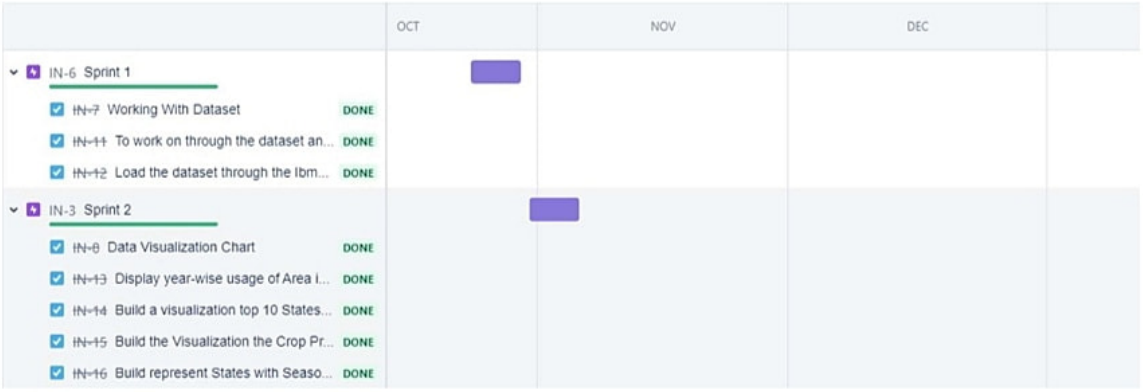
Velocity: We have a 24-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

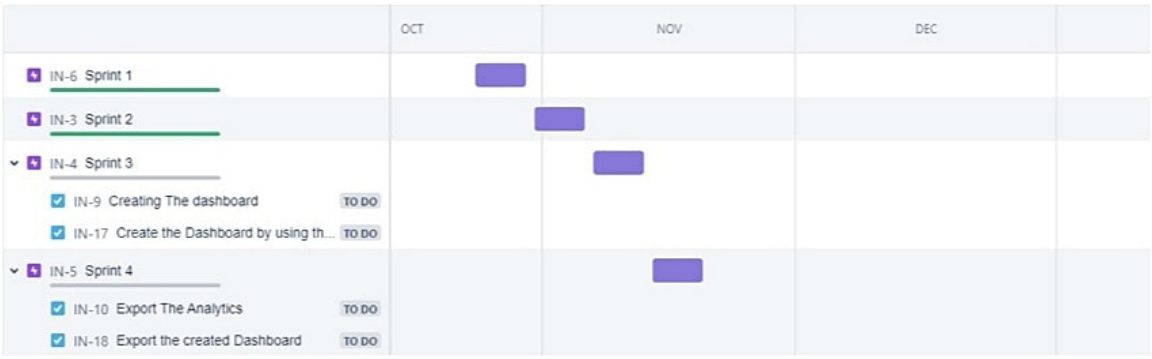
AV = Sprint Duration / Velocity = 24 / 20 = 1.2

Burndown Chart :A burn down chart is a graphical representation of work left to do versus time. It is often used in agilesoftware developmentmethodologies such as Scrum. However, burn down charts can be appliedto any project containing measurable progress over time.


20
18
16
14
12
10
8
6
4
2
0
Oct 24
Oct 25
Oct 26
Oct 27
Oct 28
Oct 29
Oct 30
Oct 31
Nov 1
Nov 2
Nov 3
Nov 4
Nov 5
Nov 6
Nov 7
Nov 8
Nov 9
Nov 10
Nov 11
Nov 12
Nov 13
Nov 14
Nov 15
Nov 16
Nov 17
Nov 18


OCT
NOV
IN-6 Sprint 1
IN-7 Working With Dataset
DONE
IN-3 Sprint 2
IN-8 Data Visualization Chart
DONE
IN-4 Sprint 3
IN-9 Creating The dashboard
TO DO
IN-5 Sprint 4
IN-10 Export The Analytics
TO DO


OCT
NOV
DEC
IN-6 Sprint 1
IN-7 Working With Dataset
DONE
IN-11 To work on through the dataset an...
DONE
IN-12 Load the dataset through the Ibm...
DONE
IN-3 Sprint 2
IN-8 Data Visualization Chart
DONE
IN-13 Display year-wise usage of Area i...
DONE
IN-14 Build a visualization top 10 States...
DONE
IN-15 Build the Visualization the Crop Pr...
DONE
IN-16 Build represent States with Seaso...
DONE



7. CODING AND SOLUTIONING

7.1 Feature 1

Login

A login page is a web page or an entry page to a website that requires user identification and authentication, regularly performed by entering a username and password combination.

Logins are used by websites, computer applications, and mobile apps. They are a security measure designed to prevent unauthorized access to confidential data. When a login fails (i.e, the username and password combination does not match a user account), the user is disallowed access.

SAMPLE CODING

```
<!DOCTYPE html>
<html>
<head>
<meta name="viewport" content="width=device-width, initial-scale=1">
<title> Login Page </title>
<style>
Body {
  font-family: Calibri, Helvetica, sans-serif;
```

```
  background-color:lightred;
}
button {
       background-color:#CBC3E3;
       width: 100%;
        color: black;
        padding: 15px;
        margin: 10px 0px;
        border: none;
        cursor: pointer;
         }
form {
        border: 3px solid #f156189;
    }
input[type=text], input[type=password] {
        width: 100%;
        margin: 8px 0;
        padding: 12px 20px;
        display: inline-block;
        border: 2px white;
        box-sizing: border-box;
    }
button:hover {
        opacity: 0.7;
    }
  .cancelbtn {
        width: auto;
        padding: 10px 18px;
        margin: 10px 5px;
    }
.container {
        padding: 25px;
```

```
        background-color: skyblue;
    }
</style>
</head>
<body>
    <center> <h1>Login Form </h1> </center>
    <form>
        <div class="container">
            <label>Username : </label>
            <input type="text" placeholder="Enter Username" name="username"
required>
            <label>Password : </label>
            <input type="password" placeholder="Enter Password" name="password"
required>
            <button type="submit">Login</button>
            <input type="checkbox" checked="checked"> Remember me
            <button type="button" class="cancelbtn"> Cancel</button>
            <a href="#"> Forgot password? </a>

        </div>
    </form>
</body>
</html>
```

Registration

The register module provides a conceptual framework for entering data on those patients in a way that: eases data entry & accuracy by matching the OpenMRS entry to the data source (usually paper files created at point of care), ties easily back to individual patient records to connect registers to patient data.

SAMPLE CODING

```
<!DOCTYPE html>
<html>
<head>
<title></title>
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<link rel="stylesheet" type="text/css"
href="{{url_for('static',filename='style.css')}}">
<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-
awesome/4.7.0/css/font-awesome.min.css">
<!-- jQuery library -->
<script
src="https://ajax.googleapis.com/ajax/libs/jquery/3.2.1/jquery.min.js"></script>
<!-- Latest compiled JavaScript -->
<script
src="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/js/bootstrap.min.js"></scrip
t>
<script src="https://www.google.com/recaptcha/api.js" async defer></script>
<style type="text/css">
.error
{
color: red;
}
</style>
</head>
<body>
<?php
include 'header.php';
?>
<div class="heading fix">
<label>REGISTRATION</label>
```

```
</div>
<div class="outerbox">
<div class="fixedbox">
<span class="content">
<h4>Hello, Friend!</h4>
<p>Enter your personal details and start journey with us</p>
</span>
</div>
<div class="scrollbox">
<div class="registerdonor">
<form action="process.php" method="POST" id="myform">
<div class="login">
<h3>Login Details</h3>
<table>
<tr>
<td colspan="2">
<label class="username">User Name:-</label>
<input type="text" name="user_name" required pattern="^[A-Za-z0-9._%+-
@]{5,10}$" title="Enter a username between 5 to 10 letter"  autocomplete="off">
</td>
</tr>
<tr>
<td>
<label>Full Name:-</label>
<input type="text" name="user_full_name" required pattern="[A-z ]+$" title="Use
only character & whitespace"  autocomplete="off">
</td>
<td>
<label>Email Id:-</label>
<input type="email" name="user_email" required pattern="[A-Za-z0-9._%+-
]+@[A-z0-9.-]+\.[a-z]{2,}$" title="Email id is not Valid"  autocomplete="off">
</td>
```

```
</tr>
<tr>
<td>
<label>Password:-</label>
<input type="password"name="password" required pattern="(?=.*\d)(?=.*[a-z])(?=.*[A-Z]).{6,}" title="Must contain at least one number and one uppercase and lowercase letter, and at least 6 or more characters" id="password" autocomplete="off">
</td>
<td>
<label>Confirm Password:-</label>
<input type="text" name="confirm_password" required pattern="(?=.*\d)(?=.*[a-z])(?=.*[A-Z]).{6,}" title="Must contain at least one number and one uppercase and lowercase letter, and at least 6 or more characters" id="confirm_password" autocomplete="off">
</td>
</tr>
</table>
</div>
<div class="contact">
<h3>Contact Details</h3>
<table>
<tr>
<td>
<label>Mobile Number:-</label>
<input type="text" name="user_number" required pattern="^[1-9]{1}[0-9]{9}$" title="Number is not valid"  autocomplete="off">
</td>
<td rowspan="2">
<label>Address:-</label>
<textarea name="Address" placeholder="---Type---" required></textarea>
</td>
```

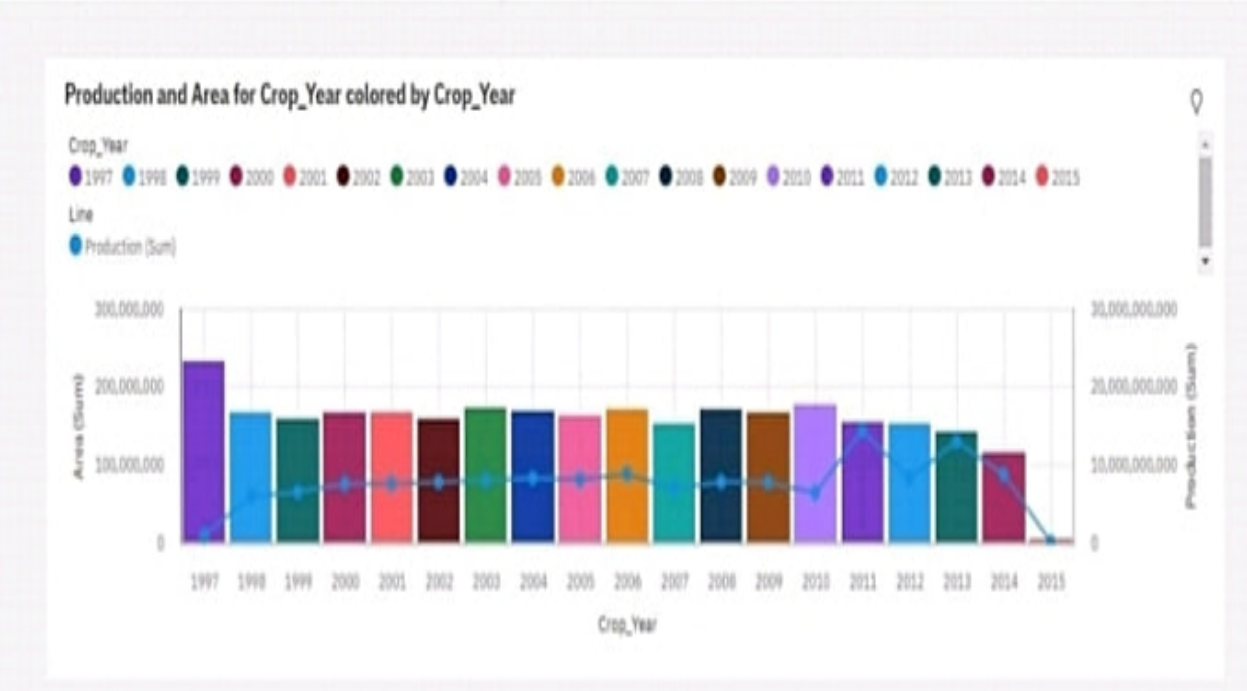
```
</tr>
<tr>
<td>
<label>Pincode</label>
<input type="text" name="pincode" required pattern="^[0-9]{6}$" title="Pincode
is not valid"  autocomplete="off">
</td>
</tr>
<tr>
<td>
<label>City:-</label>
<input type="text" name="city"  >
</td>
<td>
<label>State:-</label>
<input type="text" name="state">
</td>
</tr>
</table>
</div>
<div class="personal">
<h3>Personal Details</h3>
<table>
<tr>
<td>
<label>Date Of Birth:-</label>
<input type="date" name="date_of_birth" required  autocomplete="off">
</td>
<td>
<label>Gender:-</label>
<div class="radio">
<input type="radio" name="gender" class="radio1" value="Male"><span
```

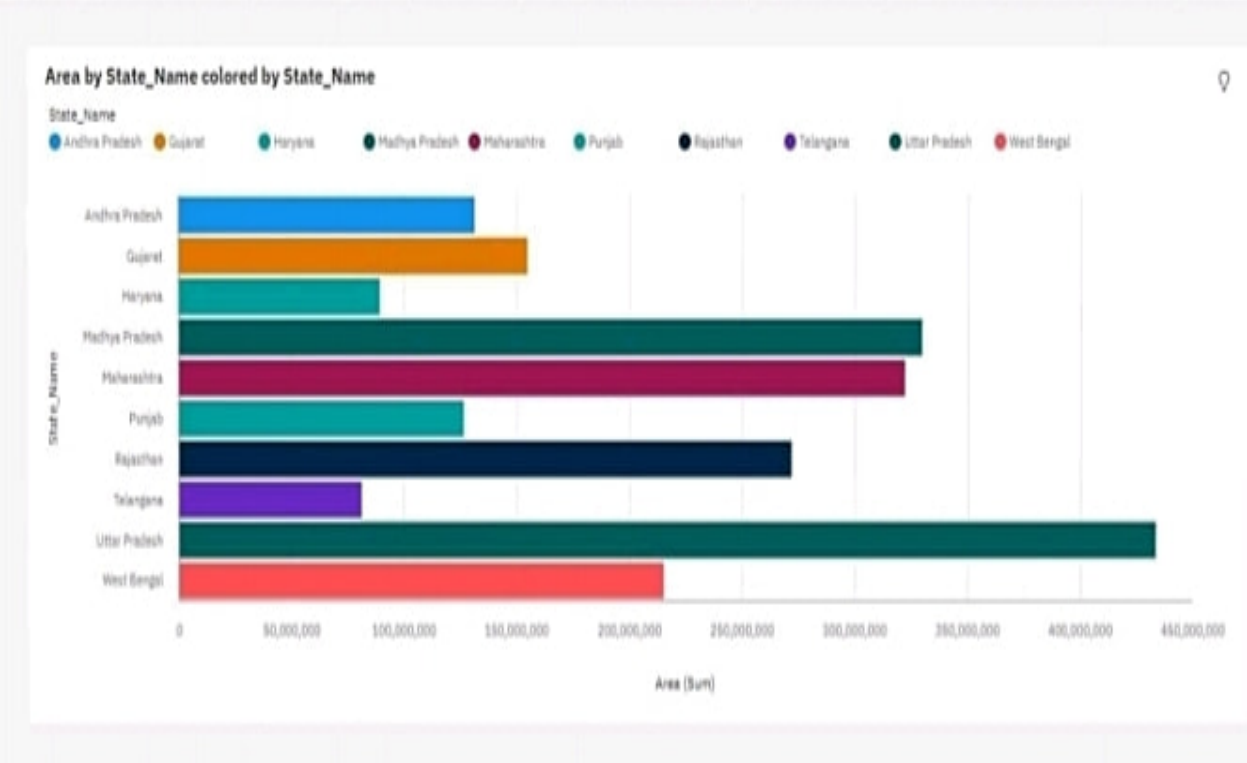
```
class="radioname" required  autocomplete="off">Male</span>
<input type="radio" class="radio2" name="gender" value="Female"><span
class="radioname" required  autocomplete="off">Female</span>
</div>
</td>
</tr>
<input type="reset" name="submit" value="Reset">
<a href="login.html"><input type="button"onclick="href='login.html';"
value="Submit"></a>
</div>
</form>
</div>
</div>
</div>
<!-- Responsive Table -->
<div class="rregisterdonor">
<form action="process.php" method="POST" id="myform">
</html>
```

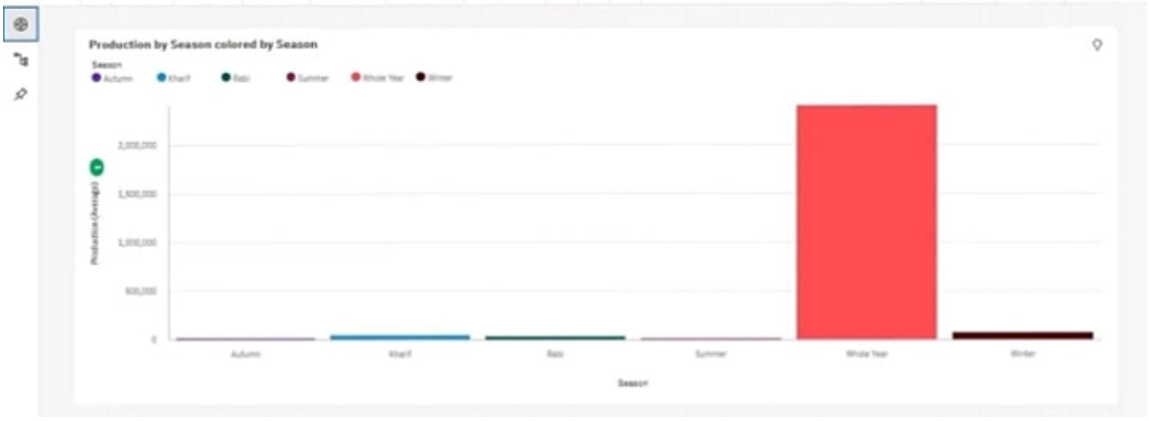
7.2 Feature 2

Data Visualization Chart

Data visualization charts are graphical representations of data that tell a story using symbols in order to improve the understanding of large amounts of data. Visual data metaphors such as charts effectively engage human perceptual processes and amplify human cognition more so than semantic data alone.


Production and Area for Crop_Year colored by Crop_Year
Crop_Year
1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
Line
Production (Sum)
300,000,000
200,000,000
100,000,000
0
Area (Sum)
30,000,000,000
20,000,000,000
10,000,000,000
0
Production (Sum)
1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
Crop_Year


Area by State_Name colored by State_Name
State_Name
Andhra Pradesh Gujarat Haryana Madhya Pradesh Maharashtra Punjab Rajasthan Telangana Uttar Pradesh West Bengal
Andhra Pradesh
Gujarat
Haryana
Madhya Pradesh
Maharashtra
Punjab
Rajasthan
Telangana
Uttar Pradesh
West Bengal
State_Name
0
50,000,000
100,000,000
150,000,000
200,000,000
250,000,000
300,000,000
350,000,000
400,000,000
450,000,000
Area (Sum)



Creating the dashboard

A data dashboard is a tool many businesses use to track, analyze, and display data—usually to gain insight into the overall wellbeing of an organization, department, or specific process.

Step 1: Import the necessary data into Excel. No data.

Step 2: Set up your workbook

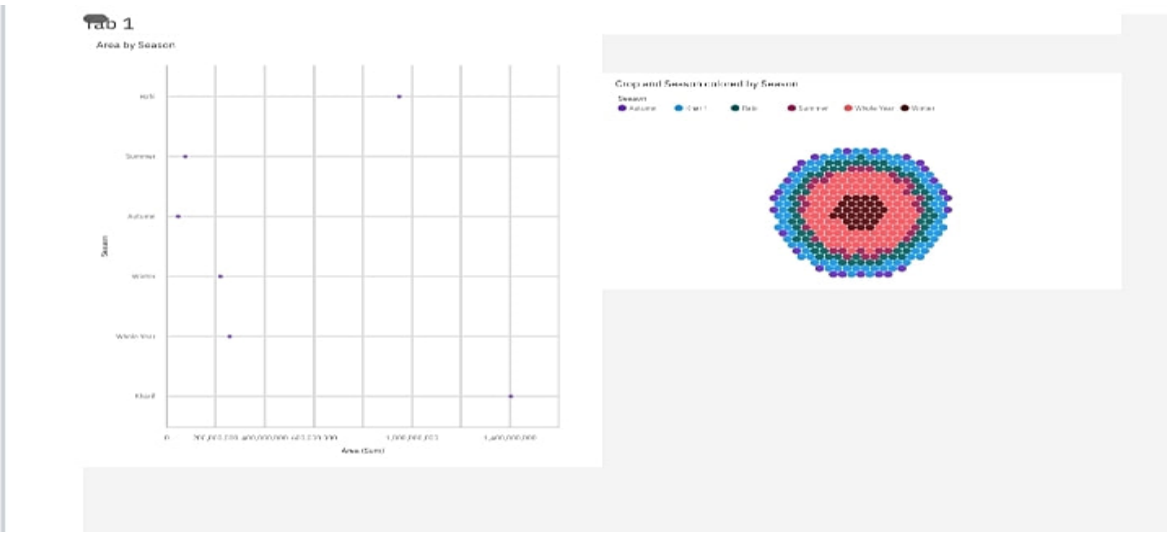
Step 3: Add raw data to a table

Step 4: Data analysis

Step 5: Determine the visuals

Step 6: Create your Excel dashboard

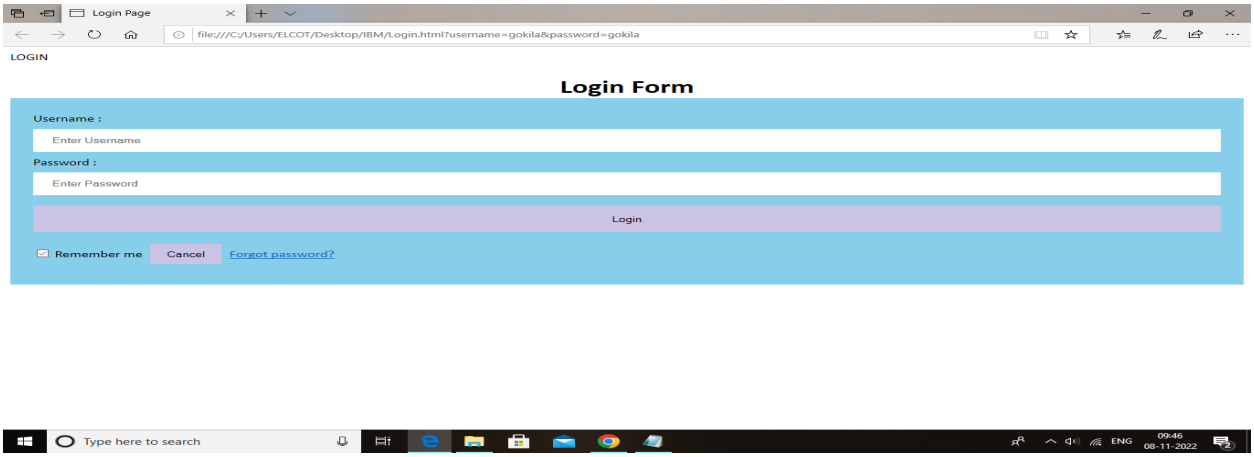
Step 7: Customize your dashboard



8.TESTING

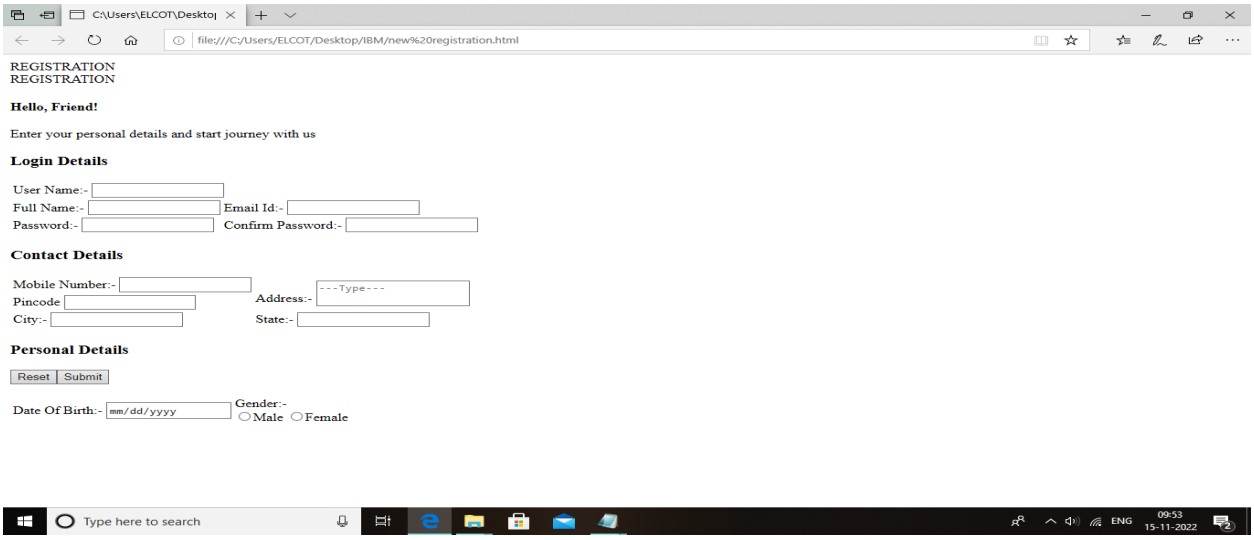
8.1 Test Cases

Test Case No	1
Module Tested	Login Credentials
Input	User Name Password
Expected Output	Entry to the website with correct credentials
Actual Output	Entry to the website with correct credentials
Comments	Successfull

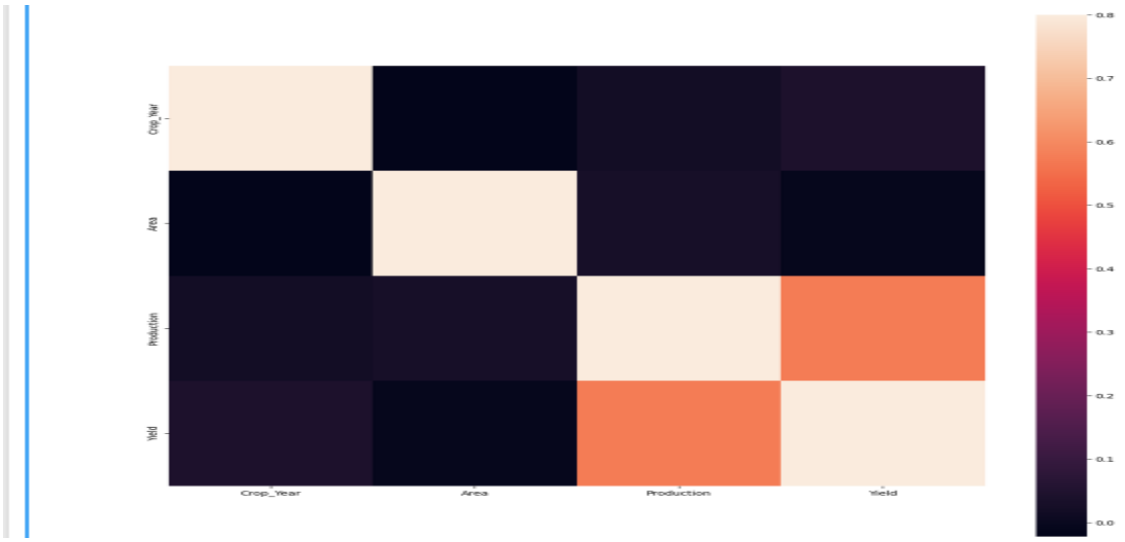


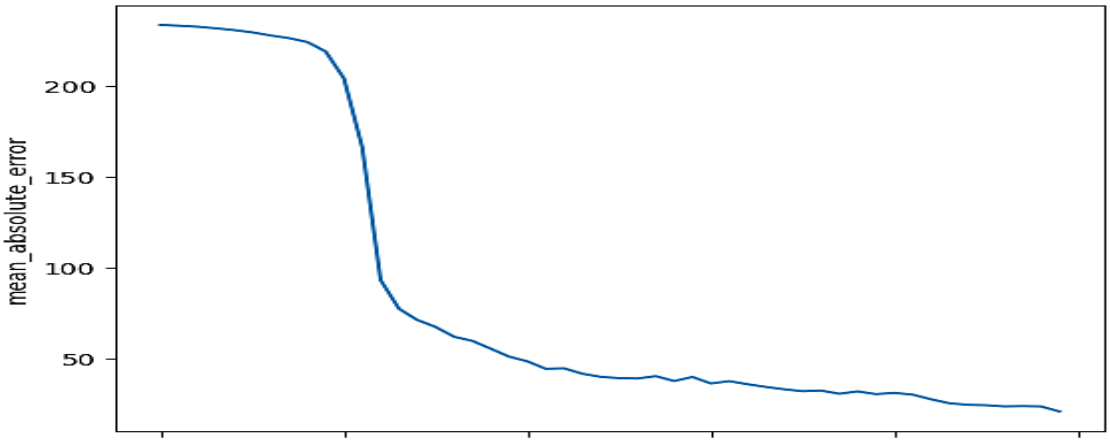
Test Case No	2
Module Tested	Registration Credentials
Input	First Name Last Name Phone Number etc.,
Expected Output	Update Profile Details

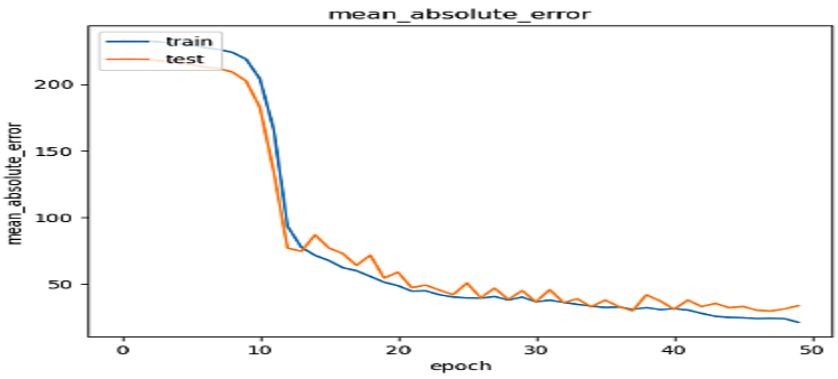
Actual Output	Update Profile Details
Comments	Successful



Test Case No	3
Module Tested	Estimation Result
Input	Dataset(crop name, season...)
Expected Output	Graph is generated based on the analysis
Actual Output	Graph is generated based on the analysis
Comments	Successful


Crop_Year
Area
Production
Yield
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0


mean_absolute_error
200
150
100
50


mean_absolute_error
train
test
mean_absolute_error
200
150
100
50
0
10
20
30
40
50
epoch

8.2 User Acceptance Testing

Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the crop yield estimation through data analytics project at the time of the release to user acceptance testing(UAT).

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	7	5	3	2	17
Duplicate	1	0	2	0	3
External	3	2	0	1	6
Fixed	11	3	5	15	34
Not Reproduced	0	0	0	1	1
Skipped	0	1	0	1	2
Won't Fix	0	3	5	1	9
Totals	22	14	15	21	72

Test Case Analysis

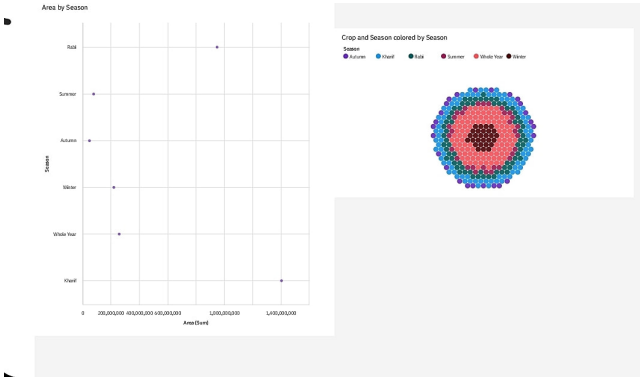
This report shows the number of test cases that have passed, failed, and untested.

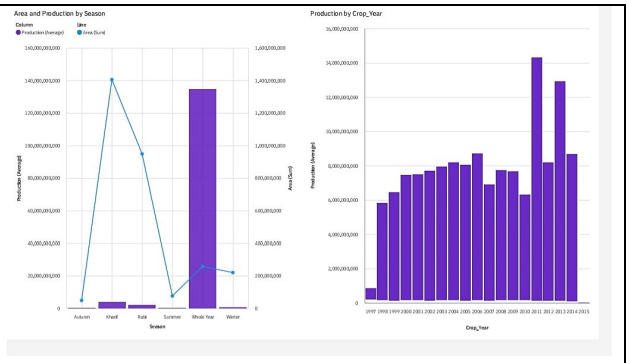
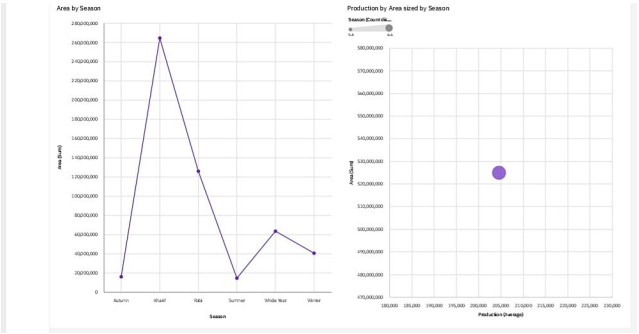
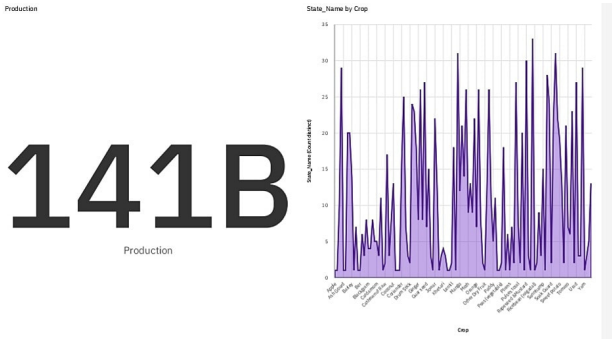
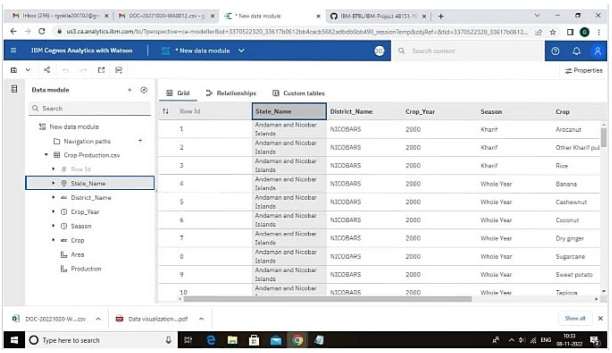
Section	Total Cases	Not Tested	Fail	Pass
Print Engine	6	0	0	6

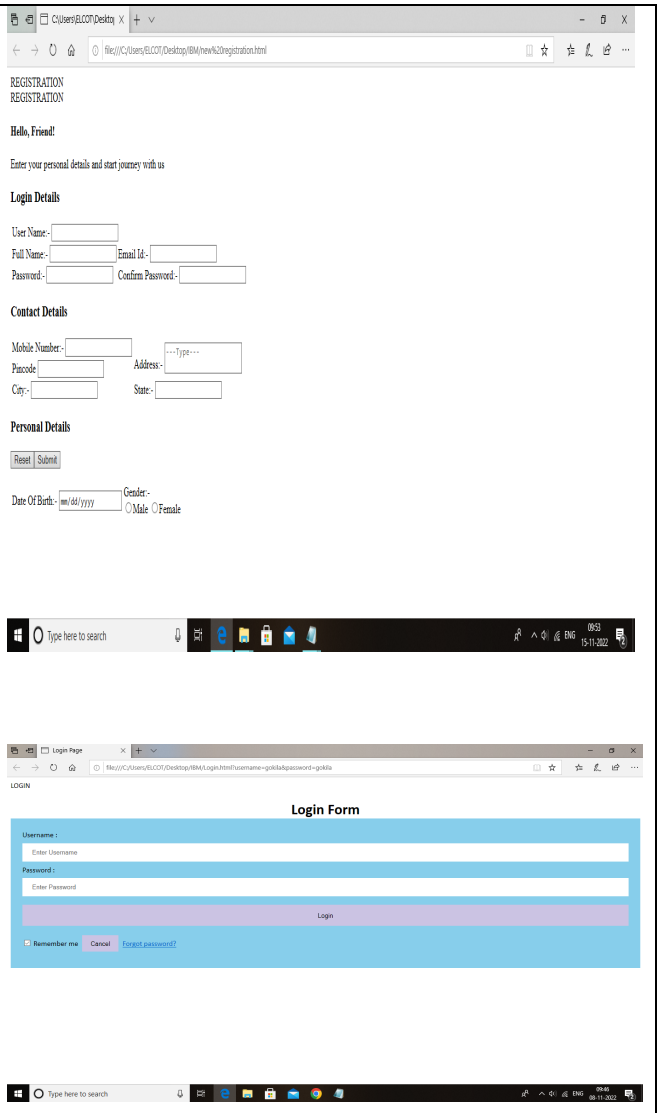
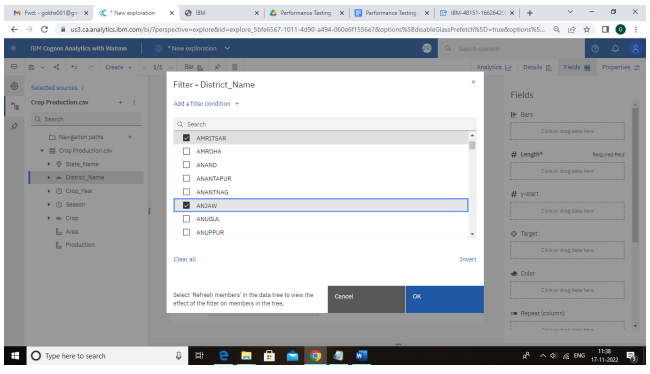
Client Application	40	0	0	40
Security	2	0	0	2
Outsource shipping	3	1	1	2
Exception Reporting	9	7	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

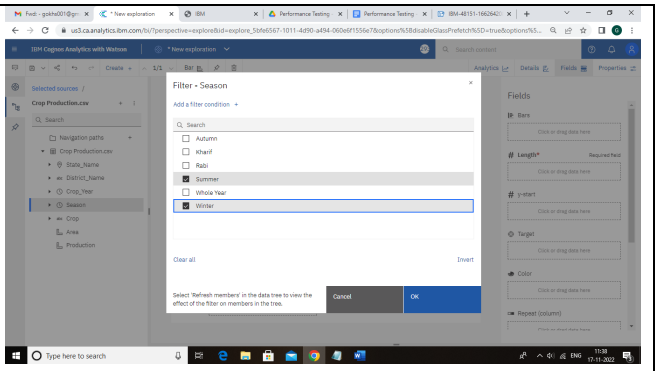
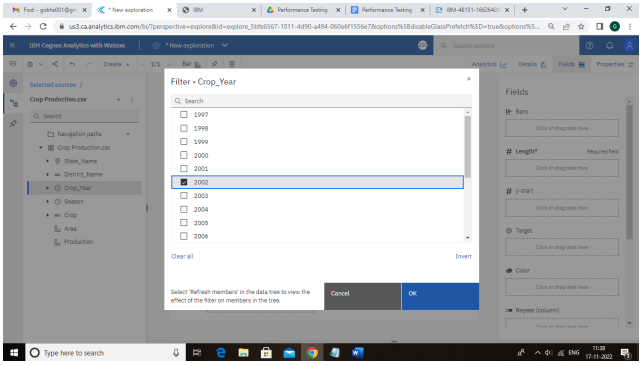
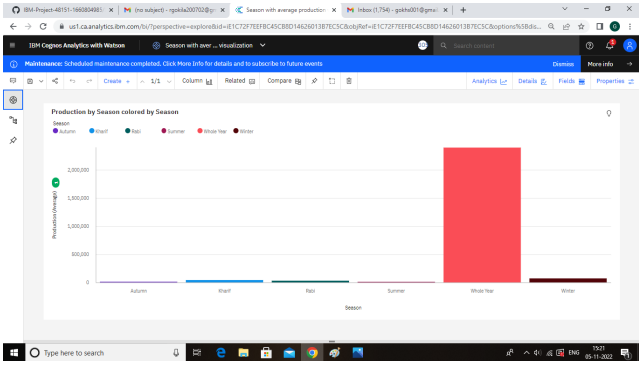
9. RESULTS

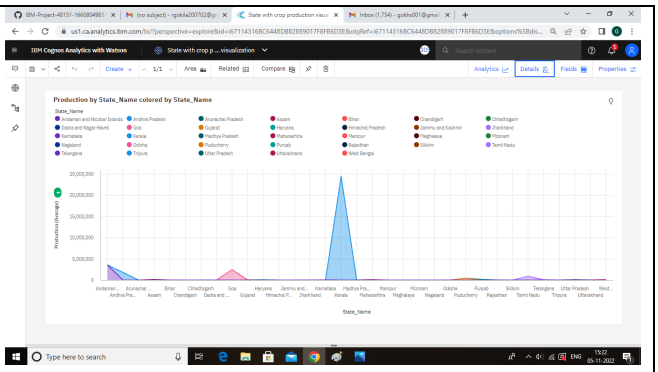
9.1 Performance Metrics

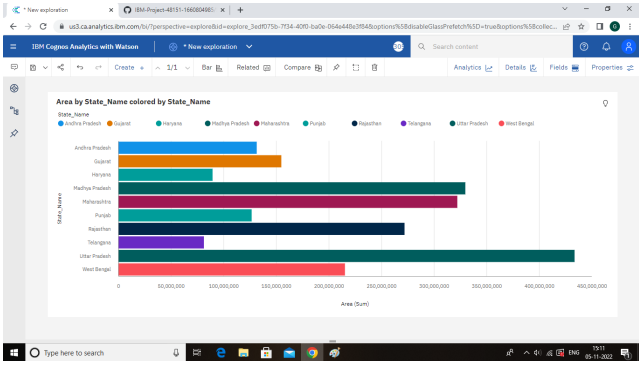
S.No	Parameter	Screenshots / values
1	Dashboard design	No of visualization / Graphs - 10 

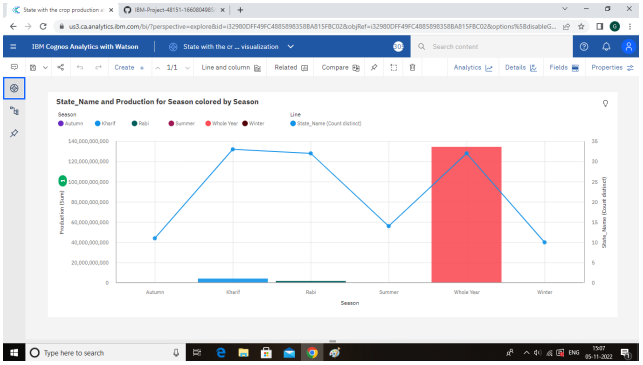
		  
2	Data Responsiveness	

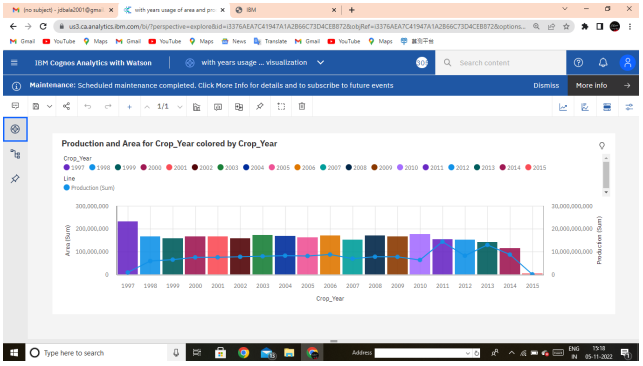
3	Amount Data to Rendered	
4	Utilization of Data Filters	

		 
5	Effective User Story	No of Scene Added - 5
6	Descriptive Reports	No of Visualization / Graphs - 5 


IBM Cognos Analytics with Watson
Production by State_Name colored by State_Name
State_Name
Production (Average)
Type here to search


IBM Cognos Analytics with Watson
Area by State_Name colored by State_Name
State_Name
Andhra Pradesh
Gujarat
Haryana
Madhya Pradesh
Maharashtra
Punjab
Rajasthan
Telangana
Uttar Pradesh
West Bengal
Area (Sum)
Type here to search


IBM Cognos Analytics with Watson
State_Name and Production for Season colored by Season
Season
Autumn
Kharif
Rabi
Summer
Whole Year
Winter
Production (Sum)
Type here to search


IBM Cognos Analytics with Watson
Maintenance: Scheduled maintenance completed. Click More Info for details and to subscribe to future events
Dismiss
More info
Production and Area for Crop_Year colored by Crop_Year
Crop_Year
Production (Sum)
Area (Sum)
Type here to search

10. ADVANTAGES & DISADVANTAGES

Advantages

- Helps in the interpretation of data pattern that assist decision-making and performance improvement.
- Helps in analysing some important visualization, creating a dashboard and by going through these we will get most of the insights of crop production.
- Helps to make the decision about when to plant and harvest crops based of rainfall, season, area.
- This will reduce the waste generated and improve the profit of the farmer in a digitialized way.

Disadvantages

Some uneducated farmers cannot able to use this applicaton.

11. CONCLUSION

As a result of penetration of technology into agriculture field, there is a marginal improvement in the productivity. The innovations have led to new concepts like digital agriculture, smart farming, precision agriculture etc. The activities of agriculture field are numerous like weather forecasting, soil quality assessment, seeds selection, crop yield prediction etc. In this project, the specific activity, crop yield estimation has been surveyed and the major trends have been identified.

It can be concluded that the research in the field of agriculture with reference to using IT trends like data analytics is in its infancy. As the food is the basic need of humans, the requirement of getting the maximum yields using optimal resource will become the necessity in near future as a result of growing population.

Based on our analysis, model will be more accurate if the more datasets are available. So as the data point increases our system will become more and more accurate. Our system accuracy is more than the existing system. Since we are displaying the results in the form of graph with actual and predicted in the graphical user interface it is easy to compare the previous year's data. This model will help farmers to grow the crop which will give
more yield so that it will be more profitable.

12. FUTURE SCOPE

In future, this model can be implemented throughout the India by adding the data points for all the region. According to our analysis model will give more accuracy as the data points increases, so to get better accuracy model data points can be increased. Our system can be integrated with messaging module so that registered farmers can get the notification of the prediction directly to their registered mobile numbers.

This project solves one of the fundamental problems that the Indian farmers are facing that is selection of which type of crop will yield the maximum results.The sole objective is to increase farmer's income. Lack of proper dataset is the major hurdle while predicting the name of the crops but we were able to manage that by merging different data sets. This project right now covers only five features that are season ,area ,temperature, rainfall and crop name but that's not the end, this project holds numerous possibilities such as the addition of vapour pressure, soil quality and market integration.

This project if compiled with a bigger data set can be a boon for the government as it may help them plan properly and in turn help our objective.

13. APPENDIX

Source code

```
import pandas as pd
df = pd.read_csv('DOC-20221020-WA0012..csv', encoding='utf-8')
df

df = df[df['State_Name'] == "Andhra Pradesh"]

df['Yield'] = df['Production']/df['Area']
df

import matplotlib.pyplot as plt
import seaborn as sb
C_mat = df.corr()
fig = plt.figure(figsize = (15,15))
sb.heatmap(C_mat, vmax = .8, square = True)
plt.show()

df = df[df['Crop_Year']>=2004]
df

df = df.join(pd.get_dummies(df['District_Name']))
df = df.join(pd.get_dummies(df['Season']))
df = df.join(pd.get_dummies(df['Crop']))
df = df.join(pd.get_dummies(df['Crop_Year']))
df = df.join(pd.get_dummies(df['State_Name']))
df

df=df.drop('District_Name', axis=1)
df = df.drop('Season',axis=1)
df = df.drop('Crop',axis=1)
```

```
df = df.drop('Crop_Year', axis=1)
df = df.drop('Production', axis=1)
df = df.drop('State_Name', axis=1)
df

from sklearn import preprocessing
x = df[['Area']].values.astype(float)
x
min_max_scaler = preprocessing.MinMaxScaler()
x_scaled = min_max_scaler.fit_transform(x)
x_scaled
df['Area'] = x_scaled
df

df.head()

df = df.fillna(df.mean())

from sklearn.model_selection import train_test_split
b = df['Yield']
a = df.drop('Yield', axis = 1)
a_train, a_test, b_train, b_test = train_test_split(a, b, test_size = 0.3, random_state
= 42)
print(a_train)
print(a_test)
print(b_train)
print(b_test)

import numpy as np
import matplotlib.pyplot as plt
import seaborn as seabornInstance
from sklearn.linear_model import LinearRegression
```

```
from sklearn import metrics
%matplotlib inline

from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
a_train = sc.fit_transform(a_train)
a_test = sc.transform(a_test)

from sklearn.ensemble import RandomForestRegressor
  regr = RandomForestRegressor(max_depth=2, random_state=0,
n_estimators=100)
  regr.fit(a_train, b_train)
  b_pred = regr.predict(a_test)
  from sklearn.metrics import mean_squared_error as mse
  from sklearn.metrics import mean_absolute_error as mae
  from sklearn.metrics import r2_score
 print('MSE =', mse(b_pred, b_test))
  print('MAE =', mae(b_pred, b_test))
  print('R2 Score =', r2_score(b_pred, b_test))

from sklearn.svm import SVR
regressorpoly=SVR(kernel='poly',epsilon=1.0)
regressorpoly.fit(a_train,b_train)
pred=regressorpoly.predict(a_test)
print(regressorpoly.score(a_test,b_test))
print(r2_score(b_test,b_pred))

from keras.callbacks import ModelCheckpoint
from keras.models import Sequential
from keras.layers import Dense, Activation, Flatten
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
```

```
from sklearn.metrics import mean_absolute_error
from matplotlib import pyplot as plt
import seaborn as sb
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import warnings
from keras.callbacks import History
warnings.filterwarnings('ignore')
warnings.filterwarnings('ignore', category=DeprecationWarning)

NN_model = Sequential()
# The Input Layer :
NN_model.add(Dense(128, kernel_initializer='normal',input_dim =
a_train.shape[1], activation='relu'))
# The Hidden Layers :
NN_model.add(Dense(256, kernel_initializer='normal',activation='relu'))
NN_model.add(Dense(256, kernel_initializer='normal',activation='relu'))
NN_model.add(Dense(256, kernel_initializer='normal',activation='relu'))
# The Output Layer :
NN_model.add(Dense(1, kernel_initializer='normal',activation='linear'))
# Compile the network :
NN_model.compile(loss='mean_absolute_error', optimizer='adam',
metrics=['mean_absolute_error'])
NN_model.summary()

from keras.callbacks import History
history = History()
History=NN_model.fit(a_train, b_train, epochs=50, batch_size=500,
validation_split = 0.2, callbacks=[history])
```

```
print(history.history.keys())
plt.plot(History.history['mean_absolute_error'])
plt.ylabel('mean_absolute_error')
plt.xlabel('epoch')

plt.plot(History.history['mean_absolute_error'])
plt.plot(History.history['val_mean_absolute_error'])
plt.title('mean_absolute_error')
plt.ylabel('mean_absolute_error')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
 # summarize history for loss
plt.plot(History.history['loss'])
plt.plot(History.history['val_loss'])
plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
```

Github & Project Demo Link

https://github.com/IBM-EPBL/IBM-Project-48151-1660804985

https://youtu.be/OW_ilBKzSd8