PROJECT REPORT

ESTIMATED CROP YIELD USING DATA ANALYTICS

TEAM ID: PNT2022TMID37320

VARUN KUMAR [TL] – 311619104079
UDHAYA KUMAR S [TM1] -311619104077
NAVEEN KUMAR SR[TM2] – 311619104046
SANTOSHKUMAR M[TM3] – 311619104061

Project Report Format

1. **INTRODUCTION**
   - 1.1 Project Overview
   - 1.2 Purpose
2. **LITERATURE SURVEY**
   - 2.1 Existing problem
   - 2.2 References
   - 2.3 Problem Statement Definition
3. **IDEATION & PROPOSED SOLUTION**
   - 3.1 Empathy Map Canvas
   - 3.2 Ideation & Brainstorming
   - 3.3 Proposed Solution
   - 3.4 Problem Solution fit
4. **REQUIREMENT ANALYSIS**
   - 4.1 Functional requirement
   - 4.2 Non-Functional requirements
5. **PROJECT DESIGN**
   - 5.1 Data Flow Diagrams
   - 5.2 Solution & Technical Architecture
   - 5.3 User Stories
6. **PROJECT PLANNING & SCHEDULING**
   - 6.1 Sprint Planning & Estimation
   - 6.2 Sprint Delivery Schedule
   - 6.3 Reports from JIRA
7. **CODING & SOLUTIONING (Explain the features added in the project along with code)**
   - 7.1 Feature 1
   - 7.2 Feature 2
   - 7.3 Database Schema (if Applicable)
8. **TESTING**
   - 8.1 Test Cases
   - 8.2 User Acceptance Testing
9. **RESULTS**
   - 9.1 Performance Metrics
10. **ADVANTAGES & DISADVANTAGES**
11. **CONCLUSION**
12. **FUTURE SCOPE**
13. **APPENDIX**

    Source Code

    GitHub & Project Demo Link

1 INTRODUCTION

1.1 PROJECT OVERVIEW

Agriculture, since its invention and inception, be the prime and pre-eminent activity of every culture and civilization throughout the history of mankind. It is not only anenormous aspect of the growing economy, but it's essential for us to survive. It's also a crucial sector for Indian economy and also human future. It also contributes an outsized portion of employment. Because the time passes the requirement for production has beenincreased exponentially. So as to produce in mass quantity people are using technology inan exceedingly wrong way. New sorts of hybrid varieties are produced day by day. However, these varieties don't provide the essential contents as naturally produced crop. These unnatural techniques spoil the soil. It all ends up in further environmental harm. Most of these unnatural techniques are wont to avoid losses. Machine learning, a fast- growing approach that's spreading out and helping every sector in making viable decisionsto create the foremost of its applications. Most devices nowadays are facilitated by modelsbeing analysed before deployment. The main concept is to increase the throughput of theagriculture sector with the Machine Learning models. Another factor that also affects theprediction is the amount of knowledge that's being given within the training period, as thenumber of parameters was higher comparatively.

1.2 PURPOSE

Data Analytics is the computing process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Itis an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Data mining (the analysis step of the "Knowledge Discovery inDatabases" process, or KDD), a field at the intersection of computer science and statistics, is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and systems. The overall goal of the data mining process is to extract information from a data set and transform it intoan understandable structure for further.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

Indian Agriculture sector requires innumerable types of data analytics in various sectors such as crop productivity prediction models, economic models, pest and crop disease prediction models, crop price forecasting models, etc. The frequent changes in climate conditions are affecting more in cotton production. Most of the forecasts are seasonal and are available around 1-2 months before the crop harvesting. Farmers are benefited if recommendation and forecast of particular crop are available before sowing of crop.

2.2 References

1. Kodimalar Palanivel, an approach for prediction of crop Yield using machine learning and big Data techniques,2020.
2. Subhadra Mishra, Adaptive boosting of weak regressors for forecasting of crop productionconsidering climatic variability: An empirical assessment,2020.
3. Tanha Talaviya, Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides,2020.

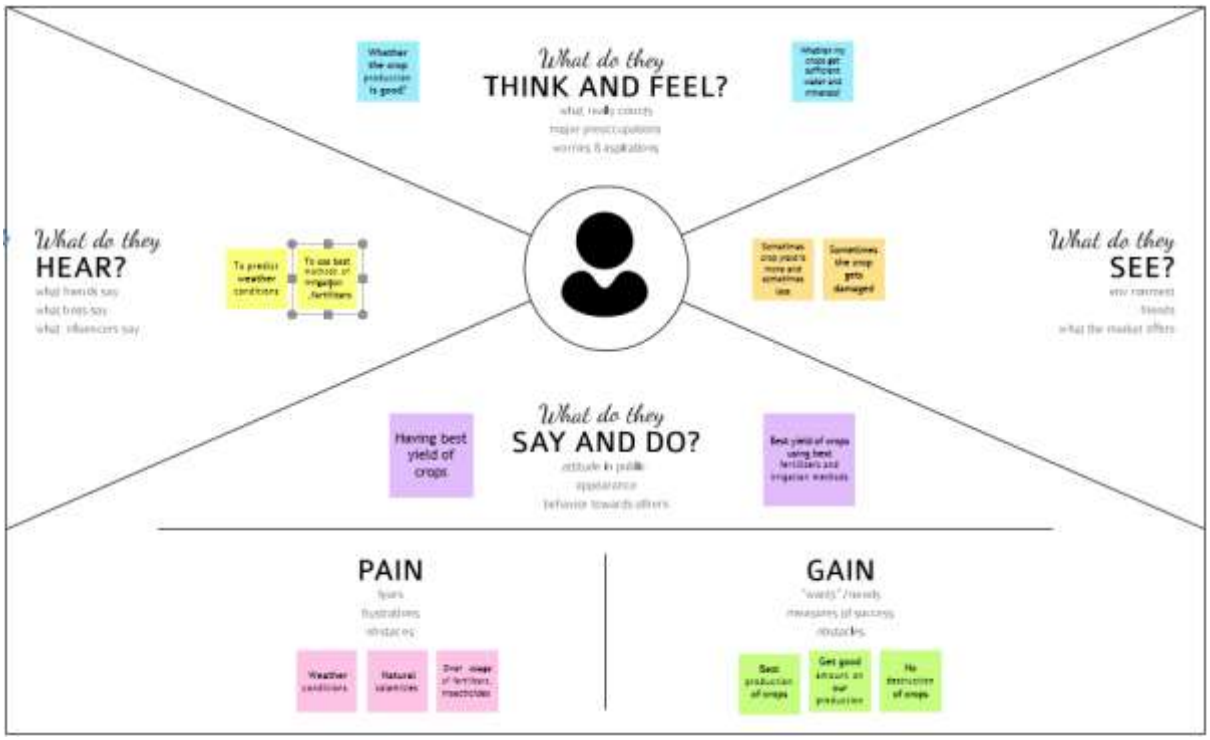
2.3 Problem Statement Definition

1. Crop production in India is one of the most important sources of income and India is one ofthe top countries to produce crops.
2. Where Digital Farming and Precision Agricultureallow precise utilization of inputs like seed, water, pesticides, and fertilizers at the right timefor the crop for maximizing productivity, healthycrop production. uality, and yields.
3. Most of farmers practice traditional farming patterns to decide on crops to be cultivated in a field.

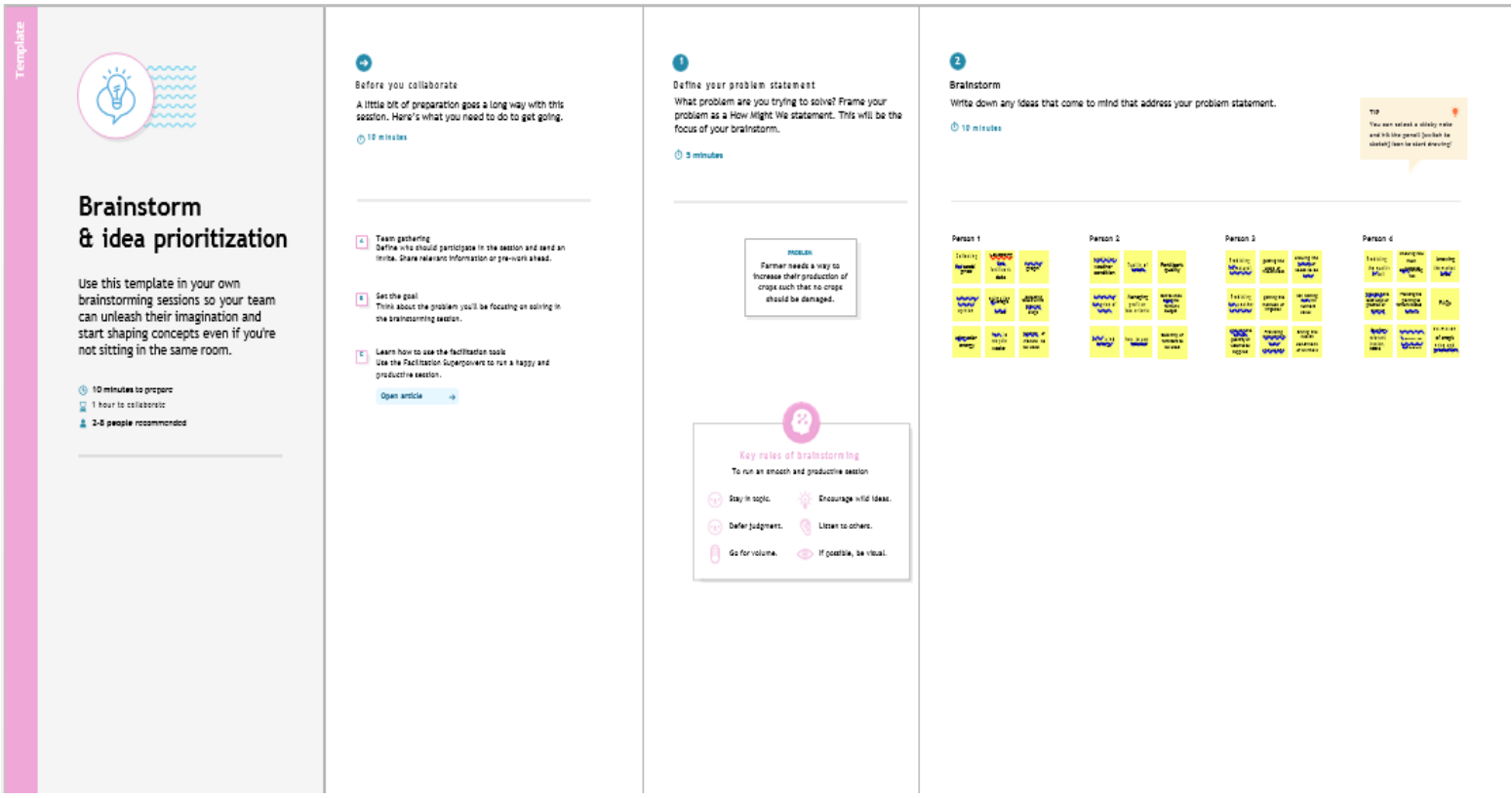
3.1 **IDEATION & PROPOSED SOLUTION**

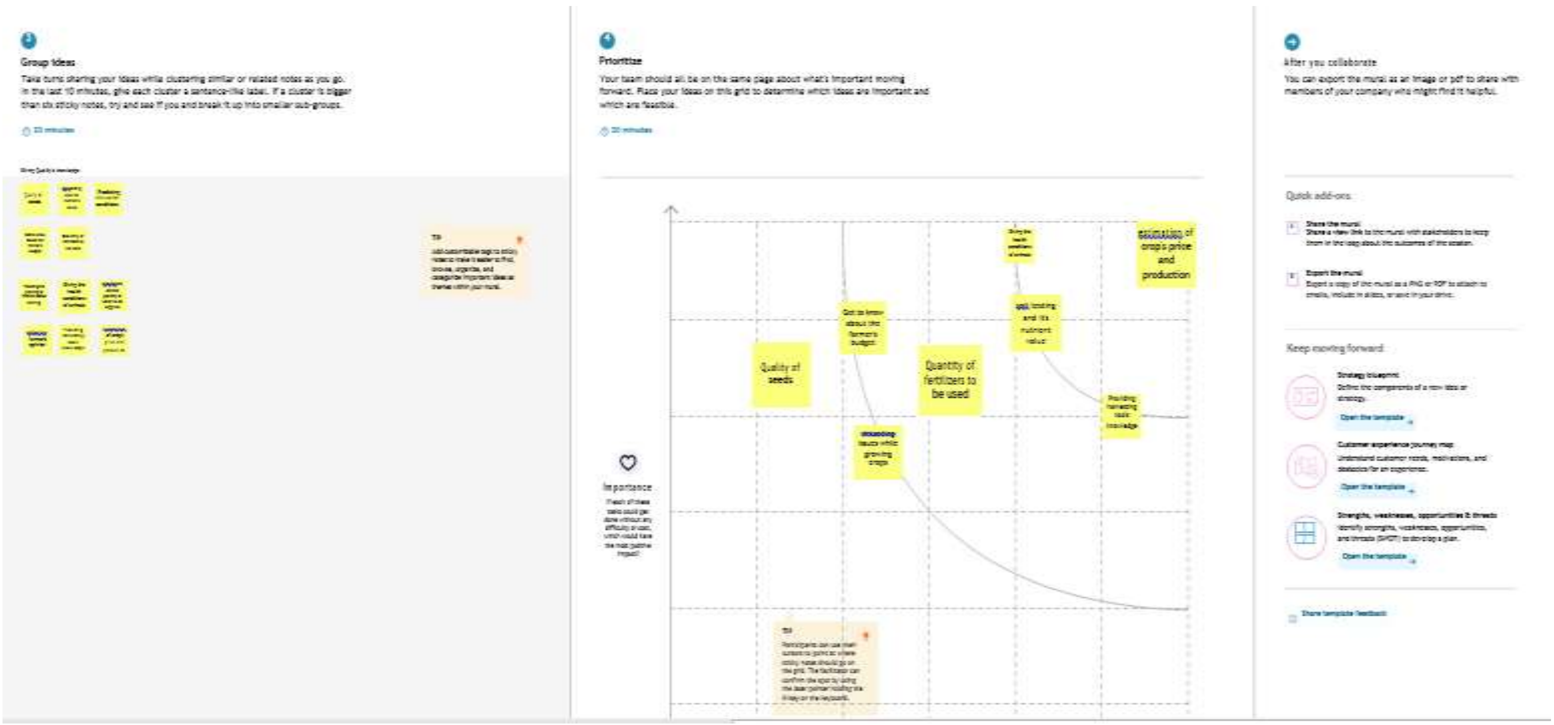
3.2

3.3 Empathy Map Canvas



3.2 Ideation & Brainstorming



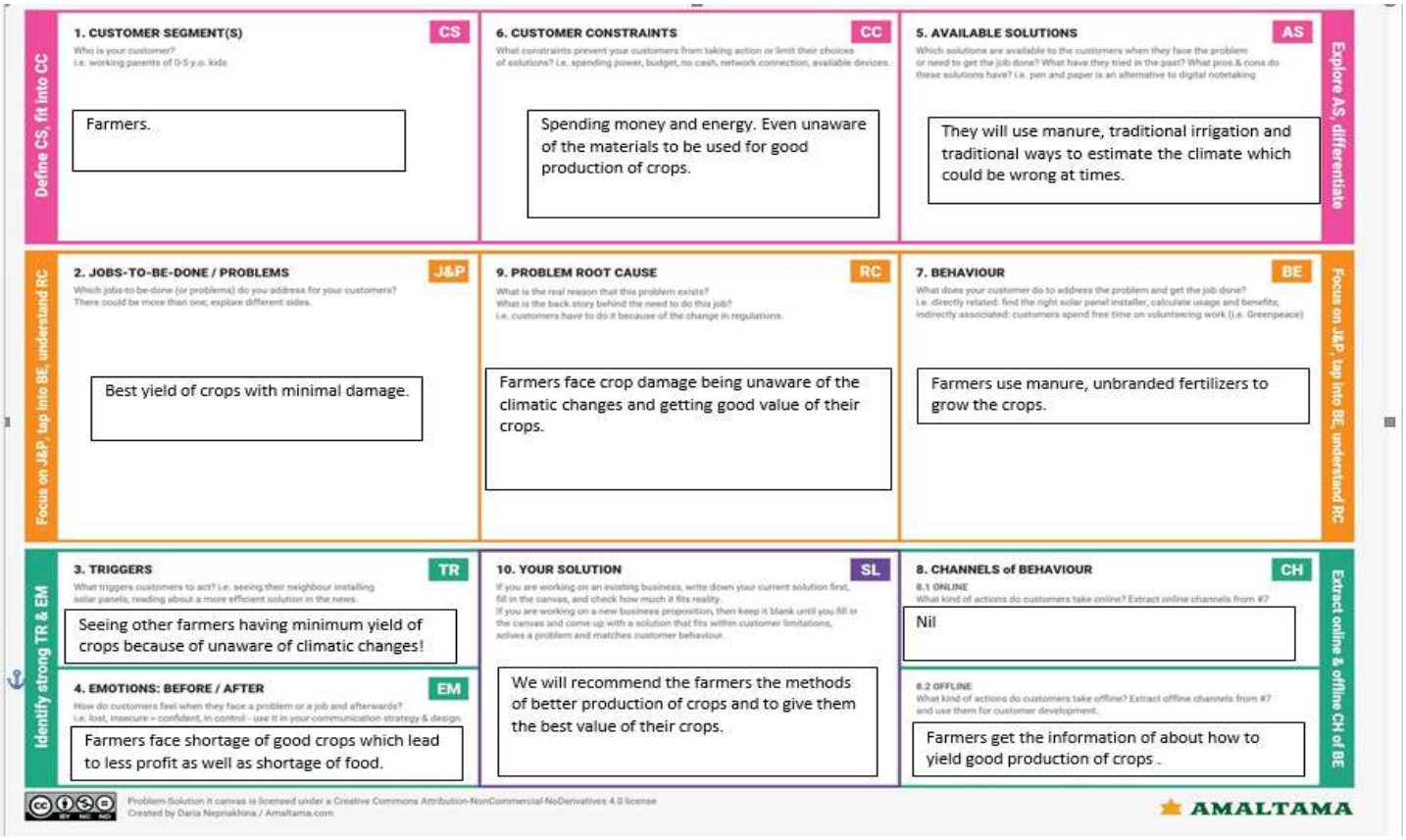


Group Ideas
Take turns sharing your ideas while clustering similar or related notes as you go. In the last 10 minutes, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.
Prioritize
Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.
30 minutes
estimation of crops price and production
Got to know about the farmer's budget
Quality of seeds
Quantity of fertilizers to be used
Providing harvesting tools knowledge
Importance
After you collaborate
You can export the mural as an image or pdf to share with members of your company who might find it helpful.
Quick add-ons
Share the mural
Export the mural
Keep moving forward
Customer experience journey map
Open the template
Strengths, weaknesses, opportunities & threats

3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Crop production in India is one of the most important sources of income and India is one of the top countries to produce crops. Where Digital Farming and Precision Agriculture allow precise utilization of inputs like seed, water, pesticides, and fertilizers at the right time for the crop for maximizing productivity, healthy crop production. uality, and yields. Most of farmers practice traditional farming patterns to decide on crops to be cultivated in a field.
2.	Idea / Solution description	Applying data Analytics methods for predicting the crop production across various areas let us to estimate the optimal crop production assisting the framers to benefit from the forecast. We can comprehend the data and make wise decisions by integrating reporting, modelling, analysis, exploration, dashboards, stories, and event management with IBM Cognos Analytics. By presenting critical insights and analyses about our data on one or more pages or screens, a dashboard enables us to keep track of events or actions at a glance. In this project, we use a dashboard to view, analyse, and extract the majority of the findings.
3.	Novelty / Uniqueness	To visualize the past crop yield data and to list out the crops that may yield poor production leading to loss of invested revenue and identify suitable areas for their production. Consideration of all factors that affect crop yield.
4.	Social Impact / Customer Satisfaction	Extreme weather conditions such as high temperature, heavy storms or droughts can severely disrupt crop production.
5.	Business Model (Revenue Model)	Increased amount of waste produced from the crop production may lead to a degrade of profit margin
6.	Scalability of the Solution	The acquired insights from the visualization of crop yield data must be durable in such a way that the production is fairly stable even in sudden change of conditions.

3.4 Problem Solution Fit



4.1Requirement Analysis

4.1 Functional Requirement

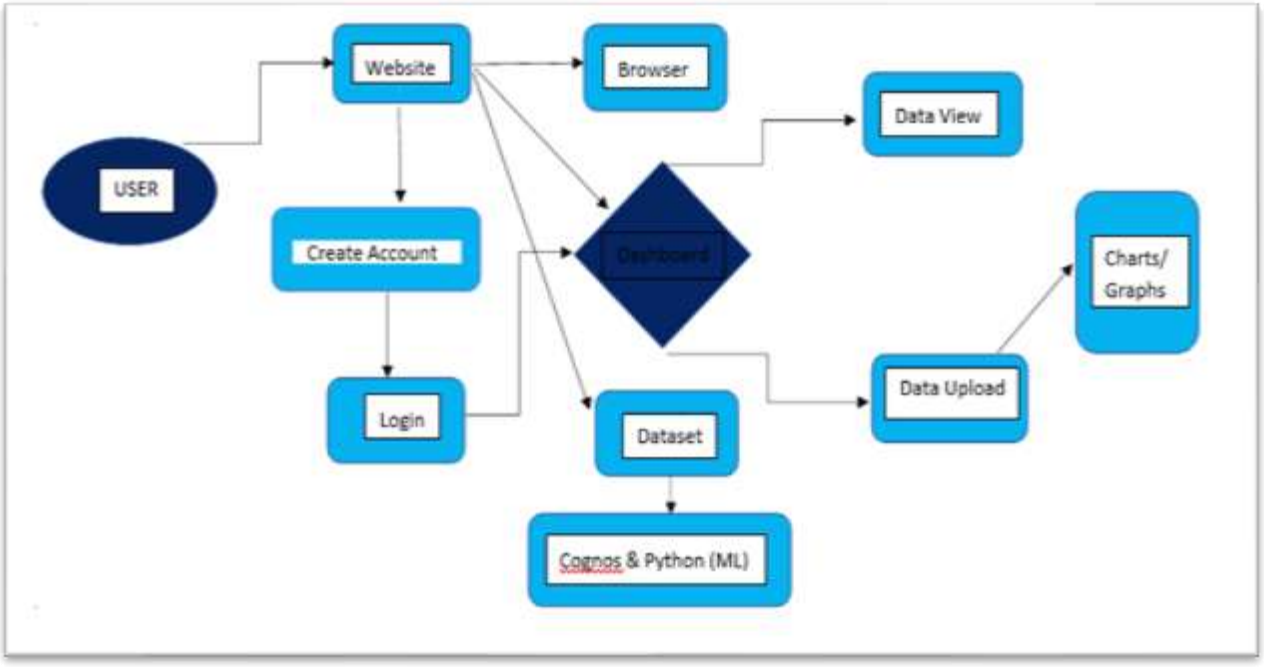
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	**User Registration**	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	**User Confirmation**	Confirmation via Email Confirmation via OTP
FR-3	**User Profile**	Log in Access the profile
FR-4	**Give the required data**	Take the data given by the user as the input for the analysis
FR-5	**Analysis**	Analyse the yield of crop from the data given by the user
FR-6	**Estimation or Predict the data**	Estimate the crop yield from the analysis,using the software from the data given by the user

4.2 Non Functional Requirement

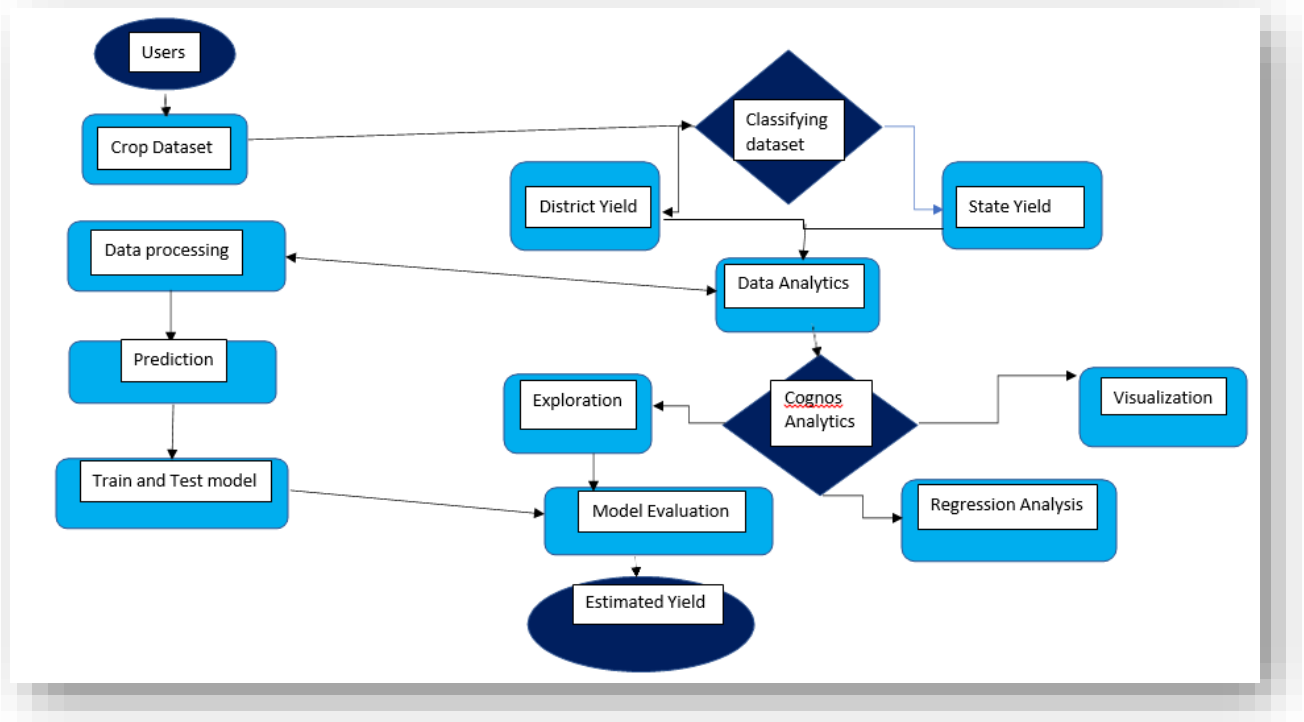
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Crop recommendations are created and saved, the these recommended crops are sown by farmers for increased crop yield.
NFR-2	**Security**	The software keeps the users information more securely.
NFR-3	**Reliability**	Creating the interactive dashboards which is easy to understand and useful for the users.
NFR-4	**Performance**	It is user friendly software and have high performance.
NFR-5	**Availability**	The software application is easily available for every user and accessing is easy for them.
NFR-6	**Scalability**	The proposed system allows the implementation of a flexible methodology that can be used to estimate the yield of crops in different types of lands.

5.PROJECT DESIGN

5.1Data Flow Diagram



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 byentering my email, password, and confirmingmy 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 confirmationemail & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with FacebookLogin	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	USN-6	I can access the dashboard of mine.		Medium	Sprint-2
Customer (Webuser)	Access resources	USN-7	I can use my credentials for accessing my resources.	Other than me, there is less chance to access myresources.	High	Sprint-1
Customer Care Executive		USN-8	As customer care executive I will always be available for the interaction with the customerto clarify the queries.	An executive will note down the customers complaints and solve theirproblems.	High	Sprint-2
Administrator	Updating data	USN-9	Collecting the data and store it	Checking and updating dataset	High	Sprint-1
Customer tools	Tools	USN-10	I can perform analysis by tools (cognos andwith ML)	I have an ease of accessing tools.	High	Sprint-1

6. PROJECT PLANNING & SCHEDULING

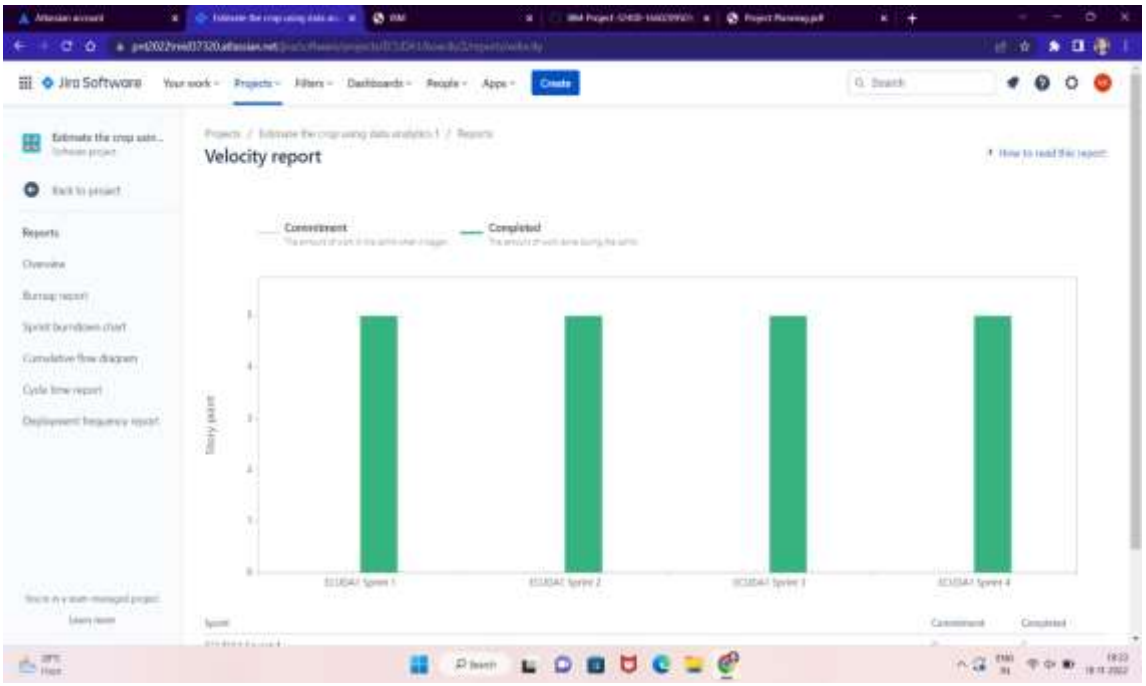
6.1 Sprint Planning & Estimation

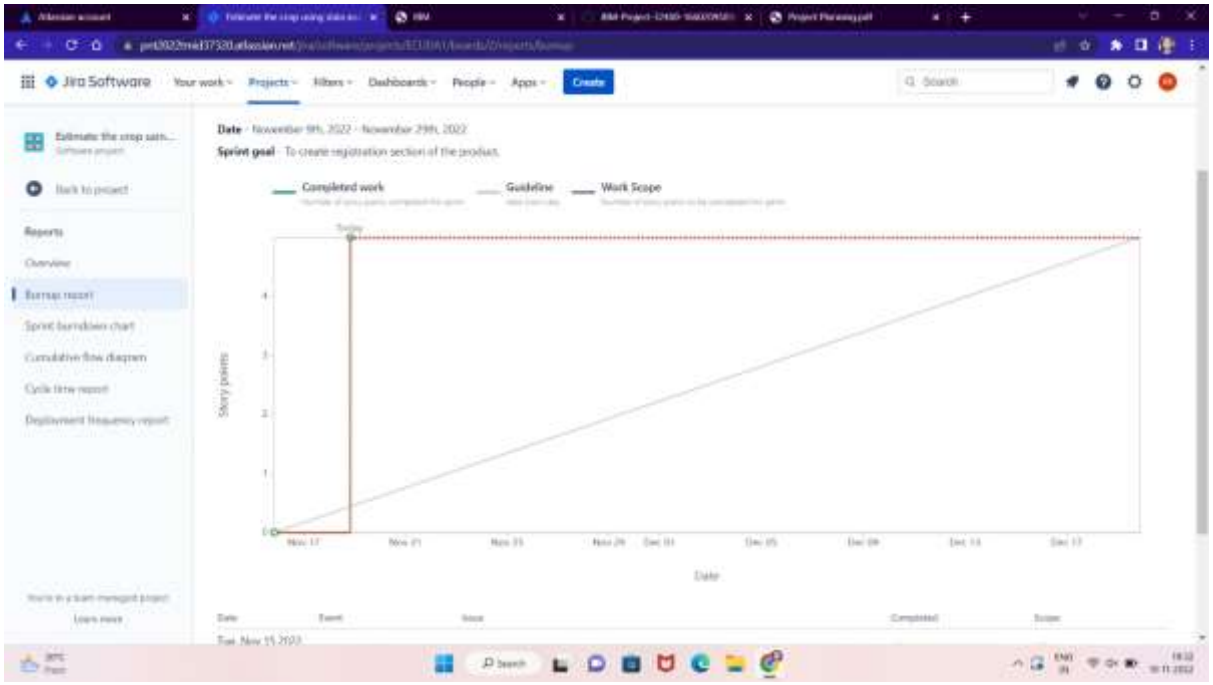
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Login	USN-1	As a user, I can register for the application byentering my phone number and name,	2	Medium	Naveen Kumar SR
Sprint-1	Analysis and Estimation(Working and Loading the dataset)	USN-4	As a user, I can view the crop forecast of thepresent and upcoming days and upload the dataset and loading the dataset	3	High	Varun Kumar
Sprint-2	Analysis and Estimation(Data Visualization Charts)	USN-5	As a user, I can visualise the data of crop production to know the insights Where Average Crop Production by Seasons, the Yearly usage of Area in Crop Production, top 10 States in Crop Yield Production by Area, theCrop Production by State and the Sates with Seasonal Crop Production can be known.	20	High	Santosh Kumar M Udhaya Kumar S
Sprint-3	Dashboard	USN-6	As a User , I can use Cognos Analytics with Watson Services, An interactive dashboard must be created and viewed.	20	High	Varun Kumar
Sprint-4	Analysis and Estimation(Exportation /Export The Analytics)	USN-7	As a user, I can view the dashboard and visualization of crop production that is being exported either through email/link/pdf.	20	High	Varun Kumar

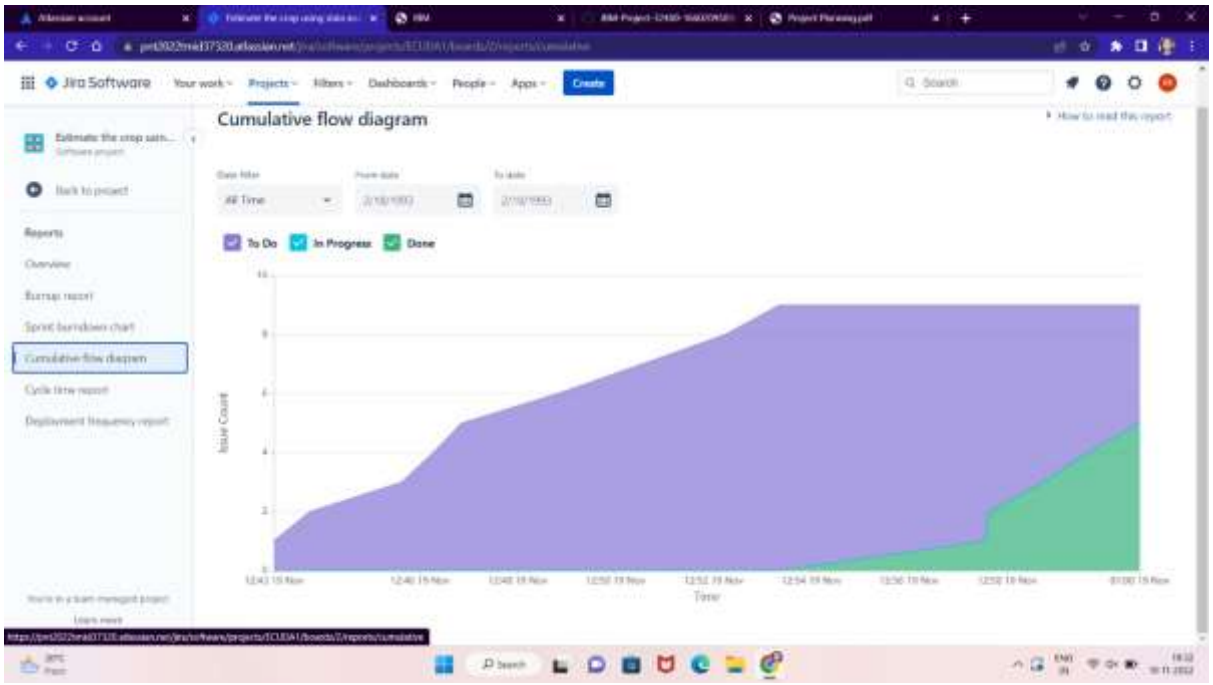
6.2 Sprint Delivery Schedule

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

6.3 Reports From Jira



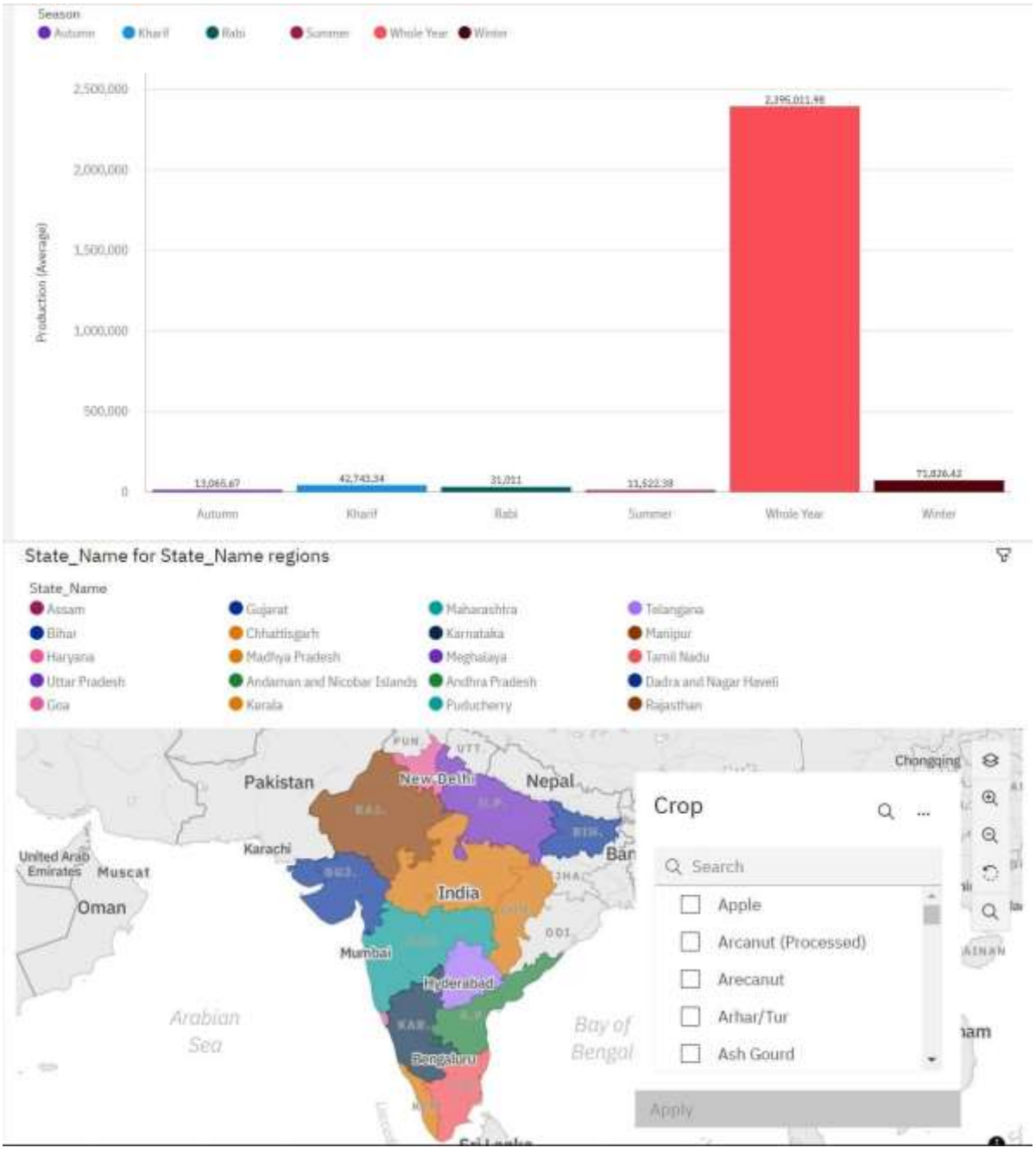



Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Cumulative flow diagram
How to read this report
Estimate the crop usin...
Software project
Back to project
Reports
Overview
Burnup report
Sprint burndown chart
Cumulative flow diagram
Cycle time report
Deployment frequency report
You're in a team-managed project
Learn more
Date filter
All Time
From date
To date
To Do
In Progress
Done
Issue Count
Time

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

7.1 Feature 1

First developing the dashboard the creating the dashboard for various

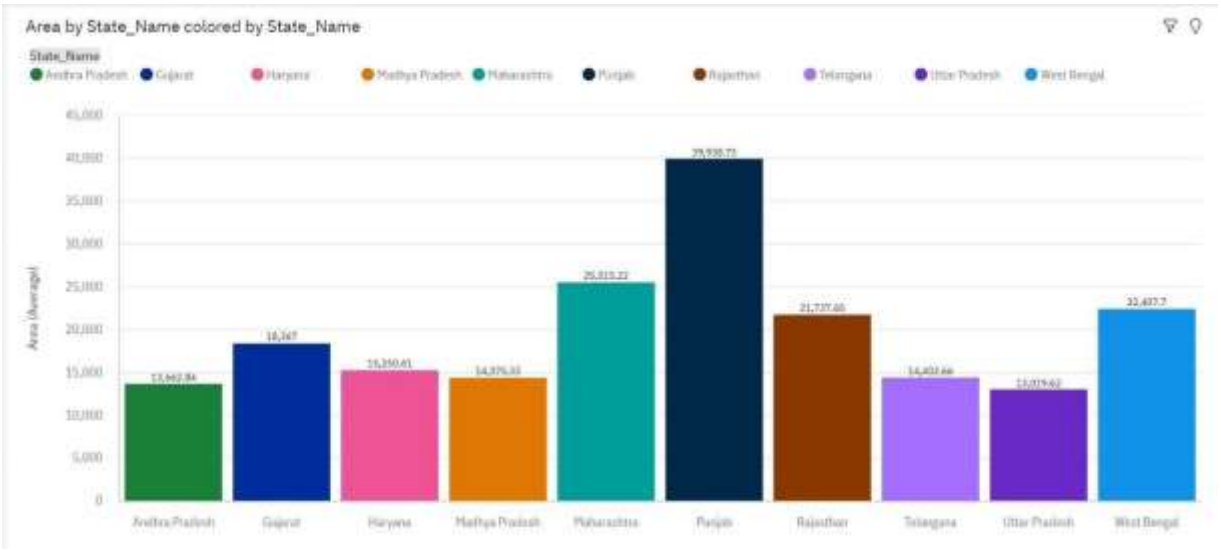


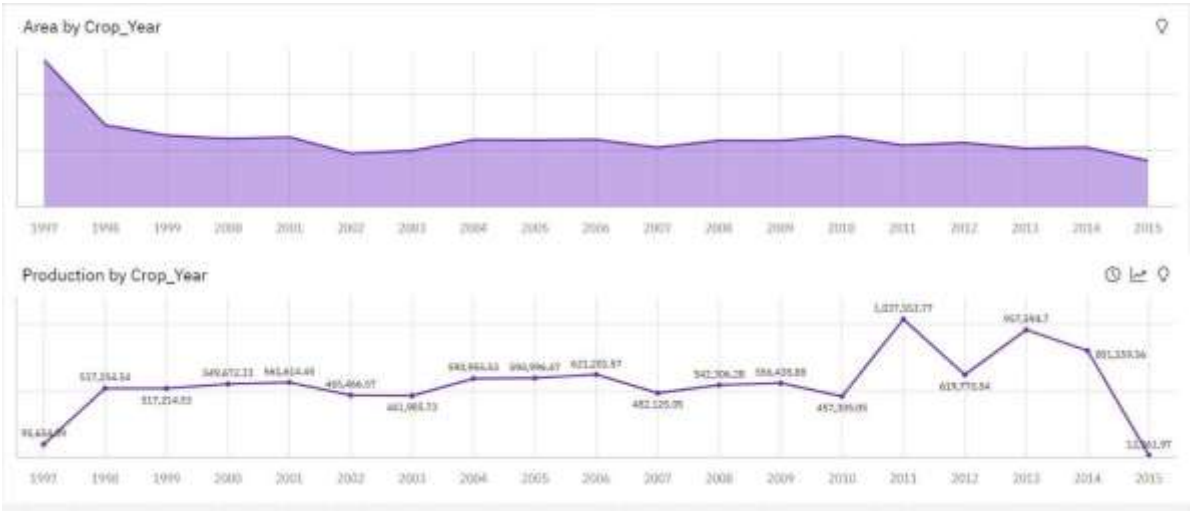
State_Name and Crop

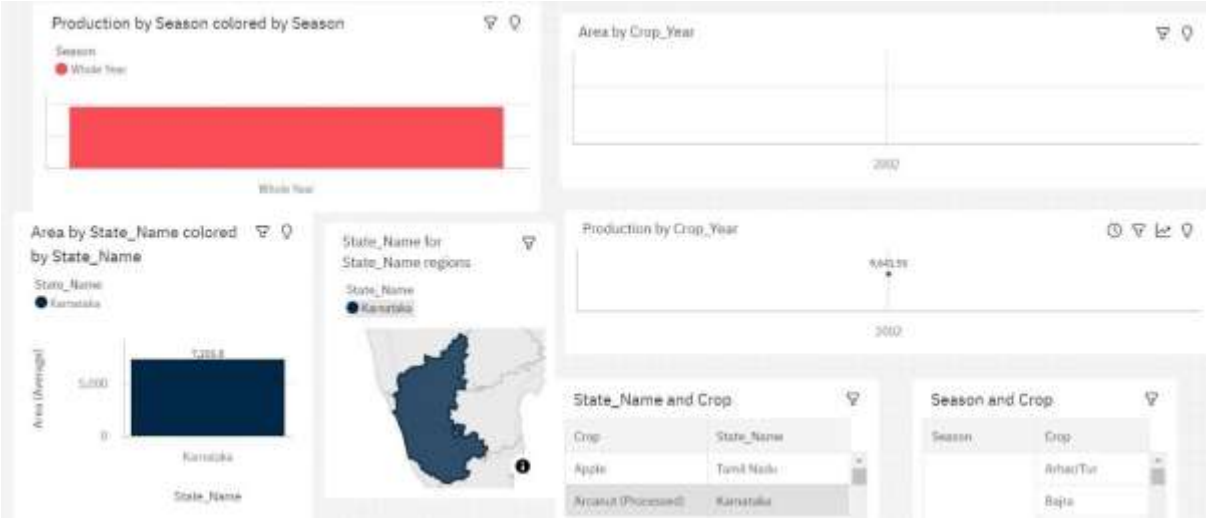
Crop	State_Name
Apple	Tamil Nadu
Ash Gourd	Tamil Nadu
Grapes	Andhra Pradesh
	Haryana
	Karnataka
	Madhya Pradesh
	Maharashtra
	Rajasthan

Season and Crop

Season	Crop
Kharif	Grapes
Whole Year	Apple
	Ash Gourd
	Grapes



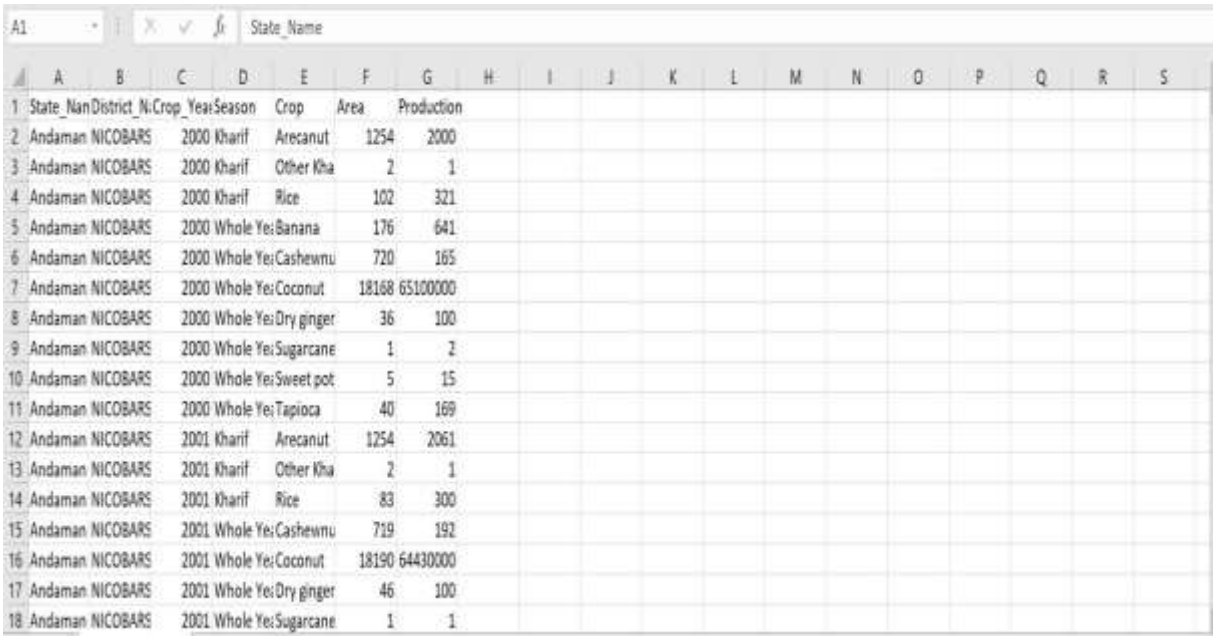




7.2 Feature 2

In this feature 2 we developed project using MLP an machine learning algorithm inorder tocreate the extra feature of creating the crop recommendation for the better yield. Multi-layer perception is also known as MLP. It is fully connected dense layers, which transform any input dimension to the desired dimension. A multi-layer perception is a neural network that has multiple layers. To create a neural network wecombine neurons together so that the outputs of some neurons are inputs of other neurons.A gentle introduction to neural networks and TensorFlow can be found here

7.3 DATABASE SCHEMA



A1 State_Name

	A	B	C	D	E	F	G
1	State_Nan	District_N	Crop_Yea	Season	Crop	Area	Production
2	Andaman	NICOBARS	2000	Kharif	Arecanut	1254	2000
3	Andaman	NICOBARS	2000	Kharif	Other Kha	2	1
4	Andaman	NICOBARS	2000	Kharif	Rice	102	321
5	Andaman	NICOBARS	2000	Whole Ye	Banana	176	641
6	Andaman	NICOBARS	2000	Whole Ye	Cashewnu	720	165
7	Andaman	NICOBARS	2000	Whole Ye	Coconut	18168	65100000
8	Andaman	NICOBARS	2000	Whole Ye	Dry ginger	36	100
9	Andaman	NICOBARS	2000	Whole Ye	Sugarcane	1	2
10	Andaman	NICOBARS	2000	Whole Ye	Sweet pot	5	15
11	Andaman	NICOBARS	2000	Whole Ye	Tapioca	40	169
12	Andaman	NICOBARS	2001	Kharif	Arecanut	1254	2061
13	Andaman	NICOBARS	2001	Kharif	Other Kha	2	1
14	Andaman	NICOBARS	2001	Kharif	Rice	83	300
15	Andaman	NICOBARS	2001	Whole Ye	Cashewnu	719	192
16	Andaman	NICOBARS	2001	Whole Ye	Coconut	18190	64430000
17	Andaman	NICOBARS	2001	Whole Ye	Dry ginger	46	100
18	Andaman	NICOBARS	2001	Whole Ye	Sugarcane	1	1

8. Testing

8.1 Test Case:

A test case has components that describe input, action and an expected response, in order to determine if a feature of an application is working correctly. A test case is a set of instructions on “HOW” to validate a particular test objective/target, which when followed will tell us if the expected behavior of the system is satisfied or not.

Characteristics of a good test case:

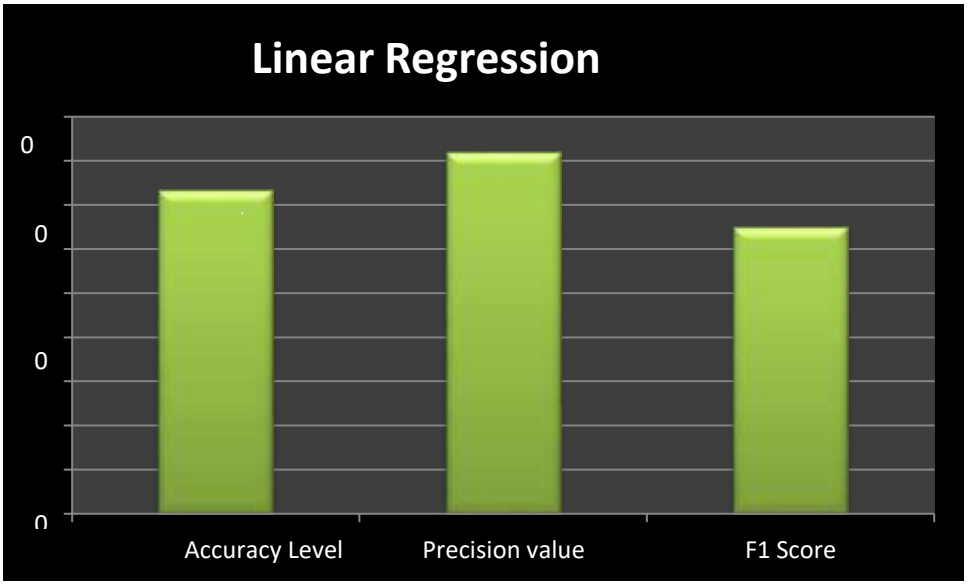
- Accurate: Exacts the purpose.
- Economical: No unnecessary steps or words.
- Traceable: Capable of being traced to requirements.
- Repeatable: Can be used to perform the test over and over.
- Reusable: Can be reused if necessary

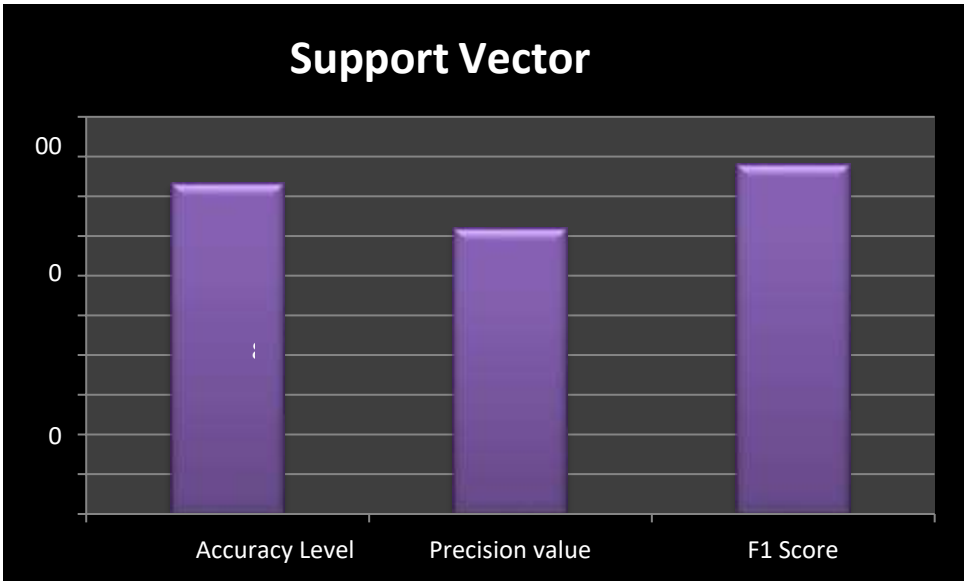
8.2 USER ACCEPTANCE TESTING

Acceptance testing can be defined in many ways, but a simple definition is the succeedswhen the software functions in a manner that can be reasonable expected by the customer.After the acceptance test has been conducted, one of the two possible conditions exists. This is to fine whether the inputs are accepted by the database or other validations. For example accept only numbers in the numeric field, date format data in the date field. Alsothe null check for the not null fields. If any error occurs then show the error messages. Thefunction of performance characteristics to specification and is accepted. A deviation from specification is uncovered and a deficiency list is created. User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

9. RESULTS

9.1 PERFORMANCE METRICS





10. ADVANTAGES & DISADVANTAGES

ADVANTAGES

- It effectively manages vast volumes of input data.
- Following training, quickly makes predictions.
- Even with less samples, the same accuracy ratio is still possible.

DISADVANTAGES

- Provide high number of false positive
- Binary classification can be occurred
- Computational complexity

11 CONCLUSION

We presented a machine learning approach for crop yield prediction, which demonstrated superior performance in Crop Challenge using large datasets of products. The approach used deep neural networks to make yield predictions (including yield, check yield, and yield difference) based on genotype and environment data. The carefully designed deep neural networks were able to learn nonlinear and complex relationships between genes, environmental conditions, as well as their interactions from historical data and make reasonably accurate predictions of yields for new hybrids planted in new locations with known weather conditions. Performance of the model was found to be relatively sensitiveto the quality of weather prediction, which suggested the importance of weather predictiontechniques. We trained two deep neural networks, one for yield and the other for check yield, and then used the difference of their outputs as the prediction for yield difference. This model structure was found to be more effective than using one single neural networkfor yield difference, because the genotype and environment effects are more directly related to the yield and check yield than their difference. In modern era, the deep neural network is the prominent tool in agricultural industry for providing support to farmers in monitoring crop yield based on multiple parameters. Thus, the machine learning model provides high accuracy in detecting the suitable crop identification compared to other methodologies

12. FUTURE SCOPE

This project describes crop yield prediction ability of the algorithm. In future we can determine the efficient algorithm based on their accuracy metrics that will helps to choose anefficient algorithm for crop yield prediction. Our research suggests that farmers' decisions about the production of non-rice crops on different plots and the use of resources other than land, such as labour and outside inputs, continue to influence their decisions regarding the production of rice. Rice production may continue to be atomistic, with many farmers producing tiny amounts of rice rather than a small number of farmers producing big amounts.

12 APPENDIX

SOURCE CODE

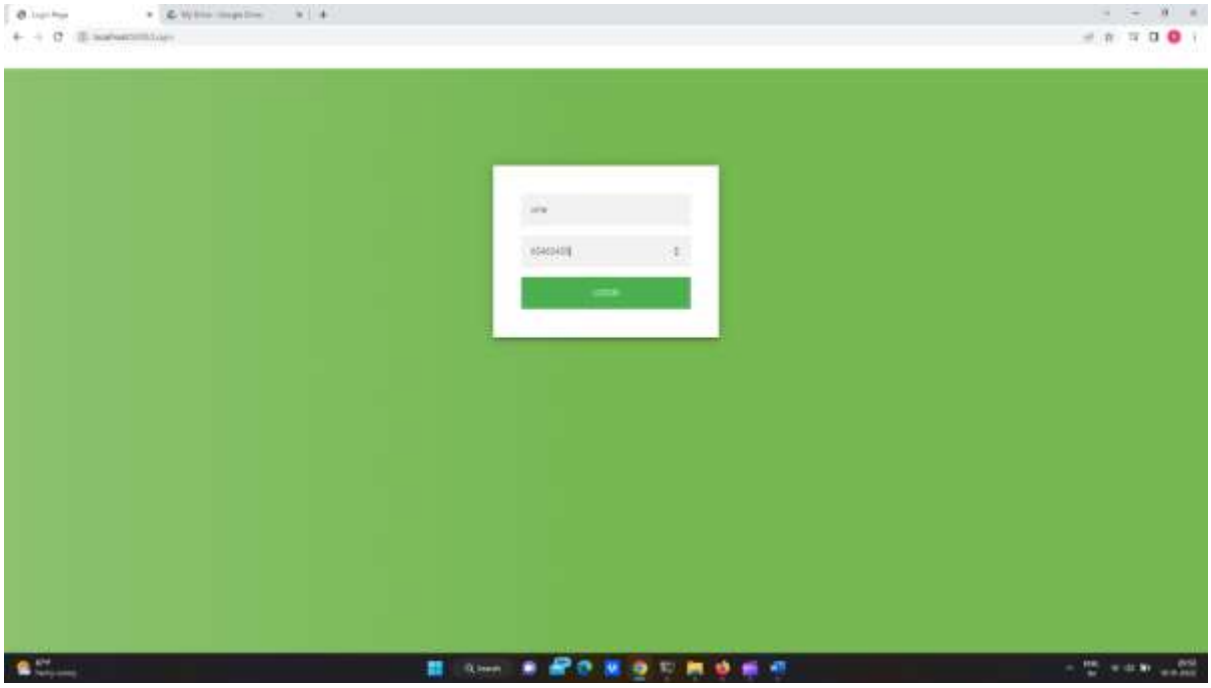
App.py

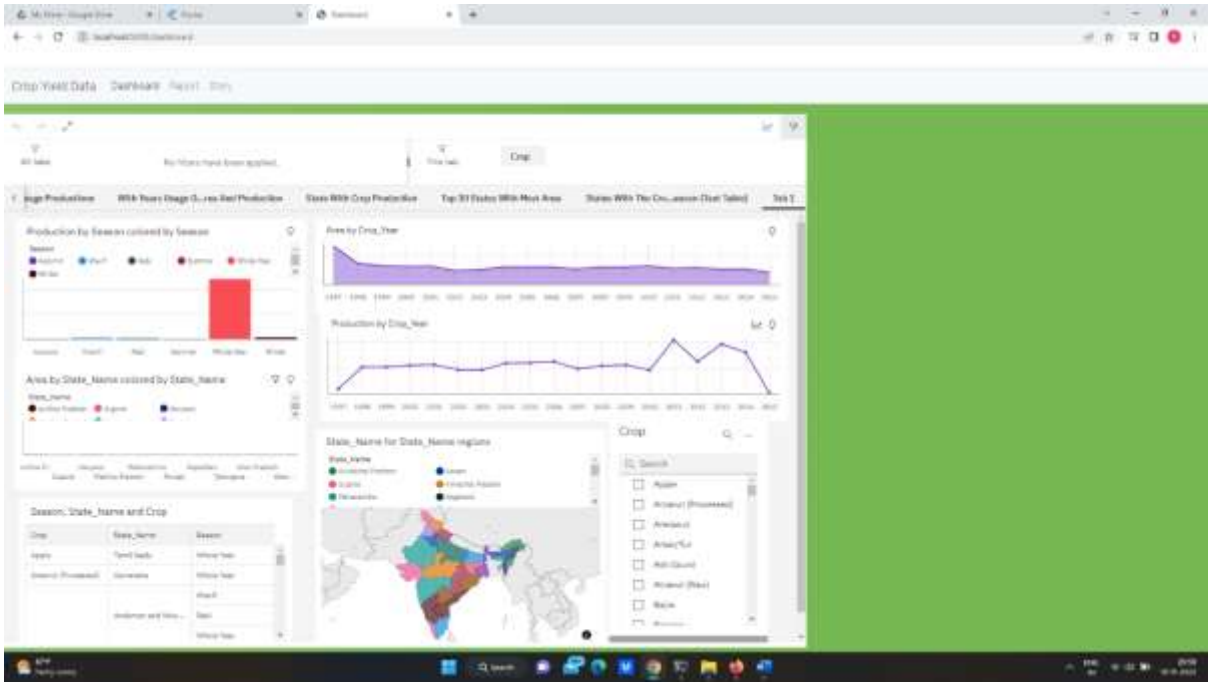
```
from flask import Flask, render_template, url_for,request,redirect, Markup,
make_response

app=Flask(__name__)



@app.errorhandler(404)
def notFound(e):
    return redirect('/dashboard')

@app.route('/dashboard')
def dashboard():
    return  render_template('dashboard.html')

@app.route('/Login', methods=['POST',  'GET'])
def Login():
    if request.method=='POST':
        return dashboard()
    else:
        return render_template('login.html')




@app.route('/story')
def story():
    return  render_template('story.html')



@app.route('/report')
def report():
    return  render_template('report.html')



if __name__=="__main__":
    app.run()
```

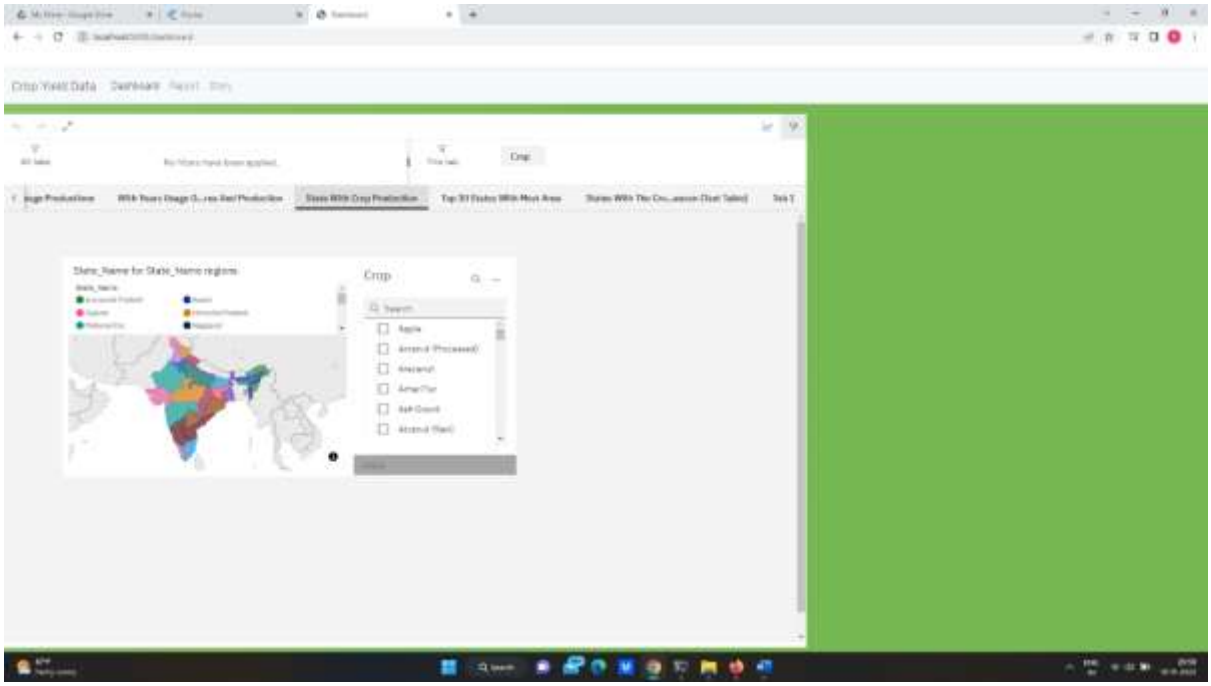
dashboard.html

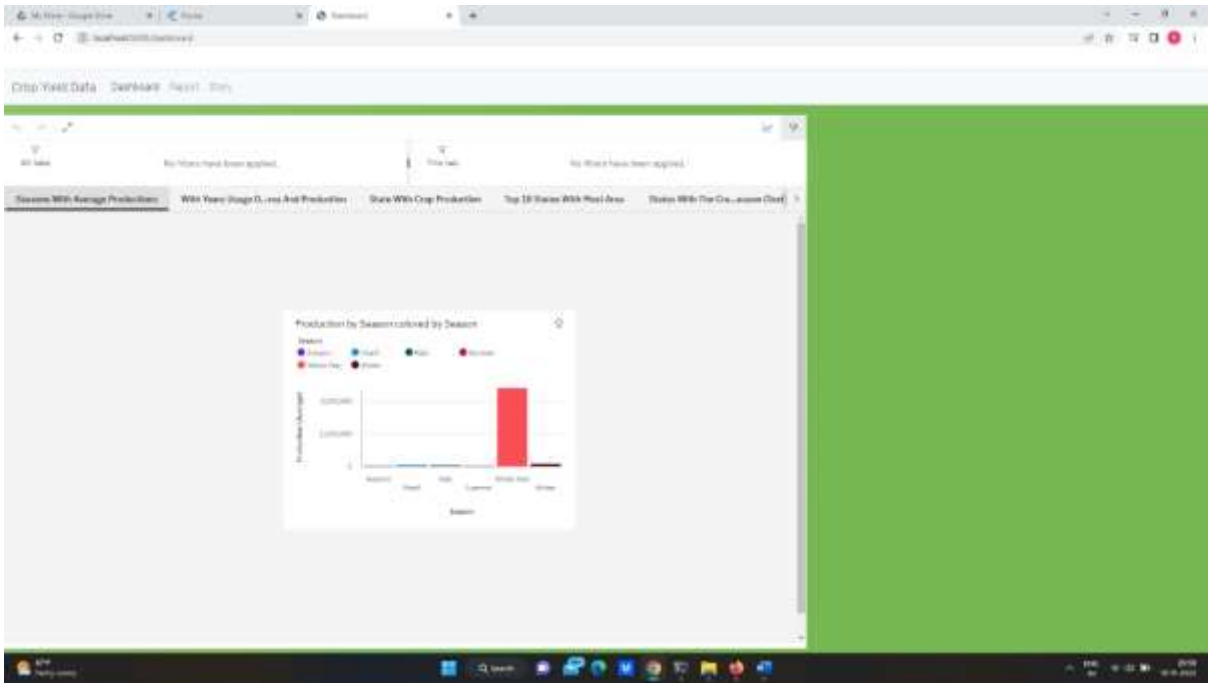
```
{% block head %}
<title>Dashboard</title>
<link rel="stylesheet" href="{{url_for('static',filename='/css/style.css')}}">


{% endblock %}

{% block body %}
<nav class="navbar navbar-expand-lg bg-light">
    <div class="container-fluid">
      <a class="navbar-brand" href="#">Crop Yield Data</a>
      <button class="navbar-toggler" type="button" data-bs-toggle="collapse"
data-bs-target="#navbarNavAltMarkup" aria-controls="navbarNavAltMarkup" aria-
expanded="false" aria-label="Toggle navigation">
        <span class="navbar-toggler-icon"></span>
      </button>
      <div class="collapse navbar-collapse" id="navbarNavAltMarkup">
        <div class="navbar-nav">
          <a class="nav-link active"
href="http://localhost:5000/dashboard">Dashboard</a>
          <a class="nav-link" href="http://localhost:5000/report">Report</a>
          <a class="nav-link" href="http://localhost:5000/story">Story</a>
        </div>
      </div>
    </div>
  </nav>
  <div class="mt-3">
    <iframe
src="https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&amp;pathRef=.p
ublic_folders%2FCrop%2Bproduction%2Brepoet%2FFinal%2Bdashboard%2BEST&amp;close
WindowOnLastView=true&amp;ui_appbar=false&amp;ui_navbar=false&amp;shareMode=em
bedded&amp;action=view&amp;mode=dashboard&amp;subView=model000001846fbe6936_00
000001" width="1280" height="850" frameborder="0" gesture="media"
allow="encrypted-media" allowfullscreen=""></iframe>
  </div>
{% endblock %}
```

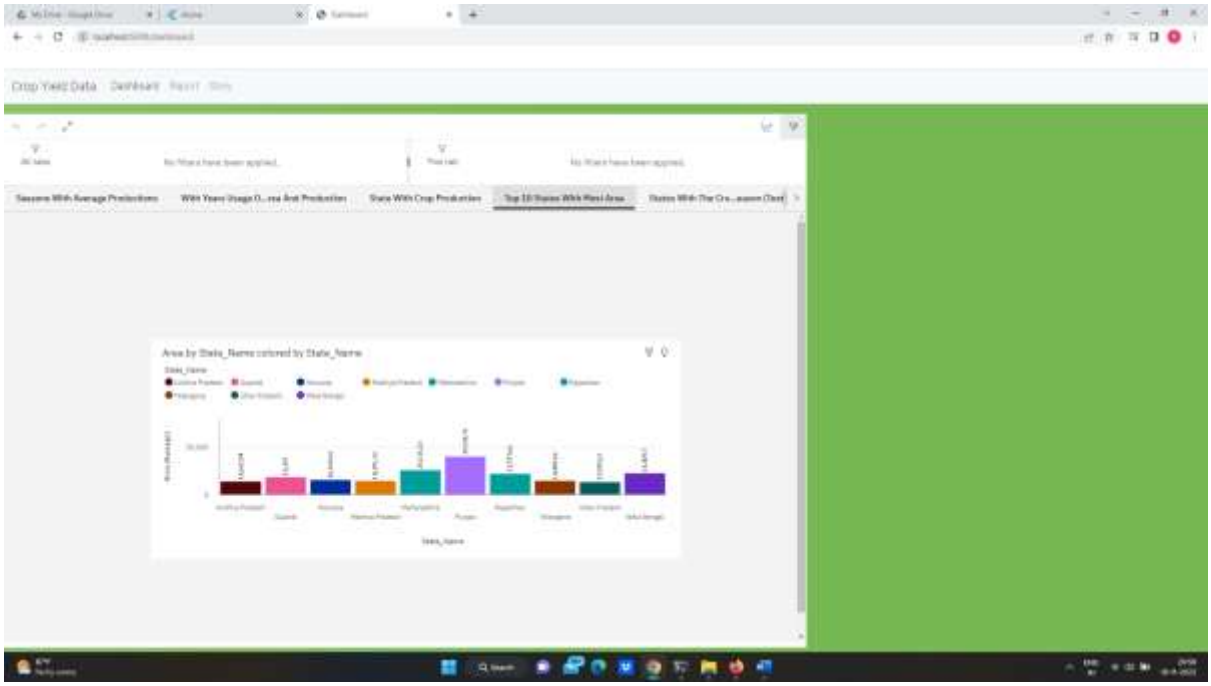
OUTPUT SCREENSHOT

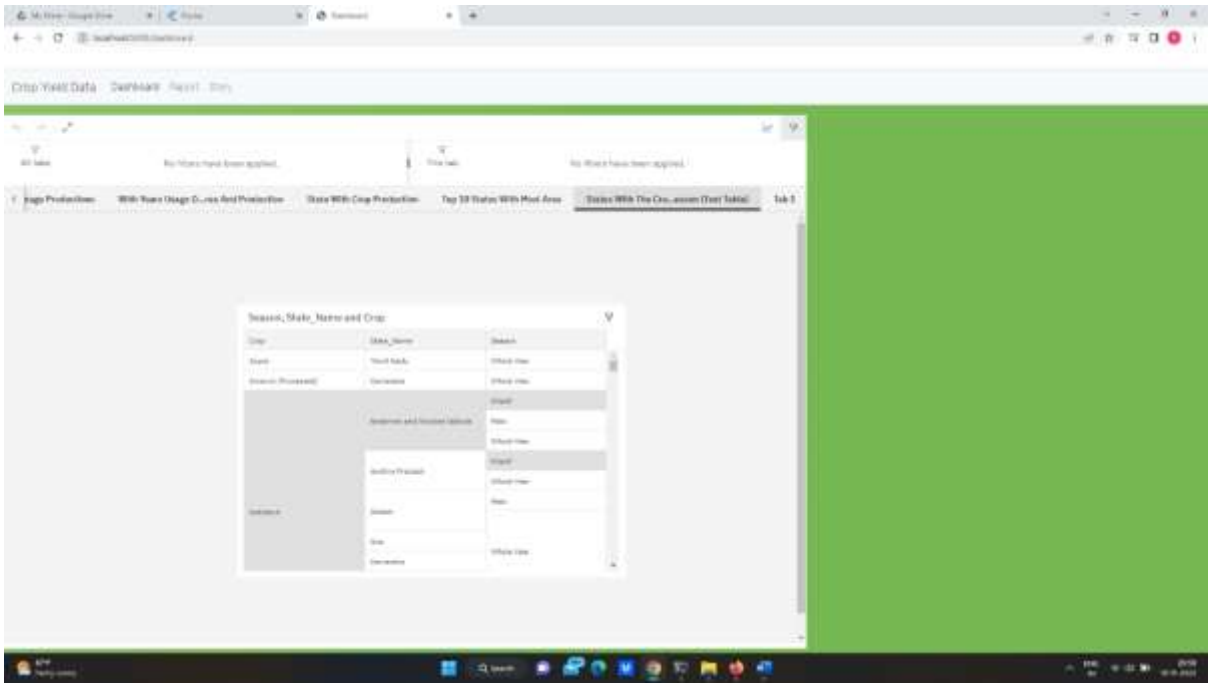


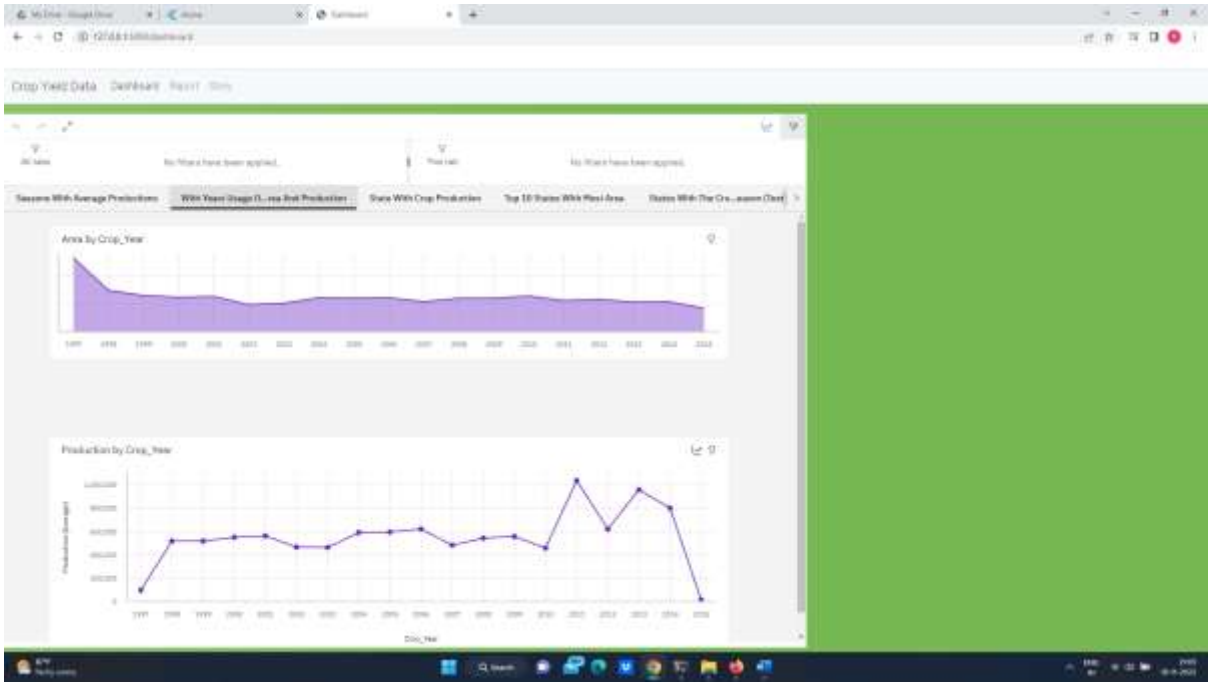


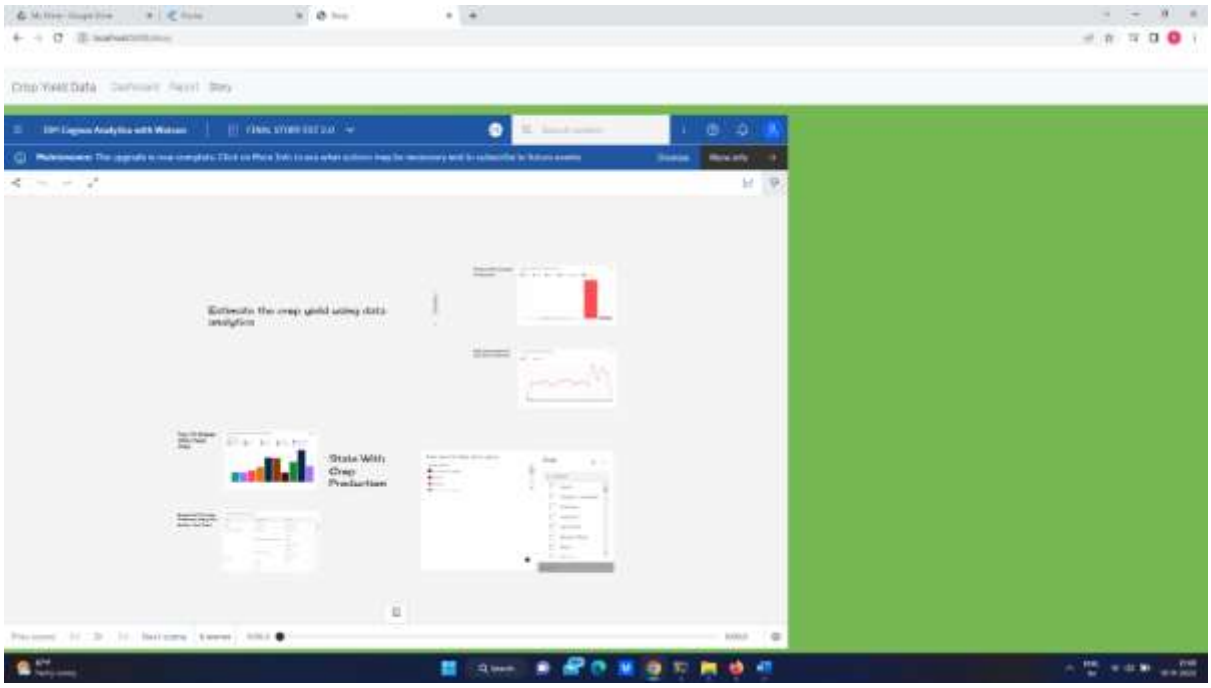

Crop Yield Data
Dashboard
Crop
State_Name for State_Name regions
Crop


Crop Yield Data
Dashboard
Production by Season colored by Season


Crop Yield Data
Dashboard
Report
Story
Area by State_Name colored by State_Name
State_Name


Crop Yield Data
Dashboard
Report
Story
Season, State_Name and Crop





GITHUB LINK

https://github.com/IBM-EPBL/IBM-Project-32400-1660209509

DEMO

https://drive.google.com/file/d/1cgZzHzvJdF1S9FCgDTkeiJYRqmcXy96E/view?usp=share_link