ESTIMATE THE CROP YIELD USING DATA ANALYTICS

IBM – DOCUMENTATION

UNDER THE GUIDANCE OF

INDUSTRY MENTOR(S) NAME : SriKanth, Mahidhar, Mohammed Azhar Uddin

FACULTY MENTOR(S) NAME : Dr. Revathi N

TEAM ID : PNT2022TMID53404

SUBMITTED BY:

SHIYAMAL K	190501124
SHREE HARI D	190501125
SATHISH RAJA G	190501118
SANJAY NAGARAJ	190501111
SARAVANA KUMAR V	190501116

DEPARTMENT OF COMPUTER SCIENCE

SRI VENKATESWARA COLLEGE OF ENGINEERING COLLEGE

1. INTRODUCTION

1.1 PROJECT OVERVIEW

Agriculture is important for human survival because it serves the basic need. A well-known fact that the majority of population (≥55%) in India is into agriculture. Due to variations in climatic conditions, there exist bottlenecks for increasing the crop production in India. It has become challenging task to achieve desired targets in Agri based crop yield. Various factors are to be considered which have direct impact on the production, productivity of the crops.

Crop yield prediction is one of the important factors in agriculture practices. Farmers need information regarding crop yield before sowing seeds in their fields to achieve enhanced crop yield. The use of technology in agriculture has increased in recent year and data analytics is one such trend that has penetrated into the agriculture field. The main challenge in using big data in agriculture is identification of effectiveness of big data analytics.

Efforts are going on to understand how big data analytics can agriculture productivity.

1.2 PURPOSE

a) The dashboard will provide the hidden insights which will help the needy to overcome the problems faced by them.
b) The dashboard can give recommendations to the users based on the past data.
c) The dashboard will predict the future crop yield and also estimate it using data analytics.

2. LITERATURE SURVEY

1.1 EXISTING PROBLEM

1. The purpose of the author is to develop a solution and help farmers to get the information regarding crop yield before sowing seeds in their fields to achieve enhanced crop yield. The farmers experience was the only way for prediction of crop yield in the past days. But now the technology penetration into agriculture field has led to automation of the activities like yield estimation, crop health monitoring etc. This survey focuses on Data mining techniques that are being widely used as a part of solution for crop yield prediction. Data Mining is the process in which the hidden patterns are discovered using analysis of large data sets. The use of technology in agriculture has increased in recent year and data analytics is one such trend that has penetrated into the agriculture field.

2. In here the author works on the crop yield prediction which benefits the farmers in reducing their losses and to get best prices for their crops. The objective of this work is to analyse the environmental parameters like Area under Cultivation, Annual Rainfall and Food Price Index that influences the yield of crop and to establish a relationship among these parameters. To implement this the author has used Regression Analysis to analyse the environmental factors and their infliction on crop yield.

3. In here the author works on the rice crop prediction which plays a vital role in food security of India. High crop production is dependent on the suitable climatic conditions. Detrimental seasonal climate conditions such as low rainfall or temperature extremes can dramatically reduce crop yield. Developing better techniques to predict crop productivity in different climatic conditions can assist farmer in important decision making in terms of agronomy and crop choice.

4. In here the author works on the implementation of the crop selection method so that this method helps in solving many agriculture and farmers problems. This improves our Indian economy by maximizing the yield rate of crop production. There are different types of land conditions so the quality of the crops is identified using ranking process. By this process the rate of the low quality and high-quality crop is also notified. With these results we will be able to find which crop gives us the maximum yield.

1.2 REFERENCES

1. Dhivya B H, Manjula R, Siva Bharathi S, Madhumathi R. A Survey on Crop Yield Prediction based on Agricultural Data, International Journal of Innovative Research in Science, Engineering and Technology. 2017; 6(3).
2. Jharna Majumdar, Sneha Naraseeyappa, Shilpa Ankalaki. Analysis of agriculture data using datamining techniques: application of big data. Journal of Big data. 2017.
3. Majumdar J, Ankalaki S. Comparison of clustering algorithms using quality metrics with invariant features extracted from plant leaves. International Conference on Computational Science and Engineering. 2016.
4. D Ramesh, B Vishnu Vardhan. Data Mining Techniques and Applications to Agricultural Yield Data. International Journal of Advanced Research in Computer and Communication Engineering. 2013; 2(9).
5. Swarupa Rani. The Impact of Data Analytics in Crop Management based on Weather Conditions. International Journal of Engineering Technology Science and Research. 2017; 4(5):299-308.

1.3 PROBLEM STATEMENT DEFINITION

Problem Statement 1:

The User Needs a way to penetrate technology into agriculture.

Problem Statement 2:

The User Needs a way to use data analytics for the crop yield prediction.

Problem Statement 3:

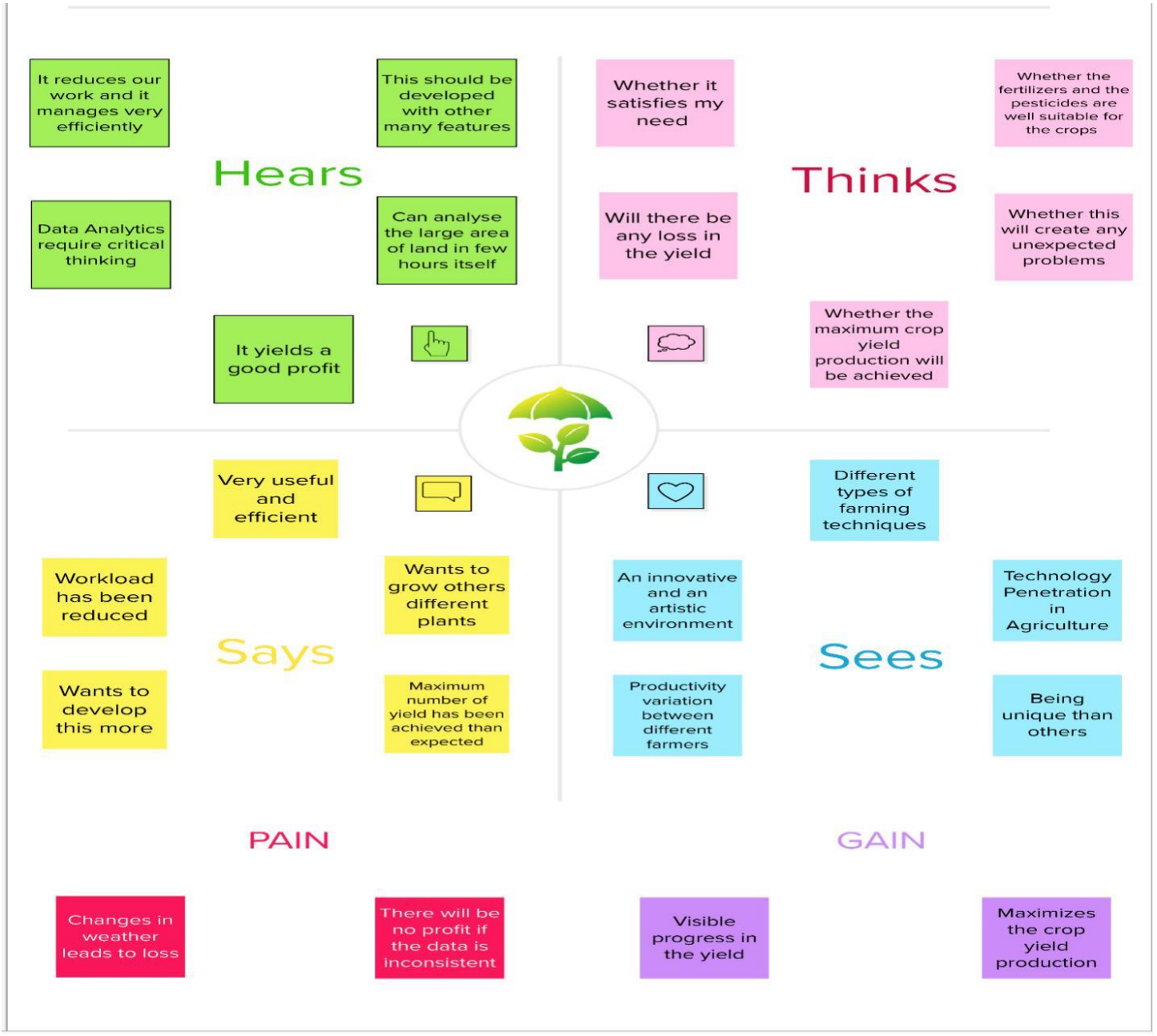
The User Needs a way to use precision agriculture.

Problem Statement 4:

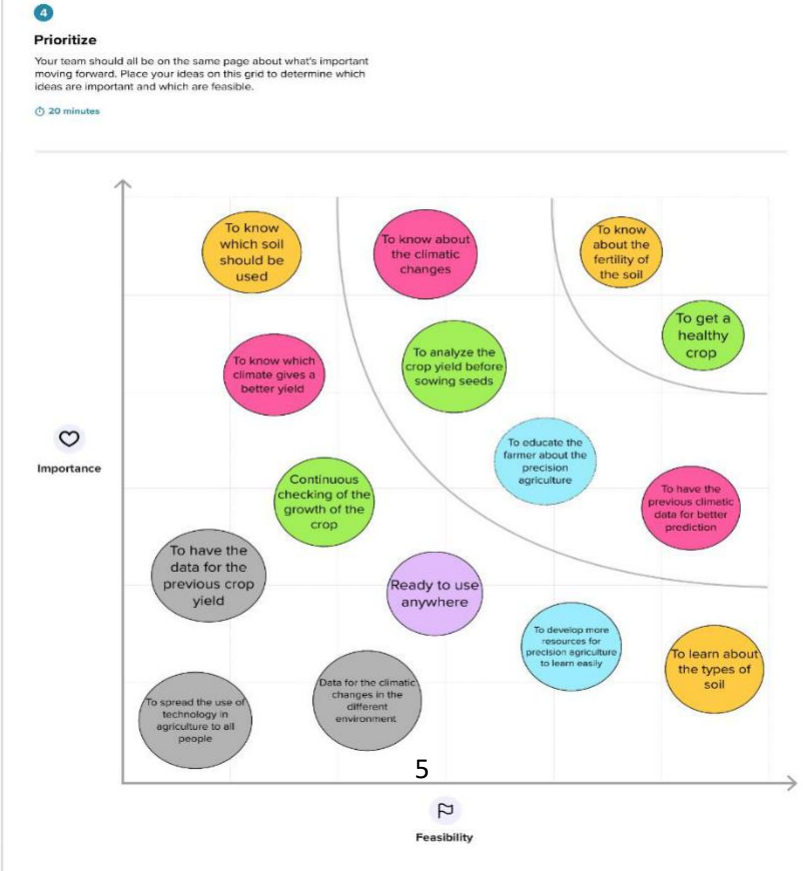
The farmer Needs a way to see visible progress in yield and increase crop yield using data analytics.

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION & BRAINSTORMING



3.3 PROPOSED SOLUTION

S. No	Parameter	Description
1.	Problem Statement (Problem to be solved)	• Crop production is one of the most important source of India. Farmers use the traditional method of agriculture which really takes a lot of time. • An efficient solution has come up through which we can directly predict the future crop yield and maximize the productivity of the crops. .
2.	Idea / Solution description	• The collected data should be explored in an efficient manner to find insights for the better decision. • The factors that will affect the crop yield should be found and the factors that are related to the crop yield should also be found. • Using this information an interactive dashboard should be created with different charts and graphs that comes out with an accurate solution to prevent the losses and maximize the production of the crops.. • This can also be compared with the past historical data to get a better knowledge.
3.	Novelty / Uniqueness	• Using data analytics for the crop production consumes a less amount of time as compared to the traditional method of farming. • It also visualizes us with the future prediction of crops and it increases the crop productivity. • It not only increases the crop productivity but it also provides us a healthy crop.
4.	Social Impact / Customer Satisfaction	• This can be used by anyone such as farmers or individuals who needs a healthy crop and maximum crop yield. • It also prevents the crops from heavy losses.
5.	Business Model (Revenue Model)	• A maximum amount of crop production will be the result of this model and the farmers can get a good revenue based on the production. • This gives the good experience to the famers and they can increase their partnerships with others to get a good profit.
6.	Scalability of the Solution	• With the past visual reports based on the area, climate, soil and water conditions we can also increase the scalability. • The data need to be explored accurately and it can also be compared with the past historical data for the better solution. • The scalability can be increased by increasing the production area, good soil conditions and climatic conditions.

3.4 PROBLEM SOLUTION FIT

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

- Farmers who needs to get a maximum and a healthy crop yield
- Persons who are interested in agriculture and cultivating new crops
- Individuals who needs to save time and maximize the yield

6. CUSTOMER CONSTRAINTS CC

- Networks and connectivity problems may arise
- Lack of Knowledge
- Lack of Awareness
- Climatic changes

5. AVAILABLE SOLUTIONS AS

- Traditional methods of farming
- Monitoring the crops using an application
- Monitoring them using a website
- Precision Farming with advanced techniques

Explore AS, differentiate

Focus on J&P, tap into BE, understand RC

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

- Proper maintenance will be difficult
- Teaching them about the precision agriculture
- Collecting the efficient data
- The correct solution needs to be provided or it would lead to heavy loss

9. PROBLEM ROOT CAUSE RC

- Extreme weather conditions
- Availability of water
- Spread of any plant disease may destroy the whole yield
- Use of harmful pesticides yields to the unhealthy crop
- Protecting the field from the animals

7. BEHAVIOUR BE

- Monitoring and evaluating the crop yield
- Comparison of the traditional and the precision method of agriculture
- Trying other new techniques by themselves
- Asking help to the other similar experts
- Identify the issues in the method

Focus on J&P, tap into BE, understand RC

Identify strong TR and EM

3. TRIGGERS TR

- Influenced by the farmers who prefers precision farming
- Seeing that their field didn't give the maximum yield
- Influenced by the future predictions of their crops

4. EMOTIONS: BEFORE / AFTER EM

Before:

- More work, Confusion, Fear, Doubt

After:

- Self confidence, Happy with the yield

10. YOUR SOLUTION SL

- To create an useful and an interactive dashboard which is very easy to understand by the farmers and the individuals
- To teach them about the future predictions of the crop yield and about the precision agriculture
- To take interesting attributes of the data to gain the insights which leads to the efficient decisions
- To provide the accurate solution for the maximum and a healthy crop yield

8. CHANNELS of BEHAVIOUR CH

Online:

- Search for other different techniques and compare it and check
- Checking whether this gives a good yield or not
- Seeing how many have tried it and got the good results
- Involving in the cross checking

Offline:

- Trying other different methods
- Changing their method of farming frequently
- Cultivating other different crops

Extract online and offline CH of BE

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENTS

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	**User Registration**	• Registration through Website • Registration through Gmail • Registration through LinkedIN
FR-2	**User Confirmation**	• Confirmation via Email • Confirmation via OTP
FR-3	**User Login**	Login using the user registered id and password
FR-4	**User Profile**	It consists of all the user information such as Name, Email, Phone number and Region.
FR-5	**Input Data**	The required input crop yield data is uploaded into the account.
FR-6	**Prepare the Data**	The crop yield data needs to be prepared before starting the analysis. In here the cleaning of the data takes place.
FR-7	**Data Exploration**	The crop yield data needs to be explored to discover and analyze the data. We can also find out the hidden relationships and identify the patterns.
FR-8	**Data Visualization**	Different types of charts, graphs can be formed with the help of the insights taken.
FR-9	**Dashboard Creation**	• With the help of the crop yield visualization charts an interactive dashboard can be created which is very easy to understand by everyone. • It helps farmers to take better decisions like which crop can be grown, at which climatic conditions it can be grown etc.
FR-10	**Present the data**	The crop yield data can be presented in different ways such as dashboard, reports and stories.

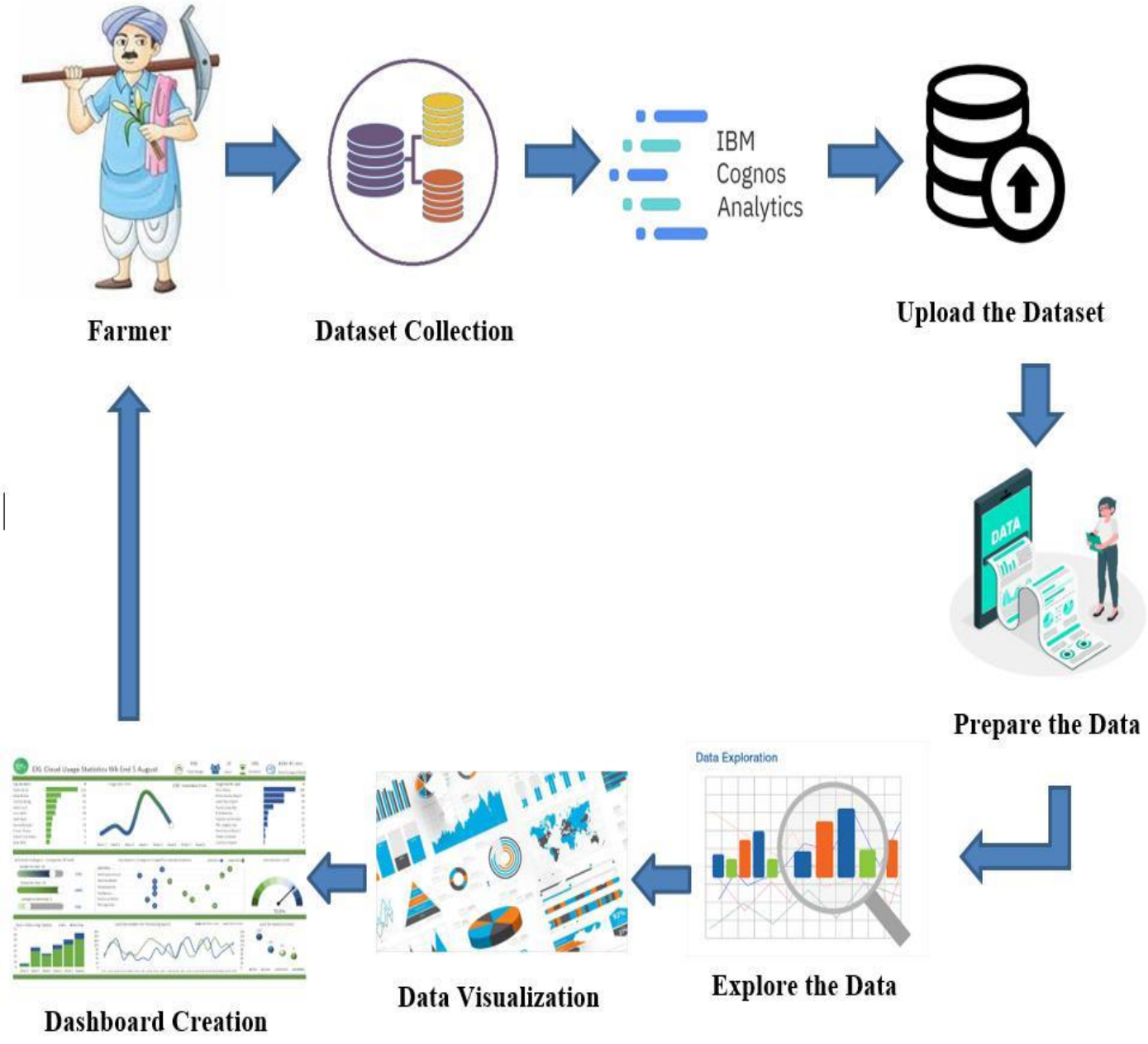
4.2 NON-FUNCTIONAL REQUIREMENTS

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	• It is very easy to navigate. • It can be used by anyone as the instructions are provided very clearly. • Different charts can be easily made i.e we can easily pick and drop these charts. • An interactive dashboard is created which provides the best decision that should be taken • It requires a very less amount of time.
NFR-2	**Security**	• The user data is securely stored in the IBM cloud. • Access to the resources through the two factor authentication. • The passwords are securely managed. • The user's information are authenticated. • To authorize and monitor the use of the anonymous accounts and to remove them.
NFR-3	**Reliability**	• The Quality of the services provided are trustworthy. • It can handle a lot of users at a single time. • It can process and initialize most functions.
NFR-4	**Performance**	• It performs very faster and it is very easy to use. • It provides the user with good interaction to make them understand the hidden patterns.
NFR-5	**Availability**	• It should be made available to access to anyone at any time. • It should be able to work at any place where the user is present. • It should be made compatible to work in all the devices.
NFR-6	**Scalability**	• To expand the server capacity, memory or disc space so that more people can use this at a time. • It should be able to hold large datasets.

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.



5.2 SOLUTION & TECHNICAL ARCHITECTURE

5.3 USER STORIES

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptancecriteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can registerfor the application by entering my email, password and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application.	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can registerfor the application through Google.	I can register and access the dashboard through google login.	Medium	Sprint-1
	Login	USN-4	As a user, I can log intothe application by entering email & Password.	I can accessmy account/ dashboard.	High	Sprint-1
	Dashboard	USN-5	As a user, I can upload, the dataset.	I can upload the dataset in the dashboard.	High	Sprint -1
		USN-6	As a user, I can work with the dataset.	I can prepare my dataset.	High	Sprint -2
		USN-7	As a user, I can make visualization charts for my dataset.	I can explore and visualize my dataset.	High	Sprint-3
		USN-8	As a user, I can create a Dashboard.	I can create a dashboard and find the insights.	High	Sprint-4
Customer Care Executive	Dashboard	USN-9	As a customer care executive, I can access customer's information and solve their queries and issues	I can ask help if I face any issues while using the webpage.	Medium	Sprint-4
Administrator	Application	USN-10	As an administrator, I can manage and maintain the database.	I can assure that the database is secured.	Medium	Sprint -4
		USN-11	As an admin, I can manage the overall process and give updates.	I can manage and give updates.	Medium	Sprint -4

6. PROJECT PLANNING & SCHEDULING

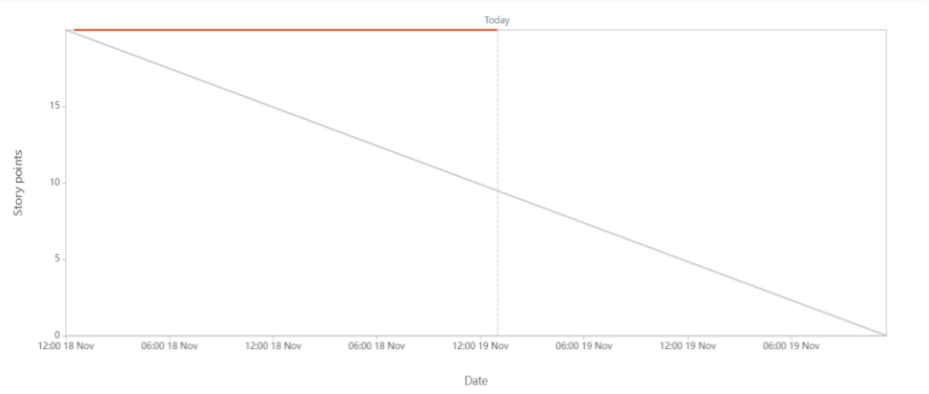
6.1 SPRINT PLANNING & ESTIMATION

SPRINTS	FUNCTIONAL REQUIREMENTS (EPIC)	USER STORY NUMBER	USER STORY / TASK	STORY POINTS	PRIORITY	TEAM MEMBERS
Sprint – 1	Working with the Dataset	USN – 1	*Understanding the Dataset, Loading the Dataset and Exploring the Dataset.*	20	High	• Aiswarya A • Jessica Susan J • Mari Priya B • Arul Sundari A
Sprint – 2	Visualization Charts	USN – 2	*Creating the Data Visualization Charts.*	20	High	• Aiswarya A • Jessica Susan J • Mari Priya B • Arul Sundari A
Sprint – 3	Dashboard	USN – 3	*Creating an Interactive Dashboard.*	20	High	• Aiswarya A • Jessica Susan J • Mari Priya B • Arul Sundari A
Sprint – 4	Export Dashboard	USN – 4	*Exporting the Dashboard*	20	High	• Aiswarya A • Jessica Susan J • Mari Priya B • Arul Sundari A

6.2 SPRINT DELIEVERY SCHEDULE

SPRINTS	TOTAL STORY POINTS	DURATION	SPRINT START DATE	SPRINT END DATE (PLANNED)	STORY POINTS COMPLETED	SPRINT RELEASE DATE (ACTUAL)
Sprint – 1	20	3 Days	*08 November 2022*	*10 November 2022*	20	*10 November 2022*
Sprint – 2	20	3 Days	*11 November 2022*	*13 November 2022*	20	*13 November 2022*
Sprint – 3	20	3 Days	*14 November 2022*	*16 November 2022*	20	*16 November 2022*
Sprint – 4	20	3 Days	*17 November 2022*	*19 November 2022*	20	*19 November 2022*

6.3 REPORTS FROM JIRA



Report: ECYUDA Sprint 1

*Issue added after sprint start

Scope changes log

Date	Key	Summary	Issue type	Epic	Details of scope change	Change in estimation

Report: ECYUDA Sprint 1

*Issue added after sprint start

Scope changes log

Date	Key	Summary	Issue type	Epic	Details of scope change	Change in estimation

There have been no scope changes this sprint

Incomplete issues

View in issue navigator

Key	Summary	Issue type	Epic	Status	Assignee	Story points
ECYUDA-6	Visualization Charts for Seasons with average productions	Story	DATA VISUALIZATI...	TO DO	AA	4
ECYUDA-7	Visualization Charts for With Years Usage Of Area and Production	Story	DATA VISUALIZATI...	TO DO	AS	4
ECYUDA-8	Visualization Charts for Top 10 states with Most Area	Story	DATA VISUALIZATI...	TO DO	M	4

Incomplete issues

View in issue navigator

Key	Summary	Issue type	Epic	Status	Assignee	Story points
ECYUDA-6	Visualization Charts for Seasons with average productions	Story	DATA VISUALIZATI...	TO DO	AA	4
ECYUDA-7	Visualization Charts for With Years Usage Of Area and Production	Story	DATA VISUALIZATI...	TO DO	AS	4
ECYUDA-8	Visualization Charts for Top 10 states with Most Area	Story	DATA VISUALIZATI...	TO DO	M	4
ECYUDA-9	Visualization Charts for State with Crop Production	Story	DATA VISUALIZATI...	TO DO	J	4
ECYUDA-10	Visualization Charts for State with Crop Production Along with S...	Story	DATA VISUALIZATI...	TO DO	AA	4

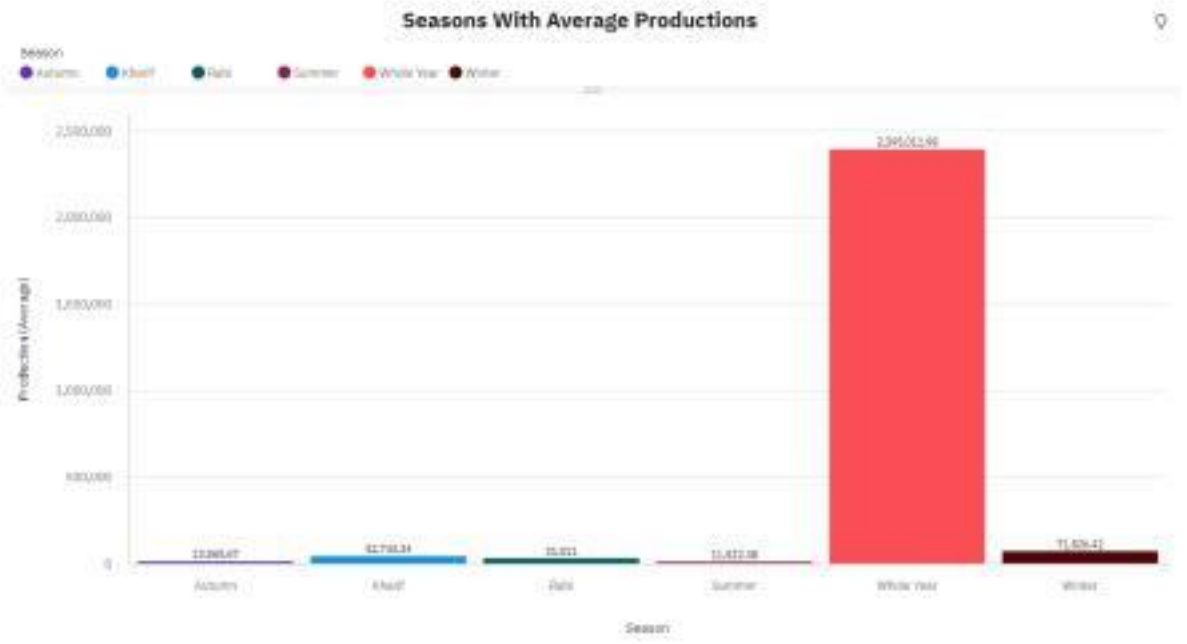
Completed issues

Key	Summary	Issue type	Epic	Status	Assignee	Story points

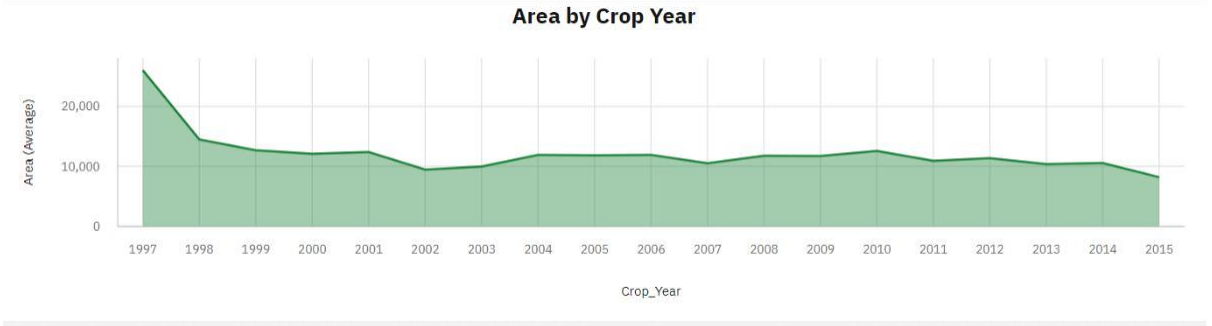
No issues in sprint have been completed

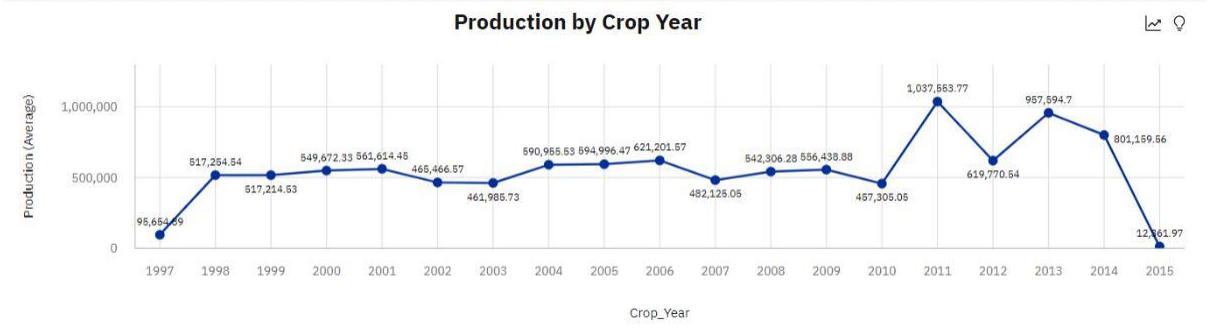
7. CODING & SOLUTIONING

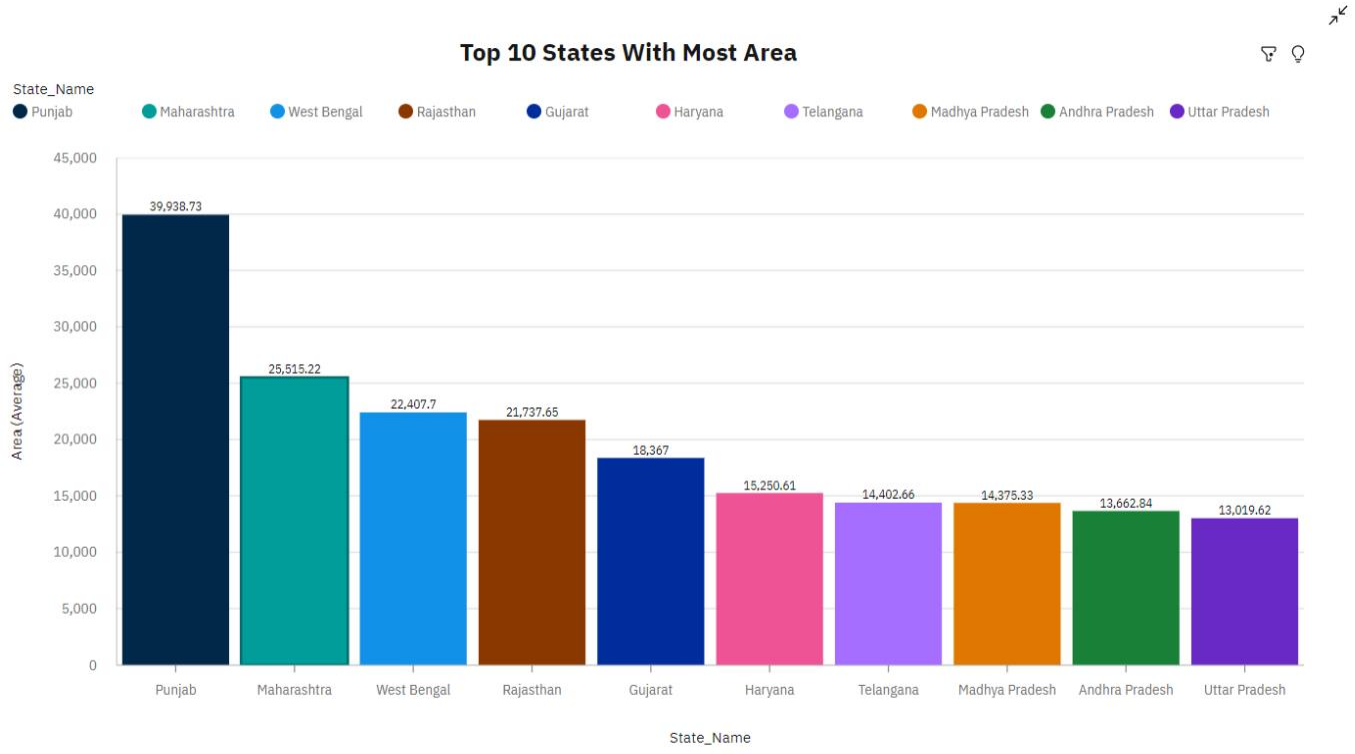
7.1 FEATURE 1

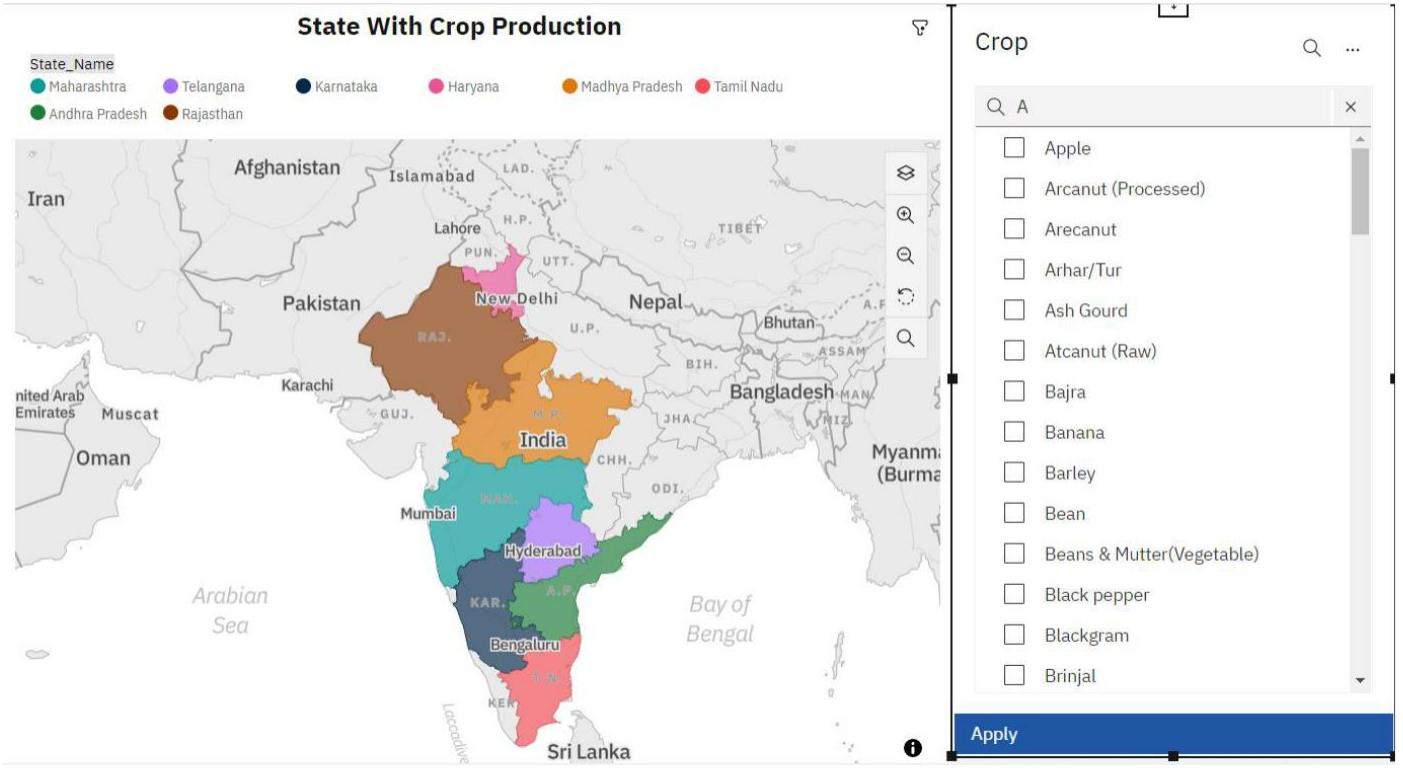


7.2 FEATURE 2






Top 10 States With Most Area
State_Name
Punjab
Maharashtra
West Bengal
Rajasthan
Gujarat
Haryana
Telangana
Madhya Pradesh
Andhra Pradesh
Uttar Pradesh
Area (Average)
45,000
40,000
35,000
30,000
25,000
20,000
15,000
10,000
5,000
0
39,938.73
25,515.22
22,407.7
21,737.65
18,367
15,250.61
14,402.66
14,375.33
13,662.84
13,019.62
Punjab
Maharashtra
West Bengal
Rajasthan
Gujarat
Haryana
Telangana
Madhya Pradesh
Andhra Pradesh
Uttar Pradesh
State_Name


State With Crop Production
State_Name
Maharashtra
Telangana
Karnataka
Haryana
Madhya Pradesh
Tamil Nadu
Andhra Pradesh
Rajasthan
Afghanistan
Islamabad
Iran
Lahore
Pakistan
New Delhi
Nepal
Bhutan
Bangladesh
India
Karachi
Muscat
Oman
Mumbai
Hyderabad
Bengaluru
Arabian Sea
Bay of Bengal
Sri Lanka
Crop
A
Apple
Arcanut (Processed)
Arecanut
Arhar/Tur
Ash Gourd
Atcanut (Raw)
Bajra
Banana
Barley
Bean
Beans & Mutter(Vegetable)
Black pepper
Blackgram
Brinjal
Apply

All tabs

Drag and drop data here to filter all tabs.

This tab

Crop
Grapes 1

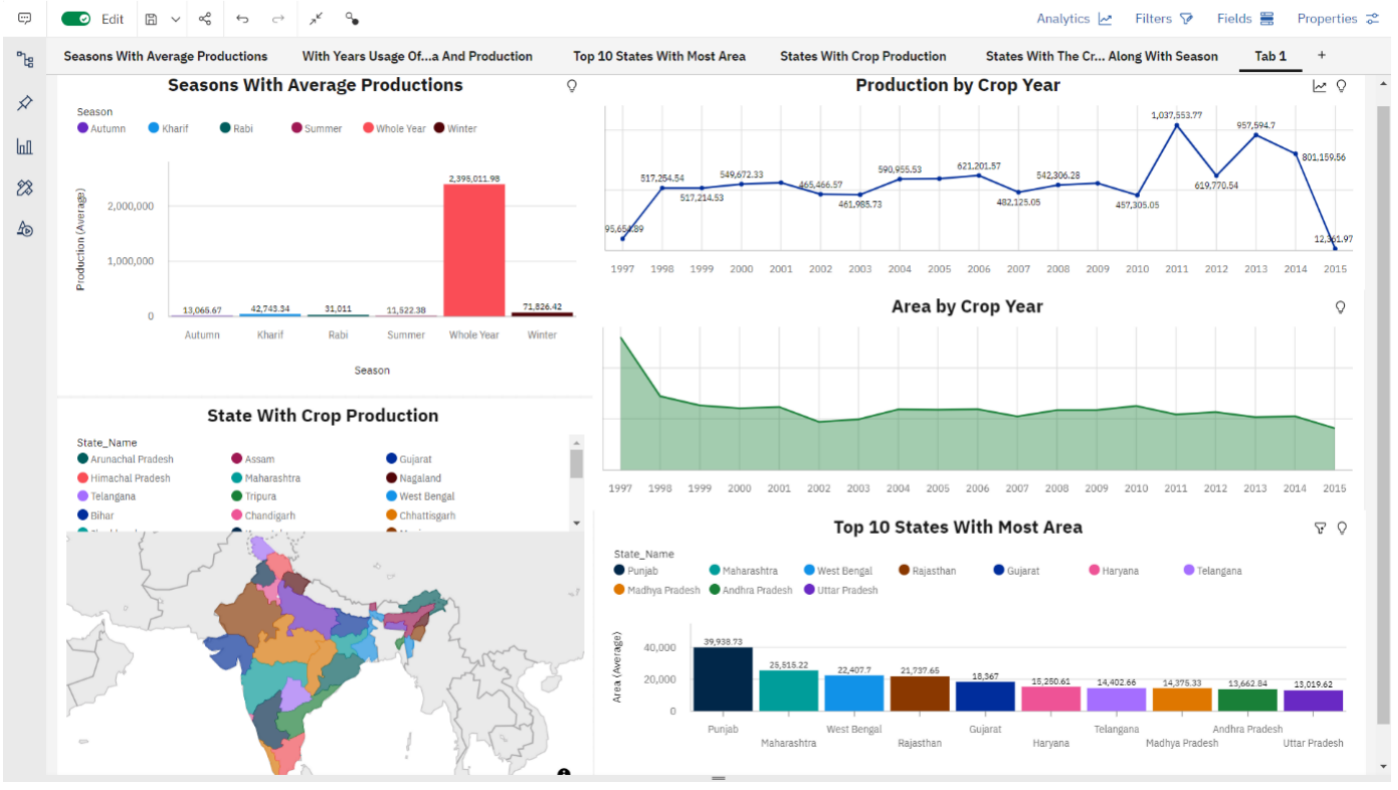
With Years Usage Of...a And Production | Top 10 States With Most Area | States With Crop Production | States With The Cr... Along With Season

State_Name and Crop

Crop	State_Name
Grapes	Andhra Pradesh
	Haryana
	Karnataka
	Madhya Pradesh
	Maharashtra
	Rajasthan
	Tamil Nadu
	Telangana

Season and Crop

Crop	Season
Grapes	Kharif
	Whole Year



8. TESTING

8.1 TEST CASES

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

Total Cases	Not Tested	Fail	Pass
5	0	0	5
7	0	0	7
2	0	0	2
3	0	0	3
9	0	0	9
4	0	0	4
2	0	0	2

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 Smart Fashion Recommender Application 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	5	5	2	3	21
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	77

9. RESULT

9.1 PERFORMANCE METRICS

Project team shall fill the following information in model performance testing.

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visualizations / Graphs – 4
2.	Data Responsiveness	1. Seasons with average production. 2. States with the crop production along with season. 3. With years usage of area and production.
3.	Amount Data to Rendered (DB2 Metrics)	The datasets that are trained and visualized are stored in IBM Cognos.
4.	Utilization of Data Filters	Classification, Prediction and Visualization are utilized to filter data.
5.	Effective User Story	No of Scene Added - 5
6.	Descriptive Reports	No of Visulizations / Graphs - 5

10. ADVANTAGES & DISADVANTAGES

Advantages:

- Risks can be measured when suitable mathematical and statistical model designs are applied on data related to soil, weather and past yield.
- The precise quantification of the rice productivity in various climatic conditions can help farmer to understand the optimum condition for the higher rice crop yield.

Disadvantages:

- The main disadvantage of this strategy is that is does not allow for assessing the impact of intercropping on crop yields.
- This strategy is that crop areas cannot be aggregated at farm or higher levels, since intercropped fields would be double counted.

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. In the literature, it has been observed that analysis has been done on agriculture soils, hidden patterns discovery using data set related to climatic conditions and crop yields data. The activities of agriculture field are numerous like weather forecasting, soil quality assessment, seeds selection, crop yield prediction etc.

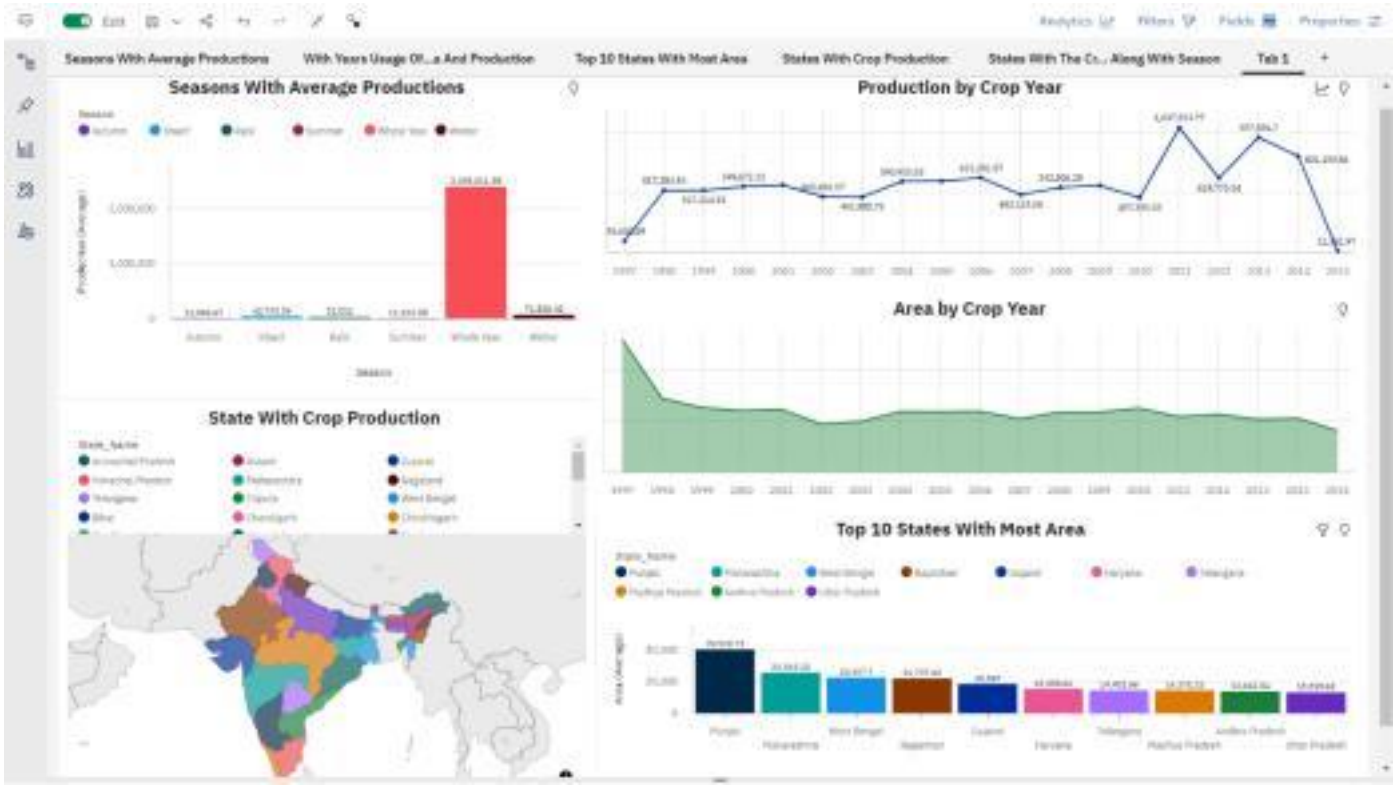
It can be concluded that the research in the field of agriculture with reference to using IT trends like data analytics is in its infancy. As the food is the basic need of humans, the requirement of getting the maximum yields using optimal resource will become the necessity in near future as a result of growing population. The survey outcomes indicate the need for improved techniques in crop yield analytics. There exists a lot of research scope in this research area.

12. FUTURE SCOPE

In India crop yield is season dependent and majorly influenced by the biological and economic causes of an individual crop. Reporting of progressive agricultural yield in all the seasons is an ample task and an advantageous task for every nation with respect to assess the overall crop yield prediction and estimation. Analyzing the yields of crop is necessary to update the policies to ensure food security.

In coming decades, two most significant and important factors found to influence crop yield is, increase in the global population and economy, which greatly demands the higher and sustainable agricultural based crop yields. The capacities of food production at global level is going to be very limited due to the less availability of cultivable land, water resources, difficulties in maintaining the sustainable crop production levels, effects of changes in the global climatic conditions and also by various biophysical parameters which influence the crop yield. The farmers need to be educated on the application of scientifically proven methods to quantify the crop yield capacities and same need to be informed to higher authorities to maintain transparency in sharing the actual information, intern helps in making the policy based, research oriented, development and investment related decisions that aim to influence future crop yield.

13. APPENDIX


Seasons With Average Productions
With Years Usage Of...a And Production
Top 10 States With Most Area
States With Crop Production
States With The Co... Along With Season
Tab 1
Seasons With Average Productions
Production by Crop Year
Area by Crop Year
State With Crop Production
Top 10 States With Most Area