Estimate the Crop Yield using Data Analytics

TEAM ID: PNT2022TMID38310

Team Leader   :   Anish Banu A

Team Members :   Kaviya S

                         Tamil Selvan R

                         Sakthivel V

Abstract

Data analytics is the process of examining data sets in order to find trends and draw conclusions about the information they contain. By using Agriculture Data Analytics in Crop Yield, we will be able to analyze some important visualizations, create a dashboard, and gain most of the insights into Crop production in India. In Cognos Analytics, we can understand our organization's data and make effective decisions based on reporting, modeling, analysis, exploration, dashboards, stories, and event management. Dashboards help us keep track of events or activities at a glance by providing key insights and analysis about our data. A dashboard is used in this project to visualize, analyze, and gain insight.

Introduction

Agriculture is an important sector of Indian economy as it contributes about 17% to the total GDP and provides employment to over 60% of the population. Several factors contribute to the low crop yields experienced by Indian farmers. The yield of agricultural crops is largely determined by weather conditions. Rainfall conditions also influences the rice cultivation. Therefore, farmers need a timely forecast of future crop productivity and an analysis is needed so that they can maximize crop production in their crops. Yield prediction is an important problem in agriculture because it directly affects the crop productivity. Every farmer is involved in knowing, how lots yield he is about expect. In the past, yield prediction was once carried out by using thinking about farmer's preceding trip on a unique crop. The data will be very useful when it is converted into an information by transforming it. IBM Cognos Business Intelligence is a web-based integrated business intelligence suite by IBM. It provides a toolset for reporting, analytics, score carding, and monitoring of events and metrics. The software consists of several components designed to meet the different information requirements in a company. Cognos event studio allows you to assign a specific event that sends a notification to the stakeholder in your organization. Cognos Metric Studio allows you to monitor and analyze business metrics of your organization by building a scorecard environment.

Literature Survey

P. Vindya "Agricultural Analysis for Next Generation High Tech Farming in Data Mining" , Anna University, Trichy, Tamilnadu, India, 5 May 2015[2]. Recent developments in Information Technology for agriculture field have become an interesting research area to predict the crop yield [1].

In today's world, the amount of information stored has been enormously increasing day by day which is generally in the unstructured form and cannot be used for any processing to extract useful information using mining technique [2]. This paper presents a brief analysis of data mining methods and agriculture techniques, farm types, soil types, prediction using Multiple Linear Regression (MLR) technique for the selected region. This work mainly focuses on analyzing the agricultural analysis of organic farming and inorganic farming, time cultivation of the plant, profit and loss of the data and analyzes the real estate business land in a specific area and comparison of irrigated and unirrigated land. It concentrates organic, inorganic and real estate data sets from which the prediction in agriculture will be achieved. The purpose is to estimate difference in efficiency and prediction between organic and inorganic farming. This work aims at finding suitable data models that achieve a high accuracy and a high generality in terms of yield prediction capabilities.

Crop Yield Prediction Using Machine Learning A research group investigated the utilization of various information mining methods which will foresee rice crop yield for the data collected from the state of Maharashtra, India. A total of 27 regions of Maharashtra were selected for the assessment and the data was collected related to the principle rice crop

yield influencing parameters such as different atmospheric conditions and various harvest parameters i.e Precipitation rate, minimum, average, maximum and most extreme temperature, reference trim cultivable area, evapotranspiration, and yield for the season between June to November referred as Kharif, for the years 1998 to 2002 from the open source, Indian Administration records. WEKA a Java based dialect programming for less challenging assistance with information data sets, assigning design outcomes tool was applied for dataset processing and the overall methodology of the study includes, (1) pre-processing of dataset (2) Building the prediction model utilizing WEKA and (3) Analyzing the outcomes. Cross validation study is carried out to scrutinize how a predictable information mining method will execute on an ambiguous dataset. Study applied 10-fold higher cross validation study design to assess the data subsets for screening and testing. Identified and collected information was randomly distributed into 10 sections where in one data section was used for testing while all other data sections were utilized for the preparation information. Study reported that the method applied was supportive in the precise estimation of rice crop yield for the state of Maharashtra, India. 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 [8].

Simulation models based on field experiment are valuable technologies for studying and understanding crop yield gaps, but one of the critical challenge remain with these methods is scaling up of these approach to assess the data collated between different time intervals from the broader geographical
regions. Satellite retrieved data have frequently been revealed to prese

nt data sets that, by itself or in grouping with other information and model designs, can precisely determine the yields of crop in agricultural lands. The yield maps developed shall provide an unique opportunity to overcome both spatial and temporal based scaling up challenges and thus improve the ideology of crop yield gaps prediction. A review was conducted to discuss the applications of remote sensing technology to determine the impact and causes of yield gaps. Even though the example discussed by the research group demonstrates the usefulness of remote sensing in the prediction of yield gaps, but also many areas of possible application with respect to the crop yield assessment, prediction and improvement remain unexplored. Study proposed two less complicated, easily assessable methods to determine and quantify the yield gaps between various agricultural fields. First method works closely with the constructive maps representing the average crop yields, it can be used directly to accesses specific crop yield influencing factors for further studies whereas the second method use the remote sensing technology to retrieve the data for providing the useful information regarding the crop yield prediction and estimation [14].

M. Chandraprabha and R. K. Dhanaraj, "Soil Based Prediction for Crop Yield using Predictive Analytics," *2021 3rd International Conference on Advances in Computing, Communication Control and Networking*. Predictive analysis is a technique of machine learning that predicts the future outcomes and analysis is based on the historical or past data. In agriculture, predictive analytics helps to predict or identify the soil nutrients level required for the crops like Paddy, Raagi, Cumbu etc.,. In this paper, the soil based dataset is

collected from TNAU website and it has 32 districts of Tamilnadu. The algorithms such as Naïve bayes, Bayes Net, and IbK have been deployed to predict the crop variety suitable for the soil based on the total production and area sown district wise. Also, its accuracy levels are compared. The accuracy is determined using true positive value, false positive value, precision, recall, f-measure and MCC.

B. Vandana and S. S. Kumar, "A Novel Approach using Big Data Analytics to Improve the Crop Yield in Precision Agriculture," *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology.* Agriculture is the main work field in India. Farming industry adopts less innovative technology compared to other industries. Information and Communication Technologies provides simple and cost effective techniques for farmers to enable precision agriculture. The work propose a state of the art model in agriculture field which will guide the rural farmers to use Information and Communication technologies (ICT) in agriculture fields. Big data analytics is used to improve the crop yield. It can be customized for precision agriculture to improve the quality of crops which improves the overall production rate.

S. Sharma, G. Rathee and H. Saini, "Big Data Analytics for Crop Prediction Mode Using Optimization Technique," *2018 Fifth International Conference on Parallel, Distributed and Grid Computing.* Agriculture is considered as the backbone of our country's economy. Big data analysis is used to discover novel solutions, which act as means for analyzing bulky data set, so that it plays a significant role for decision making in specific field such as

agriculture. In this work, soil and environment features i.e. average temperature, average humidity, total rainfall and production yield are used in predicting two classes namely: good yield and bad yield. For this purpose, a hybrid classifier model is used in optimizing the feature and the proposed approach is divided into three phase's viz pre-processing, feature selection and SVM_GWO i.e grey wolf optimizer along with Support Vector machine (SVM) classification is used to improve the accuracy, precision, recall and F-measure. The result shows that SVM_GWO approach better as compared to typical SVMs classification algorithm.

S. Bao *et al.*, "Crop yield variation trend and distribution pattern in recent ten years," *2017 IEEE International Geoscience and Remote Sensing Symposium.* In recent ten years, a perception exists that the agricultural management and crop cultivars have been improved obviously. But the crop yield variation trend due to above reason remain unknown yet. To evaluate the main food crop (maize, soybean and rice) yield trend from 2007 to 2016, the MODIS product (MCD12Q2) was used to extract the mature date of different crops. A two-band variant of the enhanced vegetation index at mature date was applied to establish empirical yield estimation model, coupling with statistical crop yield data. The validation show the estimated yield had accuracy of 90.9%, 91.7% and 83.3%, respectively. The average maize and soybean yield in study area presented increasing trend, but rice yield presented declining. However, maize yield in 22 cities and soybean yield in 19 cities show decreasing trend actually. Through statistical analysis, the crop yield distribution pattern was proved to be almost fixed. Most cities occupies approximate position on the

ranking of relevant crop yield. It was demonstrated that some cities, for example Chifeng city, was suitable to develop specific agriculture economy. This paper can be used to give suggestion for agriculture planning and management.

Ideation and Proposed Solution

A. IBM Cognos Analytics

IBM Cognos Analytics is a set of business intelligence tools available on cloud or on- premise. The primary focus is in the area of Descriptive Analytics, to help users see the information in your data through dashboards, professional reporting and self-service data exploration. In this work, we used the IBM cognos data analytics for analyzing the crop yield data.

Following are important features of IBM Cognos:

1. *Get Connected* - Connect your data effortlessly Import data from CSV files and spreadsheets. Connect to cloud or on-premises data sources, including SQL databases, Google BigQuery, Amazon, Redshift, and more.

2. *Prepare your data* – Prepare and connect data automatically Save time cleaning your data with AI-assisted data preparation. Clean and prep data from multiple sources, add calculated fields, join data, and create new tables.

3. *Build visualizations* - Create dynamic dashboards easily Quickly create compelling, interactive dashboards. Drag and drop data to create auto-

generated visualizations, drill down for more detail, and share using email or Slack.

4. *Identify Patterns* – Uncover hidden patterns Ask the AI assistant a question in plain language, and see the answer in visualization. Use time series modelling to predict seasonal trends.

5. *Generate Personalized Reports* – Create and deliver personalized reports Keep your stakeholders up-to-date, automatically. Create and share dynamic personalized, multi-page reports in the formats your stakeholders want.

6. *Gain Insights* - Make confident data decisions Get deeper insights without a data science background. Validate what you know, identify what you don't with statistically accurate time-series forecasting and pinpoint patterns to consider.

7. *Stay Connected* – Go Mobile Stay connected on the go with the new mobile app. Access data and get alerts right from your phone.

Important Components of Cognos Software:

a. Cognos Connection: Cognos connection is a web portal that allows users to access Cognos 10 and studios. Based on your assigned role, you can use this component to retrieve, view, publish, manage, and organize companies' reports, scoreboards, and agents. The Administrator also using Cognos Connection to establish roles and user permissions and manage the Cognos Connection content.

b. Cognos Business Insight: Cognos Business Insight allows users to create their dashboard using any object. All content which the user is permitted to view will be presented as an object. This can be used in your workspace to create a fully personalized dashboard.

c. Cognos Query Studio: Cognos Query Studio helps business users to get fast answers to business related queries. It helps organizations to better understand the product, customer, and organizational needs. It also helps them to react quickly and stay ahead of the competition.

d. Cognos Analysis Studio: Cognos Analysis Studio helps businesses to find and focus on things which are important to the business. It also helps to understand the latest trends, compare data, and assess business performance for multidimensional analysis.

e. Cognos Business Insight Advanced: Cognos business insight Advanced is a new module included in Cognos 10. It combines Cognos query studio and Cognos Analysis Studio. It offers robust authoring environment for business peoples.

f. Cognos Report Studio: Using the Cognos reporting tool, you can create pixel-perfect reports for your organization. It allows you to create charts, maps, lists, or any other available report type using relational or multidimensional data sources.

g. Cognos Event Studio: This tool allows you to assign a specific event that sends a notification to the stakeholder in your organization. You can create agents which enables you to your events and thresholds. Therefore, the event occurs or threshold is reached the agent sends the notification.
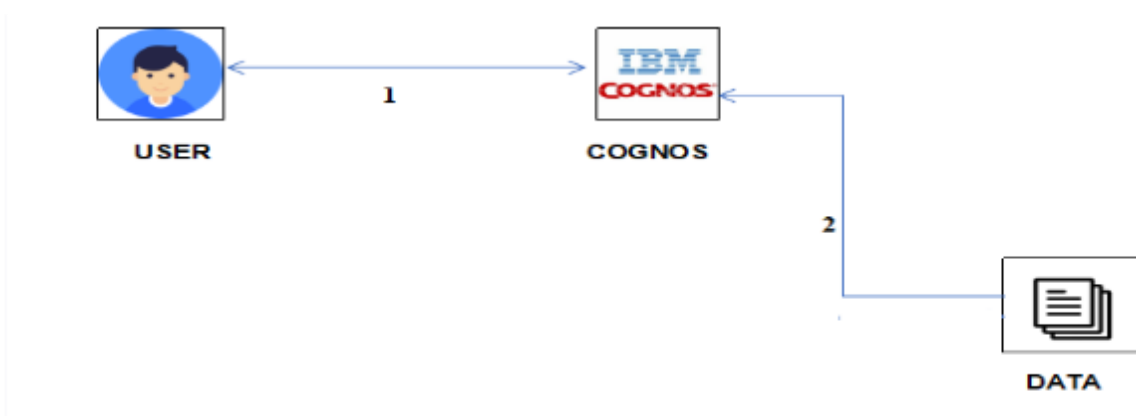
h. Cognos Metric Studio: Cognos Metric Studio allows you to monitor and analyse business metrics of your organization by building a scorecard environment. It also helps you to establish criteria and then monitor your organization to see how it is responding as the changes made in the criteria.
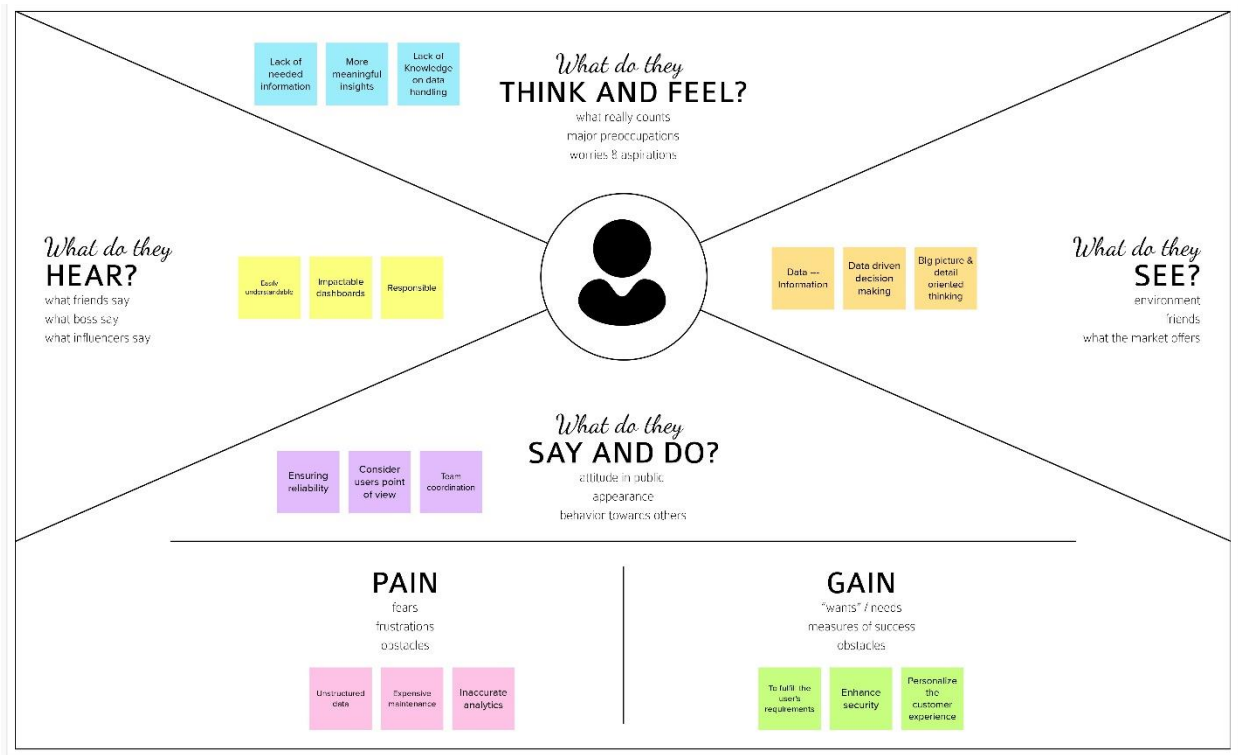
B. System Architecture

IBM Cognos Analytics provides dashboards and stories to communicate your insights and analysis. You can assemble a view that contains visualizations such as a graph, chart, plot, table, map, or any other visual representation of data. Explore powerful visualizations of your data in IBM Cognos Analytics and discover patterns and relationships that impact your business. A dashboard helps you to monitor events or activities at a glance by providing key insights and analysis about your data on one or more pages or screens.
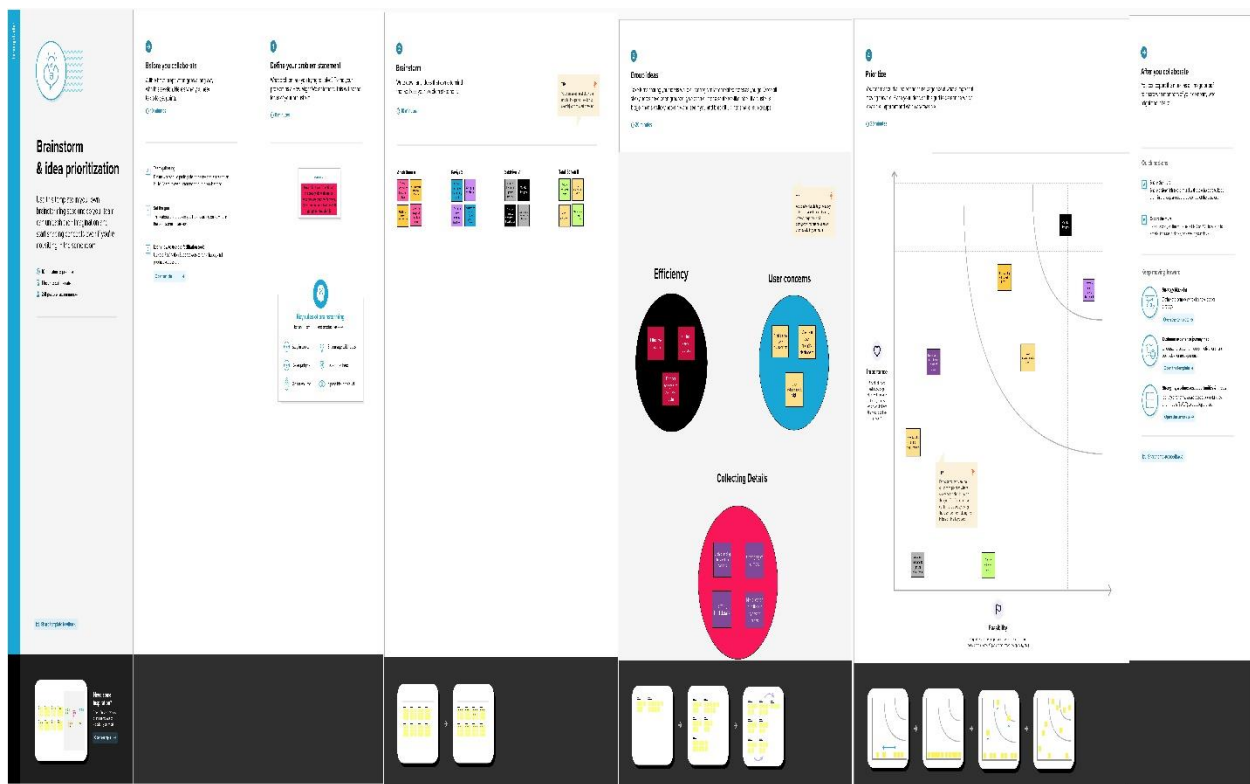
The following are the modules in our work:

1. Uploading data (dataset).
2. Cleaning data (prepare data).
3. Analyzing and interpreting (exploration).
4. Visualizing data (dashboard creation).

Empathy Map Canvas

The empathy map includes the following sections:

What do they THINK AND FEEL?
what really counts
major preoccupations
worries & aspirations
- Lack of needed information
- More meaningful insights
- Lack of Knowledge on data handling

What do they HEAR?
what friends say
what boss say
what influencers say
- Easily understandable
- Impactable dashboards
- Responsible

What do they SEE?
environment
friends
what the market offers
- Data — information
- Data driven decision making
- Big picture & detail oriented thinking

What do they SAY AND DO?
attitude in public
appearance
behavior towards others
- Ensuring reliability
- Consider users point of view
- Team coordination

PAIN
fears
frustrations
obstacles
- Unstructured data
- Expensive maintenance
- Inaccurate analytics

GAIN
"wants" / needs
measures of success
obstacles
- To fulfil the user's requirements
- Enhance security
- Personalize the customer experience

Ideation & Brainstorming

Proposed Solution

S.NO.	PARAMETER	DESCRIPTION
1.	Problem Statement (Problem to be solved)	To estimate the crop yield with atmost accuracy so that the Farmers, Government and Traders will get gain through it and it will lead them to make data - driven discision - making.
2.	Idea / Solution description	To create user understandable dashboards so that the farmers get valuable insights about various types of crop production.
3.	Novelty / Uniqueness	Using a well-researched dataset includes types of crops, regions, seasons, climatic conditions, temperature suggestions can be made on what type of crop is suitable for a particular land at a specific season to get maximum harvest.
4.	Social Impact / Customer Satisfaction	Maximum accuracy in the output will give maximum crop yield. As more revenue will be made, the customer will be satisfied.

5.	Business Model (Revenue Model)	A satisfied customer can share his experience to somebody which results in large number of customer base. More revenue can be made by providing subscription services.
6.	Scalability of the Solution	There is no problem with the dataset storage, time efficient, accurate, clear insights and an efficient way to forecast the upcoming results.

Problem Solution fit

Define CS, fit into CC

1. CUSTOMER SEGMENT(S)   CS

Who is your customer?
i.e. working parents of 0-5 y.o. kids

1. Farmers
2: Customers
3. Traders

6. CUSTOMER CONSTRAINTS   CC

What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.

1. Lack of Clarity
2. Poor Guidance
3. Network Issues
4. Less Knowledge

5. AVAILABLE SOLUTIONS   AS

Which solutions are available to the customers when they face the problem
or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking

There are methods such as interviewing farmers and field surveys which are time consuming and costly. There are some crop yield prediction models which are not too accurate.

Explore AS, differentiate

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

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

Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides.

Accurate analysis in order to minimize the crop damage and to create user understandable dashboard.

9. PROBLEM ROOT CAUSE   RC

What is the real reason that this problem exists?
What is the back story behind the need to do this job?
i.e. customers have to do it because of the change in regulations.

1. Unpredictable variation in climate
2. Various Regions
3. Past Issues
4. Soil Fertility
5. Availability of Water

7. BEHAVIOUR   BE

What does your customer do to address the problem and get the job done?
i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace)

Develop a model that has a good prediction of crop yield in accordance with season and region.

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

Identify strong TR & EM

3. TRIGGERS   TR

What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news.

An intention to get more crop production and to reduce the crop wastage, Analysts will be helpful.

4. EMOTIONS: BEFORE / AFTER   EM

How do customers feel when they face a problem or a job and afterwards? i.e. lost, insecure > confident, in control - use it in your communication strategy & design.

Before, farmers did not get the clear idea about the yield of crop and wastage of crops will be high. After, they get valuable insights through the analysis thus they have high yield which will increase the revenue.

10. YOUR SOLUTION   SL

If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality.
If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.

By creating an interactive dashboards, reports, stories using IBM Cognos Analytics the estimation of crop yield can be done with most accuracy so that, the famer will gain more revenue.

8. CHANNELS of BEHAVIOUR   CH

8.1 ONLINE
What kind of actions do customers take online? Extract online channels from #7

8.2 OFFLINE
What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.

Online : Through the insights from the dashboard correct decisions can be made.

Offline : Execute the decision which is taken from the insights.

Identify strong TR & EM

Requirement Analysis

Functional Requirement:

Following are the functional requirements of the proposed solution

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Utilizing a form for Registration and Signing up with Gmail registering via WhatsApp, Utilizing Agri - Consultancy to register.
FR-2	User Confirmation	Email confirmation required Reassurance via OTP verification via Letter.

FR-3	User Profile	1. User Information 2. Farm Information
FR-4	Required Data	The user's (farmer's) data to analyse the previous crop yield
FR-5	Analysis	Clean up, prepare and analyze the previous data from multiple users (farmer)
FR-6	Estimation	Developing the ideal data module and visuals in IBM Cognos to improve crop yield estimation.

Non-Functional requirements:

Following are the non-functional requirements of the proposed solution

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Data reports are produced based on the historical data itself. These recommendations will advise or consult on crop sowing.
NFR-2	Security	The user information is protected with IBM

		Cognos (Data Visuals)
NFR-3	Reliability	Reliability can be achieved with the saved dataset in the Cognos platform.
NFR-4	Performance	Better performance among all users is made possible through interaction, and the visual advice is impressive.
NFR-5	Availability	The dashboard is easily accessible and can be accessed on any smartphones, laptops, systems, etc.
NFR-6	Scalability	The proposed solution's flexibility in implementation makes it very simple to boost crop yield estimation in various farms.

Project Design

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.
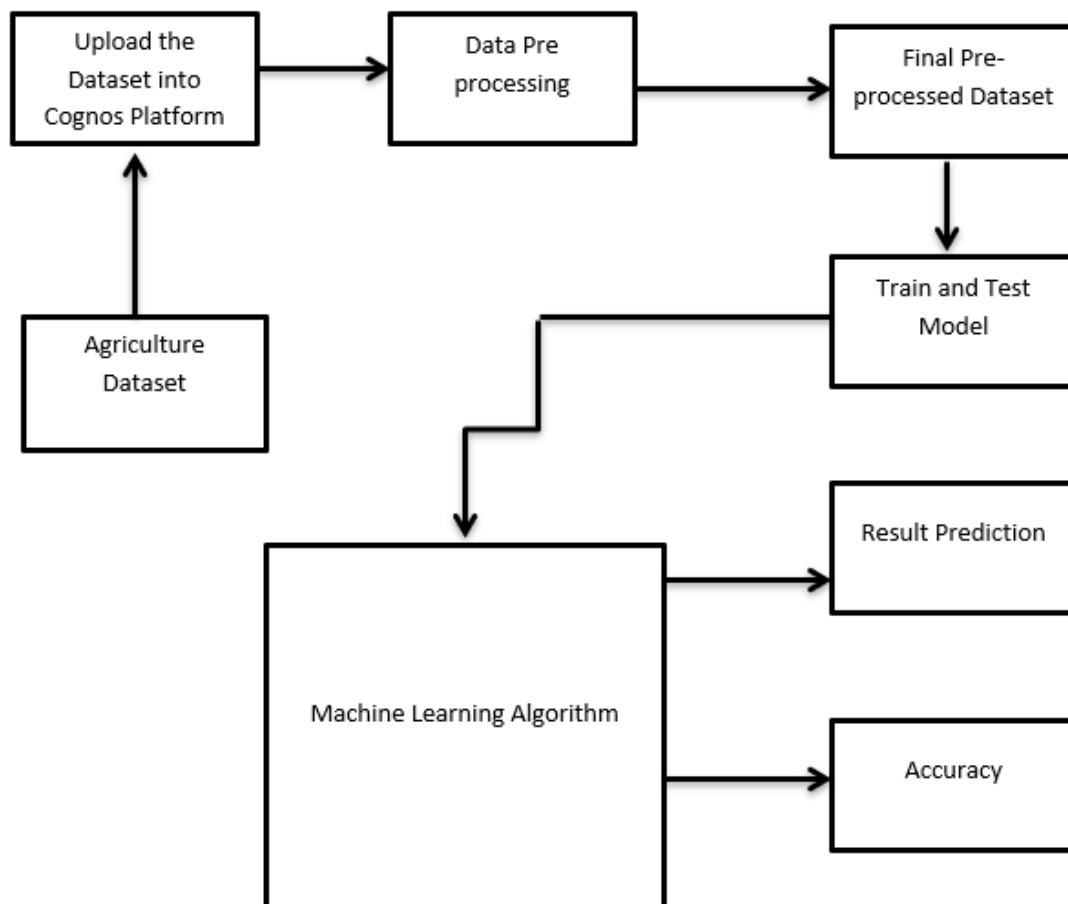

User Stories:

Use the below template to list all the user stories for the product

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 can register for the application through Facebook	I can register & access the dashboard with Facebook Login.	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	Can use the methods provided in the Dashboard.		Medium	Sprint-2
Customer (Web user)	Activity	USN-7	I can register for the application through any web browser.	I can get an notification from the browser.	Low	Sprint-1
Customer Care Executive	Access resources	USN-8	I can use my credentials For accessing my Resources.	Other than me, there is less chance to access my Resources.	High	Sprint-1
Administrator	Satellite visioning	USN-9	As, a user I can vision the geographic area.		Medium	Sprint 2
Customer tools	Tools	USN-10	I can perform analysis by tools (cognos and with ML)	I have an ease of Accessing tools.	High	Sprint 1

Technical Architecture:
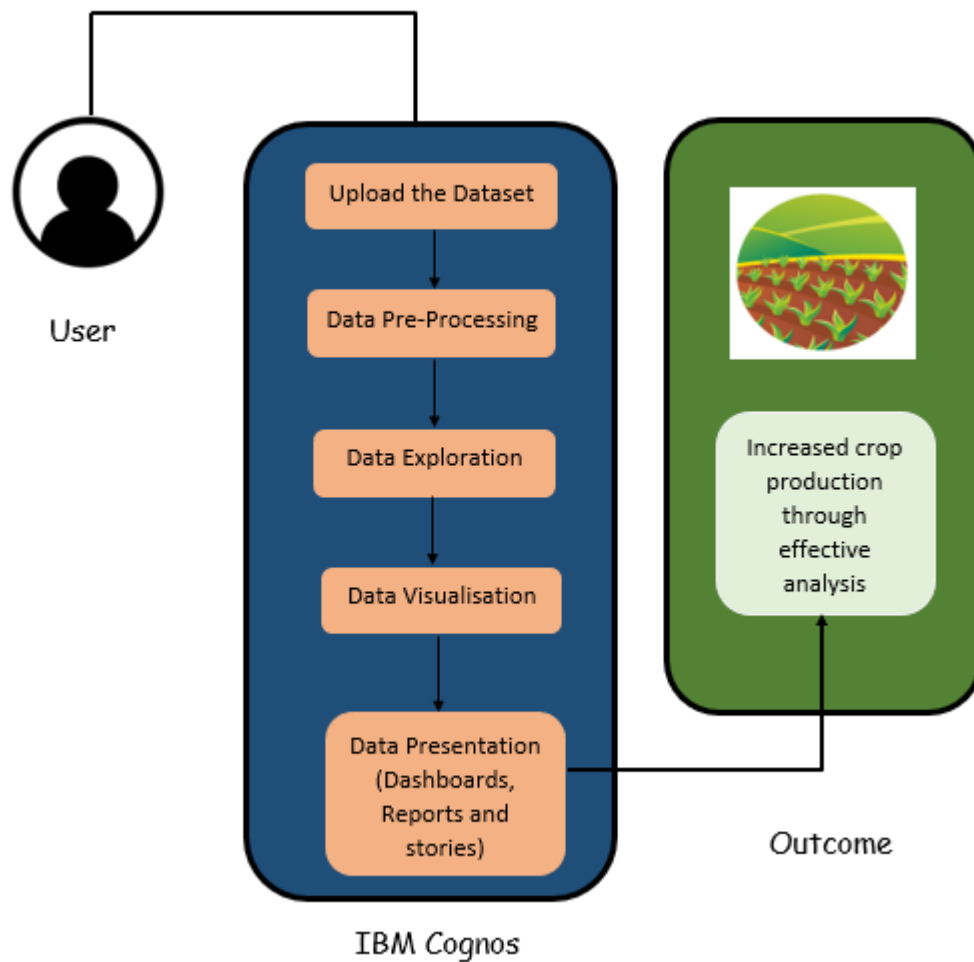
Project Flow

- ❖ Users create multiple analysis graphs/charts.
- ❖ Using the analyzed chart creation of the Dashboard is done.

❖ Saving and Visualizing the final dashboard in the IBM Cognos Analytics.

To accomplish this, we have to complete all the activities and tasks listed below:

❖ Login to Cognos Analytics.
❖ Working with the Dataset
  ➢ Understand the Dataset.
  ➢ Loading the Dataset.
❖ Data visualization charts
  ➢ Seasons with average productions.
  ➢ With years usage of Area and Production.
  ➢ Top 10 States with most area.
  ➢ State with crop production.
  ➢ States with the crop production along with season (Text Table).
❖ Dashboard Creation.
❖ Export the Analytics.

Project Planning and Scheduling
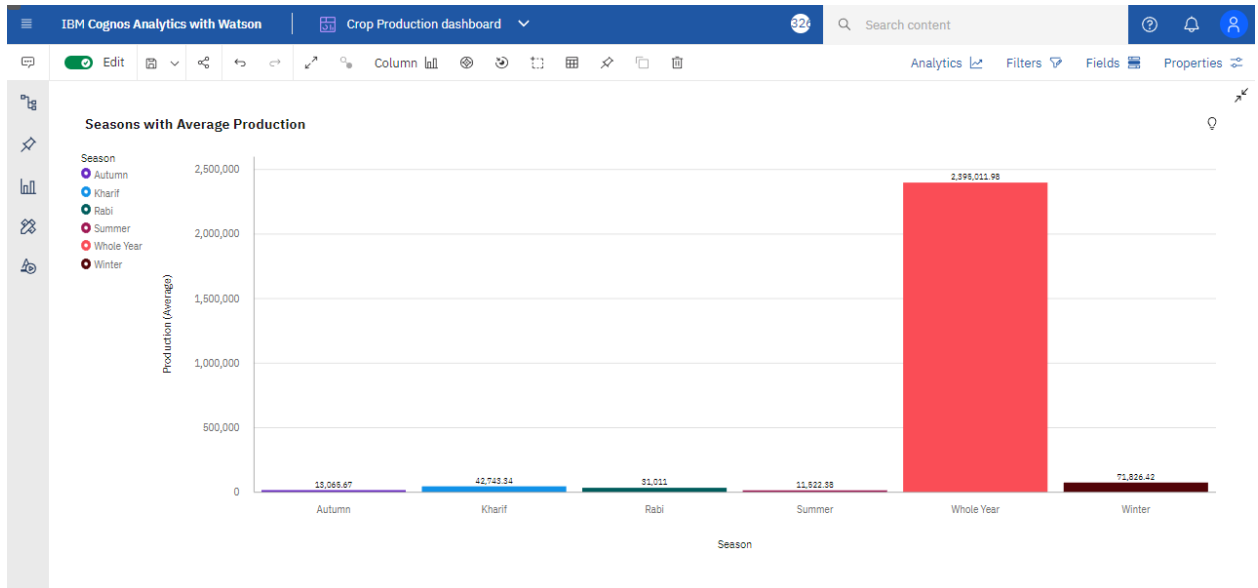
Sprint Planning and Estimation:
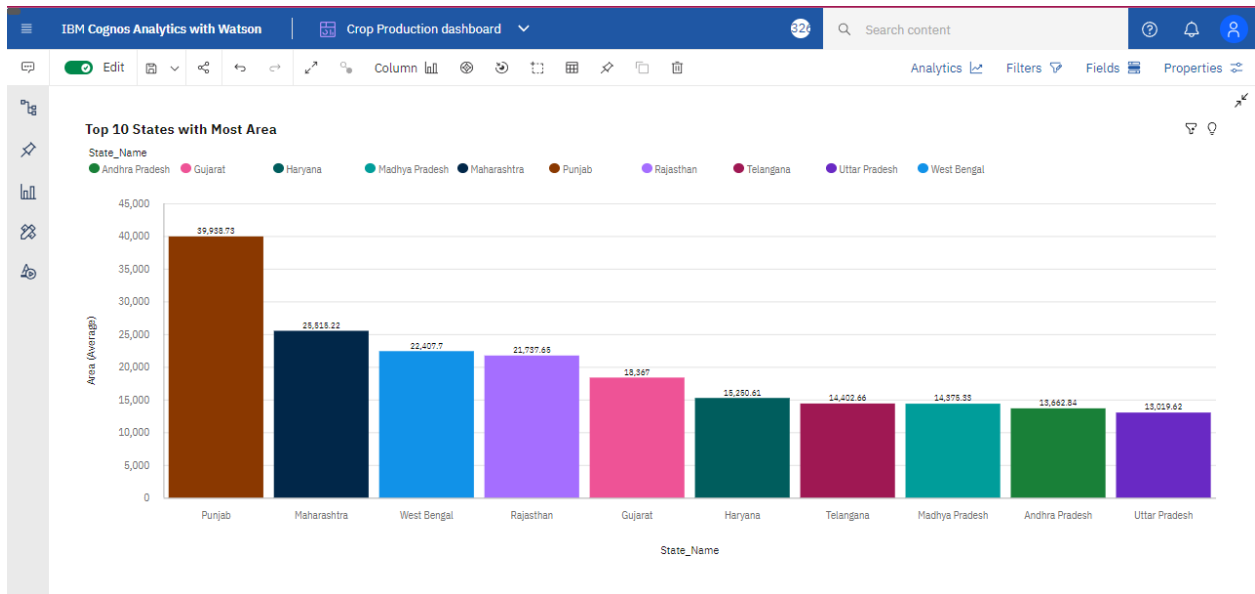
Product Backlog, Sprint Schedule, and Estimation

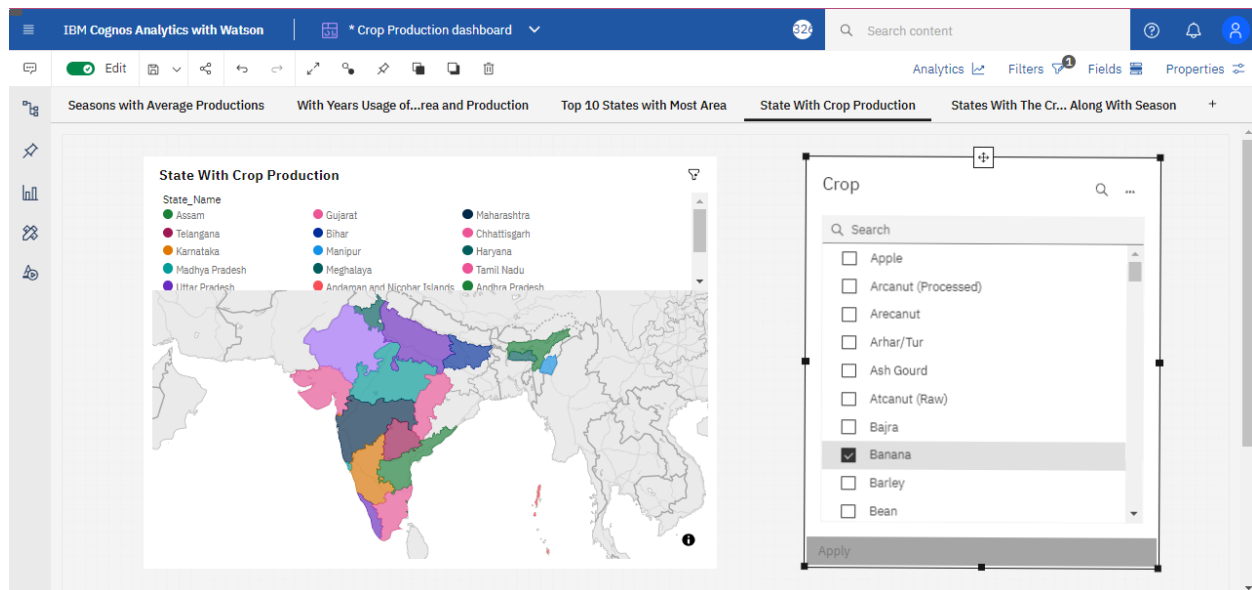| Sprint	Functional Requireme nt(Epic)	User Story Number	User Story / Task	Story Points	Prio rity	Team Members

Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	Anish Banu A Kaviya S Tamil Selvan R Sakthivel V
Sprint-1	Registration	USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	Anish Banu A Kaviya S
Sprint-2	Registration	USN-3	As a user, I can register for the application through Facebook	2	Low	Tamil Selvan R Sakthivel V
Sprint-1	Registration	USN-4	As a user, I can register for the application through Gmail	2	Medium	Kaviya S Sakthivel V
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	Anish Banu A Tamil Selvan R
Sprint-1	Dashboard	USN-6	As a user, I can use the methods provided in the dashboard	2	Medium	Kaviya S Sakthivel V
Sprint-2	Prediction	USN-7	As a user, with the results obtained, I	2	High	Anish Banu A Tamil Selvan R Sakthivel V

			can determine whether profit orloss is made			
Sprint-3	Visualisation	USN-8	Having a view with geograp hic data	2	High	Tamil Selvan R Kaviya S
Sprint-3	Customer Care	USN-9	As a Customer Care Executive, I can answer users' queries	2	Low	Sakthivel V Anish Banu A
Sprint-4	Tools	USN-10	Analysis is performed by tools like cognos analytics	1	High	Anish Banu A
Sprint-4	Administrator	USN-11	As an admin, I can make changes to the interface according the needs	3	High	Anish Banu A Kaviya S Tamil Selvan R Sakthivel V

Results

Top 10 States with Most Area

State_Name

● Andhra Pradesh ● Gujarat ● Haryana ● Madhya Pradesh ● Maharashtra ● Punjab ● Rajasthan ● Telangana ● Uttar Pradesh ● West Bengal

Chart: Top 10 States with Most Area (Area Average by State_Name)

State_Name	Area (Average)
Punjab	39,938.73
Maharashtra	25,515.22
West Bengal	22,407.7
Rajasthan	21,737.65
Gujarat	18,367
Haryana	15,250.61
Telangana	14,402.66
Madhya Pradesh	14,375.88
Andhra Pradesh	13,662.84
Uttar Pradesh	13,019.62

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 productivity, hidden patterns discovery using data set related to seasons and crop yields data. We have noticed and made analysis about different crops cultivated, area and productions in different states and districts using IBM Cognos.

Appendix

Project Demo Link: https://youtu.be/cBKL88Wgx1k