

Estimate the Crop Yield using Data Analytics

TEAM ID : PNT2022TMID10801

Team Leader : Dhivagar D

Team member : Abinesh Kumar V

Team member : Balaji Ragavendra Raj R

Team member : Gopinath S

1. ABSTRACT

Analytics is the understanding of data patterns to support performance enhancement and decision-making. Agriculture In order to analyse certain key visualisations and create a dashboard, data analytics in agricultural yield is helpful. By looking at these, we may learn the majority of the insights about crop output in India. We can comprehend the data in our organisation 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.

2. INTRODUCTION

The foundation of the Indian economy is agriculture. The majority of farmers in India are not receiving the anticipated crop output for a number of different reasons. The weather has a major impact on agricultural yield. The amount of rainfall has an impact on rice cultivation as well. The farmers in this situation unavoidably need prompt assistance to forecast future crop productivity, and an analysis must be done to assist the farmers in maximising crop production in their crops. A significant issue in agriculture is yield prediction. Every farmer wants to know how much of a yield to anticipate. In the past, farmer experience with a particular crop was taken into account when predicting production. The amount of data in Indian agriculture is huge. When data is transformed into information, it can be used for a variety of purposes. A web-based comprehensive business intelligence package from IBM is called Cognos Business Intelligence. It offers a suite of tools for analytics, scorecarding, reporting, and keeping track of events and data. The software is made up of a number of parts that are made to satisfy the various information needs of a business. For example, IBM Cognos Framework Manager, IBM Cognos ,Cube Designer, and IBM Cognos Transformer are all parts of IBM Cognos. Cognos Analysis Studio enables business users to receive prompt responses to commercially relevant questions. You may design pixel-perfect reports for your company using reporting studio.

3. LITERATURE SURVEY

At present we are at the immense need of another Green revolution to supply the food demand of growing population. With the decrease of available cultivable land globally and the decreased cultivable water resources, it is almost impossible to report higher crop yield. Agricultural based big data analytics is one approach, believed to have a significant role and positive impact on the increase of crop yield by providing the optimum condition for the plant growth and decreasing the yield gaps and the crop damage and wastage. With this aim the present paper reviews about the various advances, design models, software tools and algorithms applied in the prediction assessment and estimation of the crop yield. India is basically agriculture based country and approximately 70% our country economics is directly or indirectly related to the agricultural crops. The principle crop which occupies the highest (60-70%) percentage of cultivable land in the Indian soil is the paddy culture and it is the major crop especially in central and south parts of the India. Rice crop cultivation plays an imperative part in sustenance security of India, contributing over 40% to general yield generation. The enhanced yield of the rice crop depends largely on the water availability and climatic conditions. For example, low precipitation or temperature extremes can drastically diminish rice yield. Growing better strategies to foresee yield efficiency in a mixture of climatic conditions can help to understand the role of different principle factors that influence the rice crop yield. Big data analytic methods related to the rice crop yield prediction and estimation will certainly support the farmers to understand the optimum condition of the significant factors for the rice crop yield, hence can achieve higher crop yield.

a. 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 .

Agriculture is one of the major revenue producing sectors of India and a source of survival. Various seasonal, economic and biological factors influence the crop production but unpredictable changes in these factors lead to a great loss to farmers. These risks can be measured when suitable mathematical and statistical model designs are applied on data related to soil, weather and past yield. With the advent of data mining, crop yield can be predicted by deriving useful insights from these agricultural data that aids farmers to decide on the crop they would like to plant for the forthcoming year leading to maximum profit. There are various systems that use diverse data mining technologies to manipulate data to derive insights and help in decision making for farmers. The present data mining systems and algorithms used were focus either on one crop and predict or forecast any one parameter like either yield or price. A research presents a survey on the various algorithms used for crop yield prediction, study used to forecast the yield and price of major crops of Tamil Nadu based on historical data. The data and predicted output are accessible for the farmers through a web application. This aids farmer to decide on the crop they would like to plant for the forthcoming year. In addition, the web application also provides a forum for the farmers to goods the products without middlemen which help them to obtain maximum price for their products .

b. Crop Yield Prediction Using Data Mining Techniques

India is a country where farming and agriculture based industries are the major resource of economy. It is also one of the country which suffer from major natural calamities like drought or flood which damages the crop which cause huge financial loss for the farmers and economic stability of the country. Predicting the crop yield well in advance prior to its harvest can help the farmers and Government organizations to make appropriate planning like storing, selling, fixing

minimum support price, importing/exporting etc. Predicting a crop well in advance requires a systematic study of huge data coming from various variables like soil quality, pH, essential elements (N,P,K) quantity etc. As Prediction of crop deals with large set of database thus making this prediction system a perfect candidate for application of data mining methodologies which majorly helps in acquiring a knowledge to achieve higher crop yield. The success of any crop yield prediction system heavily relies on how accurately the features have been extracted and how appropriately classifiers have been employed. Study summarizes the results obtained by various algorithms which are being used by various authors for crop yield prediction, with their accuracy and recommendation .

Weeds and pests were the major crop damaging biotic agents and the farmers are need to be wellinformed in accessing the various data mining technologies to acquire a knowledge on applications of effective weed and pest control strategies and managing techniques to reduce crop damage. Collection of data related to the various weeds and pest, modeling of the data to prepare for the mining, selection of appropriate methodology, interpretation and sharing the information become the major challenges in weed and pest control to protect the crop damage. A study was conducted to evaluate the major challenges and noteworthy opportunities and applications of of Big Data in controlling the weed and pest damage and hence to achieve higher crop yield. Study reported that the form of the data collected, type of the assessment method and tools applied are the major influencing factors in understanding the role of crop damaging agents such as weed and pest, which provides the knowledge on using improved crop management

strategies crop yield prediction. Big Data cargo space and questioning incurs intense challenges, in respect to allocate the data across numerous technologies, and also continuously evolving data from diverse sources. When the selected data was from the different sources, semantic methodologies play a vital role in the assessment, which preliminarily detect the factors possess potential agricultural importance and developing relationships between data items in terms of meanings and units. Study presented a success story from the Netherlands in using the information from the Big Data analytics, with numerical algorithms in controlling the crop damage and reported the higher crop yield. Study concluded that, the utility and the applications and of Big data analytics for weed and pest control is very large and particularly for invasive, parasitic and herbicide-resistant weeds. Also imported the need of collaboration of agricultural scientists with data scientists to implement the methodologies for the benefit of agricultural practices .

Data mining plays a pivotal role for decision making on different concerns with respect to agriculture practices. The objective of the data mining methods is to mine knowledge from an accessible data set and convert it into a comprehensible format for some significant application of the Agri process. Crop management of certain agriculture region is depending on the climatic conditions of that region because climate can make huge impact on crop productivity. Real time weather data can help to achieve the good crop management. Effective utilization of mined agricultural based information and communications expertise enables automation of retrieving useful data in an effort to acquire knowledge, which provides opportunity to easier data acquisition from electronic sources directly, transfer to secure electronic system of documentation and

reduces manual tasks. Automation strategies reduce the overall production cost, hence support for higher crop yield and higher market price. Also identified that how the data mining helps to analyze and predict the useful pattern from huge and dynamically changed climatic data. In the field of agricultural bioengineering, scientist and engineers in collaboration have developed and discussed the application of mathematical model designs like fuzzy logic designs in optimization of the crop yield, artificial neural networks in validation studies, genetic algorithms designs in accessing the fitness of the model applied, decision trees, as well as support vector machines to assess soil, climate conditions and availability of water resources related to crop growth and pest management in agriculture. Study summarizes the application of data mining technologies i.e Neural Networks, Support Vector Machine, Big Data analysis and soft computing in the assessment of agriculture field based on weather conditions .

c. Crop yield prediction using Big Data Analytics

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 assesses the overall crop yield prediction and estimation. At present a common issue worldwide is, farmers are stressed in producing higher crop yield due to the influence of unpredictable climatic changes and significant reduction of water resource worldwide. A study was carried out to collect the data on

world climatic changes and the available water resources which can be used to encourage advanced novel approaches such as big data analytics to retrieve the information of the previous results to the crop yield prediction and estimation. Study imported that the selection and usage of the most desirable crop according to the existing conditions, support to achieve the higher and enhanced crop yield .

The accurate prediction of crop yield certainly benefits the farmers in choosing the right method to reduce the crop damage and gets best prices for their crops. 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 hectares, 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 R^2 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 .

Crop yield gaps, measured as difference between expected yields based on the potency and actual farm yield received. In order to achieve the higher crop yield, farmers must need to tackle the influencing factors such as influence of change in climate conditions on the prospects of crop yields, and change in the usage of agricultural land to assess and ultimately reduce the crop yield gaps. Several researchers reported the applications of bio simulation models to estimate the crop yield gaps in the last decade. The impact of the crop yield gaps assessment studies conducted through bio simulation based methodologies were negatively influenced by quality and resolution of climate and soil data, as well as unscientifically expectations about crop yield prediction systems and crop yield assessment modeling designs calibration method. An explicit rationale model which can effectively applied at various levels of the availability of quality information for identifying data sources to analyze crop yield and measuring yield gaps at definite geographical locations and works based on the rise in titer approach. The model is highly helpful in retrieving the useful data from the available, poor quality, less rigorous data sources or if the data is not available. A case study was discussed on the application of selected model design to quantify the yield gaps of maize crop in the state of Nebraska (USA), and also at the different geographical locations representing the nations Argentina and Kenya at national scale level. Different geographical locations such as Nebraska (USA), Argentina and Kenya were identified to symbolize the distinct scenarios of Agri based data availability and the quality for the selected variables assessed to predict and estimate the crop yield gaps. The definitive aspiration of the planned method is to afford transparent, easily accessible, reproducible and technically sound and strong guidelines for predicting the yield gaps.

The proposed guidelines were also relevant for understanding and to simulate the influence of change in climate conditions usage of cultivable land changes from national to global scales. As indicated, the better understanding of data importance and usefulness for analyzing crop yield and estimating yield gaps as illustrated can help in identifying the data gaps in the crop yield and allow focusing on the various efforts taken at the global level to address the most critical issue .

Analyzing the yields of crop is necessary to update the policies to ensure food security. A research group conducted a study with the aim in suggesting a novel data mining method to predict the yields of crop depends on agricultural big data analytics methodologies, which were progressively contrast with conventional data mining methodologies in the process of handling data and modeling designs. Study suggested that the method employed should be user friendly, work based on progressive big-data responsive processing structure, supposed to utilize the existing agricultural based significant datasets and would still be used with the larger volumes of data growing at enormous rates. Nearest neighbors modeling is one such novel data mining technique which works on the results collected based on data processing structures form the farmers and suggest a well unbiased result on the base of accuracy and prediction time in advance. Study further discussed a case study on the assessment of actual crop dataset (numerical examples on) in China from 1995-2014. Study reported that the novel model employed has publicized an improved performance and was found to be progressive in reporting prediction accuracy percentage of the compared methodologies with conventional designs [7].

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 present 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 .

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. Crop production abilities and yield gaps can be assessed and measured by comparing the possible yields at normal conditions with respect to the crop production under, respectively, irrigated and rain fed conditions by keeping the crop yield levels limited by the less availability of the water as benchmarks. Yield gaps can be defined as the difference between the expected crop yields with respect to the actual crop yield and accurate, spatially unambiguous awareness and information about the yield gaps is necessary to achieve sustainable amplification of agricultural yields. Keeping an aim of discussing the impact of the various methods practiced in measuring the yield gaps with a spotlight on the local-to-global importance of outcomes, a research group carried out a survey on the various methods applied to estimate yield gaps. Study reported few standard operation methods, employed in quantifying the crop yield potential on the data collected from the farmers of western Kenya, Nebraska (USA) and Victoria (Australia). Study recommended for the use of accurate and recent yield data assessed through calibrated crop model designs and further up scaling validated methods in the prediction of crop yield gaps The bottom-up application of this global protocol allows verification of estimated yield gaps with on-farm data and experiments .

d. M. A. Jayaram and Netra Marad, “Fuzzy interference Systems for Crop Prediction”, Journal of Intelligent Systems, 2012, 21(4), pp.363-372[1]. Prediction of crop yield is significant in order to accurately meet market requirements and proper administration of agricultural activities directed towards enhancement in yield. Several parameters such as weather, pests, biophysical and morphological features merit their consideration while

determining the yield. However, these parameters are uncertain in their nature, thus making the determined amount of yield to be approximate. It is exactly here that the fuzzy logic comes into play. This paper elaborates an attempt to develop fuzzy inference systems for crop yield prediction. Physio morphological features of Sorghum were considered. A huge database (around 1000 records) of physio morphological features such as days of 50 percent? powering, dead heart percentage, plant height, panicle length, panicle weight and number of primaries and the corresponding yield were considered for the development of the model. In order to? nd out the sensitivity of parameters, one-to-one, two-to-one and three-to-one combinations of input and output were considered. The results have clearly shown that panicle length contributes forth yield as the lone parameter with almost one-to-one matching between predicted yield and actual value while panicle length and panicle weight in combination seemed to play a decisive role in contributing for the yield with the prediction accuracy rejected by very low RMS value.

P. Vindya “Agricultural Analysis for Next Generation High Tech Farming in Data Mining” , Anna University, Trichy, Tamil Nādu, India, 5 May 2015. Recent developments in Information Technology for agriculture field have become an interesting research area to predict the crop yield.

4. IDEATION AND PROPOSED SOLUTION

System design is defined as the use of systems theory to the creation of a project. The architecture, data flow, use case, class, sequence, and activity diagrams of a project's development are defined by the system design.

IBM Cognos Analytics

A collection of business intelligence tools called IBM Cognos Analytics is offered both on premises and in the cloud. The main emphasis is on descriptive analytics, which uses dashboards, expert reporting, and self-service data exploration to help users understand the information in your data. In this study, we analysed the crop yield data using IBM cognos data analytics.

Following are important features of IBM Cognos:

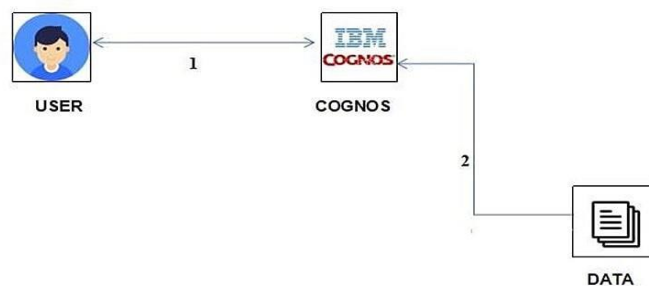
1. **Get Connected** - Connect your data effortlessly Import data from CSV files and spreadsheets. Connect to cloud or on-premises data sources, including SQL databases, Google BigQuery, Amazon, Redshift, and more.
2. **Prepare your data** – Prepare and connect data automatically Save time cleaning your 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 Personalised 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.

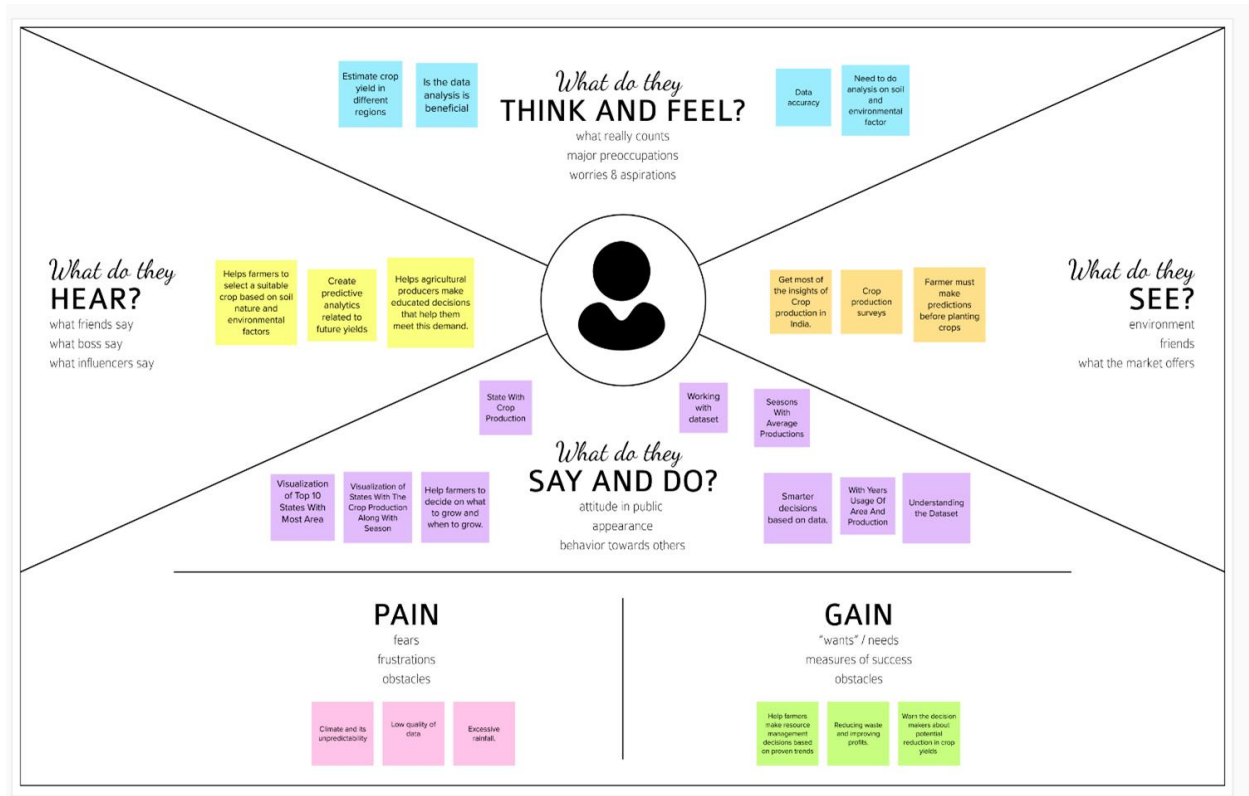
B. System Architecture

India is one of the top countries for agricultural production, making it one of the most significant sources of income. As part of this project, we will analyse some significant visualisations, build a dashboard, and then use this information to gain the majority of our understanding of crop output in India.

Technical Architecture:



EMPATHY MAP



5 . REQUIREMENT ANALYSIS

IBM Cognos Analytics

A collection of business intelligence tools called IBM Cognos Analytics is offered both on premises and in the cloud. The main emphasis is on descriptive analytics, which uses dashboards, expert reporting, and self-service data exploration to help users understand the information in your data. In this study, we analysed the crop yield data using IBM cognos data analytics.

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 Personalised 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.

B. System Architecture

India is one of the top countries for agricultural production, making it one of the most significant sources of income. As part of this project, we will analyse some significant visualisations, build a dashboard, and then use this information to gain the majority of our understanding of crop output in India.

IBM® Cognos® Analytics integrates reporting, modeling, analysis, dashboards, stories, and event management so that you can understand your organization data, and make effective business decisions.

After the software is installed and configured, administrators set up security and manage data sources. You can get started yourself by uploading local files and applying visualizations in dashboards or stories. For enterprise-level data, modelers are next in the workflow. After data modules and packages are available, report authors can then create reports for business users and analysts. Administrators maintain the system on an ongoing basis.

Whether you're an analyst, report author, data modeler, or an administrator, you start by signing in to the Welcome portal from your desktop or mobile device. There are coach marks in the user interface to help you discover what's where.

6. PROJECT DESIGN

PROJECT FLOW

1. Users create multiple analysis graphs/charts.
2. Using the analyzed chart creation of the Dashboard is done.
3. Saving and Visualizing the final dashboard in the IBM Cognos Analytics.
4. To accomplish this, we have to complete all the activities and tasks listed below
5. IBM Cloud Account
6. Login to Cognos Analytics
7. Working with the Dataset
8. Understand the Dataset
9. Loading the Dataset
10. Data visualization charts
11. Seasons with average productions
12. With years usage of Area and Production
13. Top 10 States with most area

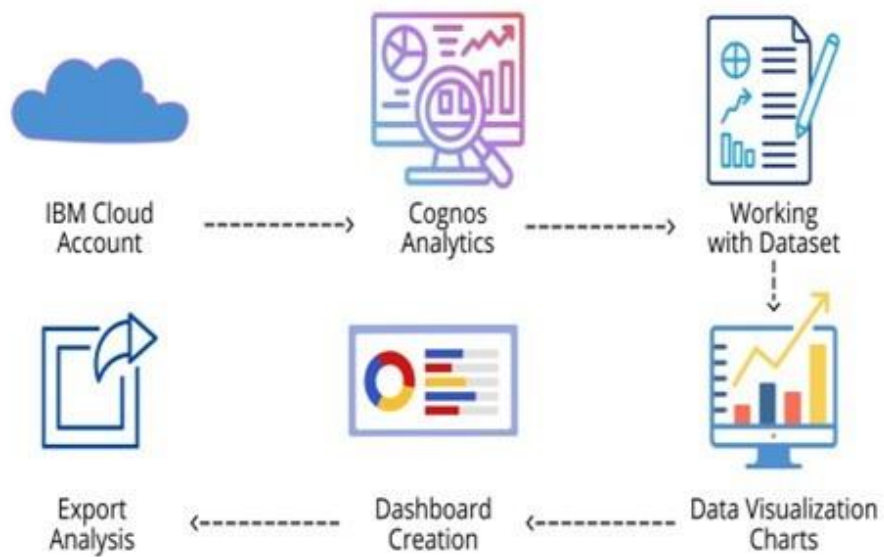
14.State with crop production

15.States with the crop production along with season (Text Table)

16.Dashboard Creation

17.Export the Analytics

PROJECT FLOW CHART



SOLUTION REQUIREMENTS

Functional Requirements:

Following are the functional requirements of the proposed solution.

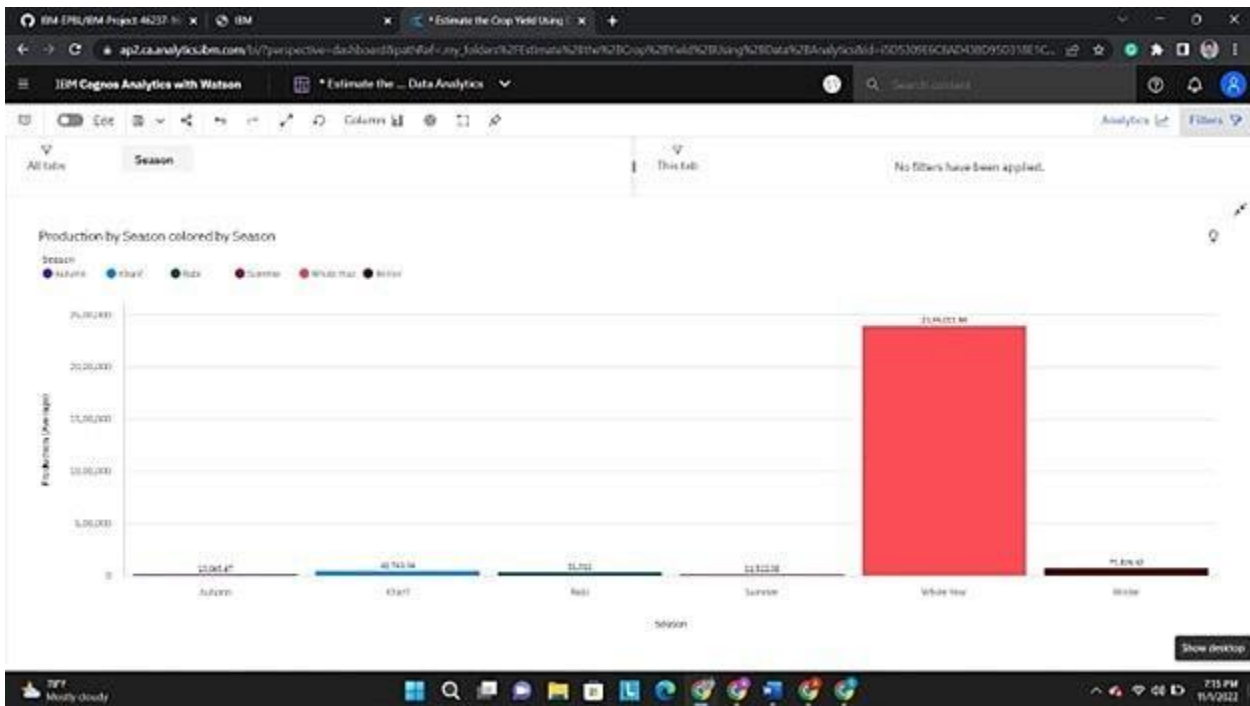
FR No.	Functional Requirement (Epic)	Sub Requirement (Sub-Task)
1	User Registration	Registration through Whatsapp Registration through Gmail Registration through Mail Registration through Agri-Consultancy modes.
2	User Confirmation	Confirmation via Email Confirmation via OTP through SMS. Confirmation via physical Letter.
3	User Profile	User Details Farm Details

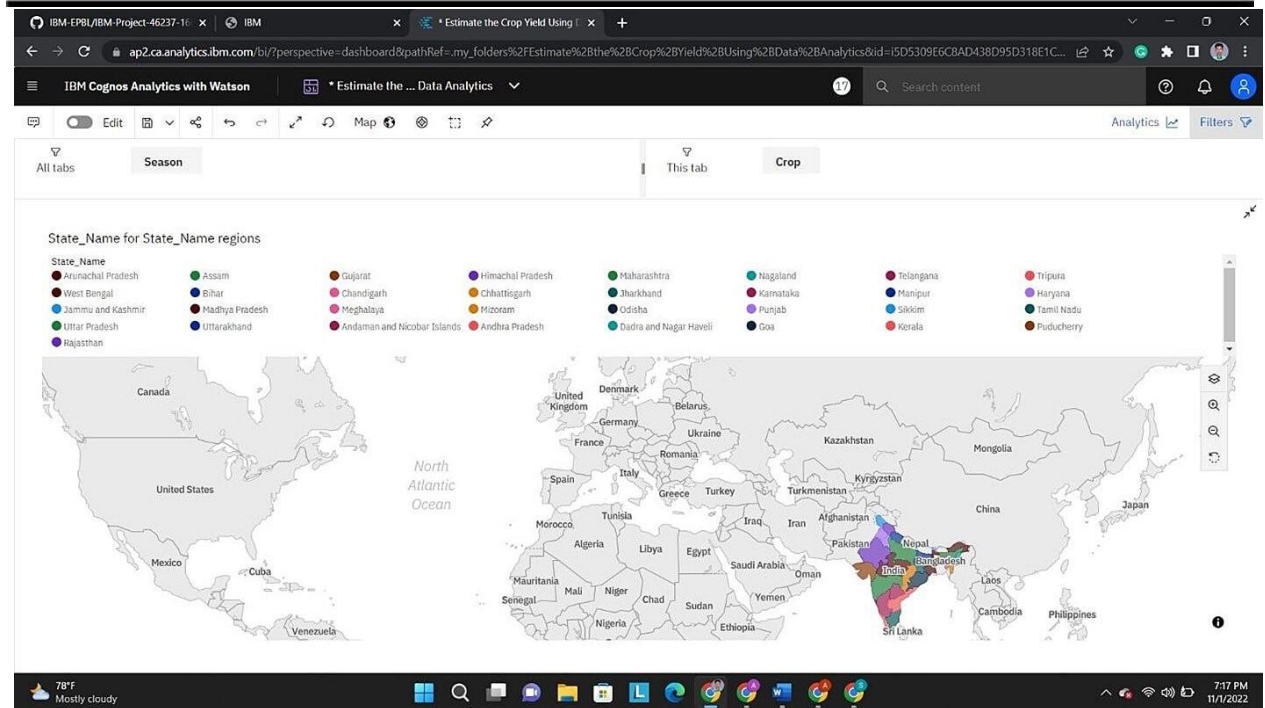
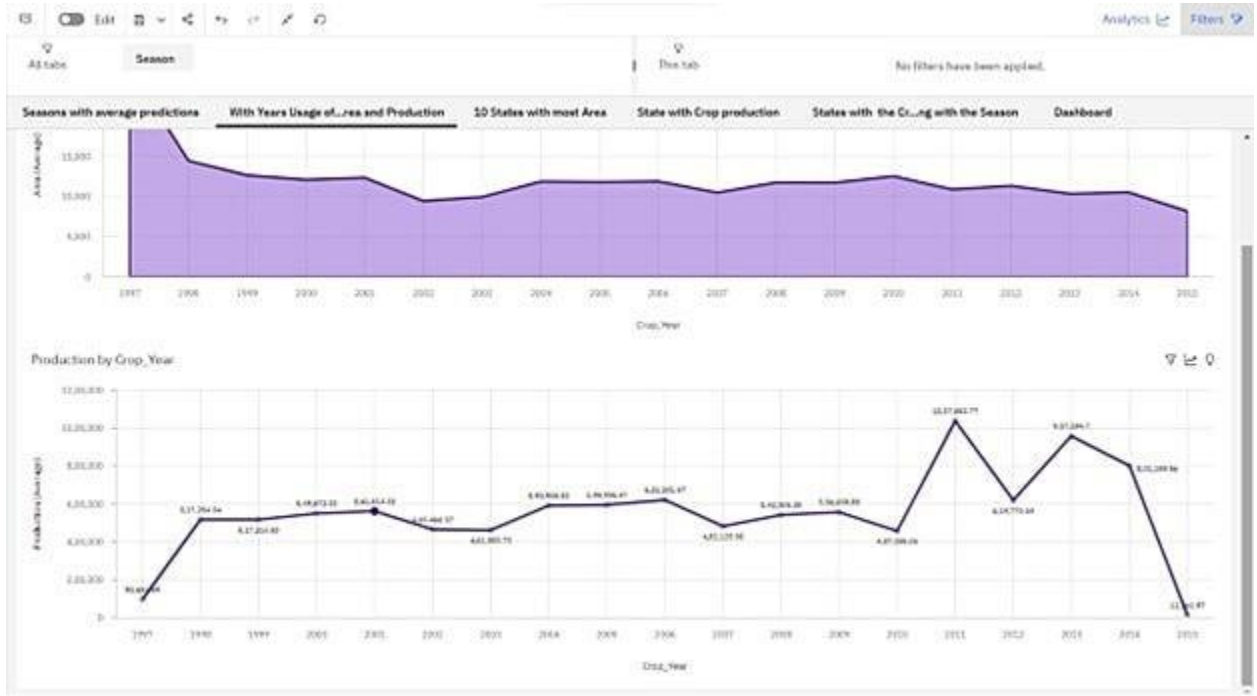
4	Required Data	The past crop yield data and data of the Farmer to analyze their yield.
5	Analysis	Clean , Analyze the data by means of set of past data of the multiple users which is farmers.
6	Estimation	Creating the perfect data module, visuals using IBM Cognos to increase the estimation of the crop yield

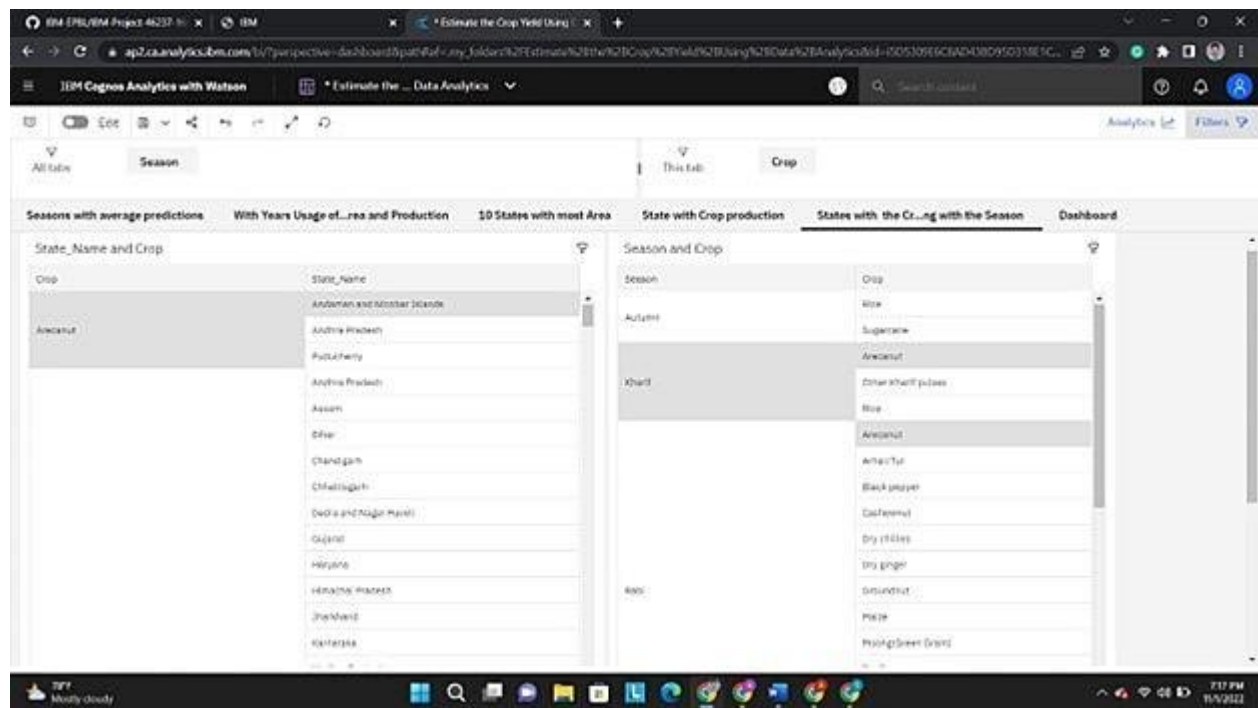
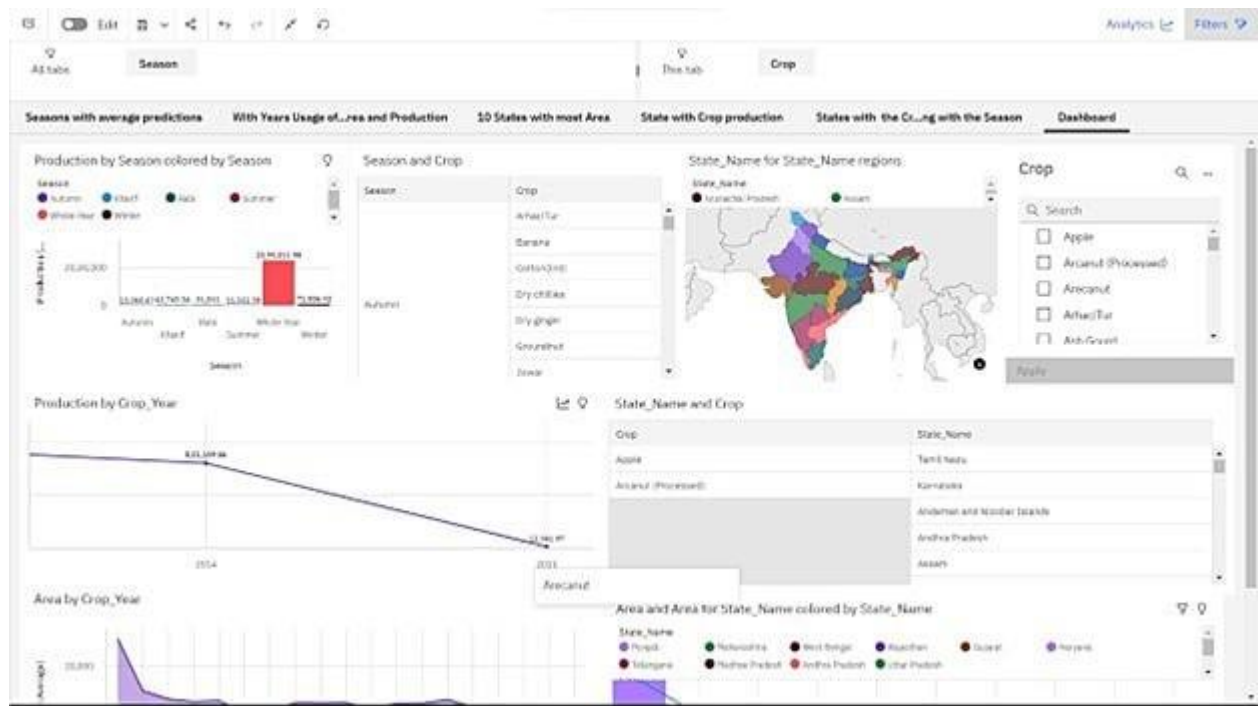
Non Functional Requirements:

FR No.	Non-Functional Requirement	Description
1	Usability	The data report is created according to the past data yield. By considering these recommendation the sowing of crops will be decided.
2	Security	IBM Cognos have a high-secure user information.
3	Reliability	The interactive data visuals of the dashboard can make easy to understand by the farmers.

7.RESULTS







8. CONCLUSION

The productivity of agriculture has slightly increased as a result of technology's introduction. New ideas like digital agriculture, smart farming, precision agriculture, etc. have been made possible by the innovations. The analysis of agricultural productivity and the uncovering of hidden patterns utilising data sets related to seasons and crop yields have been noted in the literature. Using IBM Cognos, we have observed and conducted analysis regarding various crops grown, areas, and productions in various states and districts.