Nalaiya Thiran

Professional Readiness for Innovation,

Employability,Entrepreneurship.

Estimate the Crop Yield using Data Analytics

Beni TS	962219205015
Benny Thomas	962219205016
Shaliha Safreen S	962219205036
Smyla Lucia J	962219205042

In partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

Information Technology

St.Xavier's Catholic College Of Engineering

ANNA UNIVERSITY::CHENNAI 600025

NOV/DEC-2022

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE NO
1.	**INTRODUCTION**	5
	1.1 Project Overview	
	1.2 Purpose	
2.	**LITERATURE SURVEY**	11
	2.1 Existing Problem	
	2.2 Problem Statement Definition	
3.	**IDEATION AND PROPOSED SOLUTION**	28
	3.1 Empathy Map Canvas	
	3.2 Brainstorming	
	3.3 Proposed Solution	
	3.4 Problem Solution Fit	
4.	**REQUIREMENT ANALYSIS**	32
	4.1 Functional Requirement	
	4.2 Non-Functional Requirements	

5.	**PROJECT DESIGN**	35
	5.1 Data Flow Diagram	
	5.2 Solution & Technical Architecture	
	5.3 User Stories	
6.	**PROJECT PLANNING & SCHEDULING**	40
	6.1 Sprint Planning & Estimation	
	6.2 Sprint Delivery Schedule	
	6.3 Reports from JIRA	
7 .	**CODING & SOLUTIONING**	48
	7.2 Feature 1	
	7.2 Feature 2	
8 .	**TESTING**	58
	8.1 Test Cases	
	8.2 User Acceptance Testing	
9.	**ADVANTAGES & DISADVANTAGES**	60
10.	**CONCLUSION**	61
11.	**FUTURE SCOPE**	62

12.	**APPENDIX**	63
	Source Code	
	Github & Project Demo Link	
13.	**REFERENCE**	69

Estimate the Crop Yield using Data Analytics

CHAPTER 1

INTRODUCTION

Data Analytics

Data Analytics refers to the techniques used to analyse data to enhance productivity and business gain. Data is extracted from various sources and is cleaned and categorised to analyse various behavioural patterns. The techniques and the tools used vary according to the organisation or individual.So,if you understand your Business Administration and have the capability to perform Exploratory Data Analysis, to gather the required information, then you are good to go with a career in Data Analytics.Data analytics is important because it helps businesses optimise their performances. Implementing it into the business model means companies can help reduce costs by identifying more efficient ways of doing business and by storing large amounts of data. A company can also use data analytics to make better business decisions and help analyse customer trends and satisfaction, which can lead to new and better products and services.

Once the data that's needed is in place, the next step is to find and fix data quality problems that could affect the accuracy of analytics applications. That includes running data profiling and data cleansing tasks to ensure the information in a data set is consistent, errors and duplicate entries are eliminated. Additional data preparation work is done to manipulate and organise the data for the planned analytics use. Data governance policies are then applied to ensure that the data follows corporate standards and is being used properly

What are the tools used in Data Analytics?

With the increasing demand for Data Analytics in the market, many tools have emerged with various functionalities for this purpose. Either open-source

or user-friendly, the top tools in the data analytics market are as follows.

- **R-programing**– This tool is the leading analytics tool used for statistics and data modelling. R compiles and runs on various platforms such as UNIX, Windows, and Mac OS. It also provides tools to automatically install all packages as per user-requirement.
- **Python**– Python is an open-source, object oriented programming language that is easy to read, write, and maintain. It provides various machine learning and visualisation libraries such as TensorFlow,Pandas, Keras, etc. It also can be assembled on any platform like SQL server, a MongoDB database or JSON
- **SAS**– A programming language and environment for data manipulation and analytics, this tool is easily accessible and can analyse data from different sources.
- **Microsoft Excel** – This tool is one of the most widely used tools for data analytics. Mostly used for clients' internal data, this tool analyses the tasks that summarise the data with a preview of pivot tables.

Types of Data Analytics

Data analytics is broken down into four basic types.

1. **Descriptive analytics:** This describes what has happened over a given period of time. Have the number of views gone up? Are sales stronger this month than last?
2. **Diagnostic analytics:** This focuses more on why something happened. This involves more diverse data inputs and a bit of hypothesising. Did the weather affect beer sales? Did that latest marketing campaign impact sales?
3. **Predictive analytics:** This moves to what is likely going to happen in the near term. What happened to sales the last time we had a hot summer? How many weather models predict a hot summer this year?

4. **Prescriptive analytics:** This suggests a course of action. If the likelihood of a hot summer is measured as an average of these five weather models is above 58%, we should add an evening shift to the brewery and rent an additional tank to increase output.

1.1 PROJECT OVERVIEW

Agriculture

Agriculture forms the basis for food security and hence it is important. In India, the majority of the population i.e., above 55% is dependent on agriculture as per the recent information. Agriculture is the field that enables the farmers to grow ideal crops in accordance with the environmental balance.

In India, wheat and rice are the major crops grown along with sugarcane, potatoes, oil seeds etc.Farmers also grow non-food items like rubber, cotton, jute etc. More than 70% of the household in the rural area depend on agriculture.This domain provides employment to more than 60% of the total population and has a contribution to GDPalso (about 17%).In the farm output, India ranks second considering the worldwide scenario.

This is the widest economic sector and has an important role regarding the framework of the socio-economic fabric of India. Farming depends on various factors like climate and economic factors like temperature,irrigation,cultivation, soil, rainfall, pesticide and fertilisers. Historical information regarding crop yield provides major input for companies engaged in this domain. These companies make use of agriculture products as raw materials, animal feed, paper production and so on. The estimation of production of crops helps these companies in planning supply chain decisions like production scheduling. The industries such as fertilisers, seed, agrochemicals and agricultural machinery plan production and activities like marketing based on the estimates of crop yield.

Areas in Agriculture that uses data analytics

1. Receiving Useful Data to Help Fight Food Scarcity and Empower Small Farmers

Since data scientists have tools to process and analyse gigantic amounts of data efficiently, projects are underway to determine how that information might help small-scale farmers join in the battle to solve worldwide food shortages.

In September 2018, a coalition launched a project that will run through 2030 and look at data from approximately 500 million farmers in impoverished areas from 50 countries.The people behind the project hope the data will show whether agricultural investments in various countries are paying off and help develop policies for the farmers. On a larger scale, this project aligns with the United Nations' Sustainable Development Goals to double the agricultural productivity and incomes of farmers in developing nations and help them reduce world hunger.

2. Managing Crop Diseases and Pests

Agricultural pests can quickly cut into a farmer's profits. But, misusing pesticides can have adverse effects on people, plants and other living things. Fortunately, some companies recruit data scientists to help them develop user-facing platforms that analyse when to apply pesticides and how much to use.

One of them is a Brazilian company called Agrosmart. Its technology relies on Internet of Things (IoT) sensors and artificial intelligence to determine the kind of insects on a crop and the quantity present. Farmers then get an associated report and can use it to plan their pest management approaches. The goal is to help farmers cost-effectively control pests with a minimised environmental impact.

In another case, Saillog, an Israeli startup, developed a smartphone app called Agrio that informs farmers of the diseases currently affecting their crops or ones found on surrounding farms.

3. Investigating Agricultural Niches

Data scientists know how to use tools that identify patterns and relationships that may otherwise remain hidden. As such, they can draw conclusions that push agricultural science forward through the examination of specific factors. For example, researchers know trace minerals positively affect the metabolic functions of livestock and poultry, while carotenoids play a role in increasing egg yolk quality and nutrition. The findings brought about by sifting through databases and studies to conclude things like these show how seemingly small factors in agricultural processes can bring about substantial changes.

4. To Cope With Climate Change

Climate change is a looming concern that has already affected the agriculture sector. However, data scientists are hard at work figuring out ways to compensate for the shift.

One project involves giving IoT sensors to Taiwanese rice farmers so they can collect crucial information about their crops. It'll all go into a database used to help farmers optimise their production cycles, even when climate change makes that task exceptionally challenging. Following the traditional farming calendar is no longer sufficient because of climate change. But, data analysis could forever change the future of farming.

Scientists are also scrutinising agricultural soil data to improve their understanding of how soil contributes to climate change by releasing

greenhouse gases, as well as how soil data might aid in adapting to climate change. Collecting this kind of information is tricky, but scientists believe it could fill in knowledge gaps associated with the relationship between soil and climate change.

5. To Make Yield Predictions

A poor yield can result in a devastating season for farmers, as well as all the entities that depend on the crops. IBM has a platform that estimates corn yields 2 to 3 months in advance, reducing unpleasant surprises for agricultural professionals.

1.2 Purpose

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

CHAPTER 2
LITERATURE REVIEW

1]Crop Monitoring and Crop Yield Prediction Computerised Tools in Mexico

Jesus Soria-Ruiz and Yolanda Fernandez

Crop yield prediction is an information service provided by the National Institute Research for Forestry Agriculture and Livestock (INIFAP) to the Ministry of Agriculture in Mexico where it is used as a decision making aid. A comprehensive set of state of the art tools has been developed which is currently providing accurate results. The tools consist mainly of prediction models with input data taken on site at sample plots as well as remotely sensed data. The different levels of government agencies –federal, state and municipal– are already benefiting from timely information about the condition of agricultural crops, including the relationship with climatic effects and other potentially adverse factors and the impact on crop production.

The general aims of this information service project are:

- The timely provision of information from the production plots as it relates to the general agricultural regional productivity
- To develop a decision support system of agricultural related activities to support government farmer aid programs.

The benefits of this Project are:

- Timely response and reliability of the information,
- Reliable decision support,
- More accurate and complete information
- Production risk reductions
- Savings of government funds due to a better definition of required import

volumes for crops.

- increased confidence concerning the requirements of import volumes ● Increased efficiency in crop management and production
- Availability of production volumes ahead of harvest.

2] Paddy Yield Predictor Using Temperature, Rainfall, Soil pH, and Nitrogen

Priya Rao.R

Agriculture is the backbone of India which indirectly contributes to the Indian economy. Farmers who are the drivers of agriculture are facing a lot of problems for proper identification of the crops that can be cultivated for the specific soil conditions and to maximise the crops yield. All these problems are due to lack of technology and scientific techniques being used in agriculture. Crop yield varies as a result of variations in atmospheric and soil conditions. Data mining mainly focuses on methods to elicit useful knowledge from the dataset. There are several data mining approaches that can be used for the purpose of predicting crops yield and finding association among attributes contributing for the crops yield. This paper mainly intensifies on various association algorithms, namely Apriori, Eclat, and AprioriTid to find the association among temperature, rainfall, soil pH, soil nitrogen, and paddy yield.

Methodology

- Data Collection
- Data Preprocessing
- Applying Association Rule Mining Algorithm
- Apriori Approach
- Eclat Approach
- AprioriTid Algorithm
- Analysing Association Rules

This work can be extended for predicting the type of crop that can be cultivated in a particular plot for high yield. Also this work can be enhanced to predict the better association rule by considering additional attributes like Phosphorus, Carbon, and Zinc. It can be applied to more agricultural crops of various locations. More number of association algorithms can be implemented on the agricultural dataset.

3] Soil Analysis and Crop Prediction

Shubham Prabhu and Saranya

Soil analysis is an important process to determine the available plant nutrients in the soil. Plants absorb the major nutrients through soil. In addition to soil, there are various major factors like rainfall, precipitation, fertiliser, etc that affect plant growth. Our aim is to create a prediction engine for the most suitable crop for a particular soil. As an initial step, we have focused on predicting the accurate crop yield to the user by analysing the soil fertility and rainfall in the region entered by the user as an input.

Methods

All the data analysed is continuously monitored,displayed and uploaded on the IoT cloud. Thingspeak provides a precise and accurate display of temperature and moisture data.The sensor accuracy and range is also taken care of, will collect respective data.Naive Bayes and K-NearestNeighbour (KNN) is performed .

In this project analysis of soil based on Temperature and Soil Moisture has been proposed using Arduino, Cloud Computing. The project has high efficiency and accuracy in fetching the live data of temperature and soil moisture. The project will assist the farmers in increasing the agriculture yield and take efficient care of food

production as the stick will always provide a helping hand to farmers for getting accurate live feed of environmental temperature and soil moisture with more than 99% accurate results.

The project proposes a wise agricultural model in integration with IoT. IoT has always mattered in the Agriculture domain.It is a really challenging task because of the highly localised nature of agriculture information specifically distinct conditions.Also only use soil to predict the crop yield which may not be applicable in many situations.

4]A Smart Agricultural Model by Integrating IoT, Mobile and Cloud-based Big Data Analytics

-S. Rajeswari

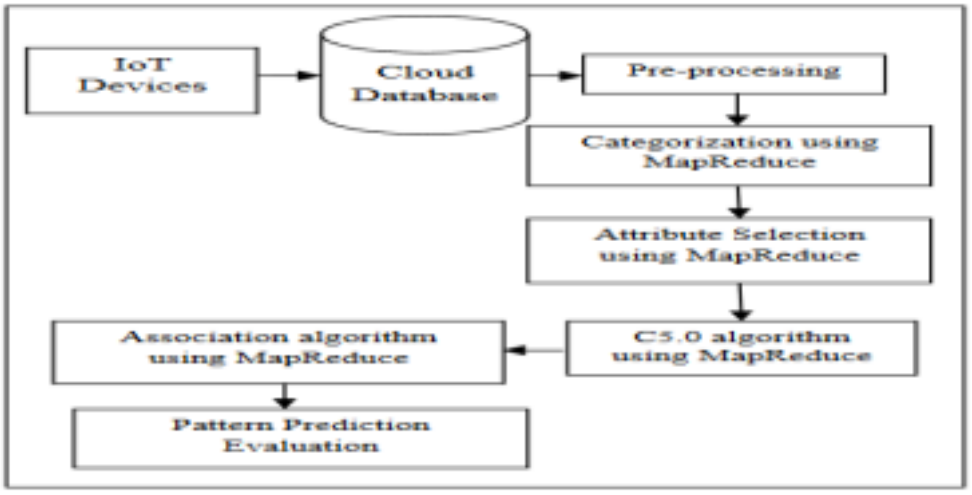
The traditional database paradigm does not have enough storage for the data produced by Internet of Things (IoT) devices leading to the need of cloud storage. These data are analysed with the help of Big Data mining techniques.Cloud based big data analytics and the IoT technology performs an important role in the feasibility study of smart agriculture.Smart or precision agricultural systems are estimated to play an essential role in improving agriculture activities. Mobile device usage is very common by everyone, including the farmers. In that, in the daily life of farmers the Information and Communication Technologies (ICT) play a vital role to get the agricultural Information. The IoT has various applications in the digital Agriculture domain like monitoring the crop growth, selection of the fertiliser, irrigation decision support system, etc. In this paper, an IoT device is used to sense agricultural data and it is stored into the Cloud database. Cloud based Big data analysis is used to analyse the data viz. fertiliser requirements, analysis the crops, market
and stock requirements for the crop. Then the prediction is performed based on data mining technique which information reaches the farmer via mobile

app. Our ultimate aim is to increase the crop production and control the agricultural cost of the products using this predicted information.

Methods:

- Cloud databases are used to store and share the crop information, prices of the fertilisers and crop prices.
- In an agriculture sector, cloud computing gives the smartness with flexibility, predictability, scalability, and optimization.
- It gives the information for farmers at an economical and reasonable cost.
- In the field of agriculture, IoT plays a very important role in collecting data.

MapReduce concept is easy to handle the data and process the multiple nodes. In this model, the process can be divided into maps and reduced. Map function is used to perform filtering and sorting and reduce function is used to perform a summary 2017 International Conference on Intelligent Computing and Control (I2C2) operation. So, the MapReduce technique is used in the predictive analysis concept to predict the data.



They propose a new prototype model for providing the sensing data as a service on the cloud.Wireless Sensor Network increasingly enables applications and services to interact with the physical world. It enables the farmer to have an effective and smart solution to improve the crop yield with less cost.

Our future work will be focussing on interfacing different soil nutrient sensors with IoT tools then collecting the data with the sensor tools and storing the data into cloud databases to analyse and predict with the data mining algorithms suitable for agricultural Big Data analysis for getting the desired outcome.

5] Crop yield prediction in agriculture using data mining predictive analytics techniques

- P.Surya

Data Mining is an emerging research field in Agriculture especially in crop yield analysis and prediction. As early into the growing season as possible, a farmer is focused on perceiving how much yield they are about to expect. As with many other sectors the amount of agriculture data is increasing on a daily source. In our proposed work, collected agriculture dataset will be used to get crop yield prediction models using various regression techniques. Regression analysis was tested for the effective prediction or forecast of the agriculture yield for various crops in Tamilnadu state particularly in the North Western zone of Tamilnadu. North western zone of Tamil Nadu state data consists of four districts. The North western zone of Tamil Nadu districts are Dharmapuri, Salem, Namakkal, Krishnagiri. The analysis depends on the results of the predictor model, in the north western zone, under the area having more cultivated crops are Tapioca, Sugar cane, Ragi, Maize, Groundnut.

Methods

- **Linear Regression**

Linear Regression is one of the commonly used well-known modelling techniques in data mining concept, in which the dependent variable is to be taken as continuous, in other independent variables will be continuous or discrete, and the regression line is linear.

- **Logistic Regression**

Logistic regression technique is used to find the probability of an event of Success and event of Failure. Logistic regression will be used when the dependent variable is binary (0/ 1, True/ False, Yes/ No) in nature.

- **Polynomial Regression**

A regression of y on x may be a polynomial regression of y on x if the ability of independent variable power is greater than one. Such an equation is known as polynomial regression equation.

- **Ridge regression**

Ridge Regression could be a technique used once the data suffers from multicollinearity (independent variables are extremely correlated).This paper dealt with various regression techniques for agriculture crop yield prediction. This mainly focused on getting a predictor model by using regression techniques. Predictor formula is most useful in the crop prediction of Agriculture crop Production in Tons.

6] Agricultural Crop Yield Prediction Using Artificial Neural Network Approach

-Miss.Snehal S.Dahikar , Dr.Sandeep V.Rode

By considering various situations of climatologically phenomena affecting local weather conditions in various parts of the world. These weather

conditions have a direct effect on crop yield. Various researches have been done exploring the connections between large-scale climatological phenomena and crop yield. Artificial neural networks have been demonstrated to be powerful tools for modelling and prediction, to increase their effectiveness. Crop prediction methodology is used to predict the suitable crop by sensing various parameters of soil and also parameters related to atmosphere. Parameters like type of soil, PH, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur, manganese, copper, iron, depth, temperature, rainfall, humidity. For that purpose we are using an artificial neural network (ANN).

Methods:

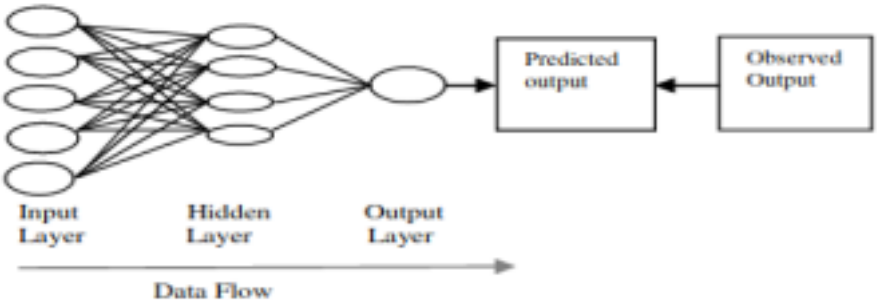
Artificial Neural Network (ANN).

The word network in the term 'artificial neural network' refers to the inter–connections between the neurons in the different layers of each system. An example system has three layers. The first layer has input neurons, which send data via synapses to the second layer of neurons, and then via more synapses to the third layer of output neurons. More complex systems will have more layers of neurons with some having increased layers of input neurons and

output neurons. The synapses store parameters called "weights" that manipulate the data in the calculations.

An ANN is typically defined by three types of parameters:

1. The interconnection pattern between different layers of neurons 2. The learning process for updating the weights of the interconnections
3. The activation function that converts a neuron's weighted input to its output activation.One type of network sees the nodes as „artificial neurons“. These are called artificial neural networks (ANNs).



Layer and connection of a feed-forward back-propagation ANN

In this way we concluded that ANN is a beneficial tool for crop prediction. This paper includes the parameters of their regional soil parameters. Then it is analysed by using feed forward back propagation ANN. Analyse in matlab ANN approach to make it more efficient. ANN Sharing Difficult and ANN Training Takes a Long Time.

7]Rice Crop Yield Prediction in India using Support Vector Machines. -Niketa Gandhi; Leisa J. Armstrong;

Food production in India is largely dependent on cereal crops including rice, wheat and various pulses. The sustainability and productivity of rice growing areas is dependent on suitable climatic conditions. Variability in seasonal climate conditions can have detrimental effects, with incidents of drought reducing production. Developing better techniques to predict crop productivity in different climatic conditions can assist farmer and other stakeholders in better decision making in terms of agronomy and crop choice.This paper discusses the experimental results obtained by applying SMO classifier using the WEKA tool on the dataset of 27 districts of Maharashtra state, India. Various climatic factors which are known to affect the rice crop yield, such as precipitation, minimum temperature, average temperature, maximum temperature, reference crop evapotranspiration,were

considered with the rice yield production for the Kharif season.

Methods:

- Acquiring each parameter (precipitation, minimum, average, maximum temperature and reference crop evapotranspiration) monthly mean records of each district is considered.
- Calculating the total precipitation, average temperature for the minimum, average and maximum temperature for each year for each district during the Kharif season.Acquiring each district's area, production and rice crop yield details of the year 1998 to 2002 from the publicly available Indian Government records.
- For preparing the data set for applying data mining techniques, unrequired columns were omitted. They were sr. no,name of the district and year.The data set was then sorted on the basis of area,which is less than 100 hectares can be omitted.
- The dataset was then sorted on the basis of yield to classify the records into low, moderate and high. The yield has been calculated on the basis of area and production hence these two columns were omitted.
- This data set was then saved in .csv format for further applying data mining techniques. This file had the following columns: precipitation, minimum temperature, average temperature, maximum temperature, reference crop evapotranspiration, crop yield and class.

Table 1 A confusion matrix

		Observed	
		True	**False**
Predicted	**True**	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

WEKA was used to construct the algorithm.The algorithm achieved the accuracy of 78.76%, sensitivity of 68.17% and specificity of 83.97%.The F1

score was computed to measure the test's accuracy and achieved a score of 0.69. Correlation Coefficient was used to measure the quality of classification which resulted in 0.54.In terms of test's accuracy and quality also BayesNet and Multilayer Perceptron showed the highest accuracy and best quality and SMO showed the lowest accuracy and worst quality.

8]Artificial Neural Network-Based Crop Yield Prediction Using NDVI, SPI, VCI Feature Vectors.

-Preeti Tiwari and Piyush Shukla

Agriculture is one of the major revenue producing sectors of India and a source of survival. The number of biological, financial, and environmental factors affects the crop yield. So, unidentifiable changes in these factors lead to failure in agriculture. This paper focuses on prediction of crop yield where different geospatial features were utilised, such as normalised difference vegetation index, standard precipitation index, vegetation condition index. In order to learn from previous weather conditions, a standard error back propagation neural network was used. Here, the training was done in such a way that all sets of features were utilised in pair with their yield value as output. To increase the reliability of the work, the whole experiment was done on a real geo-spatial dataset from the Madhya Pradesh region of India. The result shows that the proposed model has overcome various evaluation parameters on different scales as compared to previous approaches adopted by researchers.

Methods:

Normalised Difference Vegetation Index (NDVI)

For crop yield prediction, NDVI is ordinarily utilised as an input feature vector.

This feature measures vegetation cover in the land. NDVI was first recommended as a record of vegetation well-being and thickness.

$$N = (b_{NIR} - b_{RED} b_{NIR} + b_{RED})$$

where variable N represents as the normalised difference vegetation index (NDVI),

while bRED are the reflectance of red bands and bNIR stands for near infrared.

Standardised Precipitation Index (SPI)

In 1993, this index was introduced by Mckee and his team members. SPI helps in quantifying the precipitation shortage occurs in different time scales for reflecting the effect of available water resources. So, the temporal pattern was identified by estimating the standard precipitation value for different month periods. In order to estimate SPI value, the difference of the precipitation from the average of the selected time scale was taken and finally divided by the standard deviation.

Vegetation Condition Index (VCI)

It is suggested which represents correlation between the current NDVI value with the minimum NDVI value from a huge set of periods of time. It was calculated where V j stands or the jth position VCI value in the vector. Similarly,

N j stands for the jth position NDVI value in the vector. However, Nmin and Nmax are minimum and maximum values of the NDVI vector.

$$V_j = (N_j - Nmin\ Nmax - Nmin) \times 100$$

As digital data on libraries and servers are drastically increased with every new Second, researchers are attracted to work on it. So, this work focuses on geo-spatial data for crop yield prediction by utilising various features vectors of SPI, NDVI, and VCI. Classification based on the working of neural networks has also been performed by many researchers. In some of the works, document

learning is based on single information, but through this proposed work it will overcome this dependency, as well as the learning can be performed on all the available data.

Here, the result shows that proposed work has improved the prediction accuracy by reducing the RMSE, relative error value. Also through proper training and rich input vector resultant, the neural network is a less time-consuming model. It was obtained that proposed work has reduced the relative error by 33% as

compared to SNN, while RMSE also got reduced by 31.8%. In this work, average testing time for prediction of crop yield was decreased by 38.2% and overall accuracy was improved.Temporal aggregates might still be contaminated by cloud cover; the procedure will be biassed by a single false value.

9]Crop Prediction Using Predictive Analytics

-P. S. Vijayabaskar

This work is to construct a model for testing the soil fertility. It also suggests the crop which has to be planted depending upon the value obtained from the sensor. It also provides the regional wise information about the crop in the form of a graph. We have farmer chat where the farmers can share and get ideas from the expert by registering in this application. It also suggests the fertiliser which has to be added to the soil in order to increase the crop productivity. It helps the farmer to analyse the fertility of their yard and plant the better crop to increase their productivity and profit. It also provides the information about the fertiliser to be added in the soil and also provides the information about the nearby fertiliser shop.

Predictive Analytics

Predictive analytics include various statistical techniques such as predictive modelling, machine learning and data mining which will analyse the current and historical data to make predictions about the unknown event.Predictive analytics has been classified into three types. They are

- Predictive model

  The main objective of this model is to obtain the similarity in different sample will exhibit the specific performance

- Descriptive model

  A descriptive model is used to establish the different relationship between the customer or products.

- Decision model

  A decision model is used to establish the relationship between all the elements of known data, the decision, and forecasting the result of the decision.

Performance Evaluation

The applications have the following modules like Admin, user, GSM, NPK sensor, raspberry pi. By using the NPK sensors we can get the soil results whenever it's needed. This will reduce the farmer's work.

Various soil samples taken from different places can be tested. Portable , Time consumption is low. Helps the farmer to plant the right crop,High accuracy and Speed.Agro algorithm helps farmers to select a particular crop .Laborious manual statistical analysis.

The NPK sensor used will detect only the nitrogen, potassium and phosphate present in the soil. Parameters like humidity, pressure, ph are not identified. The system has to be compared at rotational level which helps us to compare the farming system with different crop composition. Climatic conditions may change hence accurate results cannot be produced.The prediction based on atmosphere is not accurate.

10] Analysis Of Crop Yield Prediction Using Data Mining Techniques -D Ramesh , B Vishnu Vardhan

Data Mining is an emerging research field in Agriculture especially in crop yield analysis and prediction. As early into the growing season as possible, a farmer is focused on perceiving how much yield they are about to expect. As with many other sectors the amount of agriculture data is increasing on a daily source. The agrarian sector in India is facing a rigorous problem to maximise crop productivity. More than 60 percent of the crop still depends on monsoon rainfall. Recent developments in Information Technology for the agriculture field has become an interesting research area to predict the crop yield. The problem of yield prediction is a major problem that remains to be solved based on available data. Data Mining techniques are the better choices for this purpose. Different Data Mining techniques are used and evaluated in agriculture for estimating the future year's crop production. This paper presents a brief analysis of crop yield prediction using Multiple Linear Regression (MLR) technique and Density based clustering technique for the selected region i.e.East Godavari district of Andhra Pradesh in India.

Methods:

In this paper the statistical method namely Multiple Linear Regression technique and Data Mining method namely Density-based clustering technique were take up for the estimation of crop yield analysis.Multiple Linear Regression A regression model that involves more than one predictor variable is called the Multiple Regression Model. Multiple Linear Regression (MLR) is the method used to model the linear relationship between a dependent variable and one or more independent variables. The dependent

variable is sometimes termed as a predictors and independent variables are called predictors. Multiple Linear Regression (MLR) technique is based on least squares and probably the most widely used method in climatology for developing models to reconstruct climate variables from tree ring services. This crop yield prediction model is presented with the use of Multiple Linear Regression (MLR) technique where the predictor is the Production and there are seven predictors namely Year,Rainfall, Area of Sowing, Yield and Fertilisers (Nitrogen,Phosphorus and Potassium).

Density-based Clustering Technique

The primary idea of Density-based clustering techniques is that, for each point of a cluster, the neighbourhood of a given unit distance contains at least a minimum number of points.However, this idea is based on the assumption that the clusters are in the spherical or regular shapes.

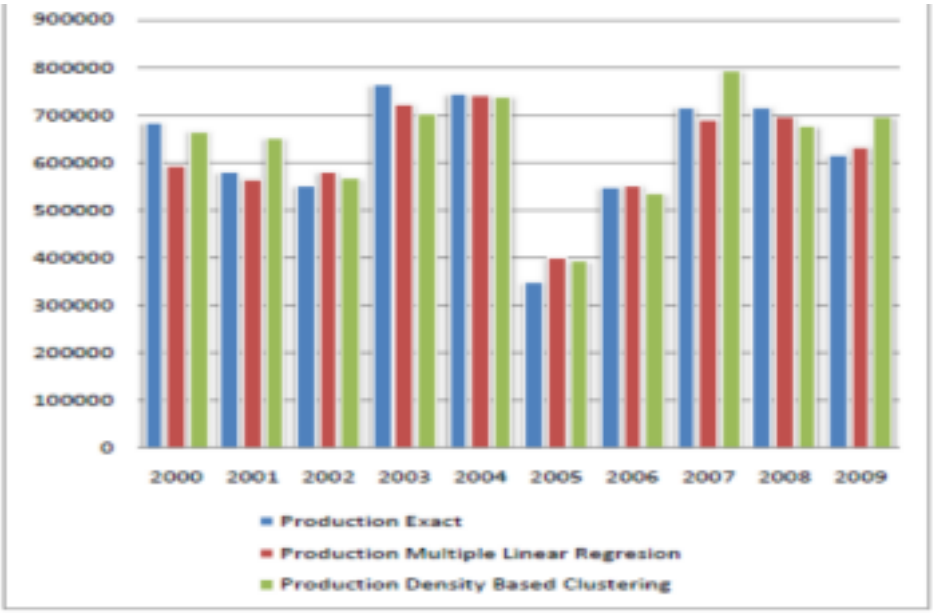


Fig-1: Comparison between Multiple Linear Regression technique and Density-based clustering technique

In this procedure the results of two methods were compared according to specific region i.e. East Godavari district of Andhra Pradesh in India. Similar process was adopted for all the districts of Andhra Pradesh to improve and authenticate the validity of yield prediction which are useful for

the farmers of Andhra Pradesh for the prediction of a specific crop.In the subsequent work a comparison of the crop yield prediction can be made with the entire set of existing available data and will be dedicated to suitable approaches for improving the efficiency of the proposed technique.

2.2 Problem Statement Definition

India is an agricultural country. Crop Yield Prediction is predicting the yield of a crop in future based on the dependent factors. Crop yield is dependent on factors like rainfall, pressure, temperature and area or the geographical location. This is achieved by

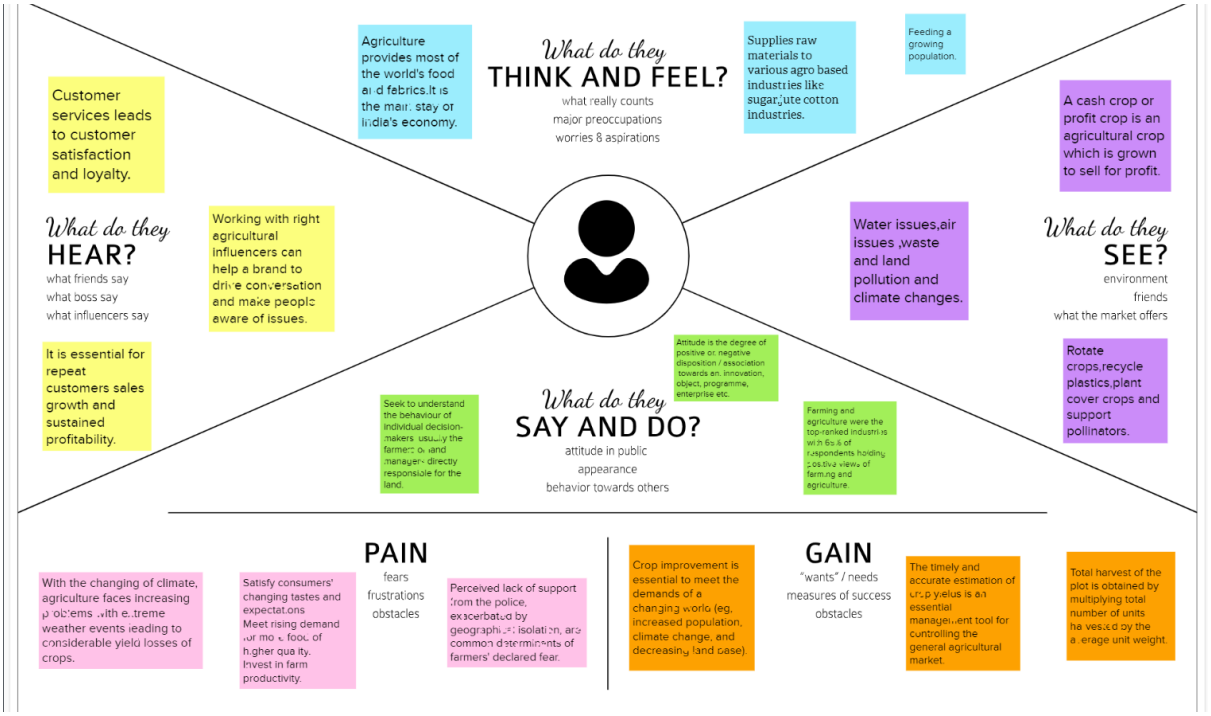
(a) Designing a system to estimate crop yield.

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

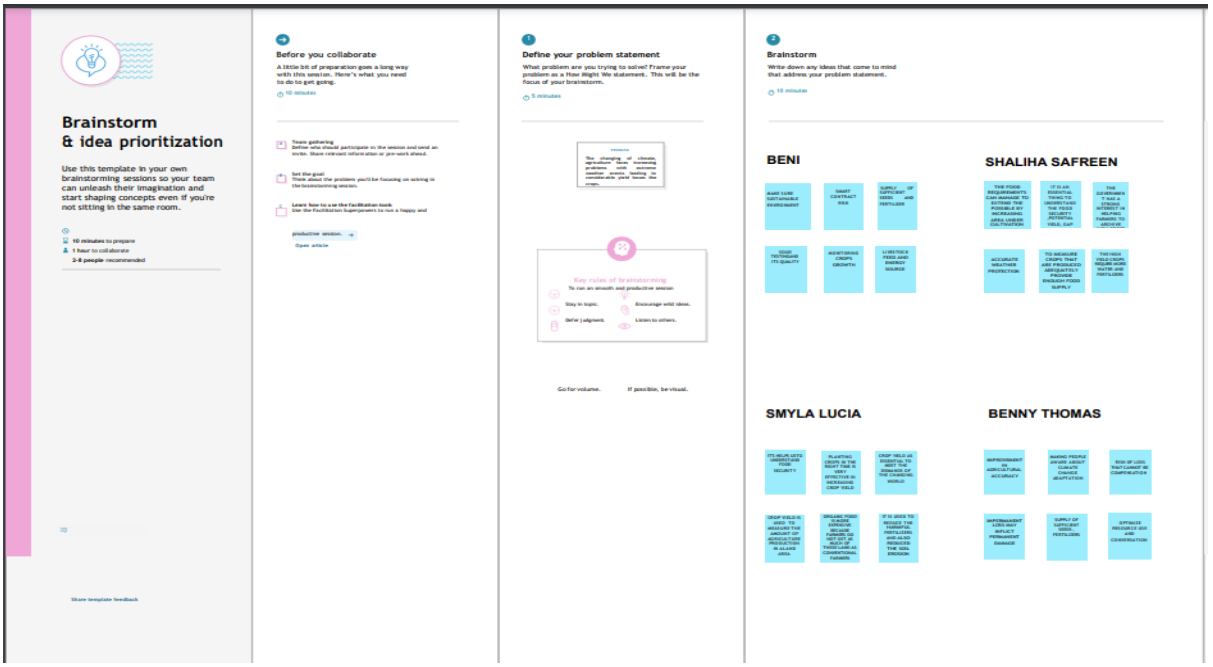
CHAPTER 3

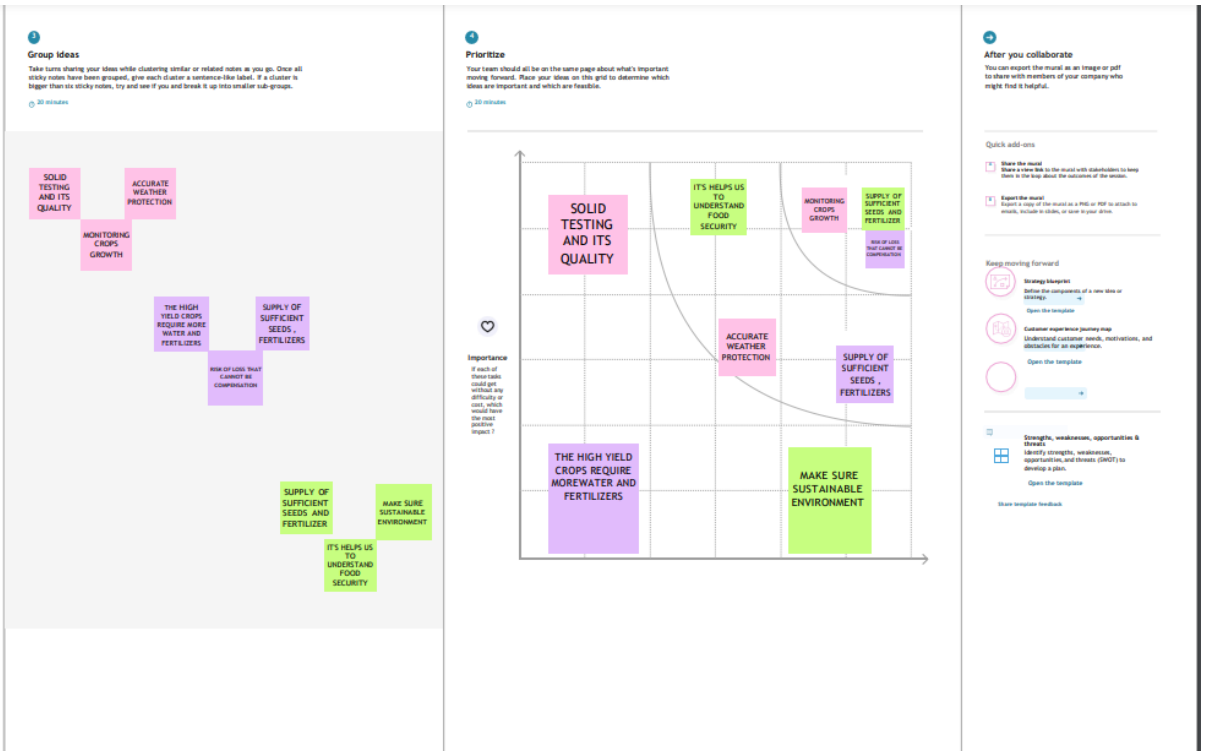
IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Brainstorming:



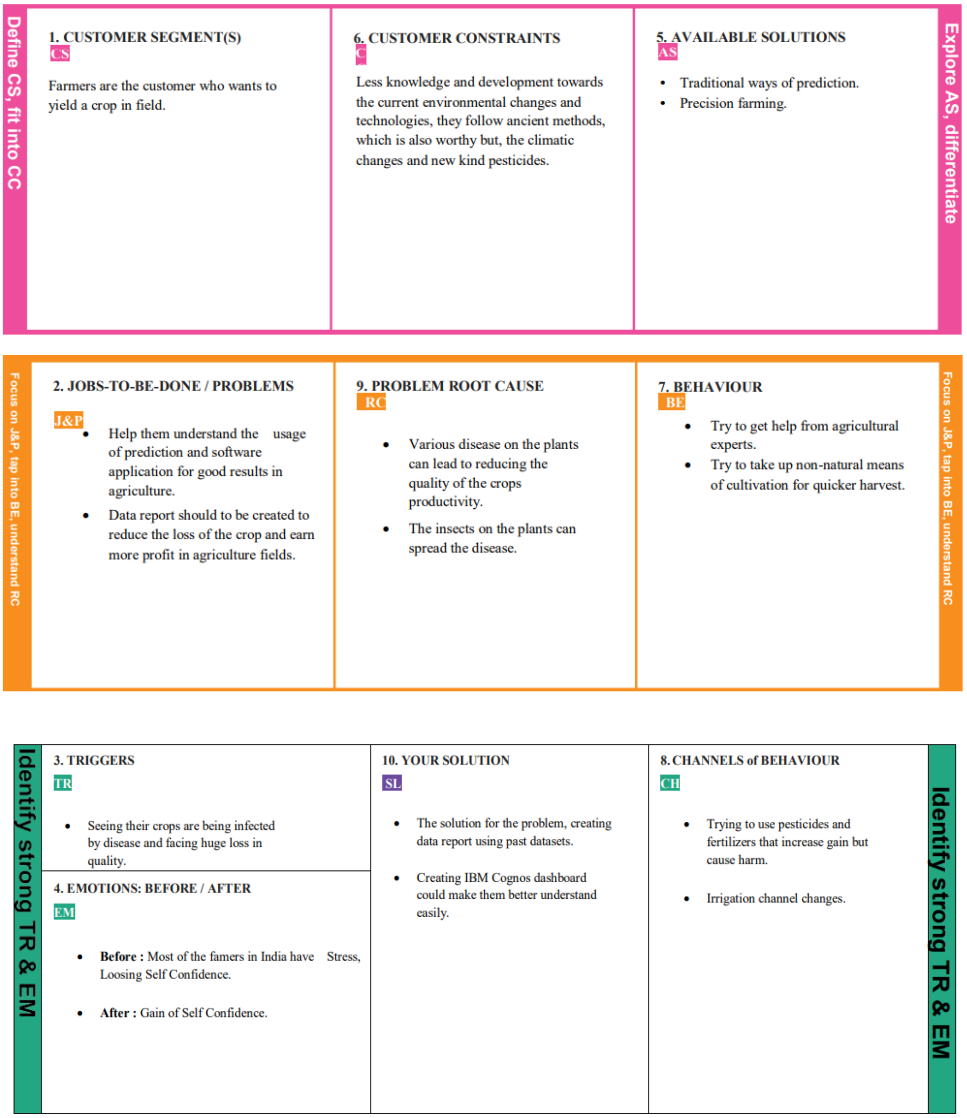


3.3 Proposed Solution

S.NO	PARAMETER	DESCRIPTION
1.	Problem Statement (Problem to be solved)	Crop yield prediction is predicting the yield of a crop in future based on the dependent factors. Crop yield is dependent on factors like weather. This is achieved by (a) Designing a system to predict yield. (b) Providing graphical user interface to view(estimate) predict results.

2.	Idea / Solution description	To predict the yield, one of the machine learning algorithms called Multiple Linear Regression algorithm is used. The result of the prediction is plotted on a graph.
3.	Novelty / Uniqueness	Using data analytics in crop yield estimation helps in analysing some important visualisation, creating a dashboard and by going through these we will get most of the insights of crop production in India.
4.	Social Impact / Customer Satisfaction	Customers(farmers) can be satisfied by increased income, reduced crop loss and high yield.
5.	Scalability of the Solution	Effectively analyze a large dataset. Easy to predict by using previous data.
6.	Business Model (Revenue Model)	Initially an raw data set was collected and subjected to analysis. From the dataset, it is subjected to feature selection to make a predictive modelling. Final representation represents the graphical result which is helpful for analysis or estimation of cross in specified rainfall.

3.4 Problem Solution Fit



4.REQUIREMENT ANALYSIS

4.1 Functional Requirement

FR No	Functional Requirement (Epic) S Requirement	(Story/sub Task)
FR-1	User login	Login through internet or app

FR-2	Login through internet or app	User can update their profile with name, mobile number and password
FR-3	Analyse the dataset	Analyse the dataset and process data preprocessing to avoid noise data.
FR-4	Choose the crop	Through which the user can choose a particular crop for their convenience.
FR-5	Predict result	The result will be predicted based on the previous year data in the way of production per hectare for particular rainfall measure in that area.
FR-6	Estimation of the result	The graphical representation showed the estimation analysis of the crop to increase more yield.

4.2 Non-Functional Requirements

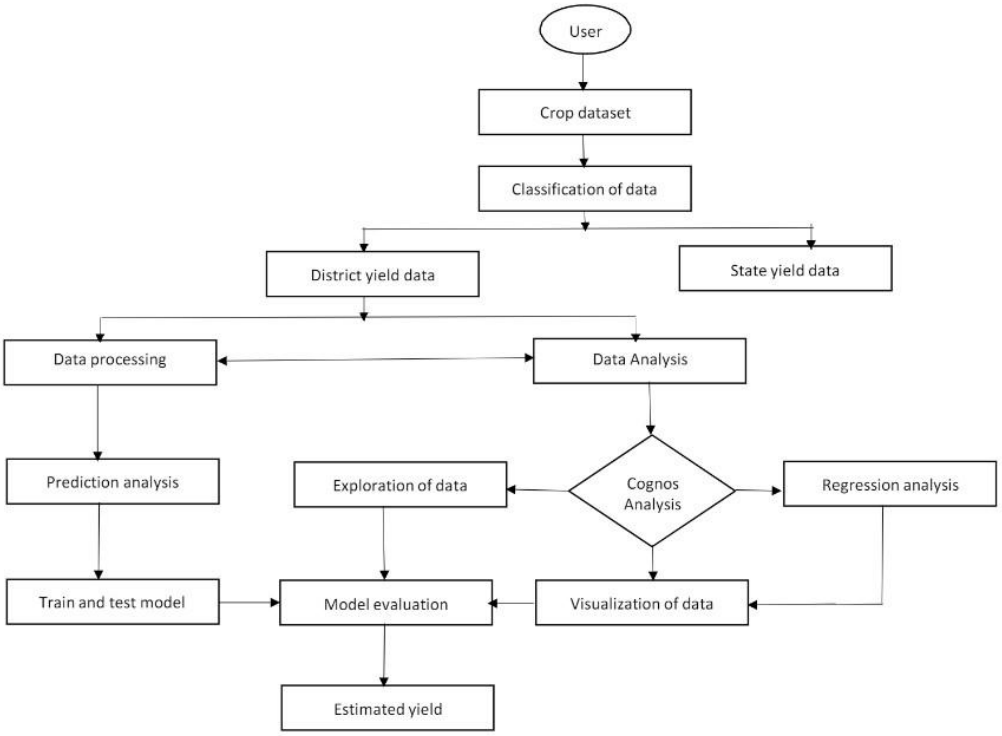
NFR No	Non FunctionalRequirement	Description
NFR-1	Usability	Contains easy user interface in order to use by uneducated people also.
NFR-2	Security	Information about the personal data is kept secure. By using data analytics, no loss or corruption of the data in the dataset. The structure of the system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

NFR-4	Performance	Performance analysis is done to find out whether the proposed system is time efficient and accurate. It is essential that the process of performance analysis and definition must be conducted in parallel.The application's load time should not be more than one second for users.
NFR-5	Availability	Users can predict and estimate the crop yield throughout the period at any time. Platforms & tools used in this project are widely used. So skilled manpower is readily available in the industry.
NFR-6	Scalability	It can also work with a large dataset without performance degradation.All changes should be in positive direction, there will be increased level of efficiency and better customer service.

5. PROJECT DESIGN

5.1 Data Flow Diagram

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



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

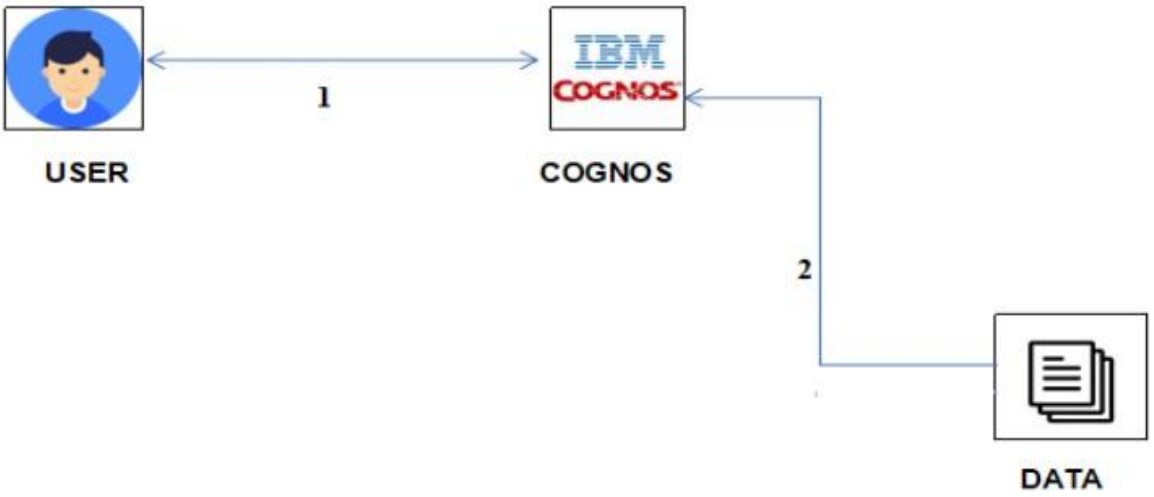
(2) Users can enter the available dataset(rainfall, area, production etc).

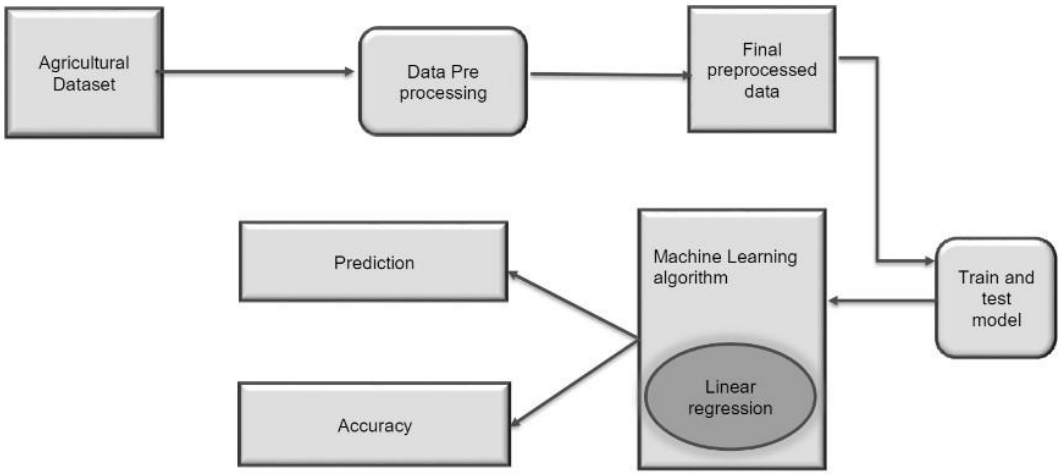
(3) The project analyze the dataset and show the prediction in graphical Format.

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

(5) Then the user can exit.

5.2 Solution & Technical Architecture





5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Prio-rity	release
Customer (Mobile user)			As a user, I can r the application by entering my email, password, and confirming my password.	I can access my account / dashboard		Sprint
			.As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	low	Sprint1
			4.a user, I can gister for the pplication through Gmail		Medium	Sprint 1
	Login		As a user ,i can log into the application by			sprint

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

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

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

S.no	Milestone	Activities	Start Date	End Date
1	Solution Requirement	Creating the IBM Cognos for creating dashboard and data visualisation charts.	22-Aug 2022	24-Aug 2022
2	Project Objectives	Prepare the project objectives.	22-Aug 2022	24-Aug 2022
3	Project Flow	Prepare the project flow.	22-Aug 2022	24-Aug 2022

4	IBM Cloud Account	Creating an IBM cloud account.	22-Aug 2022	24-Aug 2022
5	IBM Cognos Analytics	Creating IBM cognos account.	22-Aug 2022	24-Aug 2022
6	Working with the Dataset	Understanding The Dataset Loading The Dataset.	24-oct 2022	19-Nov 2022
7	Data Visualisation Charts	➢ Seasons With Average Productions ➢ With Years Usage of Area And Production ➢ Top 10 States with Most Area		
8	Creating the Dashboard	Creating The Dashboard	24-oct 2022	19-Nov 2022
9	Export the Analytics	Export The Analytics	24-oct 2022	19-Nov 2022
10	Ideation Phase	➢ Literature Survey On The Selected Project & Information Gathering Prepare ➢ Empathy Map ➢ Ideation	22-Aug 2022	17-Sept 2022

11	Project Design Phase - I	➢ Proposed Solution ➢ Problem Solution Fit ➢ Solution Architecture	22-Aug 2022	17-Sept 2022
12	Project Design Phase - II	➢ Customer Journey ➢ Functional Requirement ➢ Data Flow Diagrams ➢ Technology Architecture	22-Aug 2022	01-Oct 2022
13	Project Planning Phase	➢ Prepare Milestone & Activity List ➢ Sprint Delivery Plan	17-Oct 2022	22-Oct 2022
14	Project Development Phase	➢ Project Development - Delivery of Sprint-1 ➢ Project Development - Delivery of Sprint-2 ➢ Project Development -Delivery of Sprint-3 ➢ Project Development - Delivery of Sprint-4	24-Aug 2022	19-Nov 2022

6.2 Sprint Delivery Schedule

sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1 USN-3	As a user, I can register for by entering my Agri - id card and request.. As a user, I can register for the application through Gmail	2 2	High Medium	Beni TS Benny Thomas
	Login	USN-4	As a user, I can Call and request or Approac	2	High	Shaliha Safreen

			h for database			
	Working with the Dataset	USN-5 USN-6	To work on the given dataset, Underst and the Dataset. Load the dataset to Cloud platform then Build the required Visualisatio ns.	2 10	High High	Smyla Lucia J Beni TS
Spr	Visualisatio n Chart	USN-7	Using the Crop producti on in Indian dataset, create various graphs and charts to highlight the insights and visualisatio ns.	4 4 4	Medium Medim	Shaliha Safreen Smyla Lucia J Beni

			*Build a Visualisation to showcase Average Crop Production by Seasons *Showcase the Yearly usage of Area in Crop Production *Build a visualisation to show case top 10 States in Crop Yield Production by Area. *Build the Required Visualisation to showcase the	4	Mediu m Medium	Shaliha Safeen Benny Thomas Smyla Lucia

			Crop Production by State. *Build Visual analytics to represent the States with Seasonal Crop Production using a Text representation			
	-3 Creating ashboard	USN-8	Create the Dashboard by using the created visualisations	20	High	Shaliha Safreen
	4 Export The Analytics	USN-9	Export the created Dashboard	20	High	Beni TS Benny Thomas Shaliha Safreen Smyla Lucia J

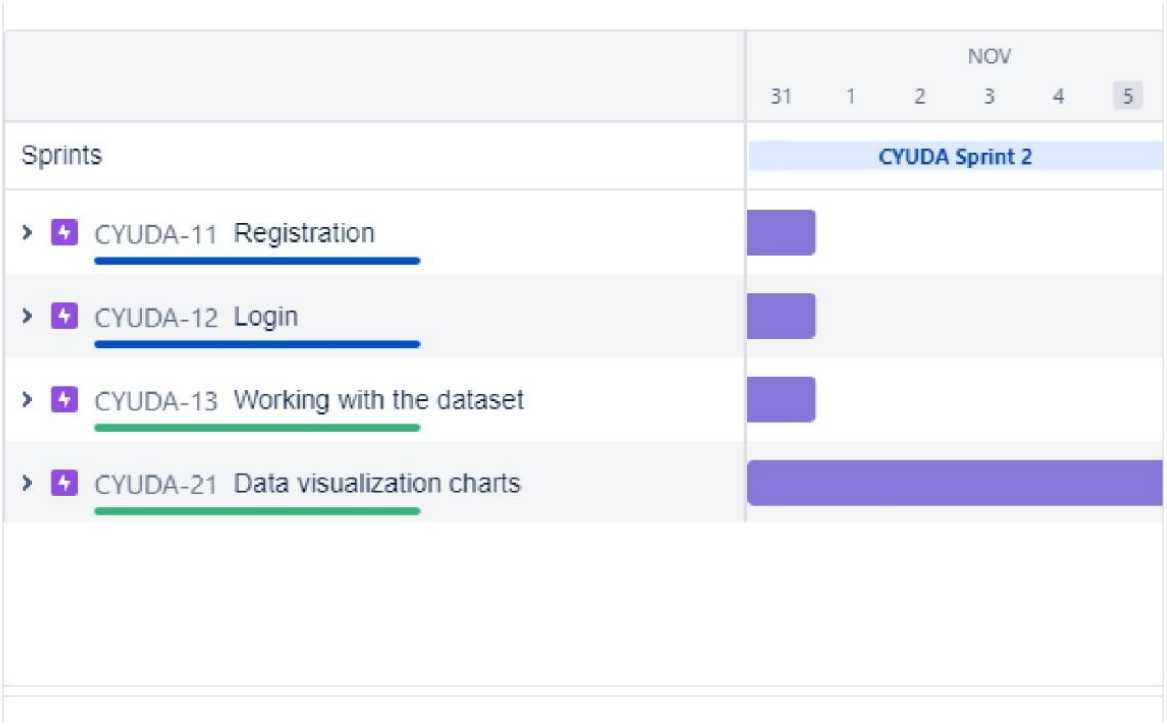
6.3 Reports from JIRA

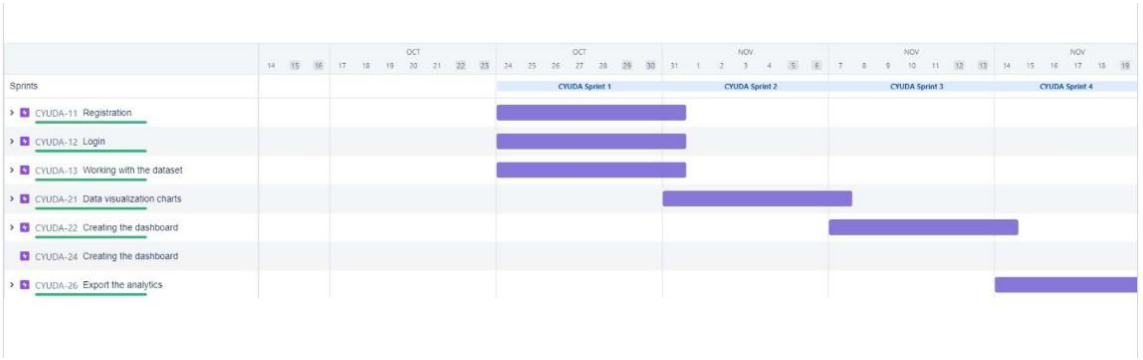
sprint	Total story points	Duration	Start Date	Sprint End Date (Planned)	Story points Completed(as on planned End Date)	Sprint Release Date(Actual)
sprint1		6 days	24oct 2022	29 oct 2022		29 oct 2022
Sprint-2 2		6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3 2		6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4 2		6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

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

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

Burndown Chart :A burndown chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burndown charts can be applied to any project containing measurable progress over time.





7. CODING AND SOLUTIONING

7.1 Feature

Login

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

Logins are used by websites, computer applications, and mobile apps. They are a security measure designed to prevent unauthorised access to confidential data. When a login fails (i.e, the username and

password combination does not match a user account), the user is disallowed access.

SAMPLE CODING

```
<!DOCTYPE html>
<html>
<head>
<meta name="viewport" content="width=device-width,
initial-scale=1"> <title> Login Page </title>
<style>
Body
  {
  font-family: Calibri, Helvetica, sans-serif;
  background-colour:light red;
  }
button
{
      background-colour:#CBC3E3;
      width: 100%;
       colour: black;
       padding: 15px;
       margin: 10px 0px;
       border: none;
       cursor: pointer;
}
form {
       border: 3px solid #f156189;
        }
input[type=text],
       input[type=password]
{
       width: 100%;
       margin: 8px 0;
```

```
        padding: 12px 20px;
        display: inline-block;
        border: 2px white;
        box-sizing: border-box;
    }
button:hover {
        opacity: 0.7;
    }
  .cancel btn {
        width: auto;
        padding: 10px 18px;
        margin: 10px 5px;
    }
.container {
        padding: 25px;


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

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

Registration

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

SAMPLE CODING

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

```
{
colour: red;
}
</style>
</head>
<body>
<?php
include 'header.php';
?>
<div class="heading fix">
<label>REGISTRATION</label>
</div>
<div class="outerbox">
<div class="fixedbox">
<span class="content">
<h4>Hello, Friend!</h4>
<p>Enter your personal details and start journey with us</p>
</span>
</div>
<div class="scrollbox">
<div class="registerdonor">
<form action="process.php" method="POST" id="myform">
<div class="login">
<h3>Login Details</h3>
<table>
<tr>
<td colspan="2">
<label class="username">User Name:-</label>
<input type="text" name="user_name" required
pattern="^[A-Za-z0-9._%+- @]{5,10}$" title="Enter a username
between 5 to 10 letter" autocomplete="off"> </td>
</tr>
<tr>
<td>
<label>Full Name:-</label>
```

```
<input type="text" name="user_full_name" required pattern="[A-z ]+$"
title="Use only character & whitespace" autocomplete="off">
</td>
<td>
<label>Email Id:-</label>
<input type="email" name="user_email" required
pattern="[A-Za-z0-9._%+- ]+@[A-z0-9.-]+\.[a-z]{2,}$" title="Email id
is not Valid" autocomplete="off"> </td>
</tr>
<tr>
<td>
<label>Password:-</label>
<input type="password"name="password" required
pattern="(?=.*\d)(?=.*[a z])(?=.*[A-Z]).{6,}" title="Must contain at
least one number and one uppercase and lowercase letter, and at
least 6 or more characters" id="password" autocomplete="off">
</td>
<td>
<label>Confirm Password:-</label>
<input type="text" name="confirm_password" required
pattern="(?=.*\d)(?=.*[a z])(?=.*[A-Z]).{6,}" title="Must contain at least
one number and one uppercase and lowercase letter, and at least 6
or more characters" id="confirm_password" autocomplete="off">
</td>
</tr>
</table>
</div>
<div class="contact">
<h3>Contact Details</h3>
<table>
<tr>
<td>
<label>Mobile Number:-</label>
<input type="text" name="user_number" required
pattern="^[1-9]{1}[0-9]{9}$" title="Number is not valid"
```

```
autocomplete="off">
</td>
<td rowspan="2">
<label>Address:-</label>
<textarea name="Address" placeholder="---Type---"
required></textarea> </td>
</tr>
<tr>
<td>
<label>Pincode</label>
<input type="text" name="pincode" required pattern="^[0-9]{6}$"
title="Pincode is not valid" autocomplete="off">
</td>
</tr>
<tr>
<td>
<label>City:-</label>
<input type="text" name="city" >
</td>
<td>
<label>State:-</label>
<input type="text" name="state">
</td>
</tr>
</table>
</div>
<div class="personal">
<h3>Personal Details</h3>
<table>
<tr>
<td>
<label>Date Of Birth:-</label>
<input type="date" name="date_of_birth" required
autocomplete="off"> </td>
```

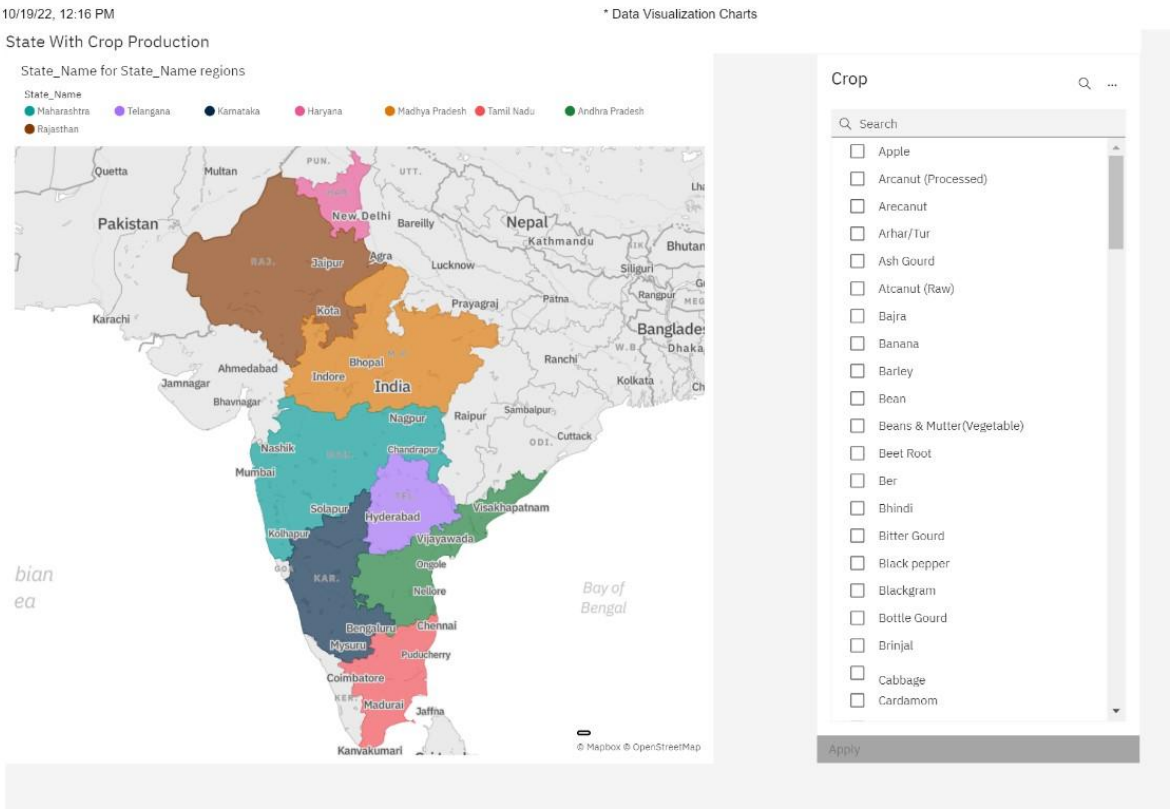
```
<td>
<label>Gender:-</label>
<div class="radio">
<input type="radio" name="gender" class="radio

value="Male"><span  class="radioname" required

autocomplete="off">Male</span> <input type="radio"

class="radio2" name="gender" value="Female"><span

class="radioname" required

autocomplete="off">Female</span> </div> </td> </tr>

<input type="reset" name="submit" value="Reset">
<a href="login.html"><input
type="button"onclick="href='login.html';"
value="Submit"></a>
</div>
</form>
</div>
</div>
</div>
<!-- Responsive Table -->
<div class="register donor">
<form action="process.php" method="POST" id="myform">
</html>
```

7.2 Feature 2

Data Visualization Chart

Data visualisation charts are graphical representations of data that tell a story using symbols in order to improve the understanding of

large amounts of data. Visual data metaphors such as charts effectively engage human perceptual processes and amplify human cognition more so than semantic data alone.



Creating the dashboard :

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

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

Step 2: Set up your workbook

Step 3: Add raw data to a table

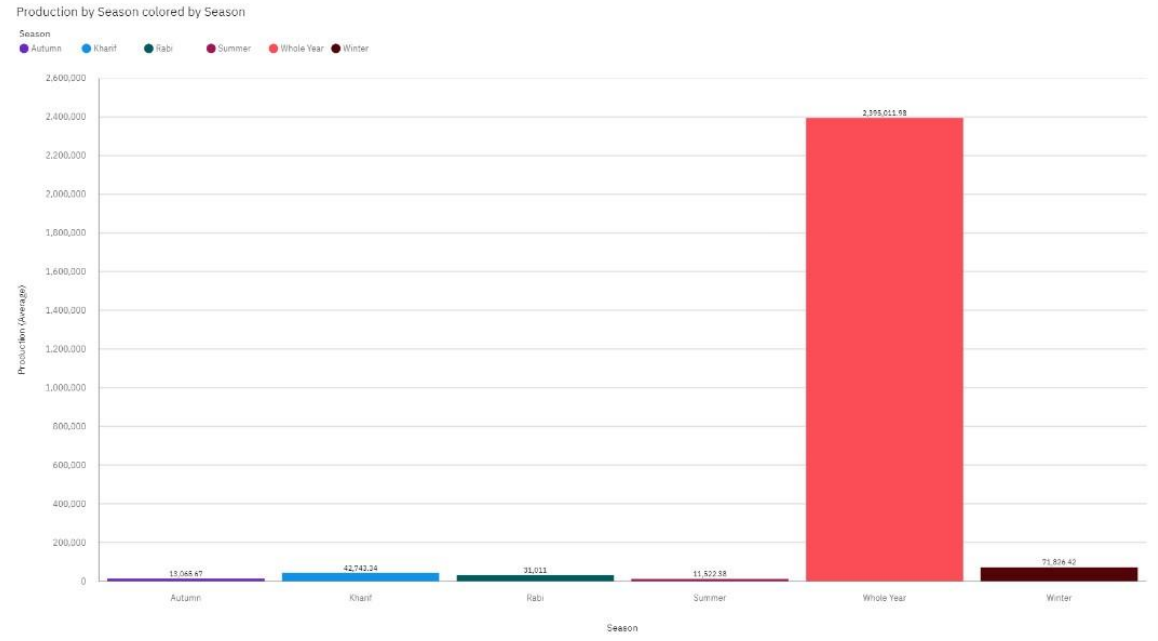
Step 4: Data analysis

Step 5: Determine the visuals

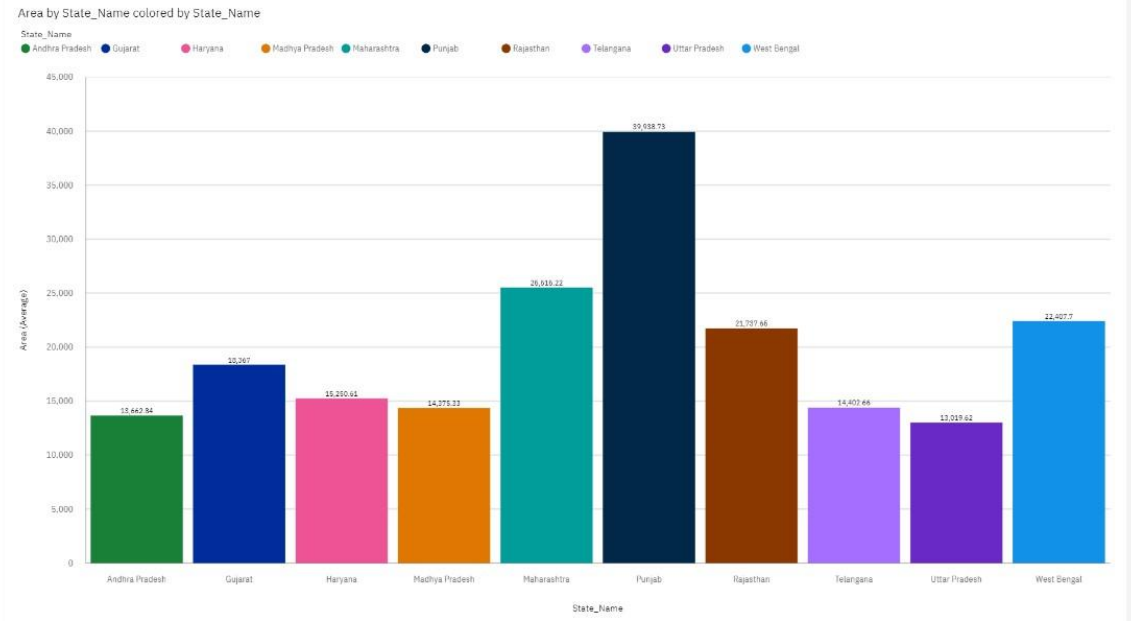
Step 6: Create your Excel dashboard

Step 7: Customise your dashboard.

Different Season with Average Production



Top 10 States With Most Area



8.TESTING

8.1 Test Cases

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

Test Case No	2
Module Tested	Registration Credentials
Input	First Name Last Name Phone Number etc.,

Actual Output	Update Profile Details
ExpectedOutput	Update Profile Details
Comments	Successful

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

8.2 User Acceptance Testing

Purpose of Document

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

Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution		**Severity 1 Severity 2 Severity 3**		**Severity 4**	**Subtotal**
By Design	7	5	3	2	17

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

Test Case Analysis

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

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	6	0	0	6
Client Application	40	0	0	40
Security	2	0	0	2
Outsource shipping	3	1	1	2
Exception Reporting	9	7	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

9. ADVANTAGES & DISADVANTAGES

Advantages

➢ Helps in the interpretation of data patterns that assist decision-making and performance improvement.

➢ Helps in analysing some important visualisation, creating a dashboard and by going through these we will get most of the insights of crop production.

➢ Helps to make the decision about when to plant and harvest crops based on rainfall, season, area.

➢ This will reduce the waste generated and improve the profit of the farmer in a digitalized way.

Disadvantages

Some uneducated farmers cannot use this application.

10. CONCLUSION

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

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

Based on our analysis, the model will be more accurate if the more datasets are available. So as the data point increases our system will become more and more accurate. Our system accuracy is more than the existing system. Since we are displaying the results in the

form of graphs with actual and predicted results in the graphical user interface it is easy to compare the previous year's data. This model will help farmers to grow the crop which will give
more yield so that it will be more profitable.

11. FUTURE SCOPE

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

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

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

12. APPENDIX

Source code

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

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


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

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

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

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

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

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

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

df.head()

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

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

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


from sk learn import metrics
%matplotlib inline
```

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

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

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

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


from sklearn.metrics import mean_absolute_error
```

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

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

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

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

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

Github & Project Demo Link

https://github.com/IBM-EPBL/IBM-Project-44480-1660724884

References

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

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

6

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

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

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

[6] Alfiya , Swaliha S, Madhumathi R.(2015)," A Survey on Crop Yield Prediction based on Agricultural Data using weather and soil ", International Journal of Innovative Research in Science, Engineering and Technology.