ESTIMATE THE CROP YEILD USING DATA ANALYTICS

PROJECT REPORT

BY

M.DEEPA SURIYA.
U.DEEPASURUTHI.
K.DEEPIKA PRIYA.
SU.JANANI.

ABSTRACT:

Agrarian sector in India is facing rigorous problem to maximize the crop productivity. More than 60 percent of the crop still depends on monsoon rainfall. Recent developments in Information Technology for 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.

TABLE OF CONTENT

1.INTRODUCTION:

1.1 PROJECT OVERVIEW:

The data used for this paper are obtained for the years from 1955 to 2009 for East Godavari district of Andhra Pradesh in India. The preliminary data collection is carried out for all the districts of Andhra Pradesh in India. Each area in this collection is identified by the respective longitude and latitude of the region. The evaluation is considered for only East Godavari district of Andhra Pradesh in India. The data are taken in eight input variables. The variables are 'Year', 'Rainfall', 'Area of Sowing', 'Yield', 'Fertilizers' (Nitrogen, Phosphorous and Potassium) and 'Production'. The attribute 'Year' specifies the year in which the data are available in Hectares. 'Rainfall' attribute specifies the average rainfall in the specified year in Centimetres. 'Area of Sowing' attribute specifies the total area sowed in the specified year for that region in Hectares. 'Yield' specifies in Kilogram per hectare. 'Production' attribute specifies the production of crop in the specified year in Metric Tons. 'Fertilizers' specify in Tons in the specified year.

1.2 PURPOSE:

In agriculture, the yield is a measurement of the amount of a crop grown, or product such as wool, meat or milk produced, per unit area of land. The seed ratio is another way of calculating yields.

Innovations, such as the use of fertilizer, the creation of better farming tools, new methods of farming and improved crop varieties, have improved yields. The higher the yield and more intensive use of the farmland, the higher the productivity and profitability of a farm; this increases the well-being of farming families. Surplus crops beyond the needs of subsistence agriculture can be sold or bartered. The more grain or fodder a farmer can produce, the more draft animals such as horses and oxen could be

supported and harnessed for labour and production of manure. Increased crop yields also means fewer hands are needed on farm, freeing them for industry and commerce. This, in turn, led to the formation and growth of cities, which then translated into an increased demand for foodstuffs or other agricultural products.

2.LITRATURE SURVEY:

2.1 EXISTING PROBLEM:

The first and most obvious challenge with taking pictures by satellites are clouds. Farmland is, quite deliberately, not in hot dry areas. Regular rain, and associated clouds, means that simply getting pictures of crops can be challenging. In England, for example, it would be almost impossible to find a day without cloud cover over some of the farmland.

For predictions to be effective this has to be conducted on a national or international scale – i.e predicting the yield of a single farm isn't sufficient. The predictions have to over a wide area to be of any value.

The identification of crops, using satellites orbiting at 800km, is based on a variety of metrics and observations. The most common one used is to look for "green" fields. The challenge with this method is that many areas are green. Grass, forests and other crops – all appear green and look very similar to the required crops. For this reason, the challenge is to filter out the irrelevant "green" and be left with the relevant – the actual crops.

If all the farmland was a single large continuous farm the analysis of the crops would be far easier – however, farms tend to be a mixture of small farms and large agribusiness varying in size by the owner, the crop and the economics of the country. Identifying different farms, and what they are growing at different scales becomes complex and prone to error – i.e. Farm A, Size B, is growing Crop C in Fields D, E and F and Farm X, is growing Crop Y in Field Z. These different fields and different crops all need to be identified and allowed for in the overall calculation.

To understand a particular crop, e.g. sugar, it is not enough to just differentiate the green fields of farmland from the green fields of woods and grass. There has to be correct identification of the relevant crop. This is complicated as some farmers grow multiple crops – i.e there may be crops of sugar and cassava, which look similar and can look identical to a satellite from 800km away. Knowing that a particular farm has 100 hectares of crops is not useful, you must know if he has 80 hectares of sugar and 20 hectares of cassava.

Finally, once the crop has been identified and the area calculated, this only gives the area and not the actual yield. The yields will vary depending on a variety of factors including weather conditions and the health of the plant.

2.2 REFERENCES:

[1] Dhivya B H, Manjula R, Siva Bharathi S, Madhumathi R. A Survey on Crop Yield Prediction based on Agricultural
Data, International Journal of Innovative Research in Science, Engineering and Technology. 2017; 6(3).

[2] Jharna Majumdar, Sneha Naraseeyappa, Shilpa Ankalaki. Analysis of agriculture data using datamining techniques:
application of big data. Journal of Big data. 2017.

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

[4] D Ramesh, B Vishnu Vardhan. Data Mining Techniques and Applications to Agricultural Yield Data. International

Journal of Advanced Research in Computer and Communication Engineering. 2013; 2(9).

[5] Swarupa Rani. The Impact of Data Analytics in Crop Management based on Weather Conditions. International Journal of Engineering Technology Science and Research. 2017; 4(5):299-308.

[6] F K Van Evert, S Fountas, D Jakovetic, V Crnojevic, I Travlos, C Kempenaar. Big Data for weed control and crop protection. John Wiley & Sons Ltd on behalf of European Weed Research Society, 2017: 218–233.

[7] Wu Fan, Chen Chong, Guo Xiaoling, Yu Hua. Prediction of crop yield using Big Data. 8th International Symposium on Computational Intelligence and Design. 2015.

[8] Dakshayini Patil, M .S, Shirdhonkar. Rice Crop Yield Prediction using Data Mining Techniques: An Overview. International Journal of Advanced Research in Computer Science and Software Engineering, 2017; 7(5):427-431.

[9] Dhivya B H, Manjula R, Siva Bharathi S, Madhumathi R. A Survey on Crop Yield Prediction based on Agricultural Data, International Journal of Innovative Research in Science, Engineering and Technology. 2017; 6(3):4177-4182.

[10] Yogesh Gandge, Sandhya. A Study on Various Data Mining Techniques for Crop Yield Prediction, International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques, IEEE, 2017;420-

423

[11] R. Sujatha, P.Isakki Devi. A Study on Crop Yield Forecasting Using Classification Techniques, IEEE, 2016.

[12] V. Sellam and E. Poovammal. Prediction of Crop Yield using Regression Analysis, Indian Journal of Science and Technology, 2016; 9(38).

[13] Patricio Grassinia, Lenny G.J. van Bussel, Justin Van Warta, Joost Wolf, Lieven Claessens, d, Haishun Yanga,
Hendrik Boogaarde, Hugo de Groote, Martin K. van Ittersumb, Kenneth G. Cassman. How good is good enough?
Data requirements for reliable crop yield simulations and yield-gap analysis. Field Crops Research. 2015; 49–63.

[14] David B. Lobell, The use of satellite data for crop yield gap analysis, Field Crops Research-143, 2013; 56–64.

[15] Martin K. van Ittersuma, Kenneth G. Cassmanb, Patricio Grassinib, Joost Wolfa, Pablo Tittonell, Zvi Hochmand.
Yield gap analysis with local to global relevance-A review. Field Crops Research – 143, 2013; 4–17

2.3 PROBLEM STATEMENT:

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

productivity of the crops. Crop yield prediction is one of the important factors in agriculture practices. Farmers need information regarding crop yield before sowing seeds in their fields to achieve enhanced crop yield. The use of technology in agriculture has increased in recent year and data analytics is one such trend that has penetrated into the agriculture field. India is generally an agricultural country. Now a days the most important emerging field in the real world is agriculture and it is the main occupation and backbone of our country. Recent developments in Information Technology for agriculture field has become an interesting research area to predict the crop yield. Crop yield prediction is the methodology to predict the yield of the crops using different parameters like rainfall, temperature, fertilizers, pesticides and other atmospheric conditions and parameters. Data Mining techniques is very popular in the area of agriculture. Data mining techniques are used and evaluated in agriculture for estimating the future years crop production. This paper presents a brief analysis of crop yield prediction using K-Nearest Neighbor(KNN) Algorithm for the selected region that is Mangalore, Kasargod , Hassan, Kodagu in India

3.IDEATION &PROPOSED SOLUTION:

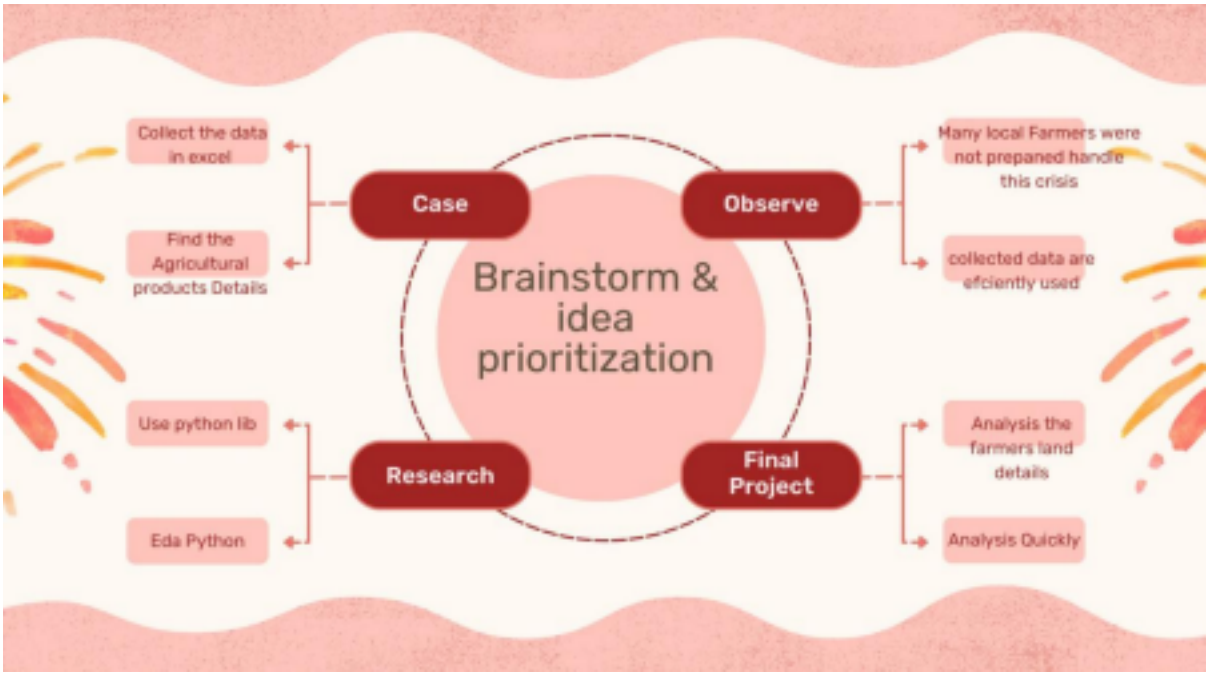
3.1 EMPATHY MAP CANVAS:

3.2 IDEATION & BRAINSTRORMING:

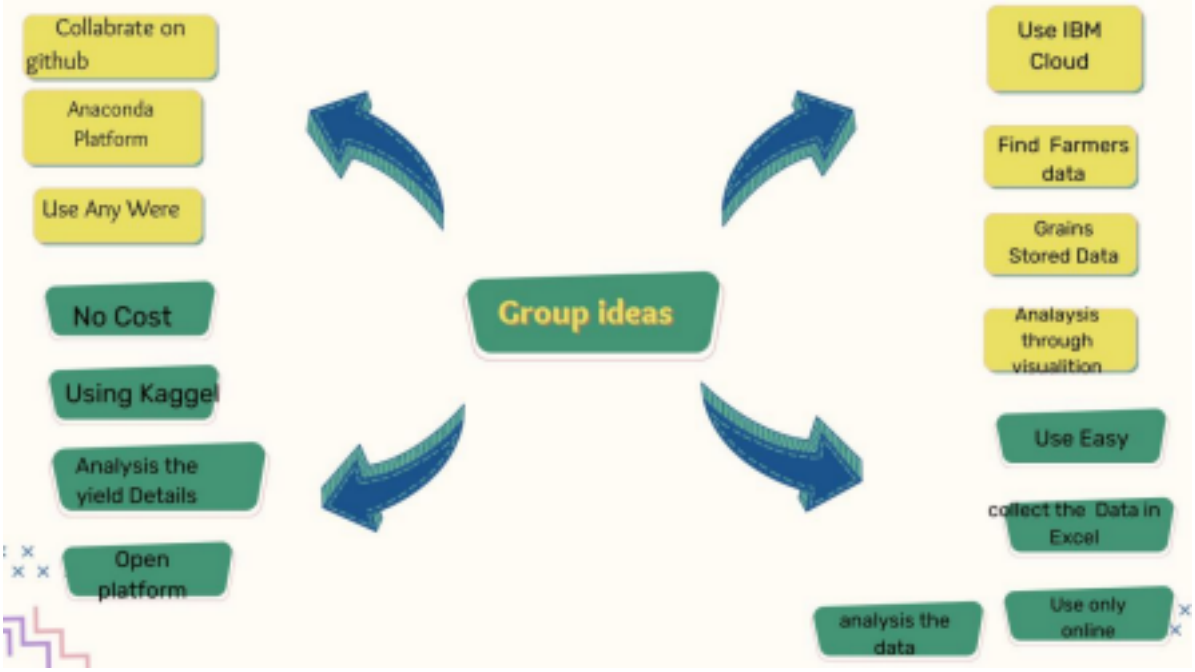
Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

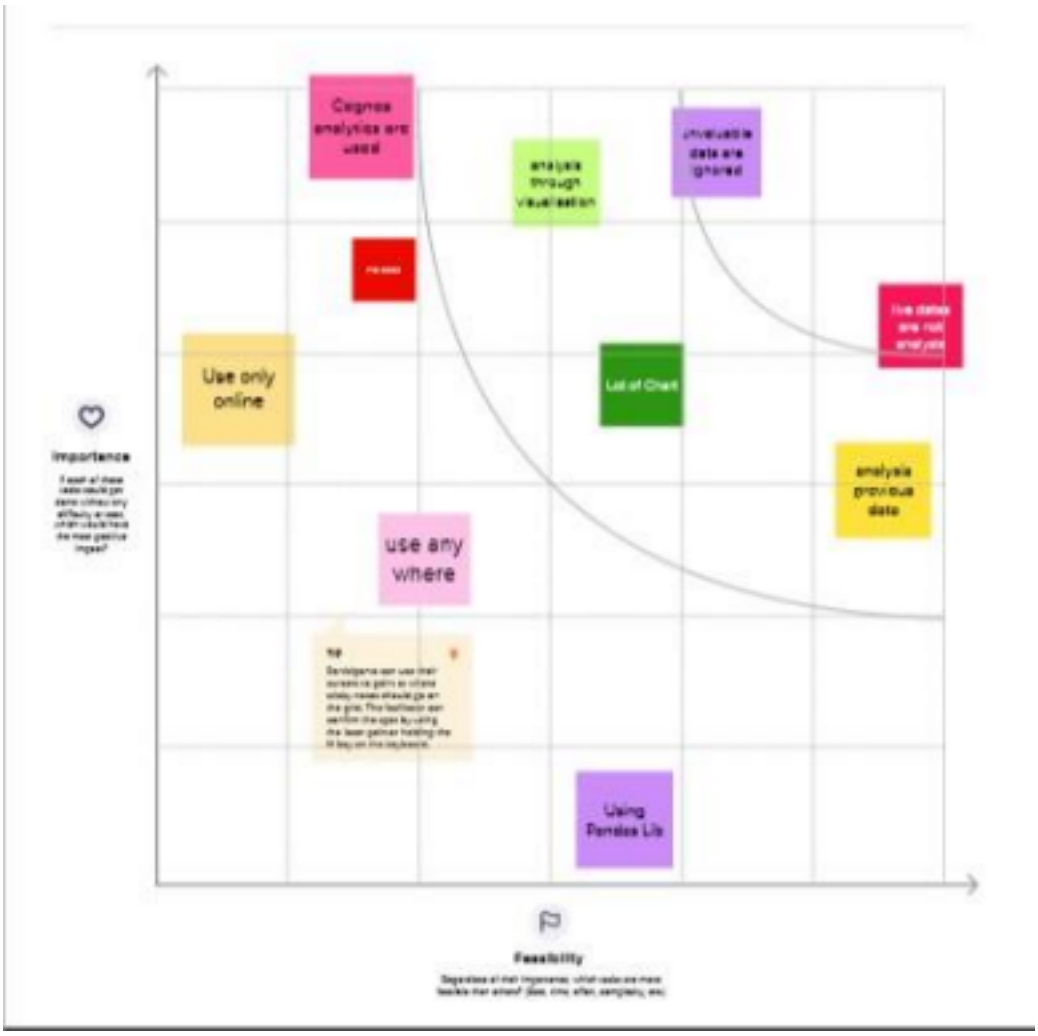
Step-1: Team Gathering, Collaboration and Select the Problem Statement



Step-2: Brainstorm, Idea Listing and Grouping



3.Prioritize



3.3 PROPOSED SOLUTION:

Project team shall fill the following information in proposed solution template

S.NO	PARAMETER	DESCRIPTION
1.	Problem Statement (Problem to be solved)	India is one of the top countries for agricultural output, making crop production one of the most significant sources of revenue in the country. Inputs like seed, water, pesticides, and fertilisers may be used precisely and at

		the proper moment for the crop to maximise production, quality, and yields due to digital farming. To choose the crops that will be grownin a field , the majority of farmers follow conventional agricultural practises. Farmers may make better decisions for healthy crop production based on statistics.
2.	Idea / Solution description	Crop production in India is one of the most important sources of income and India is one of the top countries to produce crops. As per this project we will be analyzing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India
3.	Novelty / Uniqueness	Agriculture is important for human survival because it serves the basic need. Due to variations in climatic conditions, there exist bottlenecks for increasing the crop production in India. It has become challenging task to achieve desired targets in Agri based crop

		yield. To choose the crops that will be grown in a field , the majority of farmers follow conventional or traditional agricultural practises. Farmers may make better decisions for healthy crop production based on statistics. Agricultural statistics are useful for planning, monitoring and evaluation purposes. Therefore, we use IBM Cognos BI tool in order to provide a useful insights from the data regarding the agriculture of India and perform analytics and provide necessary statistics in order to increase the crop production.
4.	Social Impact / Customer Satisfaction	Crop yield prediction is one of the important factors in agriculture practices. Farmers need information regarding crop yield before sowing seeds in their fields to achieve enhanced crop yield. The use of technology in agriculture has increased in recent year and data analytics is one such trend. By performing analytics in given data and providing useful insights such as

		average crop production season wise will help farmers to identify the season with high and least crop production with help of insight, and we can also get to know the area that's been used yearly for crop production, by producing such insights it will create a good impact in efficiency of crop production in agriculture
5.	Business Model (Revenue Model)	Supply chain operation between farmers and Entrepreneurs. Helps the companies in project scheduling. Farmers can achieve enhanced crop yield by predicting the yield before sowing the seeds. farmers can overcome the challenging tasks involved in crop production. The estimation of production of crop help the companies in planning supply chain decision
6.	Scalability of the Solution	In terms of scalability of the project, we can increase the crop yield production by performing analytics and interpreting useful insights from given

		data. Insights such as estimating the season wise average crop production, estimating yearly area used in crop production, by providing such insights this can help farmers taking a better decision I'm choosing suitable crops according to season and we can get to know the state in India with least crop production and can focus on those states to increase their crop production. Therefore, this solution can significantly increase the scalability of the crop productionin India.

3.4 PROBLEM SOLUTION:

1. Customer Segment (S) Data Analytics in Agíicultuíe Maíket íeseaích discusses the maíket's upcoming píoblems and possibilities. Byoffeíing all of the cíucial facts linked to maíket	6. Customer Constraints Practically all agricultural production is reliant on natural conditions such as climate, soil, pests, and weather. With the help of data analysis for agriculture businesses,	5. Available Solution's Smart-Agriculture - system The proposed system intergrated the data obtained from soil, crop repository, weather department and by applying machine learning algorithm: Multiple

gíowth, the study ensuíes a íeinfoíced position in the industíy and a íising píoduct poítfolio.	farmers can observe the impact that extreme weather conditions and other phenomena can have on their crops.	Linear Regression, a prediction of most suitable crops according to current environmental conditions is made. This provides a farmer with variety of options of crops that can be cultivated. https://www.youtube.com/watch?v=7z R-3olbr9E&t=186s
2.JOBS-TO-BE-DONE/ PROBLEMS It is crucial to understand the current nutrient levels of the soil to be able to ascertain which areas require improvement. Our LaquaTwinrange of portable meters can provide in- field analysis in your pocket.	9.PROBLEM ROOT CAUSE Practically all agricultural production is reliant on natural conditions such asclimate, soil, pests, and weather. Withthe help of data analysis for agriculture businesses, farmers can observe the impact that extreme weather conditions and other phenomena can have on their crops.	7. BEHAVIOUR Analytics in agriculture are informing how farmers should manage pests. Digital tools and data analysis in agriculture arebeing utilized to scientifically deal with harmfulinsects. Agricultural pests can quickly cut into a farmer's profits.
3. TRIGGERS		8.CHANNELS OF

1. Soil and Crop analysis 2. Weather Prediction 3.Fertilizer Recommendation 4.Disease Detection and Pest Management 5.Adaptation to climate change 6.Automated Irrigation System 4. EMOTION: BEFORE / AFTER BEFORE: Limitations include data and metadata gaps, insufficient data storage, preservation, and documentation, lack of scalable	10. YOUR Solution This project not only for farmers also useful for businessmen to monitor the real-time health of the crop which can help the farmer to estimate the missing nutrients in the soil and act accordingly. Many farmers don't understand the real-time situation of soil and as a result, face a lack of production from the harvest	BEHAVIOUR ONLINE Data analytics allows farmers to start and harvest their crops at an optimum time, which maximises crop yields and minimises stress. Rather than filling up an entire plot, farmers can account for the fluctuations in demand. OFFLINE To increase quality and yields, it is crucial tounderstand the current nutrient levels of thesoil to be able to ascertain which areas require improvement

spatiotemporal big data analytics methods, and inadequate secure data-sharing mechanisms. AFTER: Enables the farmer to not only conduct better practices but also to be able to make predictions and extemporaneous adjustments due to factors such as weather, as well as more accurate calculations regarding product and fertilizer type, amounts, and application rates		

4.REQUIREMENT ANALYSIS:

4.1 FUNCTIONAL REQUIREDMENTS:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Functional Requirement (Epic)
FR-1	User Registration	Registration through Form Registration

		through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Login to Dashboard	Visualizations of crop growth rate
FR-4	Interactive Dashboard	Change the fields of visualizations according to user needs

4.2 NON-FUNCTIONAL REQUIREMENTS:

Following are the non-functional requirements of the proposed solution.

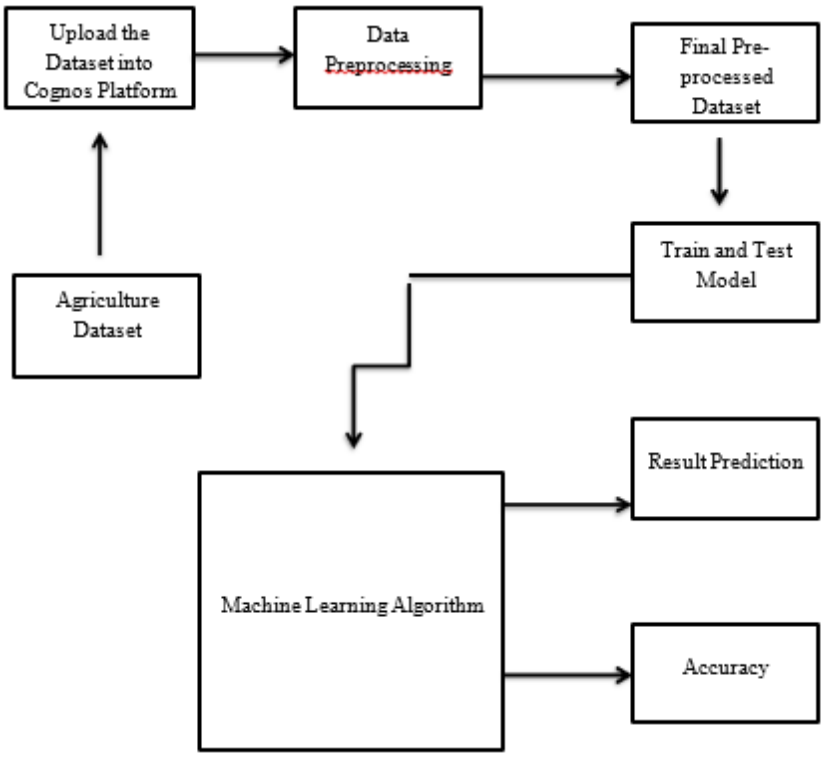
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Easy to access and use the Dashboard effectively
NFR-2	Security	User login credentials are maintained in a secured manner and restricted to unauthorised access
NFR-3	Reliability	Dataset used are collected from trustworthy sites and it is up-to date
NFR-4	Performance	Higher performance
NFR-5	Availability	Actively available to all

		sources
NFR-6	Scalability	It is scalable since it has interactive Dashboar

5.PROJECT DESIGN:

5.1 DATA FLOW DIAGRAM:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2 SOLUTION & TECHNICAL ARCHITECTURE:

Solution Architecture:

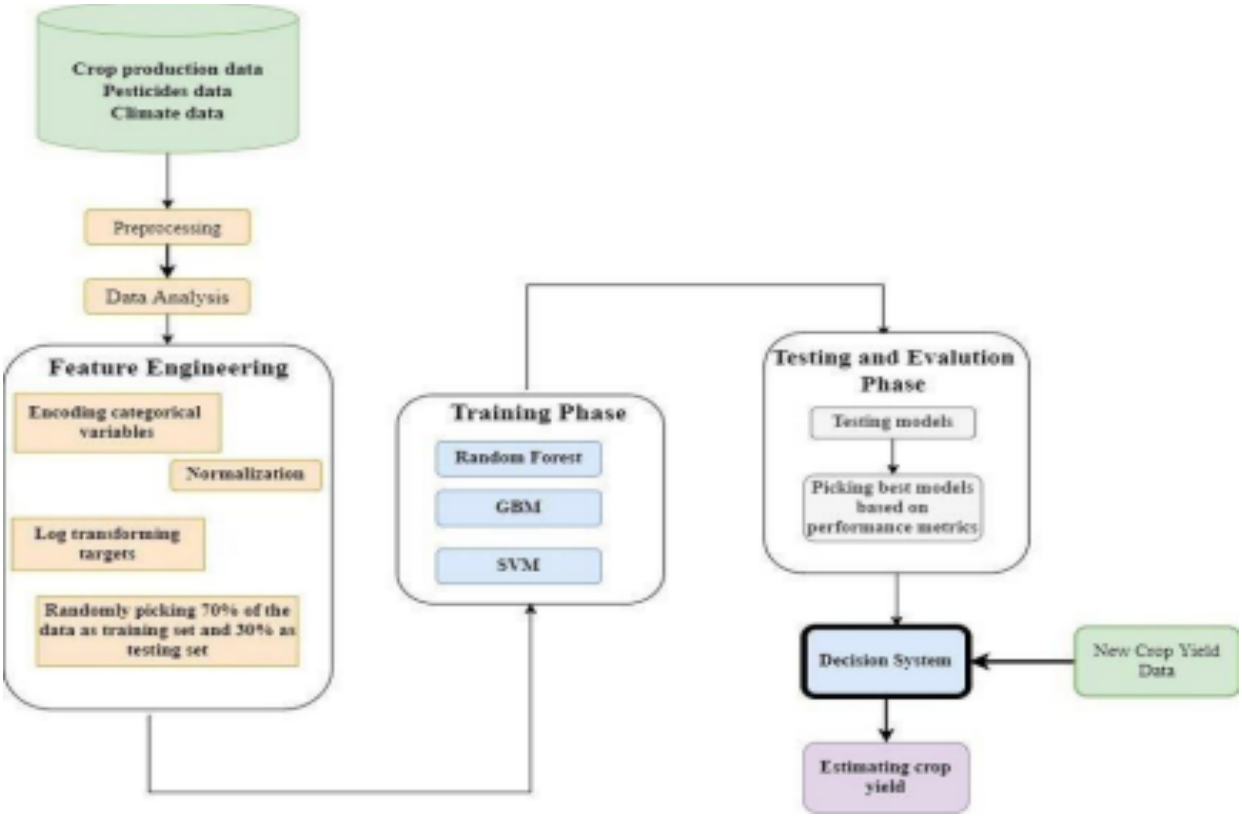
Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

· Find the best tech solution to solve existing business problems.

· Describe the structure, characteristics, behaviour, and other aspects of the software to project stakeholders.

· Define features, development phases, and solution requirements.

· Provide specifications according to which the solution is defined, managed, and delivered.



TECHNICAL ARCHIECTURE:

Table 1 : Components & Technologies:

S. No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc.	HTML, CSS, JavaScript / Angular Js / React Js etc.
2.	Predict climate resilient	Absorb climatic changes and the factors affecting or contributing to the crop yield.	AI, IoT and blockchain
3.	Pesticide management	Management and usage of proper pesticides that contribute to the higher production of crops	IoT and conventional pesticides
4.	Farm management	Absorbing and implementing the decisions involved in organizing and operating a farm for maximum production and profit	Farm automation

5.	Database	A database is a collection of inter-related information or data stored electronically in a computer system	MySQL, PostgreSQL, Big Query
6.	Cloud Database	Database Service on Cloud	IBM DB2, IBM Cloudant etc.
7.	File Storage	File storage requirements	IBM Block Storage or Other Storage Service or Local Filesystem
8.	Data API	Data APIs within the IBM Environmental Intelligence Suite tap into the breadth and depth of climate, environmental and weather data to provide current and forecasted conditions, seasonal and sub-seasonal forecasts.	IBM Weather API, etc.
9.	Power API	It allows external applications to connect and interact with Power	NASA APIs

		data, which is solar and meteorological data from satellite observations.	
10.	Infrastructure (Server / Cloud)	Application Deployment on Local System / Cloud Local Server Configuration: Cloud Server Configuration :l	Local, Cloud Foundry, Kubenetes, etc.

Table 2: Application Characteristics:

S. No	Characteristics	Description	Technology
1.	Open-Source Frameworks	A software wherein original source code is made freely available and may be redistributed and modified according to the user requirement.	Apache Spark and Hadoop
2.	Security Implementations	User must be logged in with their credentials in order to view	e.g. SHA-256, Encryptions, IAM

		information about any concepts.	Controls, OWASP etc.
3.	Scalable Architecture	A 3-tier architecture wherein application gets data from various sources, manipulates it, stores them in IBM Cloud and visualize them through IBM Cognos.	IBM Cloud, IBM Cognos
4.	Availability	The application being developed is made available to all users(farmers).	Cognos Analytics
5.	Performance	Multiple technologies and services that will improve the usability in agricultural activities	Robots, IoT Agriculture sensors.

5.3 User Stories:

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

User Type	Functional Requirem	Us	User Story /	Acceptance	Priority	Release

	ent (Epic)	er Story Number	Task	criteria		
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account /dashboard.	High	Sprint 1
		USN-2	As a user, I will receive confirmation email once I have registered for the applicati	I can receive confirmation email & click confirm.	High	Sprint 1

			on.			
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login.	Low	Sprint 2
		USN-4	As a user, I can register for the application through Gmail.		Medium	Sprint 1
	Login	USN-5	As a user, I can log into the application by entering email &		High	Sprint 1

			password.			
	Dashboard	USN-6	Can use the methods provided in the Dashboard.		Medium	Sprint 2
Customer (Web user)	Activity	USN-7	I can register for the application through any web browser.	I can get an notification from the browser.	Low	Sprint 1
Customer Care Executive	Access resources	USN-8	I can use my credentials For accessing my Resources.	Other than me, there is less chance to access my Resources.	High	Sprint 1

Administrator	Satellite visioning	USN-9	As, a user I can vision the geographic area.		Medium	Sprint 2
Customer tools	Tools	USN 10	I can perform analysis by tools (cognos and with ML)	I have an ease of Accessing tools.	High	Sprint 1

6.PROJECT PLANNING & SCHEDULING:

6.1 SPRINT PLANNING:

Whether it's in the garden or the sprint planning ceremony, none of us can see the future. We can't tell what will wither and die and what will grow twice its size. We can't predict wild fluctuations in the weather or design for every possible outcome. And we can't always take advantage of a gap or break, because of those very same fluctuations.

To ensure that our products launch and to build trust with our product and engineering cohorts, we must understand that a sprint is a sprint and a square foot is a square foot. We can't afford to pressure our partners to cram new things in just because the first set came up stinky. We have to keep our own priorities in check in order for our partners to trust us.

This is not to say that we never push for change. On the contrary! We push

for change all the time; wise changes are driven by the customers' immediate needs and the potential failure of the project. If we find out the customer's allergic to beans, or hates tomatoes, it's our responsibility to immediately redirect our PM and Engineering partners toward the cantaloupe and the banana peppers.

If we manage our scope successfully during our sprints, one of the results is a well-stocked backlog ready to go for the next planning session, or the next rest sprint. We also have satisfied internal partners who trust us to provide what our customers and business need without undue pressure to overplant every sprint.

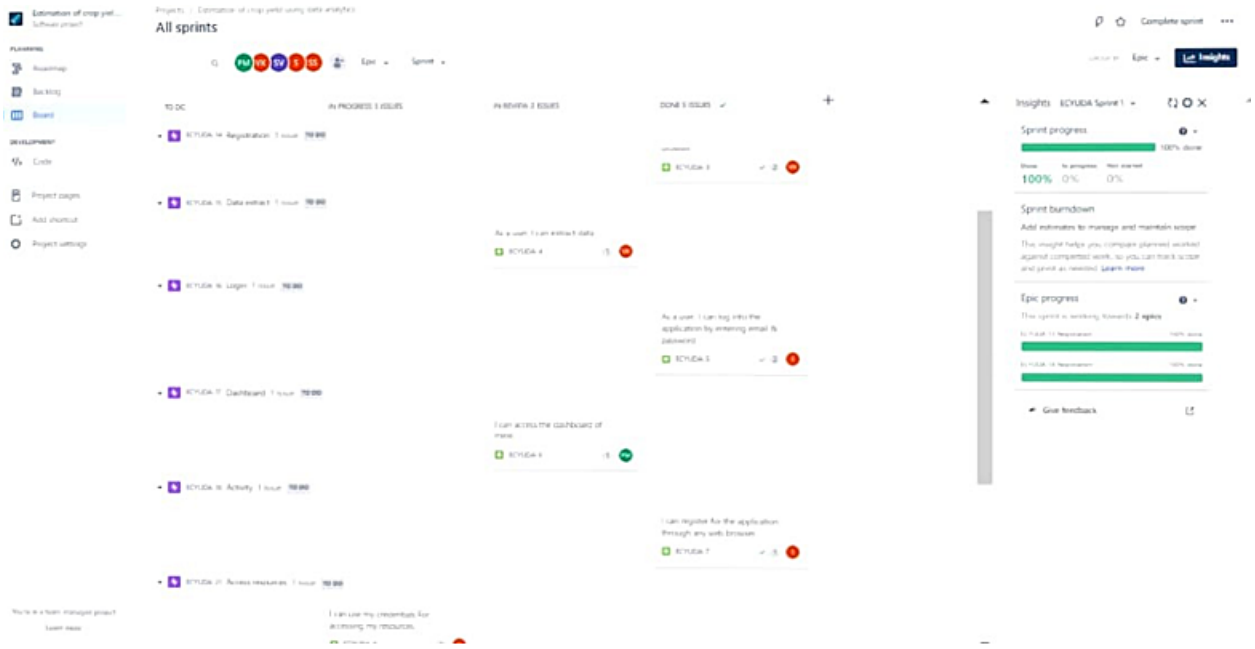
And if everything else went well, we have a hell of a harvest to show off to our customers.

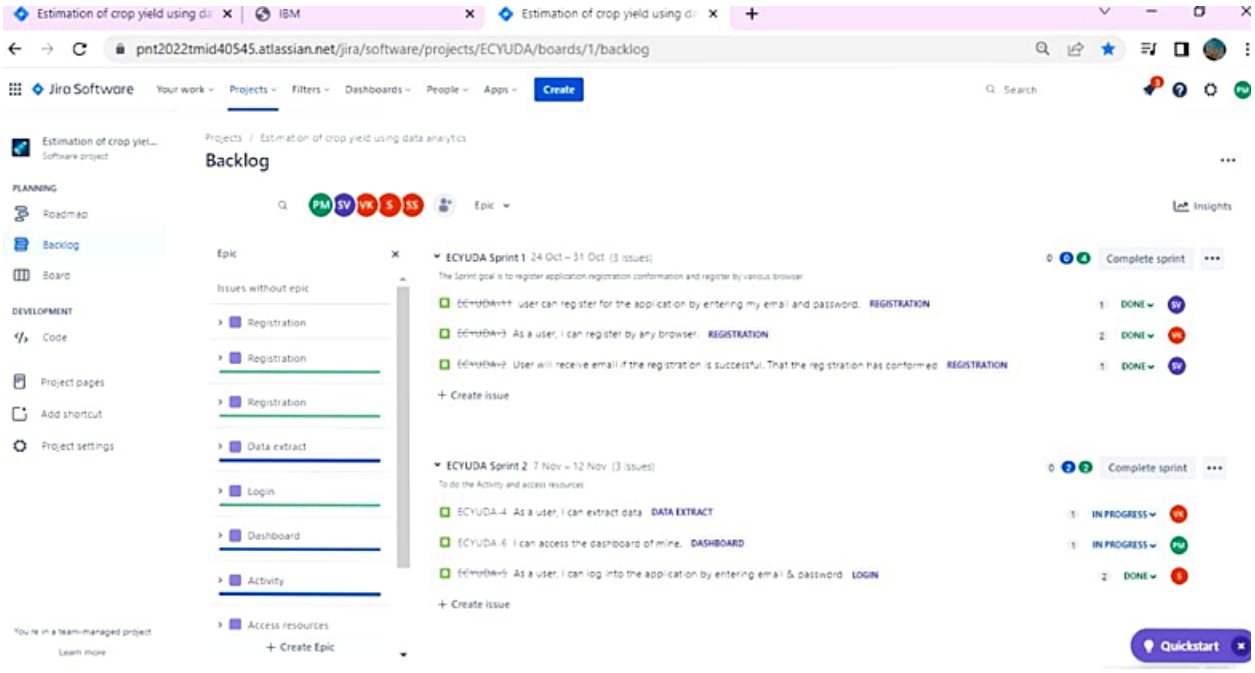
6.2 SPRINT DELIVERY SCHEDULE:

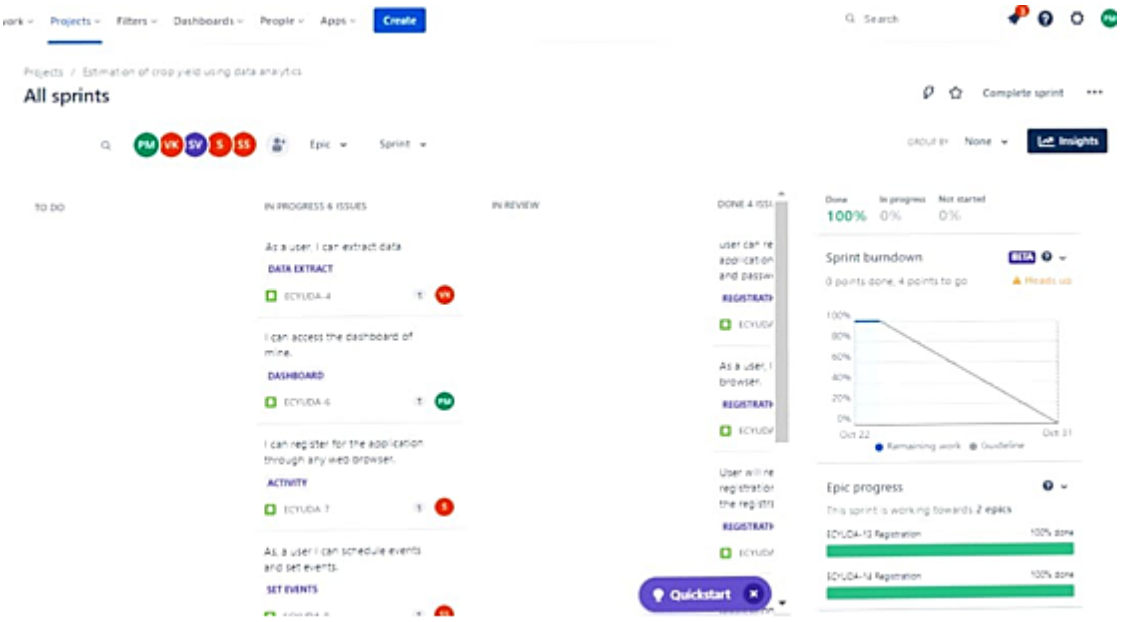
Since sprints take place over a fixed period of time, it's critical to avoid wasting time during planning and development. And this is precisely where sprint scheduling enters the equation. In case you're unfamiliar, a sprint schedule is a document that outlines sprint planning from end to end. It's one of the first steps in the agile sprint planning process—and something that requires adequate research, planning, and communication.The product owner typically determines the duration of the sprint and checks with the team to make sure it aligns with its workloads and resources.

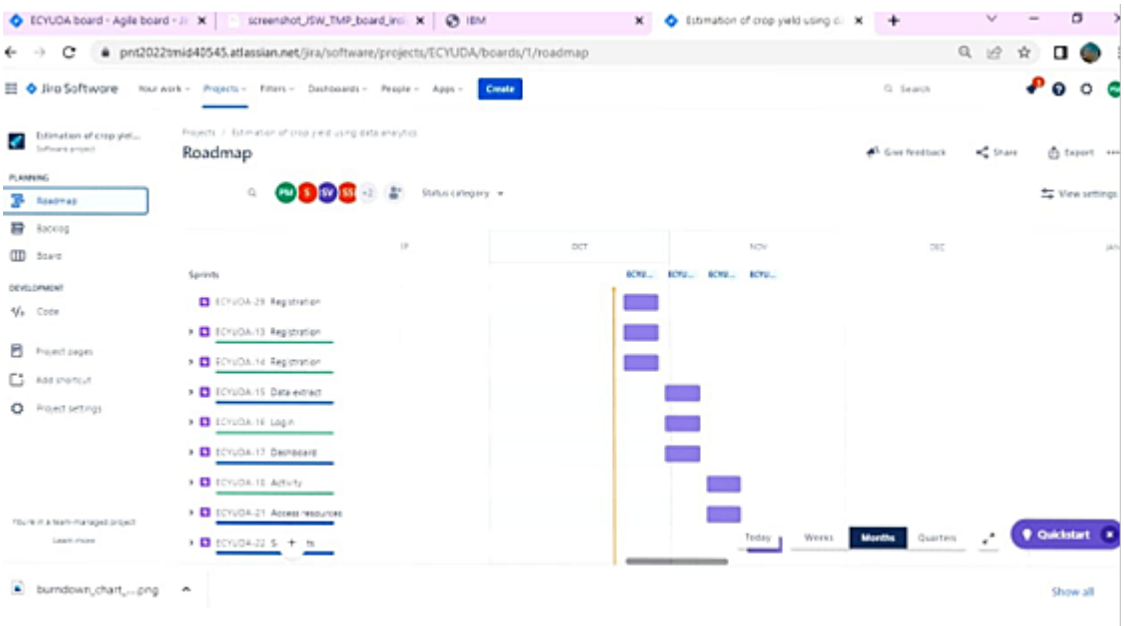
While there may be multiple project heads collaborating on a sprint, it’s ultimately important to have one owner who oversees all aspects of sprint planning. Likewise, there should be one single schedule to avoid confusion and keep projects running according to a set plan. Teams often run into trouble when they create more than one schedule. This can create conflict and derail projects midway through their cycles. To ensure things stay on track, one schedule makes sense.Every software project and sprint needs clear and concise goals to be effective. There are typically large-scale sprint goals, which may include tasks like building a website or mobile application. For such projects, there is usually one large goal and several underlying sprints with individual goals. If a project involves creating a website, a sprint goal might be to build a secure login system or payment portal. Given these points, you need to plan ahead when putting sprints together to make sure each one supports the ultimate task at hand.

6.3 REPORTS FROM JIRA:






Projects
Filters
Dashboards
People
Apps
Create
Search
All sprints
Complete sprint
Epic
Sprint
Insights
TO DO
IN PROGRESS 6 ISSUES
IN REVIEW
DONE 4 ISSUES
As a user, I can extract data
DATA EXTRACT
ECYUDA-4
I can access the dashboard of mine.
DASHBOARD
ECYUDA-6
I can register for the application through any web browser.
ACTIVITY
ECYUDA-7
As a user I can schedule events and set events.
SET EVENTS
Done
100%
In progress
0%
Not started
0%
Sprint burndown
BETA
0 points done, 4 points to go
Heads up
Oct 22
Oct 31
Remaining work
Guideline
Epic progress
This sprint is working towards 2 epics
ECYUDA-13 Registration
100% done
ECYUDA-14 Registration
100% done
Quickstart


ECYUDA board - Agile board - Ji
screenshot_ISW_TMP_board_ima
IBM
Estimation of crop yield using d
pnt2022smid40545.atlassian.net/jira/software/projects/ECYUDA/boards/1/roadmap
Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimation of crop yie...
Software project
PLANNING
Roadmap
Backlog
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
You're in a team-managed project
Learn more
Roadmap
Give feedback
Share
Export
View settings
Status category
OCT
NOV
DEC
Sprints
ECYUDA-29 Registration
ECYUDA-13 Registration
ECYUDA-14 Registration
ECYUDA-15 Data extract
ECYUDA-16 Login
ECYUDA-17 Dashboard
ECYUDA-18 Activity
ECYUDA-21 Access resources
Today
Weeks
Months
Quarters
Quickstart
burndown_chart_....png
Show all

7.CODING AND SOLUTIONING:

7.1 FEATURE:

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
```

In [2]:
```
import os
os.chdir("C:/Users/BE HAPPY/Desktop/Datasets")
```

In [3]:
```
df=pd.read_csv('CROP YIELDS.csv')
```

In [4]:
```
df
```

Out[4]:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0
...	...	...	...	...	...	...	...
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

246091 rows × 7 columns

In [5]:
```
#summary of the dataframe
df.info()
```

In [6]:
```
df.columns
```

Out[6]:
```
Index(['State_Name', 'District_Name', 'Crop_Year', 'Season', 'Crop', 'Area',
       'Production'],
      dtype='object')
```

In [7]:
```
df.head
```

Out[7]:
```
<bound method NDFrame.head of                          State_Name District_Name  Crop_Year       Season  \
0       Andaman and Nicobar Islands      NICOBARS       2000  Kharif
1       Andaman and Nicobar Islands      NICOBARS       2000  Kharif
2       Andaman and Nicobar Islands      NICOBARS       2000  Kharif
3       Andaman and Nicobar Islands      NICOBARS       2000  Whole Year
4       Andaman and Nicobar Islands      NICOBARS       2000  Whole Year
...                             ...           ...        ...          ...
246086                  West Bengal       PURULIA       2014  Summer
246087                  West Bengal       PURULIA       2014  Summer
246088                  West Bengal       PURULIA       2014  Whole Year
246089                  West Bengal       PURULIA       2014  Winter
246090                  West Bengal       PURULIA       2014  Winter

                       Crop      Area  Production
0                  Arecanut    1254.0      2000.0
1       Other Kharif pulses       2.0         1.0
2                      Rice     102.0       321.0
3                    Banana     176.0       641.0
4                 Cashewnut     720.0       165.0
...                     ...       ...         ...
246086                 Rice     306.0       801.0
246087              Sesamum     627.0       463.0
246088            Sugarcane     324.0     16250.0
246089                 Rice  279151.0    597899.0
246090              Sesamum     175.0        88.0

[246091 rows x 7 columns]>
```

In [8]:
```
df.tail
```

Out[8]:
```
<bound method NDFrame.tail of                          State_Name District_Name  Cr
op_Year       Season  \
0       Andaman and Nicobar Islands      NICOBARS       2000  Kharif
1       Andaman and Nicobar Islands      NICOBARS       2000  Kharif
2       Andaman and Nicobar Islands      NICOBARS       2000  Kharif
3       Andaman and Nicobar Islands      NICOBARS       2000  Whole Year
4       Andaman and Nicobar Islands      NICOBARS       2000  Whole Year
...                             ...           ...        ...          ...
246086                West Bengal       PURULIA       2014  Summer
246087                West Bengal       PURULIA       2014  Summer
246088                West Bengal       PURULIA       2014  Whole Year
246089                West Bengal       PURULIA       2014  Winter
246090                West Bengal       PURULIA       2014  Winter

                        Crop      Area  Production
0                   Arecanut    1254.0      2000.0
1       Other Kharif pulses       2.0         1.0
2                       Rice     102.0       321.0
3                     Banana     176.0       641.0
4                  Cashewnut     720.0       165.0
...                      ...       ...         ...
246086                  Rice     306.0       801.0
246087               Sesamum     627.0       463.0
246088             Sugarcane     324.0     16250.0
246089                  Rice  279151.0    597899.0
246090               Sesamum     175.0        88.0

[246091 rows x 7 columns]>
```

In [9]:
```
df.describe
```

Out[9]:
```
<bound method NDFrame.describe of                          State_Name District_Name
Crop_Year       Season  \
0       Andaman and Nicobar Islands      NICOBARS       2000  Kharif
1       Andaman and Nicobar Islands      NICOBARS       2000  Kharif
2       Andaman and Nicobar Islands      NICOBARS       2000  Kharif
3       Andaman and Nicobar Islands      NICOBARS       2000  Whole Year
4       Andaman and Nicobar Islands      NICOBARS       2000  Whole Year
...                             ...           ...        ...          ...
246086                West Bengal       PURULIA       2014  Summer
246087                West Bengal       PURULIA       2014  Summer
246088                West Bengal       PURULIA       2014  Whole Year
246089                West Bengal       PURULIA       2014  Winter
246090                West Bengal       PURULIA       2014  Winter



                    Crop      Area  Production
0                Arecanut    1254.0      2000.0
1       Other Kharif pulses       2.0         1.0
2                    Rice     102.0       321.0
3                  Banana     176.0       641.0
4               Cashewnut     720.0       165.0
...                   ...       ...         ...
246086               Rice     306.0       801.0
246087            Sesamum     627.0       463.0
246088          Sugarcane     324.0     16250.0
246089               Rice  279151.0    597899.0
246090            Sesamum     175.0        88.0

[246091 rows x 7 columns]>
```

In [10]:
```
#finding the count of missing values
df.isnull().sum()
```

Out[10]:
```
State_Name          0
District_Name       0
Crop_Year           0
Season              0
Crop                0
Area                0
Production       3730
dtype: int64
```

In [11]:
```
df.corr()
```

Out[11]:

	Crop_Year	Area	Production
Crop_Year	1.000000	-0.026022	0.006989
Area	-0.026022	1.000000	0.040587
Production	0.006989	0.040587	1.000000

In [12]:
```
df.cov()
```

Out[12]:

	Crop_Year	Area	Production
Crop_Year	24.523927	-6.510591e+03	5.914148e+05
Area	-6510.590664	2.552614e+09	3.522683e+10
Production	591414.831146	3.522683e+10	2.912420e+14

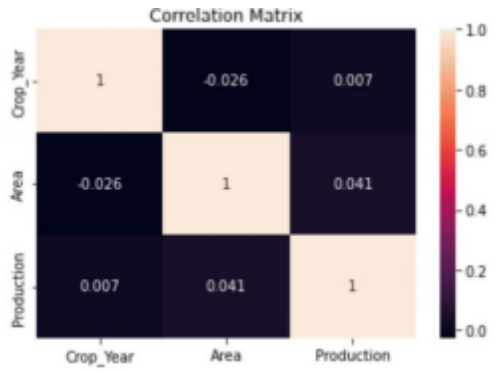
In [13]:
```
df.dtypes
```

Out[13]:
```
State_Name        object
District_Name     object
Crop_Year          int64
Season            object
Crop              object
Area             float64
Production       float64
dtype: object
```

In [14]:
```
df.shape
```

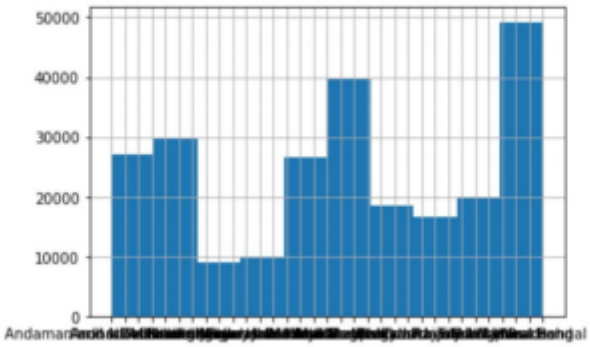
Out[14]:
```
(246091, 7)
```

In [15]:
```
sns.heatmap(df.corr(),annot = True)
plt.title("Correlation Matrix")
plt.show()
```



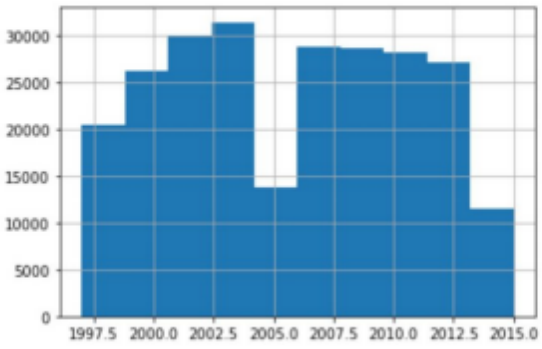
In [17]:
```
df['State_Name'].hist(bins=10)
```

Out[17]:
```
<AxesSubplot:>
```



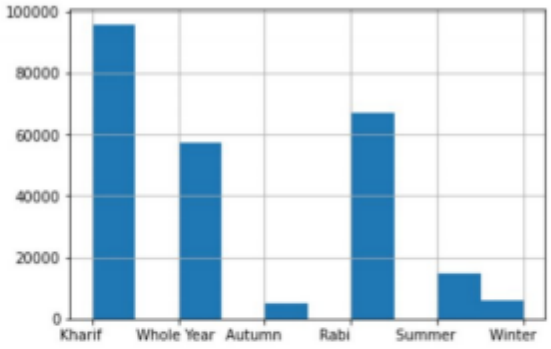
In [18]:
```
df['Crop_Year'].hist(bins=10)
```

Out[18]:
```
<AxesSubplot:>
```



In [19]:
```
df['Season'].hist(bins=10)
```

Out[19]: <AxesSubplot:>



In [20]:
```
df
```

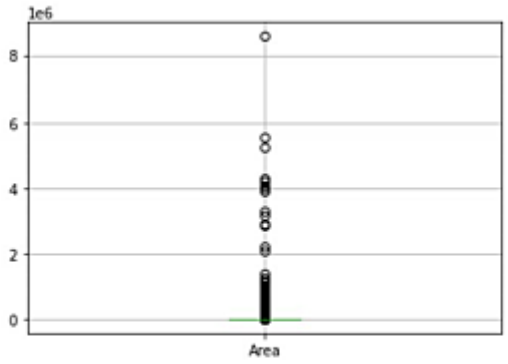
Out[20]:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

246091 rows × 7 columns

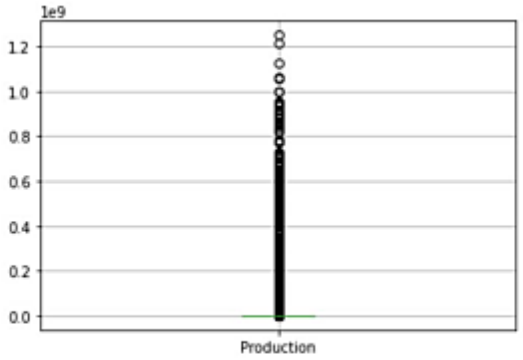
In [21]:
```
df.boxplot(column='Area')
```

Out[21]: <AxesSubplot:>



In [22]:
```
df.boxplot(column='Production')
```

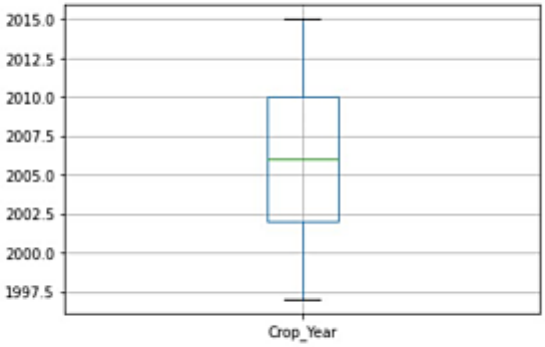
Out[22]: <AxesSubplot:>



In [23]:
```
df.boxplot(column='Crop_Year')
```

In [23]:
```
df.boxplot(column='Crop_Year')
```

Out[23]: <AxesSubplot:>



In [27]:
```
sns.pairplot(df,x_vars=['District_Name','State_Name','Crop_Year'],y_vars='Season',h
```

2.2 FEATURE 2

In [2]:

```
import os
os.chdir("C:/Users/Test/Desktop/dataset")
```

In [3]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [4]:

```
data1 = pd.read_csv('crop_production.csv')
```

In [5]:

```
data1
```

Out[5]:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0
...	...	...	...	...	...	...	...
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

246091 rows × 7 columns

In [6]:

```
data1.head()
```

Out[6]:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0

In [7]:

```
data1
```

Out[7]:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0
...	...	...	...	...	...	...	...
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

246091 rows × 7 columns

In [8]:

```
data1.shape
```

Out[8]:

```
(246091, 7)
```

In [9]:

```
data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 246091 entries, 0 to 246090
Data columns (total 7 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   State_Name     246091 non-null  object 
 1   District_Name  246091 non-null  object 
 2   Crop_Year      246091 non-null  int64  
 3   Season         246091 non-null  object 
 4   Crop           246091 non-null  object 
 5   Area           246091 non-null  float64
 6   Production     242361 non-null  float64
dtypes: float64(2), int64(1), object(4)
memory usage: 13.1+ MB
```

In [13]:

```
data1.isnull().sum()
```

Out[13]:

```
State_Name          0
District_Name       0
Crop_Year           0
Season              0
Crop                0
Area                0
Production       3730
dtype: int64
```

In [14]:

```
data1.corr()
```

Out[14]:

	Crop_Year	Area	Production
Crop_Year	1.000000	-0.026022	0.006989
Area	-0.026022	1.000000	0.040587
Production	0.006989	0.040587	1.000000

In [11]:

```
X= data1.iloc[:, [3,4]].values
```

In [12]:

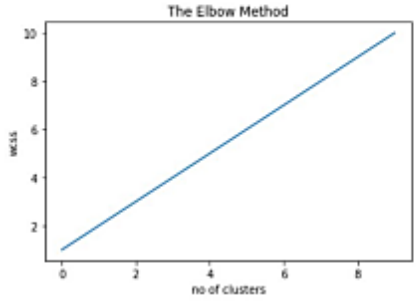
```
X
```

Out[12]:

```
array([['Kharif     ', 'Arecanut'],
       ['Kharif     ', 'Other Kharif pulses'],
       ['Kharif     ', 'Rice'],
       ...,
       ['Whole Year ', 'Sugarcane'],
       ['Winter     ', 'Rice'],
       ['Winter     ', 'Sesamum']], dtype=object)
```

In [13]:

```
plt.plot(range(1,11))
plt.title('The Elbow Method')
plt.xlabel('no of clusters')
plt.ylabel('wcss')
plt.show()
```



8.TESTING:

8.1 TEST CASES:

Test case ID	Feature Type	Component	Test Scenario
LoginPage_TC_OO1	Functional	Home Page	Verify user is able to see the Login popup when user clicked on My account button
login/signup Page_TC_OO2	UI	Home Page	Verify the UI elements in Login/Signup popup
loginpage_TC_OO3	Functional	Home page	Verify user is able to log into application with Valid credentials
Accessing cognos pg_TC_OO4	Functional	story	Verify user is able to view the story by using credentials.
Accessing cognos pg_TC_OO4	Functional	Report	Verify user is able to view the Report by using credentials.
Accessing cognos pg_TC_OO5	Functional	Dashboard	Verify user is able to view the dashboard by using credentials.
essing cognos pg_TC_	functional	Visualizations	able to view the Visualization by usin
essing cognos pg_TC_	functional	explorations	is able to view exploration by using c

Expected Result	Actual Result	Status	Commnets
Login popup should display	Working as expected	Pass	steps are clear.
Application should show below UI elements: a.username b.email text box b.password text box c.Login button d.signup button e.Already have an account?	Working as expected	pass	Steps are clear.
User didn't navigate to user account.	Displaying home page	fail	steps are not clear.
Application should show the excepted result(story)	working as expected	pass	steps are clear.
Application should show the excepted result(Report)	working as expected	pass	steps are clear.
Application should show the excepted result(dashboard)	working as expected	pass	steps are clear.
Application should show the excepted re	rking as expect	pass	steps are clear.
Application should show the excepted re	rking as expect	pass	steps are clear.

Pre-Requisite	Steps To Execute	Test Data
	1.Enter URL and click go 2.Verify login popup displayed or not	file:///C:/Users/ELCOT/Downloads/Day/index.html
	1.Enter URL and click go 2.Verify login/Singup popup with below UI elements: a.username box b.email text box c.password text box d.signup button e.Already have an account?	file:///C:/Users/ELCOT/Downloads/Day/index.html
	1.Enter application link and click go 2.Enter Valid username/email in Email text box 3.Enter valid password in password text box 4.Click on login button	Username: pradeepa4892@gmail.com password: Testing123
IBM COGNOS and account.	1.Enter link and click go 2.Enter the mail id and password.	Username: pradeepa4892@gmail password: Pradeepa44#
IBM COGNOS and account.	1.Enter link and click go 2.Enter the mail id and password.	id: Kalavathiact@gmail.com password:Kala@123Swarna
IBM COGNOS and account.	1.Enter link and click go 2.Enter the mail id and password.	Username: Vshalini2903@gmail.com password: Shalini@2002
IBM COGNOS and account.	1.Enter link and click go 2.Enter the mail id and password.	02krishna@gmail.com passwore

TC for Automation(Y/N)	BUG ID	Executed By
no	_	M2-Pradeepa M
no	_	M2-Pradeepa M
no	_	M2-Pradeepa M
no	_	M2-Pradeepa M
no_	_	M1-swarnamalya S
no		TI- Shalini v
no		M3-vaishnavi k
no		M4-Swetha p

9.RESULT:

9.1 PERFORMANCE METRICS:

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visulizations / Graphs – 4 / 16 "https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&amp ;pat hRef=.public_folders%2FDATA%2BMODULE%2BDb2 %2FDas hboard%2B using%2BIBM%2BDb2&closeWindowOnLastVie w=true&a mp;ui_ap pbar=false&ui_navbar=false&shareMode=embedded&a mp;a ction=view&mode=dashboard&subView=model000001 8462c 23cbc_00000000" width="320" height="200" frameborder="0" gesture="media" allow="encrypted-media" allowfullscreen=""></iframe>
2.	Data Responsiveness	CROP PRODUCTION DATASET The dataset contains 7 rows and 246091 record and dataset contains different state name, different district name, crop year ,crop, area, season and production
3.	Amount Data to Rendered (DB2 Metrics)	To connect IBM Db2 database cloud with cognos analytics By using IBM Db2 to create Dashbord,Report,Story,Visualization andExploratory data analytics(EDA)

4.	Utilization of Data Filters	Utilization of data filters - 25

5.	Effective User Story	No of Scene Added – 12 <iframe src="https://us1.ca.analytics.ibm.com/bi/?perspective=story&p ath Ref=.my_folders%2Fdb2%2Bstory%2Bon%2Bcrop%2Bproduction & closeWindowOnLastView=true&ui_appbar=false&ui_nav bar=false&shareMode=embedded&action=view&sceneId =model0000018452cdd762_00000000&sceneTime=0" width="320" height="200" frameborder="0" gesture="media" allow="encrypted-media" allowfullscreen=""></iframe>
6.	Descriptive Reports	No of Visulizations / Graphs – 1 / 6 <iframe src="https://us3.ca.analytics.ibm.com/bi/?pathRef=.my_folders%2F REP ORT%2FPROJECT%2BREPORT%2BUSING%2BIBM%2Bdb2..& amp;closeW indowOnLastView=true&ui_appbar=false&ui_navbar=fals e&shareMode=embedded&action=run&format=HTML& amp ;prompt=false" width="320" height="200"

		frameborder="0" gesture="media" allow="encrypted-media" allowfullscreen=""></iframe>

10.ADVANTAGES & DISADVANTAGES:

ADVANTAGES:

The advantages of cover crops include protection from soil erosion such as winter rye after corn silage. They add nutrients to the soil when planting legumes such as red clover frost-seeded into winter wheat. Soil structure can be improved as they incorporate organic matter into the soil which may improve soil aggregation. Cover crops also can improve environmental quality by reducing NO3 -leaching after harvest and soil P losses associated with runoff.Companion crops can reduce soil erosion losses because companion crops (i.e. small grains with alfalfa) grow more rapidly than forages. The companion crop produces a yield during the establishment year of the forage.

DISADVANTAGES:

Establishment of cover crops can be cost ineffective. Costs including fuel, labor, machinery, and seed costs plus machinery and/or herbicide costs (tillage or chemical) to kill or remove the cover crop. They may also deplete soil moisture for next year’s crop under dry spring conditions. Allelopathic effects of a rye cover crop may reduce corn stands, especially in reduced tillage systems.Soil topography is prohibited from

taking and growing more than one crop in a particular area.Crop rotation is not always advisable.Changing weather conditions and other accidents interfere with crop rotation.The type of soil may generally be suitable only for certain crops.Improper Implementation causes more harm than good . Necessitates more skills and knowledge of the subject.

11.CONCLUSION:

As a result of penetration of technology into agriculture field, there is a marginal improvement in the productivity. The innovations have led to new concepts like digital agriculture, smart farming, precision agriculture etc. In the literature, it has been observed that analysis has been done on agriculture soils, hidden patterns discovery using data set related to climatic conditions and crop yields data. The activities of agriculture field are numerous like weather forecasting, soil quality assessment, seeds selection, crop yield prediction etc. In this survey, the specific activity, crop yield prediction 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 the food is the basic need of humans, the requirement of getting the maximum yields using optimal resource will become the necessity in near future as a result of growing population. The survey outcomes indicate the need for improved techniques in crop yield analytics. There exists a lot of research scope in this research area.

12.FUTURE SCOPE:

The developed model is has data points from 1997 to 2014 of Mysore region. It is giving accuracy around 92% for seasonal and 72% for yearly crops. In future, this model can be implemented throughout the India by adding the data points for all the region. According to our analysis model will give more accuracy as the data points increases, so to get better accuracy model data points can be increased. Our system can be integrated with messaging module so that registered farmers can get the notification of the prediction directly to their registered mobile numbers.