**A PROJECT REPORT ON**

# ESTIMATE THE CROP YIELD USING DATA ANALYTICS

## DOMAIN: DATA ANALYTICS

DONE BY

**TEAM ID: PNT2022TMID54302**



**Team Size :** 4

**Team Leader :** Aishwarya M

**Team member :** Amruthashreya G

**Team member :** Gopika J

**Team member :** Atshaya R

# TABLE OF CONTENTS

# 1.ABSTRACT

One of the most important occupations for human survival is Agriculture and especially the majority of the population in India is into this. Due to variations in climatic conditions and various other challenges, the Agrarian sector in India is facing rigorous problems to maximize crop productivity. It has become an arduous task to achieve the desired targets in crop yield. Various factors like monsoon rains, rodents, water scarcity are to be considered, which do have a direct impact on the production and productivity of the crops. And thus, crop yield prediction with recent advancements in technology can be one of the important factors in agriculture practices as the use of technology in agriculture has comparatively increased in the recent years. The present study gives insights on data analytics methods and various machine learning techniques applied to find the crop yield prediction

# 2.INTRODUCTION

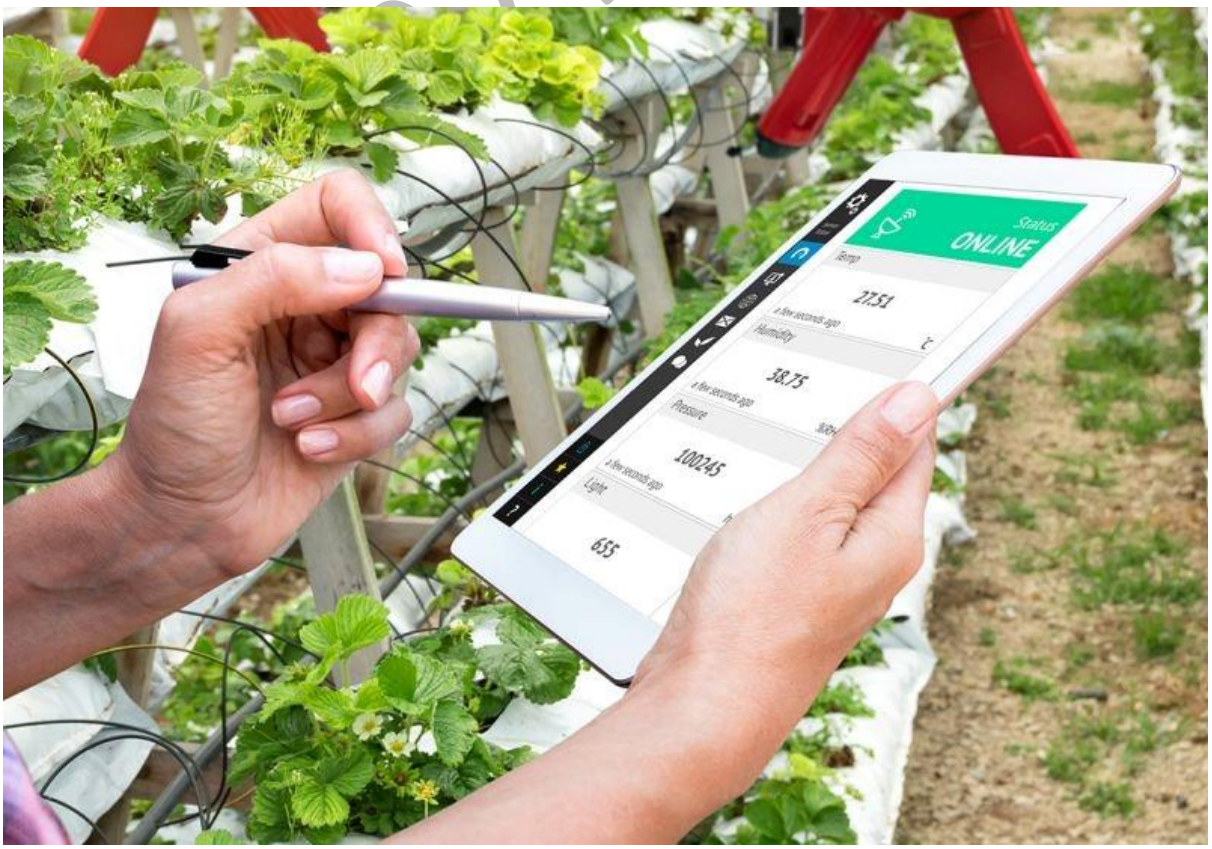
## 2.1 Project overview

Weather plays an important role in agriculture production. Thus there is no aspect of crop culture that is immune to impact of weather. Weather factor contribute to optimal crop growth, development and yield. For rainfall variability needs to be expressed in terms of percentage so that minimum assured rainfall amounts at a certain level of probability. For optimal productivity at a given location crops must be such that their weather requirements match the temporal match of relevant weather elements. A detailed knowledge of rainfall regime at a place is an important prerequisite for agriculture planning and management. Soil fertility refers to the inherent capacity of soil to supply nutrients in adequate amount and in suitable proportion for crop growth and crop yield.

## 2.2 Purpose

- ✓ To prepare the fields for sowing of crops with adequate availability of seed zone.
- ✓ To contribute optimal crop growth, development, and yield.
- ✓ To predict appropriate crop yield using data analytics and machine learning techniques.





# 3.LITERATURE SURVEY

## 3.1 Existing problem

Existing Problem: Agriculture, being the most important and predominant occupation of our country, is facing a lot of hurdles for the past few years.

Increase in global warming and pollution leading to abrupt climate change and depletion of soil fertility is affecting the crop yield for the farmers. Wrong seed for the soil, unexpected flooding rains, etc are some of the issues faced by the farmers that can be rectified by Data Analytics.

## 3.2 References

[1] CROP YIELD PREDICTION USING DATA ANALYTICS AND HYBRID APPROACH by Ms. Shreya V. Bhosale, Ms. Ruchita A. Thombare, Mr. Prasanna G. Dhemey, Students and Ms. Anagha N. Chaudhari

Various techniques and algorithms are discussed in this paper for predicting the crop yield. Using one such method or interrelating two or more methods results in better productivity and efficiency.

The three approaches are listed as follows

1. K-means clustering
   Clustering is the process of grouping the data. Basically the K in K-means denotes the number of clusters. Each cluster is represented by the mean value of the objects in the cluster. Data objects which are similar to each other belong to the same cluster. The database about the clusters are stored in a computer, then the studied k-means algorithm is applied and the frequent item is obtained as the output.
2. Apriori algorithm
   Here in apriori the support for the crop is found. With the help of it, the crops which have maximum support from the people are selected for estimation.
3. Naive Bayes algorithm

Generally, this algorithm is based on Bayesian classification. The probability of the crops grown in an acre of land is considered for calculation and the majority is taken for final accumulation of the result.

[2] CROP YIELD PREDICTION USING MACHINE LEARNING

By B M Sagar

A research group investigated the utilization of various information mining methods which will foresee rice crop yield for the data collected from the state of Maharashtra, India. A total of 27 regions of Maharashtra were selected for the assessment and the data was collected related to the principle rice crop yield influencing parameters such as different atmospheric conditions and various harvest parameters i.e Precipitation rate, minimum, average, maximum and most extreme temperature, reference trim cultivable area, evapotranspiration, and yield for the season between June to November referred as Kharif, for the years 1998 to 2002 from the open source, Indian Administration records. WEKA a Java based dialect programming for less challenging assistance with information data sets, assigning design outcomes tool was applied for dataset processing and the overall methodology of the study includes, (1) pre-processing of dataset (2) Building the prediction model utilizing WEKA and (3) Analysing the outcomes. Cross validation study is carried out to scrutinize how a predictable information mining method will execute on an ambiguous dataset. Study applied 10-fold higher cross validation study design to assess the data subsets for screening and testing. Identified and collected information was randomly distributed into 10 sections where one data section was used for testing while all other data sections were utilized for the preparation information. Study reported that the method applied was supportive in the precise estimation of rice crop yield for the state of Maharashtra, India. The precise quantification of the rice productivity in various climatic conditions can help farmers to understand the optimum condition for the higher rice crop yield.

[3] ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING TECHNIQUES by D Ramesh, B Vishnu Vardhan

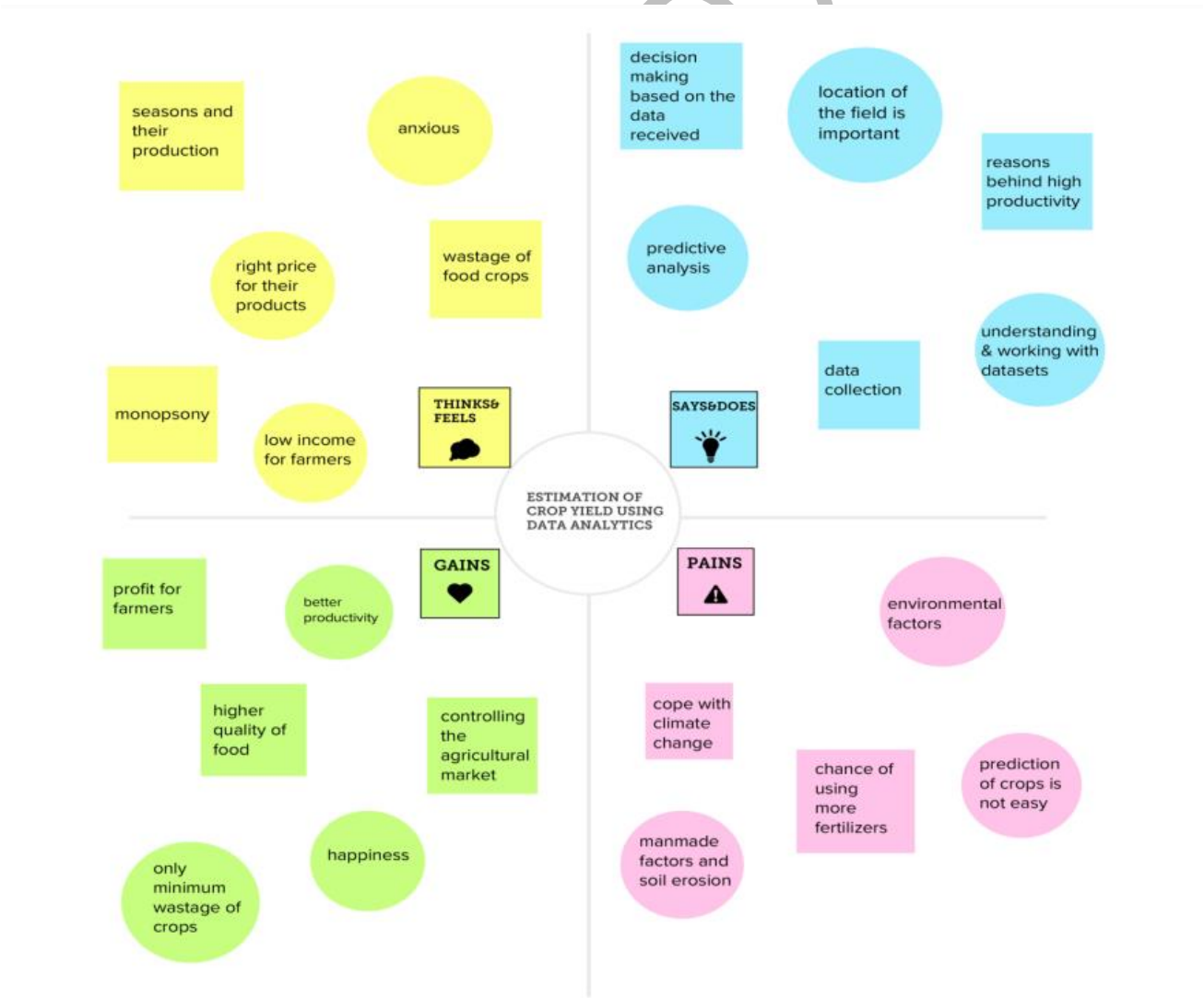
Generally, in this paper, the data are taken into account to analyse the production and yield. The data are taken in eight input variables. The variables are 'Year', 'Rainfall', 'Area of Sowing', 'Yield', 'Fertilizers' (Nitrogen, Phosphorus and Potassium) and 'Production'. The attribute 'Year' specifies the year in which the data are available in Hectares. 'Rainfall' attribute specifies the average rainfall in centimetres. 'Area of Sowing' attribute specifies the total area sowed in the specified year for that region in Hectares. 'Yield' specifies a kilogram per hectare. 'Production' attribute specifies the production of crops in the specified year in Metric Tons. 'Fertilizers' specify in Tons. Similarly statistical model like Multiple Linear Regression (MLR) techniques can be applied on existing data and analysed using data mining techniques like density-based clustering technique which is a method of considering the data points in the region separated by two clusters of low point density were taken into account for the estimation crop yield analysis.

### 3.3 Problem statement definition

The project Crop Yield Estimation using Data Analytics uses millions of datasets previously collected and stored from farmers, soil experts, agriculturists, environmentalist, and meteorologists to visualise the data for the customer when he enters the required details. This helps farmers predict about the profit they might incur on the requested seasonal harvest.

## 4.IDEATION&PROPOSED SOLUTION

### 4.1 Empathy Map Canvas



## 4.2 Ideation & Brainstorming

- One of the most important occupations for human survival is Agriculture and especially the majority of the population in India is into this. Due to variations in climatic conditions and various other challenges, the Agrarian sector in India is facing rigorous problems to maximize crop productivity. It has become an arduous task to achieve the desired targets in crop yield. Various factors like monsoon rains, rodents, water scarcity are to be considered, which do have a direct impact on the production and productivity of the crops. And thus, crop yield prediction with recent advancements in technology can be one of the important factors in agriculture practices as the use of technology in agriculture has comparatively increased in the recent years.

- With the decrease of available cultivable land globally and the decreased cultivable water resources, it is almost impossible to report higher crop yield. Agricultural big data analytics is one approach, believed to have a significant role and positive impact on the increase of crop yield by providing the optimum condition for the plant growth and decreasing the yield gaps and the crop damage and wastage. India is basically agricultural country and approximately 70% our country economics is directly or indirectly related to the agricultural crops. The crop which occupies the highest (60-70%) percentage of cultivable land in the Indian soil is the paddy culture and it is the major crop especially in central and south parts of the India. Rice crop cultivation plays an imperative part in sustenance security of India, contributing over 40% to general yield generation. The enhanced yield of the rice crop depends largely on the water availability and climatic conditions.

- The accurate prediction of crop yield certainly benefits the farmers in choosing the right method to reduce the crop damage and gets best prices for their crops. A research group conducted a work with an objective of accurate prediction of crop yield through big data analytics to assess various crop yield influencing factors such as Area under Cultivation (AUC) interims of hectors, Annual Rainfall (AR) rates and Food Price Index (FPI) and to develop relationship among these parameters. Regression Analysis (RA) methodology was applied to examine the selected factors and their impact on crop prediction and final yield. RA methodology is a multivariable investigation practice which can categorize the factors in to groups such as

explanatory and response variables and helps to assess their interaction to obtain a resolution.

- Analysing the yields of crop is necessary to update the policies to ensure food security. A research group conducted a study with the aim in suggesting a novel data mining method to predict the yields of crop depends on agricultural big data analytics methodologies, which were progressively contrast with conventional data mining methodologies in the process of handling data and model the designs. Study suggested that the method employed should be user friendly, work based on progressive big-data responsive processing structure, supposed to utilize the existing agricultural significant datasets and would still be used with the larger volumes of data growing at enormous rates.

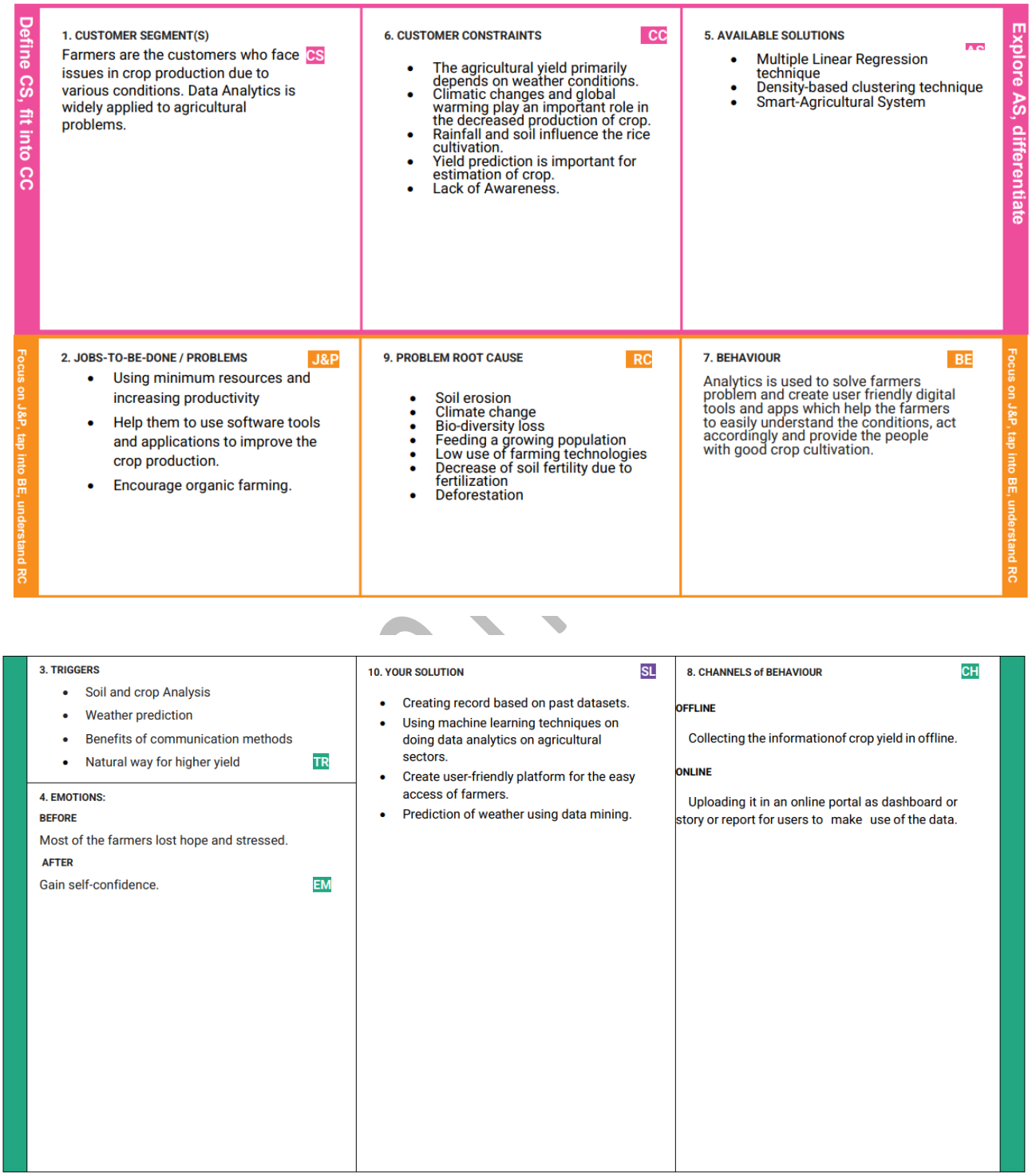
## 4.3 Proposed Solution

| S.No. | Parameter | Description |
|---|---|---|
| 1. | Problem Statement (Problem to be solved) | Agriculture is one of the most important sources of income which eventually affect everyone's life. Indian agriculture is affected by various circumstances. This may directly or indirectly affect the life of a farmer. Feeding the growing population, loss of agricultural land and low production in the varieties of crops and livestock may be the root cause. |
| 2. | Idea / Solution description | Increase in crop yield, better productivity of crops through effective techniques and efficient methods. A comparison of the subsequent crop yield predictions can be made with the entire set of existing available data and can be dedicated to suitable approaches for improving the efficiency of the proposed technique. |
| 3. | Novelty / Uniqueness | Increase in knowledge of agriculture in terms of crop production from sowing the seeds till harvesting. It gives an overview on the total crop yield before even sowing the seed from the previous datasets. |
| 4. | Social Impact / Customer Satisfaction | Agriculture sector contributes approximately 14% of the country's total Gross Domestic Product. Even though the agriculture sector plays an important role in the Indian Economy, there is a constant drop in this sector compared to the other sectors. |

| | | |
|---|---|---|
| 5. | Business Model (Revenue Model) | Regression analysis: we use this method for mainly two purposes. One is to predict the values of the dependent variables for individuals or to estimate the effects of few explanatory variables on the dependent variables. Regression analysis is a powerful statistical method which allows us to analyze the relationship between two or more variables of our own. The dependent and independent variables show a linear relationship between the slope and the intercept. |

| | | |
|---|---|---|
| 6. | Scalability of the Solution | Data analytics help the farmers to predict the market conditions, climatic changes, usage of fertilizers, factor-in inflation, and other variables that will helpthem plan the entire process even before sowing the seeds. All of this information assists farmers in making accurate and effective decisions that maximize the productivity |

# 4.4 Problem Solution fit



# 5.REQUIREMENT ANALYSIS

## 5.1 Functional requirements

| FR No. | Functional Requirement (Epic) | Sub Requirement (Story / Sub-Task) |
|---|---|---|
| FR-1 | **User Registration** | Registration through Gmail<br>Registration through IBM Cognos analytics |
| FR-2 | **User Confirmation** | Email confirmation & OTP verification |
| FR-3 | **Data collection** | Various datasets are collected and processed |
| FR-4 | **Data processing** | Working with datasets collected |
| FR-5 | **Data visualization** | Data visualization charts can be done through IBM Cognos analytics |

## 5.2 Non-Functional requirements

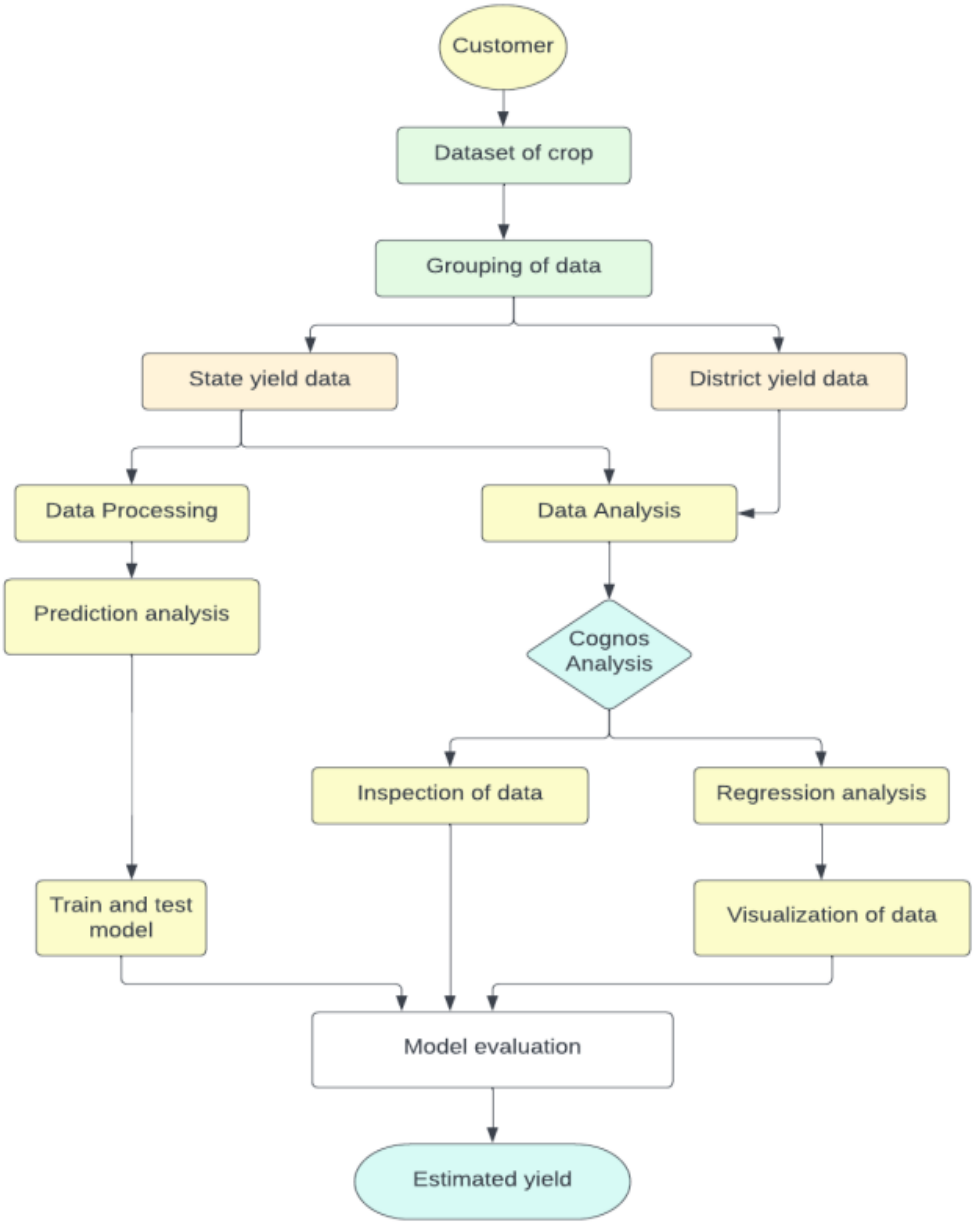
Following are the non-functional requirements of the proposed solution.

| FR No. | Non-Functional Requirement | Description |
|---|---|---|
| NFR-1 | **Usability** | It is completely user friendly. As it is designed to provide support for agriculturalists and farmers, the user experience is simple and efficient. |

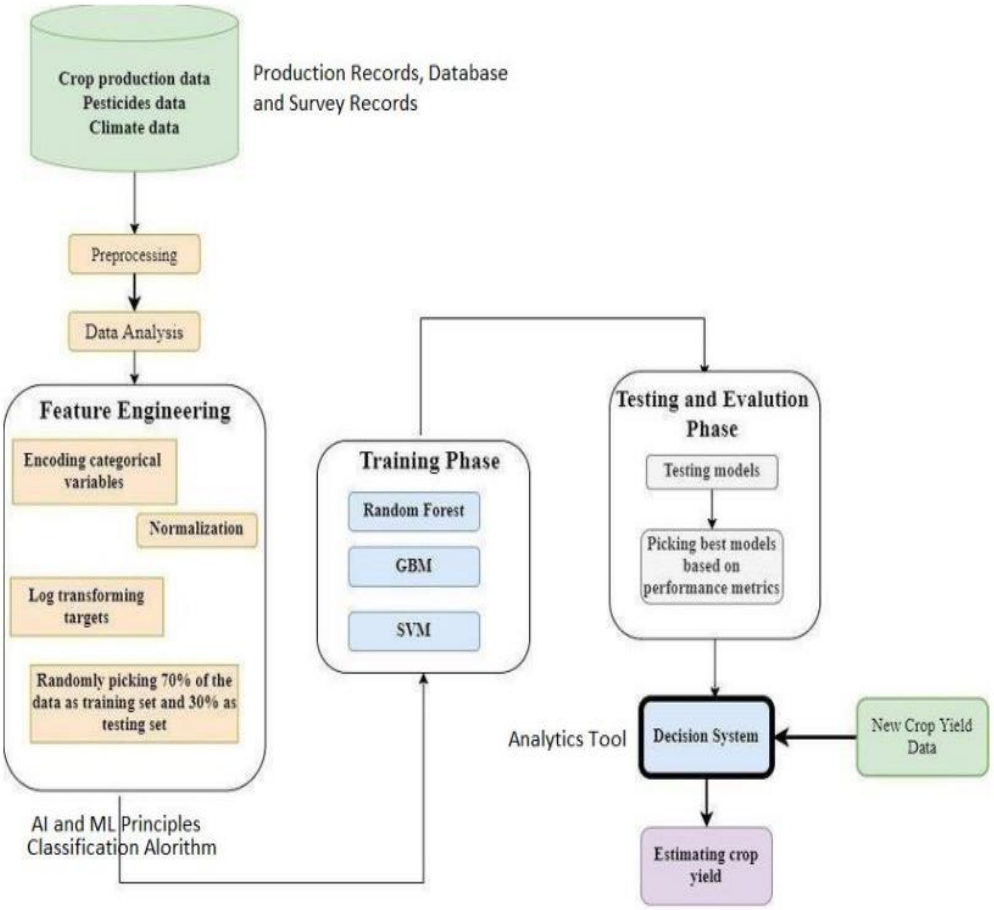
| | | |
|---|---|---|
| NFR-2 | **Security** | Should be secured with unique and standard user id and password with integration with mail id. |
| NFR-3 | **Performance** | With efficient usability and increased security, the performance in turn is increased. |
| NFR-4 | **Reliability** | This deals with lots of previous data undergoing machine learning techniques and thus provides reliable output. |
| NFR-5 | **Scalability** | Will have more datasets loaded in and thus performance does not vary with the increased need/demand. |
| NFR-6 | **Availability** | Can be made available in all platforms with minimum constraints and a stable speed network connection. |

# 6.PROJECT DESIGN

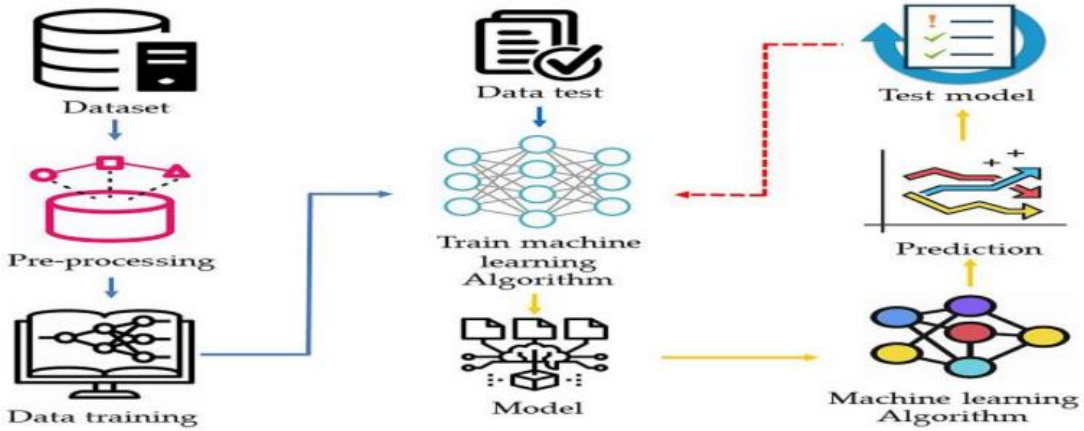
## 6.1 Data Flow Diagram



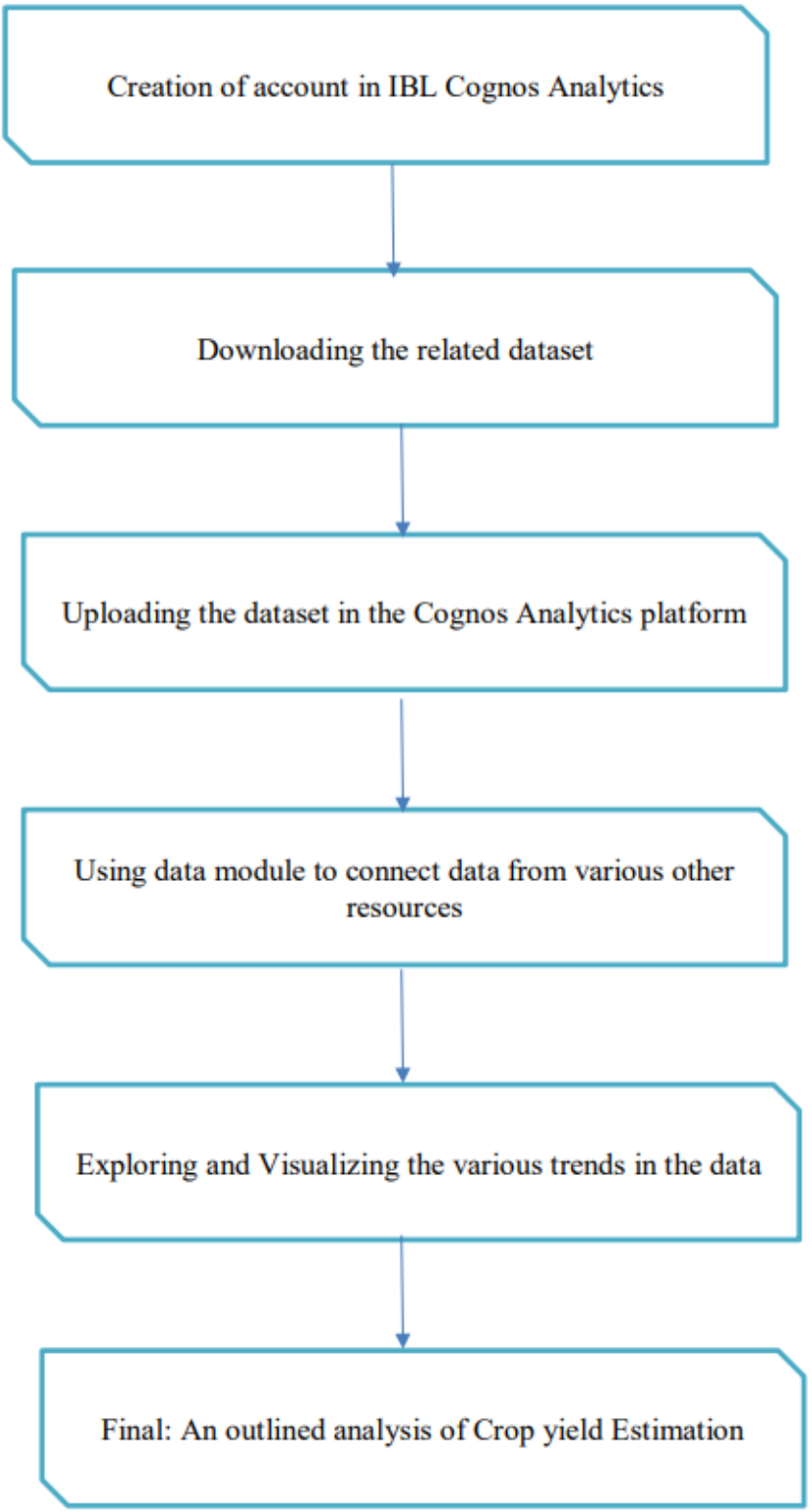
## 6.2 Solution & Technical Architecture



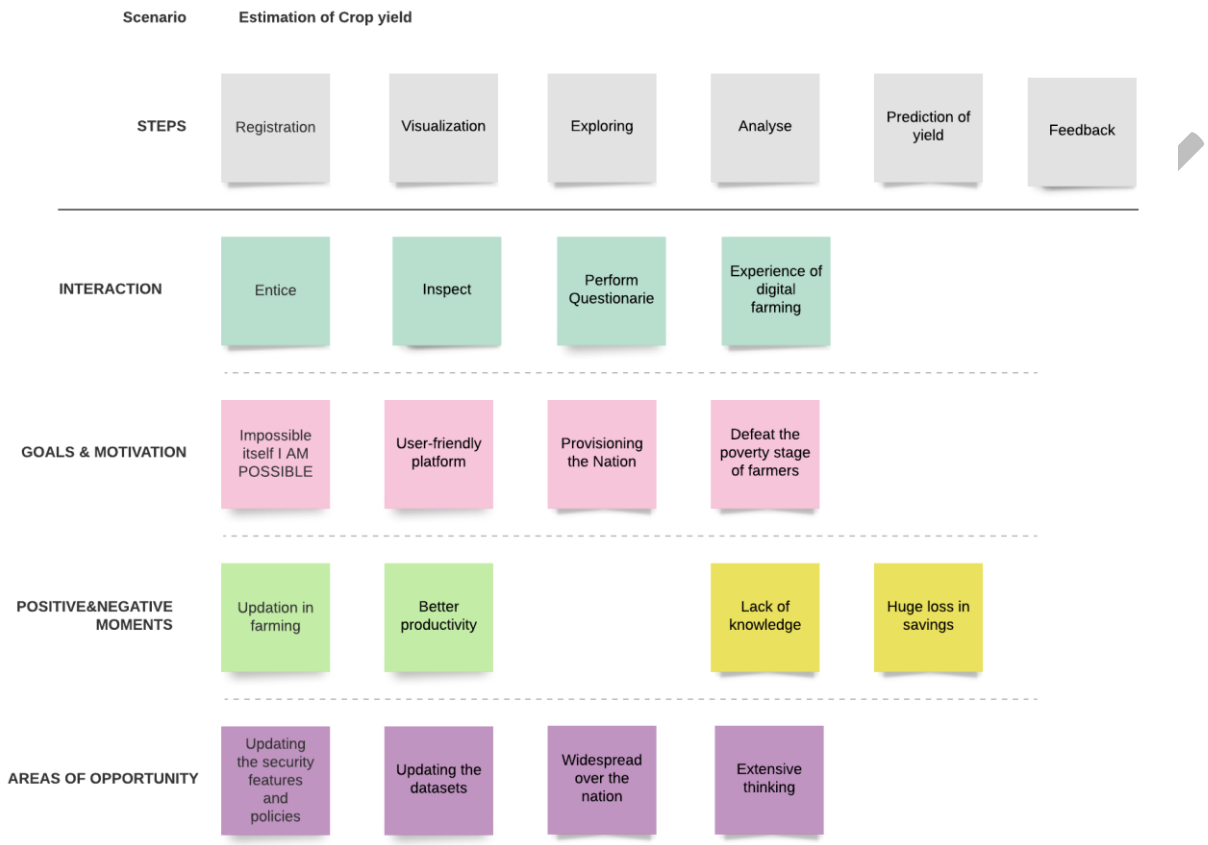

**Technical Architecture:**



## Problem solution steps



## 6.3 User Stories




# 7.PROJECT PLANNING & SCHEDULING

## 7.1 Sprint Planning & Estimation

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

| Sprint | Functional Requirement (Epic) | User Story Number | User Story & task | Story Points | Priority | Team Members |
|---|---|---|---|---|---|---|
| Sprint-1 | Registration | USN-1 | As a user,I can sign up for the application by providing my email ID,password and confirmation. | 2 | High | Aishwarya Gopika |

| | | | | | | |
|---|---|---|---|---|---|---|
| Sprint-1 | | USN-2 | After registering for the application, I as a user will receive a confirmation email. | 1 | High | Atshaya Amruthashreya |
| Sprint-2 | | USN-3 | I can sign up for the application as a user using Google. | 2 | Low | Atshaya Gopika |
| Sprint-1 | | USN-4 | I can sign up for the application as a user using Gmail. | 2 | Low | Amruthashreya Aishwarya |
| Sprint-1 | Login | USN-5 | I can access the application as a user by providing my email address and password. | 1 | High | Aishwarya |
| Sprint-3 | Dashboard | USN-6 | I am free to use my dashboard and explore the features as a user. | 2 | High | Gopika Amruthashreya |
| Sprint-2 | | USN-7 | I can access using the credentials as a user the assets I'm applying for. | 2 | High | Aishwarya Amruthashrey aAtshaya Gopika |
| Sprint-3 | | USN-8 | Data manipulation operations carried out by the application. | 1 | High | Atshaya Aishwarya |
| Sprint-3 | Visualization | USN-9 | Can use certain datasets to generate dashboards. | 2 | Medium | Gopika Aishwarya |
| Sprint-4 | | USN-10 | One can perform predictive analysis. | 1 | High | Amruthashreya Atshaya |
| Sprint-3 | | USN-11 | With certain datasets, I can produce stories. | 2 | High | Aishwarya Atshaya |

| | | | | | | |
|---|---|---|---|---|---|---|
| Sprint-4 | | USN-12 | Can export and send reports in accordance with the built-in dashboards and stories. | 2 | High | Amruthashreya Gopika |

## 7.2 Sprint Delivery Schedule

Project Tracker, Velocity & Burndown Chart:

| Sprint | Total Story Point | Duration (days) | Sprint Start Date | Sprint End Date(Planned) | Story Points Completed (as on planned end date) | Sprint Release date(actual) |
|---|---|---|---|---|---|---|
| Sprint-1 | 20 | 6 | 24-10-2022 | 29 Oct 2022 | 20 | 29 Oct 2022 |
| Sprint-2 | 20 | 6 | 31-10-2022 | 05 Nov 2022 | 20 | 05 Nov 2022 |
| Sprint-3 | 20 | 6 | 07-10-2022 | 12 Nov 2022 | 20 | 12 Nov 2022 |
| Sprint-4 | 20 | 6 | 14 Oct 2022 | 19 Nov 2022 | 20 | 19 Nov 2022 |

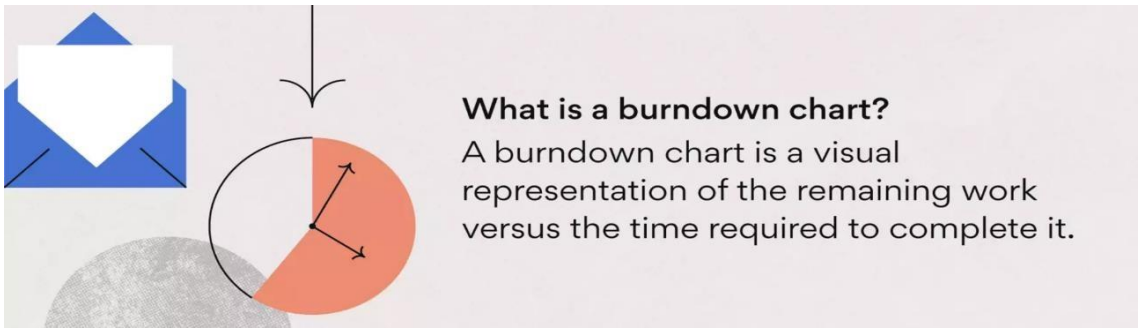
**Velocity:**

Consider a scenario in which the sprint will last 10 days and the team's velocity is 20. (Points per sprint). Let us determine the group's average velocity (AV) for each iteration (story points per day).

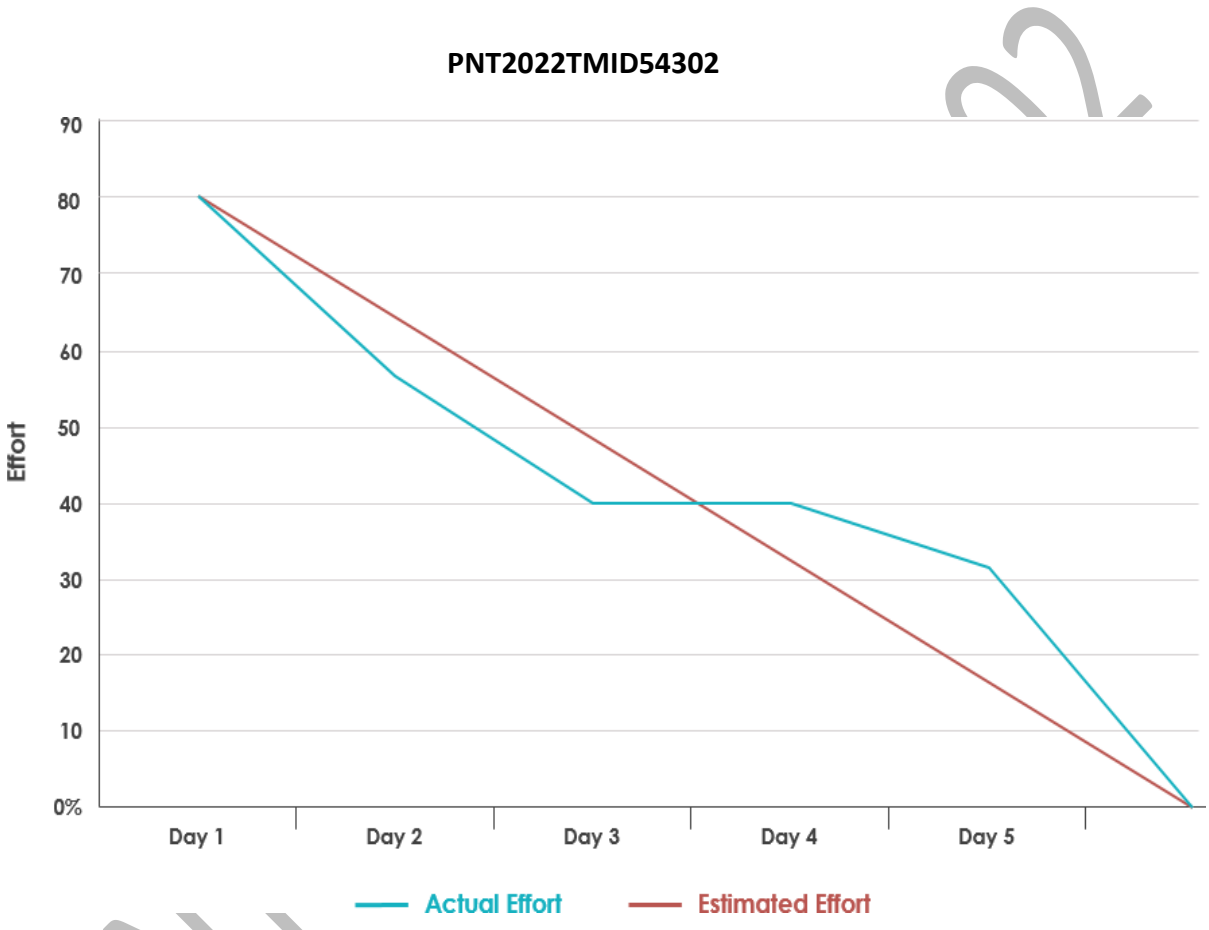
**AV = Sprint Duration/Velocity**

AV = Sprint Duration/Velocity = 20/10=2

**Burndown Chart:**



A burn down chart is a graphical representation of work left to do versus time. It is often usedin agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over.



## 8. CODING & SOLUTIONING

**(Explain the features added in the project along with code)**

### 8.1 Feature 1

# Registration Form

Name

Email

Password

Register

# Login Form

Email

Password

Login

Register



**DATABASE CONNECTIVITY:CODING**

**CSS COMPONENT**

```
mat-card {
    max-width: 600px;
    margin: 2em auto;
    text-align: center;
    max-height: 600px;
```

```
}

.header{
  text-align: center;

}
.full-width {
  width: 80%;
}

.button-row {
  padding-top: 5px;
}

.button-row a {
  margin-right: 8px;
  text-align: center;
}

.forget-password{
  padding-left: 0px;
}

.emailInput{
  padding-top: 10px;
}
```

```
.contentBody {
  padding: 60px 1rem;
  background :#006064;
  display: block;
}

.aLink{
  float: right;
  padding-right: 60px;
  text-decoration: none;
}
```

**HTML COMPONENT:**

```
<mat-card>
    <mat-card-content>
      <div class="header">
        <P>Sign Into Your Account </P>
      </div>
      <form (ngSubmit)="onLogin()" name="loginForm" [formGroup]="loginForm">
        <div class="emailInput">
          <mat-form-field class="full-width" appearance="outline">
            <mat-label>Email</mat-label>
            <input
              formControlName="email"
              matInput
              placeholder="Enter email address" required
            />
            <mat-error *ngIf="!loginForm.controls['email'].valid">
```

```
                Email is required
              </mat-error>
          </mat-form-field>
        </div>

        <div>
          <span>
            <a class="text-link" class="aLink" routerLink="/auth/forgot-
password">Forgot Password?</a>
          </span>
          <mat-form-field class="full-width" appearance="outline">
            <mat-label>Password</mat-label>
            <input formControlName="password" matInput [type]=" hide ?
'password' : 'text'" required />
            <button  mat-icon-button matSuffix (click)="hide = !hide"
[attr.aria-label]="'Hide Password'"
            [attr.aria-pressed]="hide">
            <mat-icon>
                {{hide ? 'visibility_off' : 'visibility'}}
            </mat-icon>
        </button>
        <mat-error *ngIf="!loginForm.controls['password'].valid">
            Password is required
          </mat-error>
        </mat-form-field>
        </div>
        <button mat-flat-button color="primary">Login</button>
      </form>

      <div class="button-row">
        <p>Create New Account</p>
      </div>
```

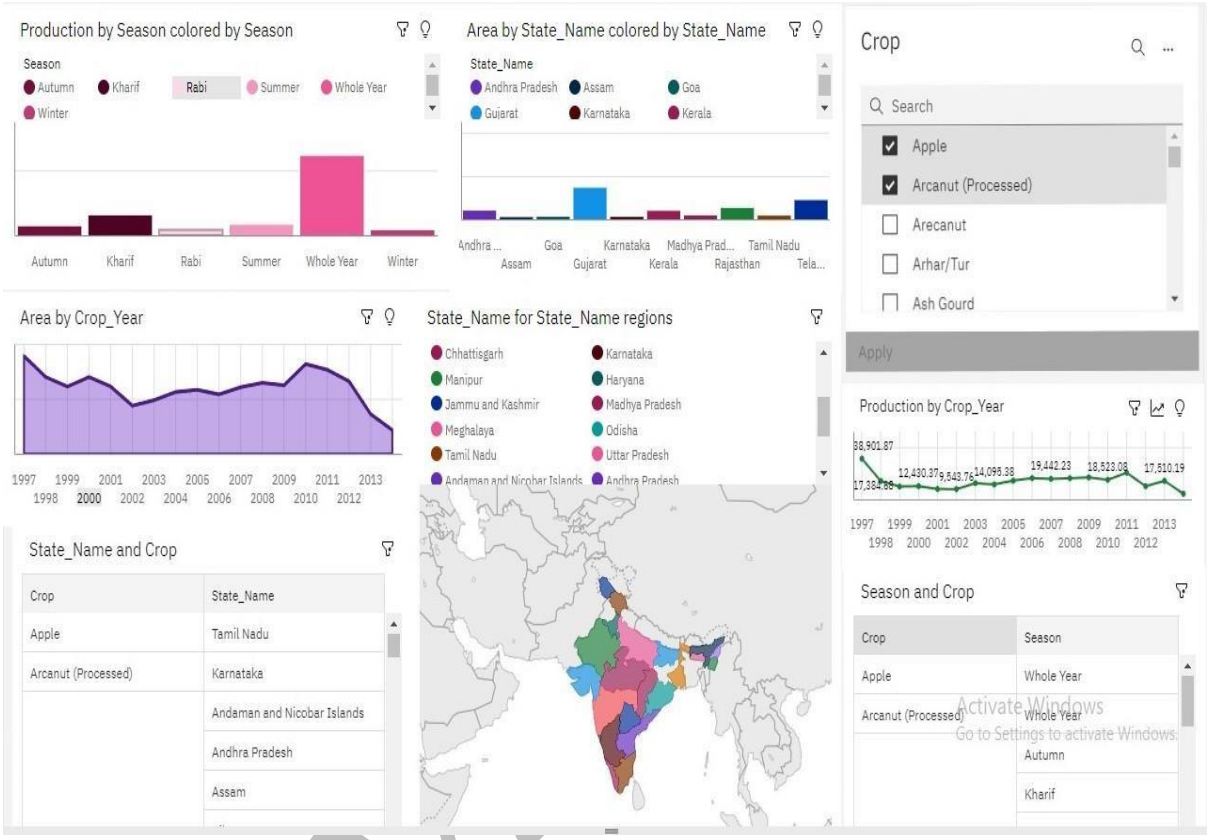
```
    </mat-card-content>
  </mat-card>
```

## .SPEC.TS COMPONENT:

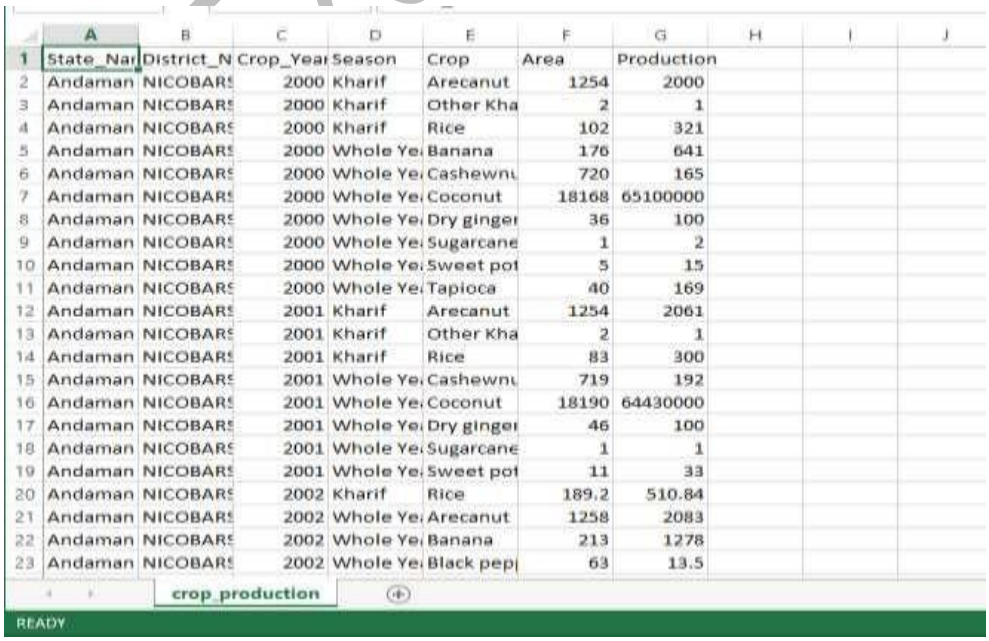
```
import { ComponentFixture, TestBed } from '@angular/core/testing';

import { LoginComponent } from './login.component';

describe('LoginComponent', () => {
  let component: LoginComponent;
  let fixture: ComponentFixture<LoginComponent>;

  beforeEach(async () => {
    await TestBed.configureTestingModule({
      declarations: [ LoginComponent ]
    })
    .compileComponents();

    fixture = TestBed.createComponent(LoginComponent);
    component = fixture.componentInstance;
    fixture.detectChanges();
  });

  it('should create', () => {
    expect(component).toBeTruthy();
  });
});
```

## 8.2 Feature 2

## CREATING DASHBOARD USING IBM COGNOS




## 8.3 Database Schema (if Applicable)



| | State_Nar | District_N | Crop_Year | Season | Crop | Area | Production |
|---|---|---|---|---|---|---|---|
| 2 | Andaman | NICOBARS | 2000 | Kharif | Arecanut | 1254 | 2000 |
| 3 | Andaman | NICOBARS | 2000 | Kharif | Other Kha | 2 | 1 |
| 4 | Andaman | NICOBARS | 2000 | Kharif | Rice | 102 | 321 |
| 5 | Andaman | NICOBARS | 2000 | Whole Ye | Banana | 176 | 641 |
| 6 | Andaman | NICOBARS | 2000 | Whole Ye | Cashewnu | 720 | 165 |
| 7 | Andaman | NICOBARS | 2000 | Whole Ye | Coconut | 18168 | 65100000 |
| 8 | Andaman | NICOBARS | 2000 | Whole Ye | Dry ginger | 36 | 100 |
| 9 | Andaman | NICOBARS | 2000 | Whole Ye | Sugarcane | 1 | 2 |
| 10 | Andaman | NICOBARS | 2000 | Whole Ye | Sweet pot | 5 | 15 |
| 11 | Andaman | NICOBARS | 2000 | Whole Ye | Tapioca | 40 | 169 |
| 12 | Andaman | NICOBARS | 2001 | Kharif | Arecanut | 1254 | 2061 |
| 13 | Andaman | NICOBARS | 2001 | Kharif | Other Kha | 2 | 1 |
| 14 | Andaman | NICOBARS | 2001 | Kharif | Rice | 83 | 300 |
| 15 | Andaman | NICOBARS | 2001 | Whole Ye | Cashewnu | 719 | 192 |
| 16 | Andaman | NICOBARS | 2001 | Whole Ye | Coconut | 18190 | 64430000 |
| 17 | Andaman | NICOBARS | 2001 | Whole Ye | Dry ginger | 46 | 100 |
| 18 | Andaman | NICOBARS | 2001 | Whole Ye | Sugarcane | 1 | 1 |
| 19 | Andaman | NICOBARS | 2001 | Whole Ye | Sweet pot | 11 | 33 |
| 20 | Andaman | NICOBARS | 2002 | Kharif | Rice | 189.2 | 510.84 |
| 21 | Andaman | NICOBARS | 2002 | Whole Ye | Arecanut | 1258 | 2083 |
| 22 | Andaman | NICOBARS | 2002 | Whole Ye | Banana | 213 | 1278 |
| 23 | Andaman | NICOBARS | 2002 | Whole Ye | Black pep | 63 | 13.5 |

crop_production

READY

# 9. TESTING

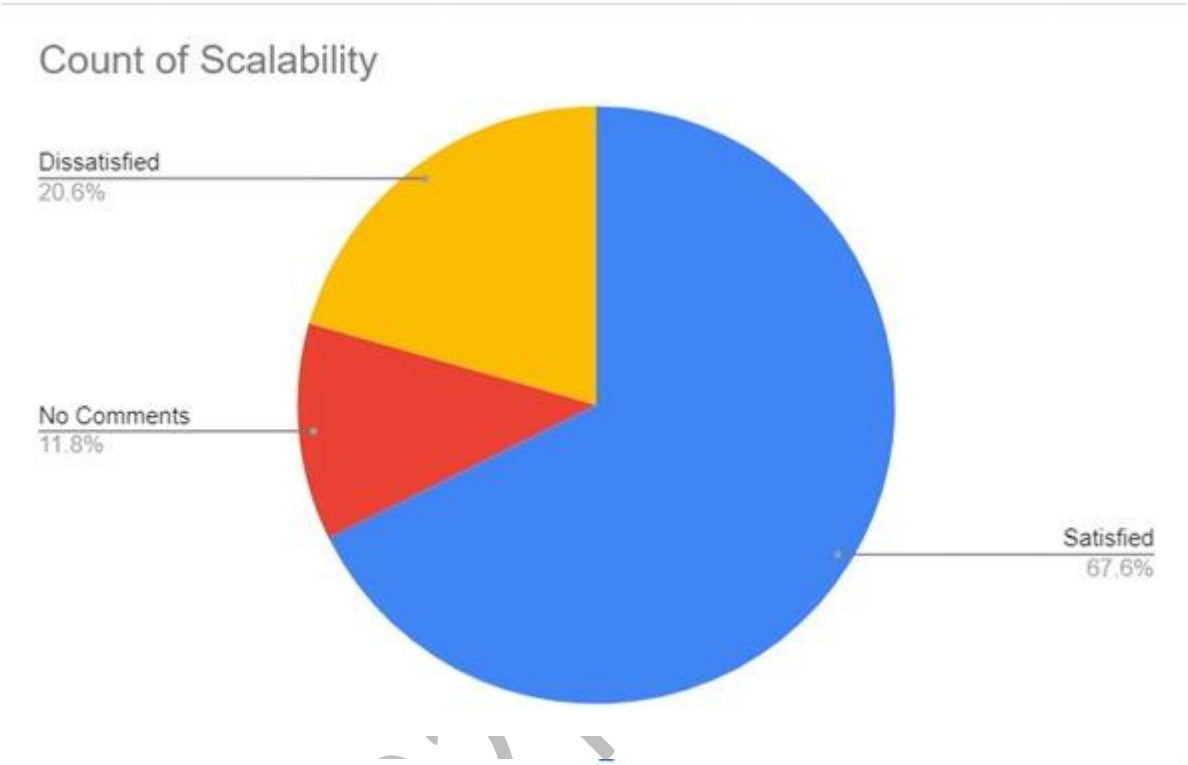
## 9.1 Test Cases

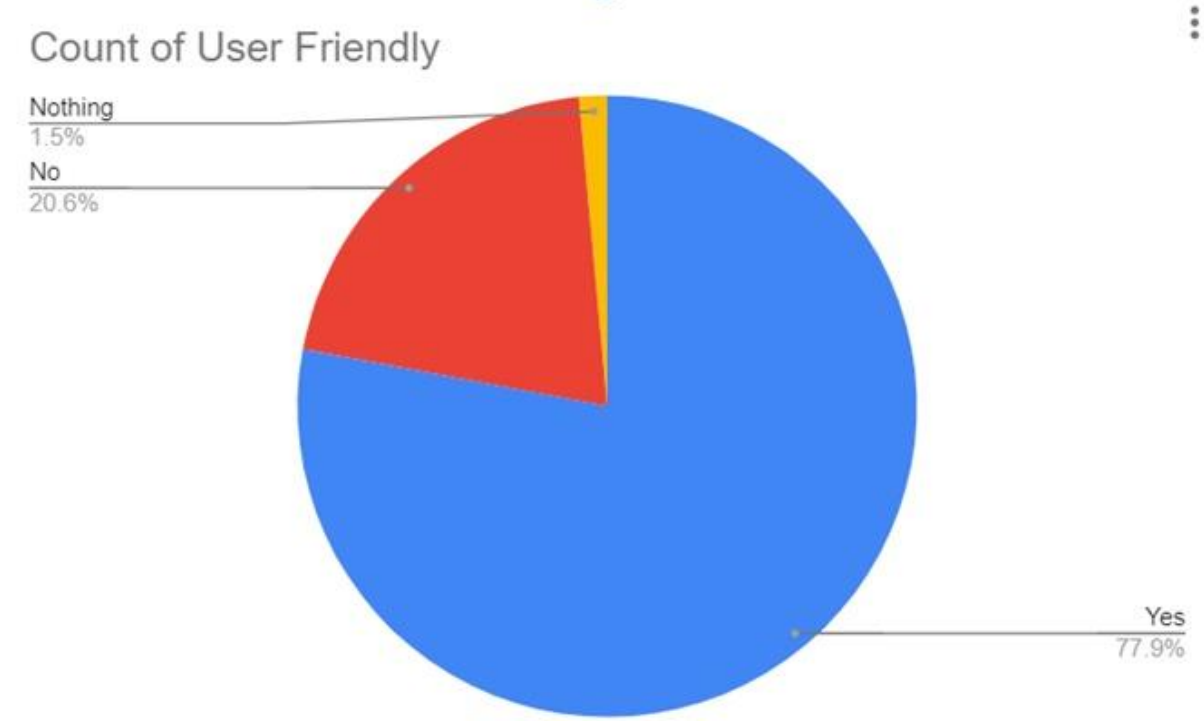
| Section | Total Cases | Not Tested | Fail | Pass |
|---|---|---|---|---|
| Overall process | 7 | 0 | 0 | 7 |
| Security | 51 | 0 | 0 | 51 |
| Client side | 2 | 0 | 0 | 2 |
| Outsource shipping | 3 | 0 | 0 | 3 |
| Exception Reporting | 9 | 0 | 0 | 9 |
| Final Report Output | 4 | 0 | 0 | 4 |
| Version Control | 2 | 0 | 0 | 2 |

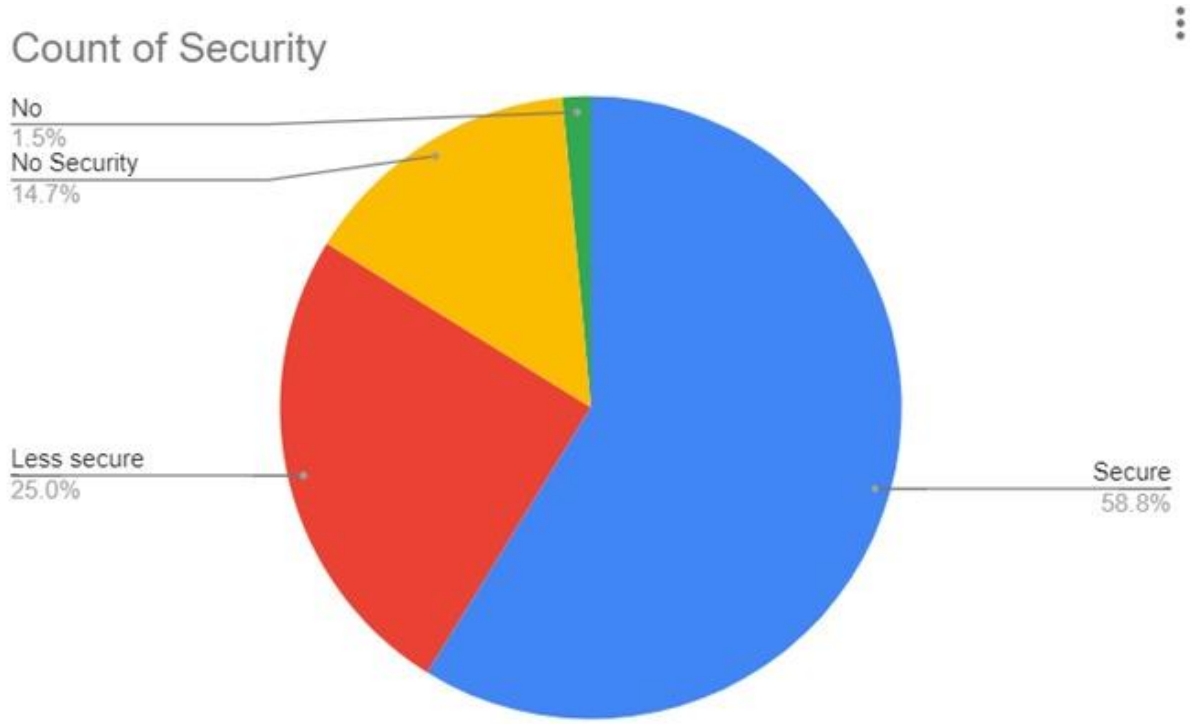
| Resolution | Severity 1 | Severity 2 | Severity 3 | Severity 4 |
|---|---|---|---|---|
| Use-Easiness | 7 | 5 | 4 | 4 |
| Reliability | 8 | 5 | 9 | 4 |
| Security | 4 | 5 | 7 | 6 |
| Availability | 5 | 7 | 6 | 2 |
| Performance | 5 | 4 | 5 | 8 |
| Scalability | 2 | 4 | 3 | 3 |

## 9.2 User Acceptance Testing

This is a survey result on our project's Scalability, User friendliness and Security.

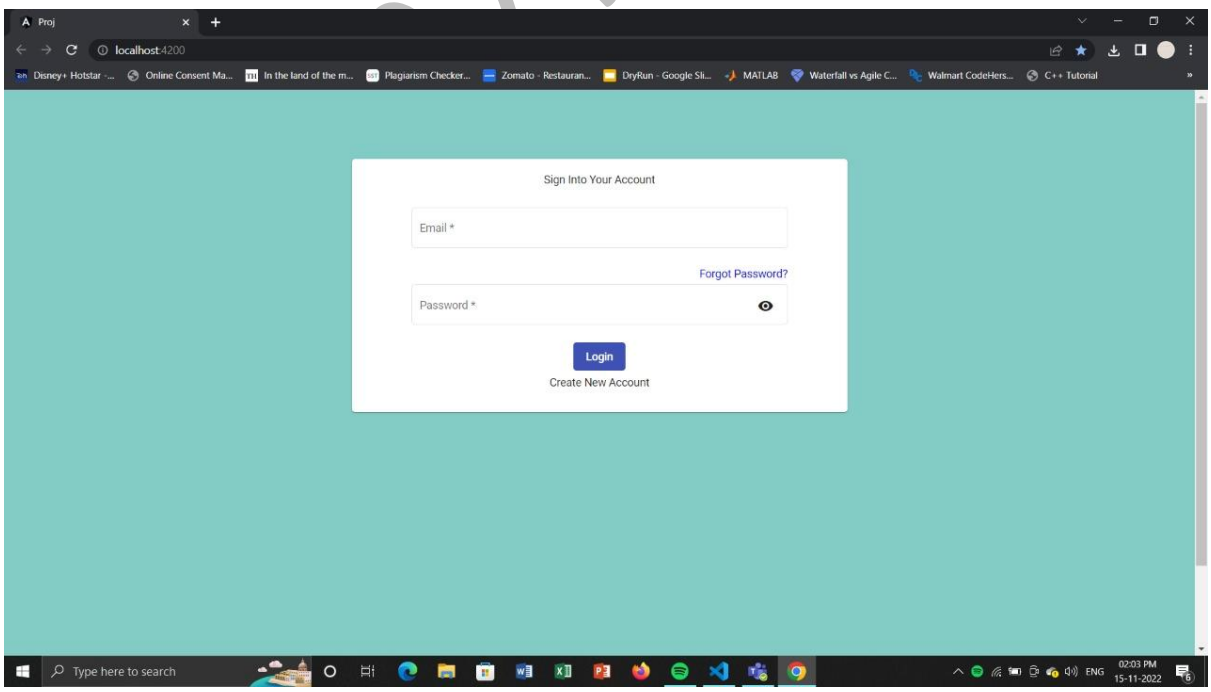





# 10. RESULTS

## 10.1 Performance Metrics



## 11.MERITS AND DEMERITS

Merits

- ✓ Weather aberrations can cause physical damage to crops.
- ✓ Help in cut costs.
- ✓ Product higher crop yield.
- ✓ Prevent over or under watering.

Demerit

- ✓ Sudden change in weather cause crop damage

## 12.CONCLUSION

Weather aberrations can cause physical damage to crops. With help of this project, we can predict in certain environmental conduction which crop should be taken. From the graph of % of production we can determine sowing and harvesting period of particular crop in given temperature and rainfall. This data will continue to enhance farmer efficiency by further enabling them to monitor each plot of land and determine the precise input needed for their crops.

## 13.FUTURE SCOPE

1. Predict appropriate crop and maximum yield in the climate change.

2. Create an android app.

3. Collection of data, Analysis of it and modification of the algorithm.

4. IOT application in agriculture, automation in production line and man free agriculture which is the future of the world, this is the first step of it.

5. Find the percentage yield to happen from the match given percentage in terms of % error.

## 14. APPENDIX

**GitHub Link:**

**Our team ID PNT2022TMID54302**

**https://github.com/IBM-EPBL/IBM-Project-41480-1660642404**

**Project Demo Link :**

**https://drive.google.com/file/d/1SFHlknxQxJHT9T6ICsiIXSHHohpZFCt_/view?usp=sharing**