

## ESTIMATION OF CROP YIELD USING DATA ANALYSTICS



By

| M.NITHEESH KUMAR | (810419104071) |
|------------------|----------------|
|------------------|----------------|

M.RICH DEVOS (810419104096)

A.SUDHAKAR (810419104110)

B.SURIYA (810419104111)

S.PRAKASH (810419104085)

Of

(DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING)

## DHANALAKSHMI SRINIVASAN ENGINEERING COLLEGE (AUTONOMOUS)

**PERAMBALUR - 621 212** 

## A PROJECT REPORT

**Submitted to** 

NALAIYA THIRAN PROJECT BASED

on

# PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND

**ENTERPRENEURSHIP** 



## DHANALAKSHMI SRINIVASAN ENGINEERING COLLEGE (AUTONOMOUS)



## **PERAMBALUR - 621 212**

## **INTERNAL MENTOR**

Miss. R. Aruna priya AP/AI&DS,

Assistant professor,

Department of Artificial intelligence and data science,

Dhanalakshmi Srinivasan engineering college(Autonomous)

## **INDUSTRY MENTORS**

Srikanth Sir,

Mahidhar Sir,

Mohammed Azhar Uddin Sir.

## ESTIMATE THE CROP YIELD USING DATA ANALYTICS ABSTRACT

The primary goal of agricultural production is to increase crop yield. For the agriculture management industry, crop producers, food trade policy, and studies on the carbon cycle, crop prediction on a global and regional scale is crucial. The government places a significant focus on the prediction of crop production in order to sustain the high demand and secure level of food supply to the population. One of the main objectives in agriculture is to maximise crop yield while minimising cost. Early detection and resolution of issues relating to crop yield indicators in the rural field can result in increased yield and earnings. Massive-scale meteorological phenomena will have a completely green impact on agricultural production by reading local weather patterns. Farmers can use the predicted crop yields to cut losses when unfavourable circumstances may arise. While the conditions are favourable for farming, projections may also be employed to maximise crop prediction. We are now working on a system for automatically estimating crop yield in orchards at different growth stages using optical methods. We build on recent developments in the multi-layer perceptron method instead of utilising conventional machine vision to deliver data regarding crop features with a higher accuracy rate.

#### 1. INTRODUCTION

#### 1.1 PROJECT OVERVIEW

Agriculture, since its invention and inception, be the prime and pre-eminent activity of every culture and civilization throughout the history of mankind. It is not only an enormous aspect of the growing economy, but it's essential for us to survive. It's also a crucial sector for Indian economy and also human future. It also contributes an outsized portion of employment. Because the time passes the requirement for production has been increased exponentially. So as to produce in mass quantity people are using technology in an exceedingly wrong way. New sorts of hybrid varieties are produced day by day. However, these varieties don't provide the essential contents as naturally produced crop. These unnatural techniques spoil the soil. It all ends up in further environmental harm. Most of these unnatural techniques are wont to avoid losses. Machine learning, a fastgrowing approach that's spreading out and helping every sector in making viable decisions to create the foremost of its applications. Most devices nowadays are facilitated by models being analysed before deployment. The main concept is to increase the throughput of the agriculture sector with the Machine Learning models. Another factor that also affects the prediction is the amount of knowledge that's being given within the training period, as the number of parameters was higher comparatively.

#### 1.2 PURPOSE

Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD), a field at the intersection of computer science and statistics, is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and systems.

#### 2. LITERATURE REVIEW

#### 2.1 EXISTING PROBLEM

2.1.1 K.H. Anantha et al.., ensure prospects for a living for a population that is expanding quickly. Given that the world's population is anticipated to reach 9.5 billion people by 2050, it becomes even more important to do so with a small water footprint.

Crop yields are being impacted by changes in land use and increasing land degradation all over the world. Global annual soil loss was estimated to be around 75 billion tonnes in 1995, costing the world economy about US\$ 600 billion annually, or US\$ 80 per person. Due to current utilisation levels exceeding permissible thresholds, there is limited room to explore the natural resources that are still available; however, there are opportunities to increase resource use efficiency in order to meet future food and fodder demands. Around the world, 80% of the land is used for rainfed agriculture, which also provides roughly 60% of the food. About 40% of the world's food supply is supported by the remaining 20% of irrigated agricultural land, which also helps certain emerging nations become self-sufficient in food production. However, current resource usage efficiency is well below the potential that can be achieved on a global scale, both in rainfed and irrigated systems. Both systems face particular difficulties. Land degradation and physical water scarcity have been problems for rainfed systems. The Indian subcontinent's many agro-ecologies employ a variety of best water management strategies, which are reviewed in this study in peer-reviewed literature.

important in situ conservation technologies' published works .The evaluations included raised bed, zero tillage, direct seeded rice, laser land levelling, and mulching. Analyzed similarly were studies describing ex situ rainwater harvesting methods on meso scale landscapes.

2.1.2 Kodimalar Palanivel et al..., The principal industry that supports our nation's economy is agriculture. Farmers need to be prepared for the current issues of water scarcity, uncontrollable cost owing to demand-supply, and weather volatility. Low in particular. It is necessary to solve crop output issues brought on by erratic climatic changes, subpar irrigation systems, declining soil fertility, and conventional farming methods. One such method used in agriculture to forecast crop yield is machine learning. Crop yield predictions are made using a variety of machine learning approaches, including prediction, classification, regression, and clustering. Some of the algorithms used to implement prediction include artificial neural networks, support vector machines, linear and logistic regression, decision trees, and Nave Bayes. To the

researchers' task, choosing the right algorithm from the pool of possible algorithms with regard to the crop of choice. The purpose of this research is to investigate how various algorithms for machine learning are helpful in predicting crop yield. The big data computing paradigm has a method for crop yield prediction using machine learning techniques. Crop production is a vital aspect of agriculture, which contributes significantly to the Indian economy. There are two types of crops: food crops and commercial crops. Paddy, wheat, maize, gramme, millets, and other food crops are included, but sugarcane cotton, peanuts, such as cashew. Weather conditions have a substantial impact on crop productivity. Therefore, predicting yields accurately is a serious issue that needs to be solved. Early yield forecasting would help farmers take preventative measures to boost productivity. Early forecasting is made possible by compiling farmers' prior experience, weather data, and other influencing factors; storing it in a sizable database. Rainfall, temperature, humidity, sun radiation, and crop are frequent input characteristics. A farm's capacity, depth, population density, fertiliser use, irrigation, tillage, and soil organic matter. Early decisions are possible by using data mining techniques like prediction, classification, and clustering.

- 2.1.3 Subhadra Mishra et al.., Forecasting crop yields for coastal locations based on various meteorological conditions is a crucial procedure. The datasets for the Kharif and Rabi seasons, as well as the climatic characteristics of three coastal districts that are part of the state of Odisha, are used in this work to propose a regression-based adaptive boosting prediction model India. This study explores and tests several weak regressors, including linear, lasso, ridge, and SVR regression, and suggests strong predictors by avoiding the drawbacks of individual weak regressors and disseminating the advantages of AdaBoost to enhance prediction accuracy. The degree of learning problem prediction accuracy. AdaBoost assists in combining the outputs of the weak regressors into a weighted sum that represents the combined outputs of the weak regressors and the boosted strong regressor, which is likely to be more accurate. To adaptively be warped in favour of occurrences that were incorrectly predicted. The trials have shown that weak regressors make different decisions as a result of frequent, innate characteristics of climatic circumstances for crop production. The effectiveness of the proposed strong regressors forecasting methods for crop production has been highlighted by obtained numerical simulation results in terms of errors, various performance indicators, and statistical analysis.
- 2.1.4 Tanha Talaviyaet al.., The agricultural industry contributes significantly to the economy. The primary issue and a hot topic worldwide is the automation of agriculture. The population is growing rapidly, and with with it, so are the need for food and work. Farmers adopted

conventional techniques, which were not adequate to satisfy these needs. New automated techniques were consequently introduced. These innovative techniques supplied the world's food needs while simultaneously giving billions of people access to jobs. A revolution in agriculture has been sparked by artificial intelligence. The crop yield has been shielded by this technology from a number of factors, including population growth, employment issues, and food security concerns. The lack of efficient irrigation systems, weeds, difficulties with plant monitoring owing to crop height, and harsh weather conditions are only a few of the difficulties the agricultural sector encounters. But with the help of technology, performance may be improved and these issues can be resolved. It can be improved with a variety of AI-driven techniques, such as automated irrigation using GPS and remote sensors to detect soil moisture content. Farmers had the issue that precision weeding techniques outweighed the substantial amount of crops lost during the weeding process.

2.1.5 Vaishali Pandith et al.., An key factor influencing crop production prediction is soil. By creating advance plans, farmers and soil analyzers can use analysis of soil nutrients to increase crop productivity. In this study, a variety of machine learning techniques have been applied to forecast mustard crop yield from soil analysis in advance. The Department of Agriculture, Talab Tillo, Jammu, provided the information for the experimental setup, which included soil samples from various districts in the Jammu region for the mustard crop. Five supervised machine learning methods have been used in the current study: K-Nearest Neighbor (KNN), Naive Bayes, Multinomial Logistic Regression, Artificial Neural Network (ANN), and Random Forest. Five factors—accuracy, recall, precision, specificity, and f-score—have been studied to gauge each technique's effectiveness. The most precise method for predicting mustard crop yield has been determined through experimentation. According to experimental findings, KNN and ANN (among the ML approaches used for the study) were discovered to be the most accurate techniques for predicting mustard crop output. Machine learning is a technique that gives systems the capacity to continuously train themselves to automatically learn from experience and improve. It consists of a group of well defined models that gather particular data and use particular algorithms to produce desired outputs. In order to increase crop output and quality, machine learning techniques have been used to the agricultural sector. To ascertain for a certain crop the conditions that would generate the best yield, machine learning techniques are applied.

#### 2.2 REFERENCES

- [1] K.H. Anantha, Impact of best management practices on sustainable crop production and climate resilience in smallholder farming systems of South Asia, 2021.
- [2] Kodimalar Palanivel, an approach for prediction of crop Yield using machine learning and big Data techniques, 2020.
- [3] Subhadra Mishra, Adaptive boosting of weak regressors for forecasting of crop production considering climatic variability: An empirical assessment,2020.
- [4] Tanha Talaviya, Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides,2020.
- [5] Vaishali Pandith, Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis,2020.

#### 2.3 PROBLEM STATEMENT DEFINITION

India's economy is heavily dependent on agriculture, which is one of the country's key industrial sectors and is essential for rural sustainability. The level of agriculture in India is declining as a result of various issues, including climate change, unpredictable rainfall, falling water levels, excessive pesticide use, etc. On the agricultural data, we used descriptive analytics to determine the level of production. The main goal of this research project is to offer a methodology that can be used to effectively perform descriptive analytics on crop yield production. Deep neural networks feature numerous stacked non-linear layers, each of which transforms the raw input data into a higher-level representation. As a result, as the network becomes deeper, more complicated features are extracted, increasing the results' accuracy. Although it may be quite difficult to determine the proper parameters, deep neural networks are known to be universal approximation functions, which means that they can approximate practically any function.

#### 3.IDEATION AND PROPOSED SOLUTION

#### 3.1 EMPATHY MAP CANVAS

An empathy map is a collaborative tool teams can use to gain a deeper insight into their customers. Much like a user persona, an empathy map can represent a group of users, such as a customer segment. The empathy map was originally created by Dave Gray and has gained much popularity within the agile community

The *Says* quadrant contains what the user says out loud in an interview or some other usability study. Ideally, it contains verbatim and direct quotes from research.

- "I am allegiant to Delta because I never have a bad experience."
- "I want something reliable."
- "I don't understand what to do from here."

The *Thinks* quadrant captures what the user is thinking throughout the experience. Ask yourself (from the qualitative research gathered): what occupies the user's thoughts? What matters to the user? It is possible to have the same content in both *Says* and *Thinks*. However, pay special attention to what users think, but may not be willing to vocalize. Try to understand why they are reluctant to share — are they unsure, self-conscious, polite, or afraid to tell others something?

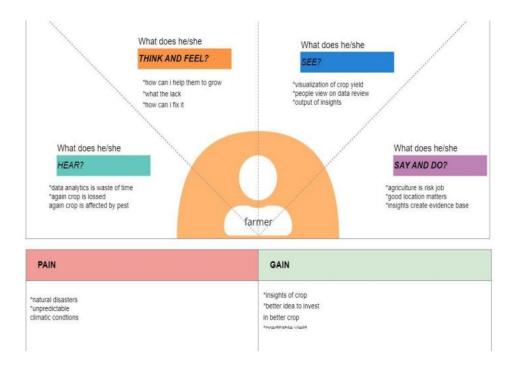
- "This is really annoying."
- "Am I dumb for not understanding this?"

The *Does* quadrant encloses the actions the user takes. From the research, what does the user physically do? How does the user go about doing it?

- Refreshes page several times.
- Shops around to compare prices.

The *Feels* quadrant is the user's emotional state, often represented as an adjective plus a short sentence for context. Ask yourself: what worries the user? What does the user get excited about? How does the user feel about the experience?

- *Impatient: pages load too slowly*
- Confused: too many contradictory prices
- Worried: they are doing something wrong
- Our users are complex humans. It is natural (and extremely beneficial) to see
  juxtaposition between quadrants. You will also encounter inconsistencies for
  example, seemingly positive actions but negative quotes or emotions coming from the
  same user. This is when empathy maps become treasure maps that can uncover
  nuggets of understanding about our user. It is our job as UX professionals to
  investigate the cause of the conflict and resolve it.
- Some of these quadrants may seem ambiguous or overlapping for example, it may be difficult to distinguish between *Thinks* and *Feels*. Do not focus too much on being precise: if an item may fit into multiple quadrants, just pick one. The 4 quadrants exist only to push our knowledge about users and to ensure we don't leave out any important dimension. (If you don't have anything to put into a certain quadrant, it's a strong signal that you need more user research before proceeding in the design process.)



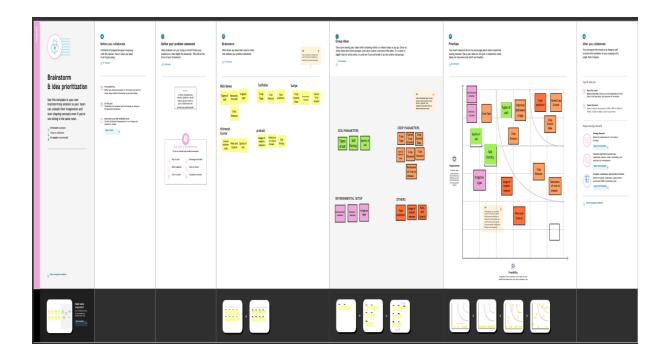
#### 3.2 IDEATION AND BRAINSTROMING

Brainstorming is a method of generating ideas and sharing knowledge to solve a particular commercial or technical problem, in which participants are encouraged to think without interruption. Brainstorming is a group activity where each participant shares their ideas as soon as they come to mind.

When planning a brainstorming session it is important to define clearly the topic to be addressed. A topic which is too specific can constrict thinking, while an ill-defined topic will not generate enough directly applicable ideas. The composition of the brainstorming group is important too. It should include people linked directly with the subject as well as those who can contribute novel and unexpected ideas. It can comprise staff from inside or outside the organisation.

To ensure a productive session and one to which all present contribute, there are several brainstorming 'rules' -

- Encourage novel and innovative ideas, however odd they may first appear
- The quantity of ideas is more important than quality, so while ideas are shared with the group they are not discussed or criticised in detail; this is reserved for a later stage
- Build on the ideas put forward by others
- Every person and every idea has equal worth



In this each 5 members in the team gave their ideas as each 5 idea and the best 5 idea's among those ideas is selected and taken as the idea to be implemented in the project.

#### 3.3 PROPOSED SOLUTION

| 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.                                     |

#### PROBLEM STATEMENT

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.

#### IDEA AND 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.

## **NOVELTY AND 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

#### 3.4 PROBLEM SOLUTION FIT



Problem-Solution Fit - this occurs when you have evidence that customers care about certain jobs, pains, and gains. At this stage you've proved the existence of a problem and have designed a value proposition that addresses your customers' jobs, pains and gains

These change from time to time. And, when that happens, you need to find yourself in a position that will allow you to *detect* them, early on. To put it another way, you'll need to be looping through a continuous testing process. However, the first *validation*, the first test you need to conduct — and make sure you get some good conclusions out of it — is that of the *problem-solution fit*. That's, actually, a prerequisite for the next steps

The problem-solution fit is what we're going to discuss below. Together with the problems that may arise if you don't conduct the experiment; or fail to get some useful results out of it. Will that hold you back? Of course! And that's simply because if you don't produce enough hard evidence that you do meet the standards — or, much worse, if you've carelessly omitted this critical step — you'll inevitably be forced to retrace your steps all the way **back** to defining the problem

Once you get into building a company from scratch, that alone means that you've endeavored into uncertainty and hardship. There are so many parameters you need to take into account, long before you can make the right decisions and put your plan into action. And, for startups, things are a bit trickier. To use an example, it's as if you're *exploring the backcountry* with no map. No matter the <u>product</u> type you're aiming to build and no matter the <u>market type</u> you plan to target, the data you have in your hands is yet to be validated. You only start with a set of assumptions. That's what you have. That you should treat them as such.

A problem-solution fit helps make sure that your initial hypotheses have been validated (proven right); and that your solution — the one you're about to build— does solve a problem worth solving. A problem you initially assumed your customers have, one that you need to verify whether it's one they care to solve. That's what's going to bring paying customers in. And that's what you'll be using to build your business on. To be more precise, a problem-solution fit helps you get closer to these facts

#### **4.REQUIREMENT ANALYSIS**

#### **4.1FUNCTIONAL REQUIREMENT**

#### **Datasets Acquisition:**

In this module, we can upload the crop yield datasets in the form of CSV file format. And also store the data in database for future purpose. The dataset includes the temperature, rain fall, pH value, nitrogen, phosphorus, potassium values. These values are extracted from Kaggle website and values are stored in the form of integer values.

#### **Preprocessing:**

Data pre-processing is an important step in the data mining process. The phrase <u>"garbage in, garbage out"</u> is particularly applicable to data mining and <u>machine learning</u> projects. Data preparation and filtering steps can take considerable amount of processing time. In this module, we can eliminate the irrelevant values and also estimate the missing values of data. Finally provide structured datasets.

#### **Features Extraction:**

Feature selection refers to the process of reducing the inputs for processing and analysis, or of finding the most meaningful inputs. In this module, select the multiple features from uploaded datasets. And train the datasets with various crop labels with multiple attribute values. We can train the datasets to multiple crops such as rice, maize and so on.

#### **Yield Estimation:**

In this module, we can implement support vector machine algorithm to classify the uploaded datasets. Support vector machine algorithm is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyper plane.

## **Crop Details:**

Appropriate prediction of crop productivity is required for efficient planning of land usage and economic policy. In recent times, forecasting of crop productivity at the within-field level has increased. The most influencing factor for crop productivity is weather conditions. In this module, we can test the datasets by using Machine learning algorithm. Finally provide the details about crops and yield information with improved accuracy rate. The accuracy of the project is calculated in terms of Precision, Recall and F-measure values.

## **4.2 NON FUNCTIONAL REQUIREMENTS**

#### **Usability**

The system shall allow the users to access the system with pc using web application. The system uses a web application as an interface. The system is user friendly which makes the system easy

## **Availability**

The system is available 100% for the user and is used 24 hrs a day and 365 days a year. The system shall be operational 24 hours a day and 7 days a week.

## **Scalability**

Scalability is the measure of a system's ability to increase or decrease in performance and cost in response to changes in application and system processing demands.

## **Security**

A security requirement is a statement of needed security functionality that ensures one of many different security properties of software is being satisfied.

#### **Performance**

The information is refreshed depending upon whether some updates have occurred or not in the application. The system shall respond to the member in not less than two seconds from the time of the request submittal. The system shall be allowed to take more time when doing large processing jobs.

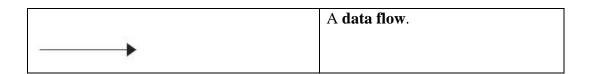
#### **5.PROJECT DESIGN**

#### 5.1 DATA FLOW DIAGRAMS

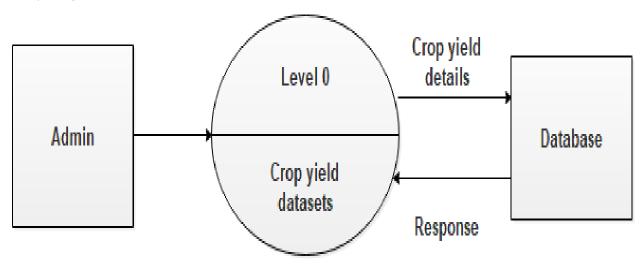
A two-dimensional diagram explains how data is processed and transferred in a system. The graphical depiction identifies each source of data and how it interacts with other data sources to reach a common output. Individuals seeking to draft a data flow diagram must identify external inputs and outputs, determine how the inputs and outputs relate to each other, and explain with graphics how these connections relate and what they result in. This type of diagram helps business development and design teams visualize how data is processed and identify or improve certain aspects.

## **Data flow Symbols:**

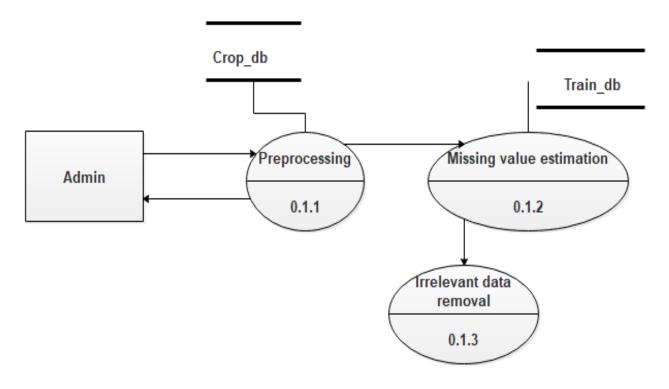
| Symbol | Description                              |
|--------|--|
|        |  |
|        | An <b>entity</b> . A source of data or a |
|        | destination for data.                    |
|        |  |
|        |  |
|        |  |
|        | A <b>process</b> or task that is         |
|        | performed by the system.                 |
|        |  |
|        |  |
|        |  |
|        |  |
|        |  |
|        | A data store, a place where data is      |
|        | held between processes.                  |
|        |  |
|        |  |
|        |  |



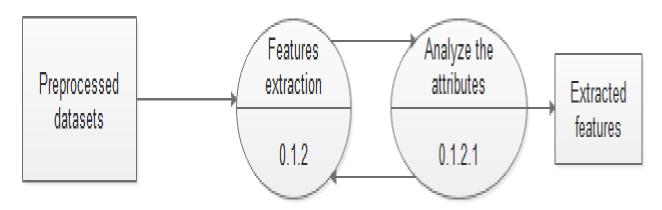
LEVEL 0



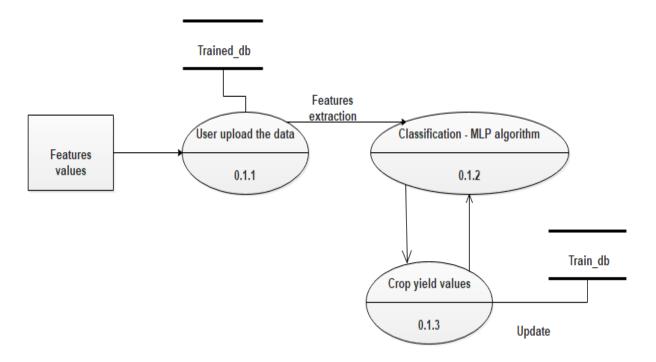
## LEVEL 1



## LEVEL 2

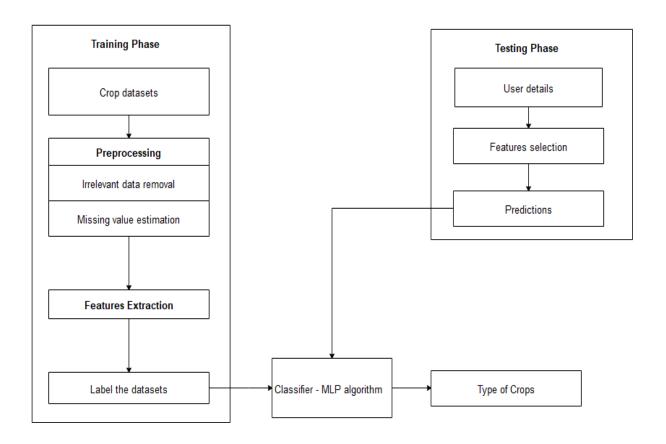


## LEVEL 3



#### 5.2 SOLUTION & TECHNICAL ARCHITECTURE

In this architecture, image can be uploaded from datasets. Then perform preprocessing steps to eliminate the noises in image and perform median filtering algorithm remove the noises in image. Then extract the features from images. Finally classify the pixels whether it is fire or not. Based on features, send alert about fire.



## **5.3 USER STORIES**

Farmers must meet the changing needs of our planet and the expectations of regulators, consumers, and food processors and retailers.

There are increasing pressures from climate change, soil erosion and biodiversity loss and from consumers' changing tastes in food and concerns about how it is produced. And the natural world that farming works with – plants, pests and diseases – continue to pose their own challenge

#### 6.PROJECT PLANNING & SCHEDULING

#### **6.1 SPRINT PLANNING & ESTIMATION**

| Milestone<br>Name         | Milestone<br>Number | Description  | Milestone<br>Type | Team Members         |
|---------------------------|---------------------|--|-------------------|----------------------|
| Registration<br>and Login |                     | Registration for the application by entering my email, password, and confirming my password. | Mandatory         | Rich<br>nitheesh     |
|                           |                     | Confirmation email once I have registered for the application                                | Mandatory         | sudhakar<br>rich     |
|                           | MS-1                | Registration for the application through Google  | Optional          | Prakash<br>rich      |
|                           |                     | Registration for the application through Gmail   | Optional          | surya<br>nitheesh    |
|                           |                     | Log into the application by entering email & password  | Mandatory         | Sudhakar<br>rich     |
|                           |                     | Usage of dashboard and exploration of the features   | Mandatory         | Rich<br>Prakash      |
| Dashboard<br>Usage        | MS-2                | Usage of the credentials to access the resources of my application                           | Mandatory         | surya<br>prakash     |
|                           |                     | Performance of Data manipulations on the application   | Mandatory         | Rich<br>nitheesh     |
| Creation of               | MS-3                | Creation of dashboards with particular datasets  | Mandatory         | Rich<br>surya        |
|                           | MS-4                | Predictive analysis can be done  | Mandatory         | Prakash<br>rich      |
| Deliverables              | MS-5                | Creation of stories with particular datasets   | Mandatory         | nitheesh<br>sudhakar |
|                           | MS-6                | Deliver and export reports according to the dashboards and stories created                   | Mandatory         | Rich<br>surya        |

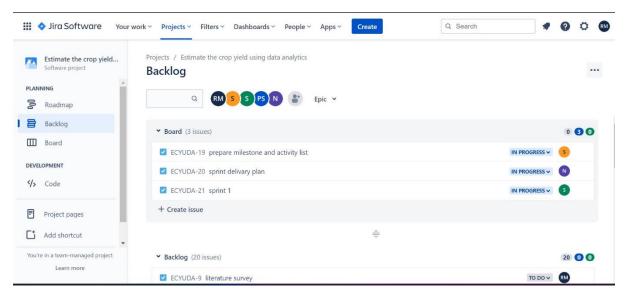
#### **REGISTRATION AND LOGIN**

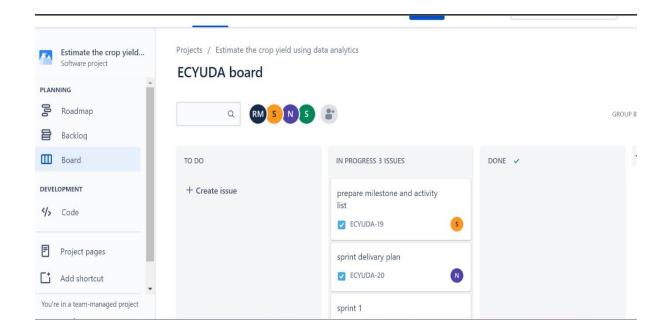
Registration for the application by entering my email, password, and confirming my password. Mandatory Rich nitheesh Confirmation email once I have registered for the application Mandatory sudhakar rich Registration for the application through Google Optional Prakash rich Registration for the application through Gmail Optional surya nitheesh Log into the application by entering email & password Mandatory Sudhakar rich Usage of dashboard and exploration of the features Mandatory Rich Prakash Usage of the credentials to access the resources of my application Mandatory surya prakash Performance of Data manipulations on the application Mandatory Rich nitheesh

#### **6.2 SPRINT DELIVERY SCHEDULE**

| Sprint    | Functional Requirement (Epic) | User Story<br>Number | User Story / Task   | Story Points | Priority | Team Members         |
|-----------|-------------------------------|----------------------|---|--------------|----------|----------------------|
| Sprint-1  | Registration                  | USN-1                | As a user, I can register for the application by entering my email, password, and confirming my password. | 2            | High     | Rich nitheesh        |
| Sprint-1  |                               | USN-2                | As a user, I will receive confirmation email once I have registered for the application                   | 1            | High     | Sudhakar<br>rich     |
| Sprint-2  |                               | USN-3                | As a user, I can register for the application through Google  | 2            | Low      | Prakash rich         |
| Sprint-1  |                               | USN-4                | As a user, I can register for the application through Gmail   | 2            | Low      | surya<br>nitheesh    |
| Sprint-1  | Login                         | USN-5                | As a user, I can log into the application by entering email & password                                    | 1            | High     | sudhakar<br>rich     |
| Sprint-3  | Dashboard                     | USN-6                | As a user, I can freely use my dashboard and explore the features   | 2            | High     | Rich<br>prakash      |
| Sprint-2  |                               | USN-7                | As a user, I can use the credentials to access the resources of my application                            | 2            | High     | surya<br>prakash     |
| Sprint-3  |                               | USN-8                | Performance of Data manipulations on the application  | 1            | High     | rich<br>nitheesh     |
| Sprint-3  | Visualizations                | USN-9                | I can create dashboards with particular datasets  | 2            | High     | rich<br>surya        |
| Sprint-4  |                               | USN-10               | Predictive analysis can be done   | 1            | High     | Prakash rich         |
| Sprint- 3 |                               | USN-11               | I can create stories with particular datasets   | 2            | High     | nitheesh<br>sudhakar |
| Sprint-4  |                               | USN-12               | I can deliver and export reports according to the dashboards and stories created                          | 2            | High     | Surya<br>rich        |

## **6.3 REPORTS FROM JIRA**

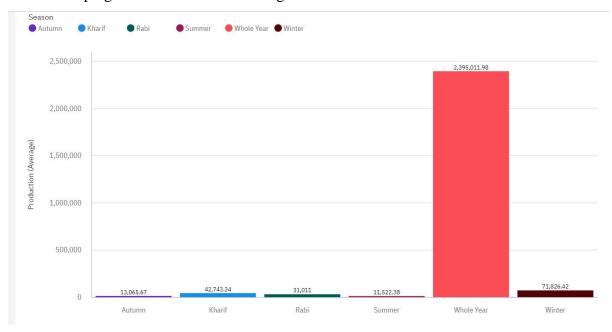


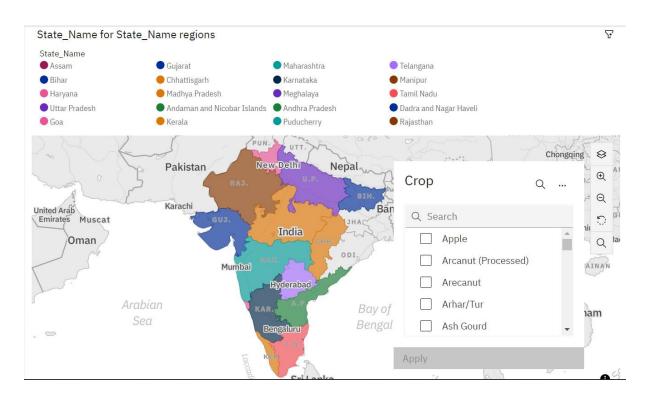


## **7.CODING & SOLUTION**

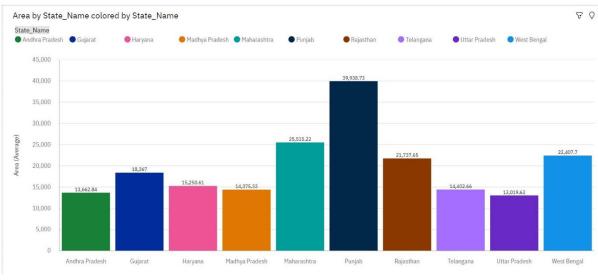
#### **7.1 FEATURE 1**

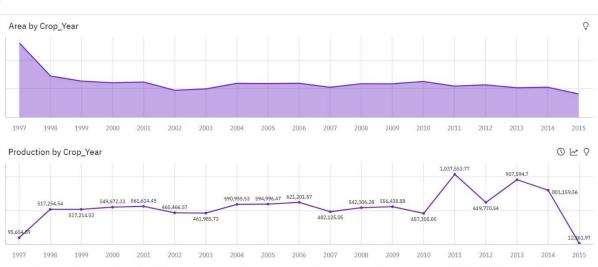
First developing the dashboard the creating the dashboard for various data

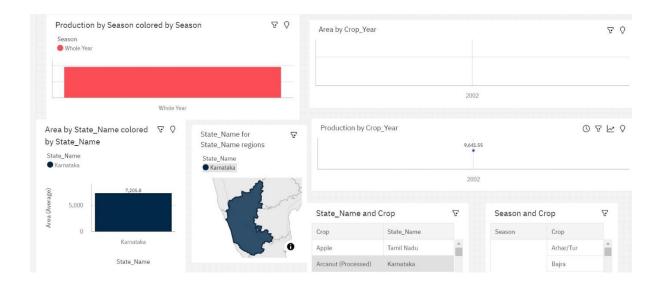




| Crop      | State_Name     |          | Season     | Crop      |  |
|-----------|----------------|----------|------------|-----------|--|
| Apple     | Tamil Nadu     | <u> </u> | Kharif     | Grapes    |  |
| Ash Gourd | Tamil Nadu     |          |            | Apple     |  |
|           | Andhra Pradesh |          | Whole Year | Ash Gourd |  |
| Grapes    | Haryana        |          |            | Grapes    |  |
|           | Karnataka      |          |            |           |  |
|           | Madhya Pradesh |          |            |           |  |
|           | Maharashtra    |          |            |           |  |
|           | Rajasthan      | -        |            |           |  |







#### **7.2 FEATURE 2**

In this feature 2 we developed project using MLP an machine learning algorithm inorder to create the extra feature of creating the crop recommendation for the better yield

Multi-layer perception is also known as MLP. It is fully connected dense layers, which transform any input dimension to the desired dimension. A multi-layer perception is a neural network that has multiple layers. To create a neural network we combine neurons together so that the outputs of some neurons are inputs of other neurons.

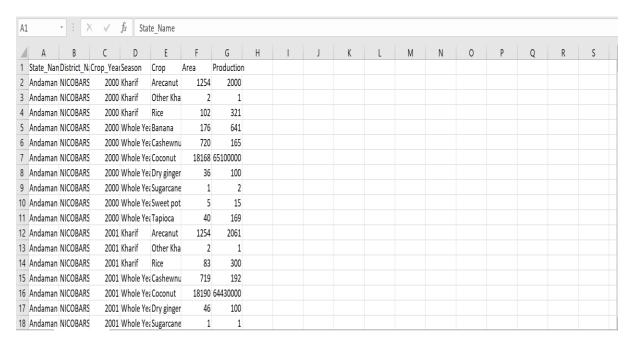
A gentle introduction to neural networks and TensorFlow can be found here:

Neural Networks

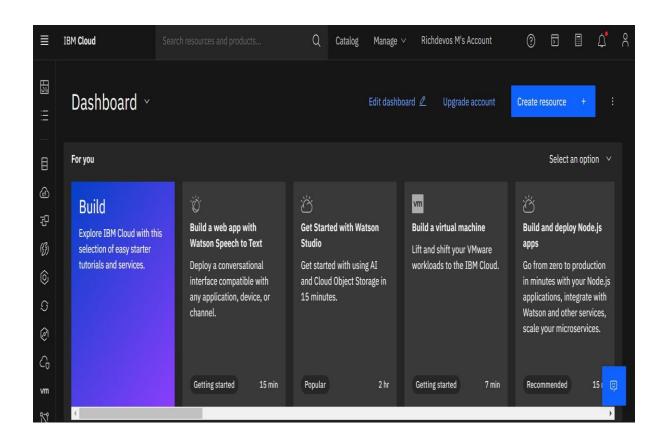
Introduction to TensorFlow

A multi-layer perceptron has one input layer and for each input, there is one neuron(or node), it has one output layer with a single node for each output and it can have any number of hidden layers and each hidden layer can have any number of nodes. A schematic diagram of a Multi-Layer Perceptron (MLP) is depicted below

#### 7.3 DATABASE SCHEMA



In this we used the crop production dataset that is provided by Kaggle website and trained that dataset in ibm cognos analytics and then we uploaded our data at ibm cloud and used cloud as the backend to our project



#### 8.TESTING

#### 8.1 TEST CASES

A test case has components that describe input, action and an expected response, in order to determine if a feature of an application is working correctly. A test case is a set of instructions on "HOW" to validate a particular test objective/target, which when followed will tell us if the expected behavior of the system is satisfied or not.

Characteristics of a good test case:

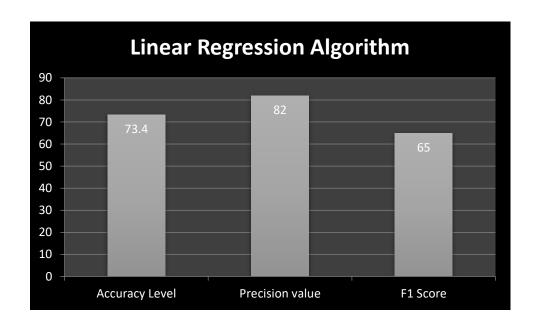
- Accurate: Exacts the purpose.
- Economical: No unnecessary steps or words.
- Traceable: Capable of being traced to requirements.
- Repeatable: Can be used to perform the test over and over.
- Reusable: Can be reused if necessary

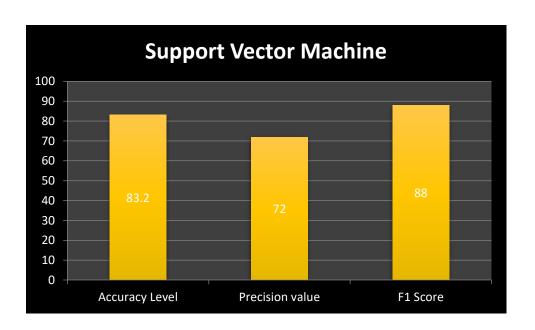
#### 8.2 USER ACCEPTANCE TESTING

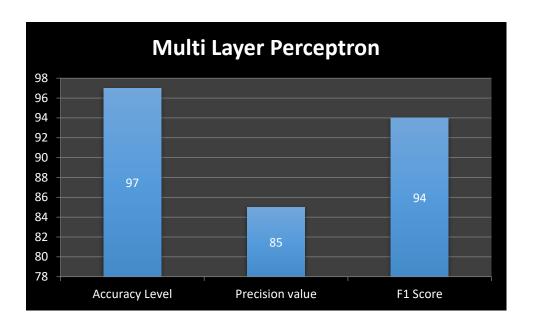
Acceptance testing can be defined in many ways, but a simple definition is the succeeds when the software functions in a manner that can be reasonable expected by the customer. After the acceptance test has been conducted, one of the two possible conditions exists. This is to fine whether the inputs are accepted by the database or other validations. For example accept only numbers in the numeric field, date format data in the date field. Also the null check for the not null fields. If any error occurs then show the error messages. The function of performance characteristics to specification and is accepted. A deviation from specification is uncovered and a deficiency list is created. User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

## 9.RESULTS

## 9.1 PERFORMANCE METRICS







## 10.ADVANTAGES & DISADVANTAGES

## **ADVANTAGES**

- It can be applied to resolve challenging nonlinear issues.
- It effectively manages vast volumes of input data.
- Following training, quickly makes predictions.
- Even with less samples, the same accuracy ratio is still possible.

## **DISADVANTAGES**

- Labeled data based classification
- Provide high number of false positive
- Binary classification can be occurred
- Computational complexity

# 11.CONCLUSION

We presented a machine learning approach for crop yield prediction, which demonstrated superior performance in Crop Challenge using large datasets of products. The approach used deep neural networks to make yield predictions (including yield, check yield, and yield difference) based on genotype and environment data. The carefully designed deep neural networks were able to learn nonlinear and complex relationships between genes, environmental conditions, as well as their interactions from historical data and make reasonably accurate predictions of yields for new hybrids planted in new locations with known weather conditions. Performance of the model was found to be relatively sensitive to the quality of weather prediction, which suggested the importance of weather prediction techniques. We trained two deep neural networks, one for yield and the other for check yield, and then used the difference of their outputs as the prediction for yield difference. This model structure was found to be more effective than using one single neural network for yield difference, because the genotype and environment effects are more directly related to the yield and check yield than their difference. In modern era, the deep neural network is the prominent tool in agricultural industry for providing support to farmers in monitoring crop yield based on multiple parameters. Thus, the machine learning model provides high accuracy in detecting the suitable crop identification compared to other methodologies

# 12.FUTURE SCOPE

This project describes crop yield prediction ability of the algorithm. In future we can determine the efficient algorithm based on their accuracy metrics that will helps to choose an efficient algorithm for crop yield prediction. Our research suggests that farmers' decisions about the production of non-rice crops on different plots and the use of resources other than land, such as labour and outside inputs, continue to influence their decisions regarding the production of rice. Rice production may continue to be atomistic, with many farmers producing tiny amounts of rice rather than a small number of farmers producing big amounts.

#### 13.APPENDIX

# **SOURCE CODE**

```
from flask import render_template, redirect, url_for, request
import sys
import pickle
import numpy as np
import ibm_db
import pandas
import ibm_db_dbi
from sqlalchemy import create_engine
engine = create_engine('sqlite://',
            echo = False)
dsn hostname = "125f9f61-9715-46f9-9399-
c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud"
dsn_uid = "jft01867"
dsn_pwd = "UkVOMQdSAOdR9NmF"
dsn_driver = "{IBM DB2 ODBC DRIVER}"
dsn_database = "BLUDB"
dsn_port = "30426"
dsn_protocol = "TCPIP"
dsn_security = "SSL"
dsn = (
  "DRIVER={0};"
  "DATABASE={1};"
  "HOSTNAME={2};"
  "PORT={3};"
  "PROTOCOL={4};"
  "UID={5};"
  "PWD={6};"
```

```
"SECURITY={7};").format(dsn_driver, dsn_database, dsn_hostname, dsn_port,
dsn_protocol, dsn_uid, dsn_pwd,dsn_security)
try:
  conn = ibm_db.connect(dsn, "", "")
  print ("Connected to database: ", dsn_database, "as user: ", dsn_uid, "on host: ",
dsn_hostname)
except:
  print ("Unable to connect: ", ibm_db.conn_errormsg() )
app = Flask(__name__)
app.config['DEBUG']
app.config['SECRET_KEY'] = '7d441f27d441f27567d441f2b6176a'
@app.route("/")
def homepage():
  return render_template('index.html')
@app.route("/AdminLogin")
def AdminLogin():
  return render_template('AdminLogin.html')
@app.route("/UserLogin")
def UserLogin():
  return render_template('UserLogin.html')
@app.route("/NewUser")
def NewUser():
  return render_template('NewUser.html')
@app.route("/NewQuery1")
def NewQuery1():
  return render_template('NewQueryReg.html')
```

```
@app.route("/AdminHome")
def AdminHome():
  conn = ibm_db.connect(dsn, "", "")
  pd_conn = ibm_db_dbi.Connection(conn)
  selectQuery = "SELECT * from regtb "
  dataframe = pandas.read_sql(selectQuery, pd_conn)
  dataframe.to_sql('Employee_Data',
            con=engine,
           if_exists='append')
  # run a sql query
  data = engine.execute("SELECT * FROM Employee_Data").fetchall()
  return render_template('AdminHome.html',data=data)
@app.route("/UserHome")
def UserHome():
  user = session['uname']
  conn = ibm db.connect(dsn, "", "")
  pd_conn = ibm_db_dbi.Connection(conn)
  selectQuery = "SELECT * FROM regtb where UserName= "" + user + "" "
  dataframe = pandas.read_sql(selectQuery, pd_conn)
  dataframe.to_sql('booktb1', con=engine, if_exists='append')
  data = engine.execute("SELECT * FROM booktb1").fetchall()
  return render_template('UserHome.html',data=data)
@app.route("/UQueryandAns")
def UQueryandAns():
  uname = session['uname']
  conn = ibm_db.connect(dsn, "", "")
```

```
pd_conn = ibm_db_dbi.Connection(conn)
  selectQuery = "SELECT * FROM querytb where UserName= "" + uname + "" "
  dataframe = pandas.read_sql(selectQuery, pd_conn)
  dataframe.to_sql('booktb1', con=engine, if_exists='append')
  data = engine.execute("SELECT * FROM booktb1").fetchall()
  return render_template('UserQueryAnswerinfo.html', data=data)
@app.route("/adminlogin", methods=['GET', 'POST'])
def adminlogin():
  error = None
  if request.method == 'POST':
   if request.form['uname'] == 'admin' or request.form['password'] == 'admin':
      conn = ibm_db.connect(dsn, "", "")
      pd_conn = ibm_db_dbi.Connection(conn)
      selectQuery = "SELECT * FROM regtb "
      dataframe = pandas.read_sql(selectQuery, pd_conn)
      dataframe.to_sql('booktb1', con=engine, if_exists='append')
      data = engine.execute("SELECT * FROM booktb1").fetchall()
      return render_template('AdminHome.html', data=data)
    else:
    return render_template('index.html', error=error)
@app.route("/userlogin", methods=['GET', 'POST'])
def userlogin():
  if request.method == 'POST':
```

```
username = request.form['uname']
    password = request.form['password']
    session['uname'] = request.form['uname']
    conn = ibm db.connect(dsn, "", "")
    pd_conn = ibm_db_dbi.Connection(conn)
    selectQuery = "SELECT * from regtb where UserName="" + username + "" and
password="" + password + """
    dataframe = pandas.read_sql(selectQuery, pd_conn)
    if dataframe.empty:
       data1 = 'Username or Password is wrong'
       return render_template('goback.html', data=data1)
    else:
       print("Login")
      selectQuery = "SELECT * from regtb where UserName="" + username + "' and
password="" + password + """
       dataframe = pandas.read_sql(selectQuery, pd_conn)
       dataframe.to_sql('Employee_Data',
                con=engine,
                if_exists='append')
       # run a sql query
       print(engine.execute("SELECT * FROM Employee_Data").fetchall())
       return render_template('UserHome.html', data=engine.execute("SELECT * FROM
Employee_Data").fetchall())
@app.route("/newuser", methods=['GET', 'POST'])
def newuser():
```

```
if request.method == 'POST':
     name1 = request.form['name']
     gender1 = request.form['gender']
     Age = request.form['age']
     email = request.form['email']
     pnumber = request.form['phone']
     address = request.form['address']
     uname = request.form['uname']
     password = request.form['psw']
     conn = ibm_db.connect(dsn, "", "")
     insertQuery = "INSERT INTO regtb VALUES ("" + name1 + "","" + gender1 + "","" +
Age + "'," + email + "'," + pnumber + "'," + address + "'," + uname + "'," + password + "')"
     insert_table = ibm_db.exec_immediate(conn, insertQuery)
     print(insert_table)
     # return 'file register successfully'
  return render_template('UserLogin.html')
@app.route("/newquery", methods=['GET', 'POST'])
def newquery():
  if request.method == 'POST':
     uname = session['uname']
     nitrogen = request.form['nitrogen']
     phosphorus = request.form['phosphorus']
     potassium = request.form['potassium']
     temperature = request.form['temperature']
     humidity = request.form['humidity']
     ph = request.form['ph']
     rainfall = request.form['rainfall']
```

```
location = request.form['select']
     nit = float(nitrogen)
     pho = float(phosphorus)
     po = float(potassium)
     te = float(temperature)
     hu = float(humidity)
     phh = float(ph)
     ra = float(rainfall)
     # age = int(age)
     filename = 'crop-prediction-rfc-model.pkl'
     classifier = pickle.load(open(filename, 'rb'))
     data = np.array([[nit, pho, po, te, hu, phh, ra]])
     my_prediction = classifier.predict(data)
     print(my_prediction)
     crop = "
     fertilizer = "
     if my_prediction == 0:
       Answer = 'Predict'
       crop = 'rice'
       fertilizer = '4 kg of gypsum and 1 kg of DAP/cent can be applied at 10 days after
sowing'
     elif my_prediction == 1:
       Answer = 'Predict'
       crop = 'maize'
```

```
fertilizer = 'The standard fertilizer recommendation for maize consists of 150 kg ha-1
NPK 14-23-14 and 50 kg ha-1 urea'
    elif my_prediction == 2:
       Answer = 'Predict'
       crop = 'chickpea'
       fertilizer = 'The generally recommended doses for chickpea include 20–30 kg nitrogen
(N) and 40–60 kg phosphorus (P) ha-1. If soils are low in potassium (K), an application of 17
to 25 kg K ha-1 is recommended'
    elif my_prediction == 3:
       Answer = 'Predict'
       crop = 'kidneybeans'
       fertilizer = 'It needs good amount of Nitrogen about 100 to 125 kg/ha'
    elif my_prediction == 4:
       Answer = 'Predict'
       crop = 'pigeonpeas'
       fertilizer = 'Apply 25 - 30 kg N, 40 - 50 k g P 2 O 5, 30 kg K 2 O per ha area as Basal
dose at the time of sowing.'
    elif my_prediction == 5:
       Answer = 'Predict'
       crop = 'mothbeans'
       fertilizer = 'The applications of 10 kg N+40 kg P2O5 per hectare have proved the
effective starter dose'
    elif my_prediction == 6:
       Answer = 'Predict'
       crop = 'mungbean'
       fertilizer = 'Phosphorus and potassium fertilizers should be applied at 50-50 kg ha-1'
    elif my_prediction == 7:
       Answer = 'Predict'
```

```
crop = 'blackgram'
       fertilizer = 'The recommended fertilizer dose for black gram is 20:40:40 kg NPK/ha.'
    elif my_prediction == 8:
       Answer = 'Predict'
       crop = 'lentil'
       fertilizer = 'The recommended dose of fertilizers is 20kg N, 40kg P, 20 kg K and 20kg
S/ha.'
    elif my_prediction == 9:
       Answer = 'Predict'
       crop = 'pomegranate'
       fertilizer = 'The recommended fertiliser dose is 600-700 gm of N, 200-250 gm of
P2O5 and 200-250 gm of K2O per tree per year'
    elif my_prediction == 10:
       Answer = 'Predict'
       crop = 'banana'
       fertilizer = 'Feed regularly using either 8-10-8 (NPK) chemical fertilizer or organic
composted manure'
    elif my_prediction == 11:
       Answer = 'Predict'
       crop = 'mango'
       fertilizer = '50 gm zinc sulphate, 50 gm copper sulphate and 20 gm borax per
tree/annum are recommended'
    elif my_prediction == 12:
       Answer = 'Predict'
       crop = 'grapes'
       fertilizer = 'Use 3 pounds (1.5 kg.) of potassium sulfate per vine for mild deficiencies
or up to 6 pounds (3 kg.)'
    elif my_prediction == 13:
```

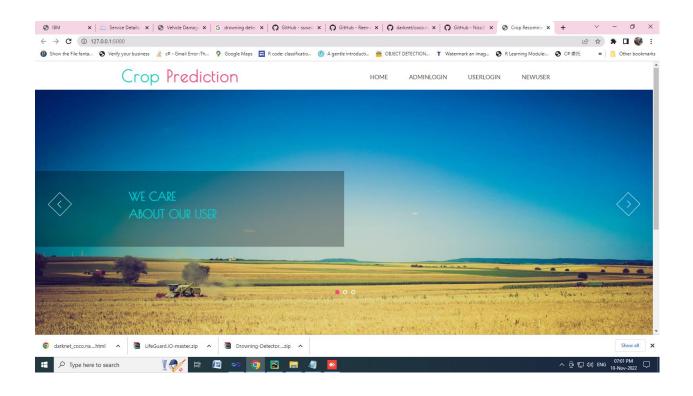
```
Answer = 'Predict'
       crop = 'watermelon'
       fertilizer = 'Apply a fertilizer high in phosphorous, such as 10-10-10, at a rate of 4
pounds per 1,000 square feet (60 to 90 feet of row)'
     elif my_prediction == 14:
       Answer = 'Predict'
       crop = 'muskmelon'
       fertilizer = 'Apply FYM 20 t/ha, NPK 40:60:30 kg/ha as basal and N @ 40 kg/ha 30
days after sowing.'
     elif my_prediction == 15:
       Answer = 'Predict'
       crop = 'apple'
       fertilizer = 'Apple trees require nitrogen, phosphorus and potassium, Common granular
20-10-10 fertilizer is suitable for apples'
     elif my_prediction == 16:
       Answer = 'Predict'
       crop = 'orange'
       fertilizer = 'Orange farmers often provide 5.5 - 7.7 lbs (2.5-3.5 \text{ kg}) P2O5 in every
adult tree for 4-5 consecutive years'
     elif my_prediction == 17:
       Answer = 'Predict'
       crop = 'papaya'
       fertilizer = 'Generally 90 g of Urea, 250 g of Super phosphate and 140 g of Muriate of
Potash per plant are recommended for each application'
     elif my_prediction == 18:
       Answer = 'Predict'
       crop = 'coconut'
```

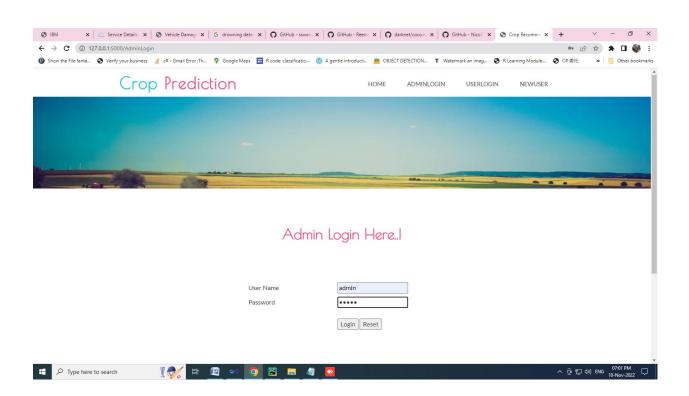
fertilizer = 'Organic Manure @50kg/palm or 30 kg green manure, 500 g N, 320 g P2O5 and 1200 g K2O/palm/year in two split doses during September and May'

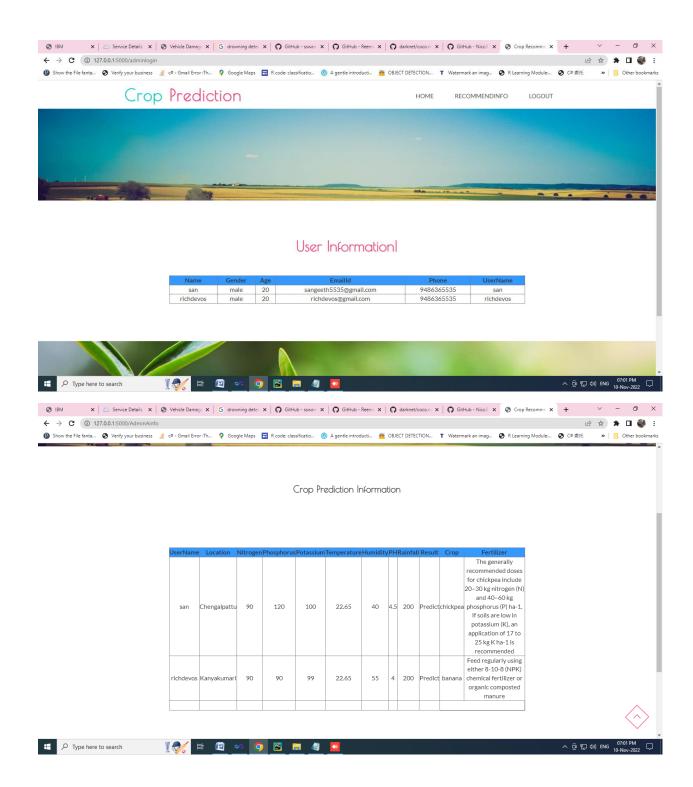
```
elif my_prediction == 19:
       Answer = 'Predict'
       crop = 'cotton'
       fertilizer = 'N-P-K 20-10-10 per hectare during sowing (through the sowing machine)'
     elif my_prediction == 20:
       Answer = 'Predict'
       crop = 'jute'
       fertilizer = 'Apply 10 kg of N at 20 - 25 days after first weeding and then again on 35
- 40 days after second weeding as top dressing'
     elif my_prediction == 21:
       Answer = 'Predict'
       crop = 'coffee'
       fertilizer = 'Coffee trees need a lot of potash, nitrogen, and a little phosphoric acid.
Spread the fertilizer in a ring around each Coffee plant'
     else:
       Answer = 'Predict'
       crop = 'Crop info not Found!'
     conn = ibm_db.connect(dsn, "", "")
     pd_conn = ibm_db_dbi.Connection(conn)
     insertQuery = "INSERT INTO Querytb VALUES ('" + uname + "','" + nitrogen + "','" +
phosphorus + "',"" + potassium + "',""+temperature+"',""+humidity +"',""+ ph+"',""+ rainfall
+"', 'Predict', '"+ crop +"', '"+fertilizer +"', ""+location+"')"
     insert_table = ibm_db.exec_immediate(conn, insertQuery)
     print(insert_table)
     uname = session['uname']
```

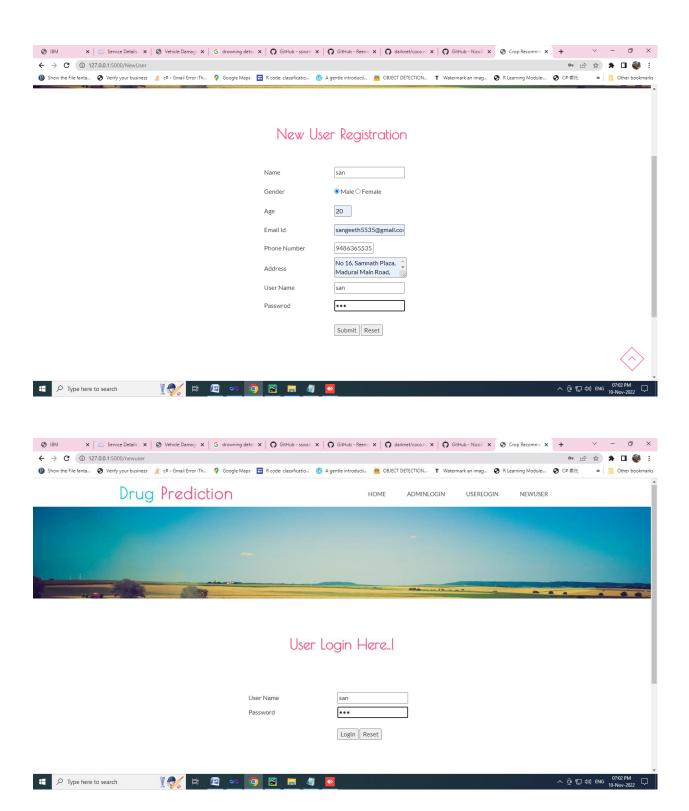
```
selectQuery = "SELECT * FROM Querytb where UserName=""+ uname+"" "
    dataframe = pandas.read_sql(selectQuery, pd_conn)
    dataframe.to_sql('booktb1', con=engine, if_exists='append')
    data = engine.execute("SELECT * FROM booktb1").fetchall()
    return render_template('UserQueryAnswerinfo.html', wait=data)
@app.route("/AdminAinfo")
def AdminAinfo():
  conn = ibm_db.connect(dsn, "", "")
  pd_conn = ibm_db_dbi.Connection(conn)
  selectQuery = "SELECT * FROM Querytb "
  dataframe = pandas.read_sql(selectQuery, pd_conn)
  dataframe.to_sql('booktb1', con=engine, if_exists='append')
  data = engine.execute("SELECT * FROM booktb1").fetchall()
  return render_template('AdminAnswer.html', data=data
if __name__ == '__main__':
  app.run(debug=True, use_reloader=True)
```

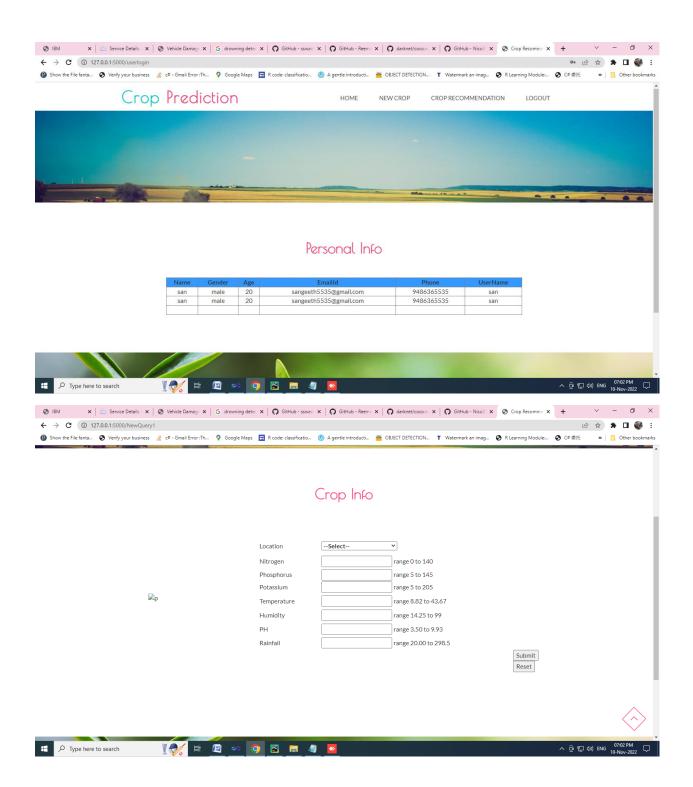
# **SCREENSHOTS:**

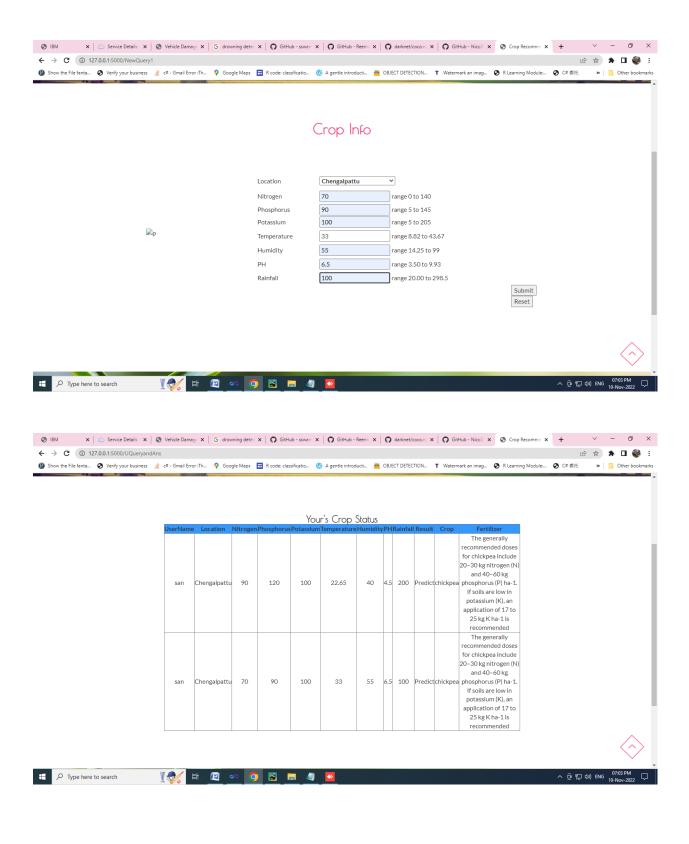












# GITHUB & PROJECT DEMO LINK

| GITHUB LINK  |
|--|
| https://github.com/IBM-EPBL/IBM-Project-15221-1659595129.git |
| PROJECT DEMO LINK  |
| https://drive.google.com/file/d/1aLb3VXwowBj6WTlNiqH         |
| PRm-3mwmCNInz/view?usp=drivesdk                              |