

**PROJECT REPORT**  
**ON**  
**“Crop Recommendation System”**  
**SUBMITTED TO**  
**ROURKELA INSTITUTE OF MANAGEMENT STUDIES**  
(As a partial fulfillment of the requirement for the award of degree)  
**FOR**  
**“MASTER IN COMPUTER APPLICATIONS”**  
(2022-2024)  
**SUBMITTED BY**  
**CHANDAN KUMAR NAYAK**  
Registration Roll No: 2205260005  
MCA 4<sup>th</sup> SEMESTER

**ROURKELA INSTITUTE OF MANAGEMENT STUDIES**  
(Affiliated to Biju Patnaik University of Technology, Odisha)  
Rourkela-769015



**Rourkela Institute of Management Studies**

Rourkela

Department of Computer Science

Rourkela Institute of Management Science

Chhend, Rourkela-15, Odisha

Phone: 0661 2480482

Fax: 91-0661-1480665

Mail: [rkl\\_rimsgrol@sancharnet.in](mailto:rkl_rimsgrol@sancharnet.in)

Visit: <https://rimsedu.ac.in/>

**CERTIFICATE OF EXAMINATION**

This is to certify that this project report entitled “**Crop Recommendation System**” submitted by **CHANDAN KUMAR NAYAK** of 4<sup>th</sup> Semester, **Rourkela Institute of Management Studies, Rourkela**, is accepted as partial fulfillment of requirements for the degree in **Master in Computer Applications**, under **Biju Patnaik University of Technology, Rourkela**, this has been verified by us and found be original up to our satisfaction.

**External Examiner**



**Rourkela Institute of Management Studies**

Rourkela

Department of Computer Science

Rourkela Institute of Management Science

Chhend, Rourkela-15, Odisha

Phone: 0661 2480482

Fax: 91-0661-1480665

Mail: [rkl\\_rimsgrol@sancharnet.in](mailto:rkl_rimsgrol@sancharnet.in)

Visit: <https://rimssedu.ac.in/>

**CERTIFICATE OF EXAMINATION**

This is to certify that this project report entitled “**Crop Recommendation System**” submitted by **CHANDAN KUMAR NAYAK** of 4<sup>th</sup> Semester, **Rourkela Institute of Management Studies, Rourkela**, is accepted as partial fulfillment of requirements for the degree in **Master in Computer Applications**, under **Biju Patnaik University of Technology, Rourkela**, this has been verified by us and found to be original up to our satisfaction.

**Internal Examiner**



**Rourkela Institute of Management Studies**

Rourkela

Department of Computer Science

Rourkela Institute of Management Science

Chhend, Rourkela-15, Odisha

Phone: 0661 2480482

Fax: 91-0661-1480665

Mail: [rkl\\_rimsgrol@sancharnet.in](mailto:rkl_rimsgrol@sancharnet.in)

Visit: <https://rimssedu.ac.in/>

**CERTIFICATE**

This is to certify that this project report entitled “**Crop Recommendation System**” has been submitted by **CHANDAN KUMAR NAYAK** of M.C.A 4<sup>th</sup> Semester of session 2022-2024, **Rourkela Institute of Management Studies, Rourkela**, has been examined by us.

He is found fit and approved for the award of “**Master in Computer Applications**” Degree.

To the best my knowledge this work has not been submitted for the award of any other degree.

I wish all success in his life.

Dean Academic  
RIMS, Rourkela



Prof. Bibhudendu Panda  
Head of the Department, IT  
**Rourkela Institute of Management Science**  
Rourkela Institute of Management Science  
Chhend, Rourkela-15, Odisha  
Phone: 0661 2480482  
Fax: 91-0661-1480665  
Mail: [rkl\\_rimsgrol@sancharnet.in](mailto:rkl_rimsgrol@sancharnet.in)  
Visit: <https://rimsedu.ac.in/>

### CERTIFICATE

This is to certify that **CHANDAN KUMAR NAYAK** student of **M.C.A,**  
**Rourkela Institute of Management Studies, Rourkela, and Odisha** of  
session 2022-2024 has completed the project successfully.

I wish all success in his life.

**(Prof. Bibhudendu Panda)**



Prof. Bibhudendu Panda  
Head of the Department, IT  
**Rourkela Institute of Management Science**  
Rourkela Institute of Management Science  
Chhend, Rourkela-15, Odisha  
Phone: 0661 2480482  
Fax: 91-0661-1480665  
Mail: [rkl\\_rimsgrol@sancharnet.in](mailto:rkl_rimsgrol@sancharnet.in)  
Visit: <https://rimssedu.ac.in/>

### **DECLARATION**

I, **CHANDAN KUMAR NAYAK**, hereby declare that the project report entitled “**Crop Recommendation System**” is of my work. The above work I submitted to “**Biju Patnaik University of Technology Rourkela**” for the award of “**Master in Computer Applications**” Degree.

To the best of my knowledge, this work has not been submitted or published anywhere for the award of any degree.

**CHANDAN KUMAR NAYAK**

**Rourkela Institute of Management Studies**  
**Rourkela**

## **ABSTRACT**

The Crop Recommendation System for agriculture is based on various inputs parameters. This proposes a hybrid model for recommending crops to South Indian states by considering various attributes such as soil type, rainfall, groundwater level, temperature, fertilizers, pesticides and season.

The concept of this paper is to implement the crop selection method so that this method helps in solving many agriculture and farmers problems. This improves our Indian economy by maximizing the yield rate of crop production. Different types of land condition. So, the quality of the crops is identified using ranking process. By this process the rate of the low quality and high-quality crop is also notified







## **CONTRIBUTION OF INDIVIDUAL TEAM**

### **MEMBERS**

<b>Name of the Student(s)</b>	<b>Registration Number</b>	<b>Contributions</b>
NIRLIPT DAS	2205260014	Database and some backend part along with report contributions.
SATYABRATA PANDA	2205260020	Presentation part(ppt) and some report part.
GEETIKA GARNAIK	2205260008	Documentation and some frontend part along with report contributions.
CHANDAN KUMAR NAYAK	2205260005	Database part and Documentation part with report contribution.



## **GANTT CHART**

<b>ID</b>	<b>TASK NAME</b>	<b>NO. OF DAYS</b>	<b>BAR REPRESENTATION</b>
<b>1</b>	<b>Project Management</b>		
<b>1.1</b>	<b>Project Initiation</b>	5	
<b>1.2</b>	<b>Project Planning</b>	9	
<b>2</b>	<b>Analysis</b>	15	
<b>3</b>	<b>Design</b>	25	
<b>4</b>	<b>Implementation</b>	35	
<b>5</b>	<b>Testing</b>	4	

<b>6</b>	<b>Evaluation</b>	2	
----------	-------------------	---	---

## Table of Contents

	<b>Title Page</b>	<b>01</b>
	<b>Certificate of examination</b>	<b>02</b>
	<b>Certificate of examination</b>	<b>03</b>
	<b>Certificate</b>	<b>04</b>
	<b>Certificate</b>	<b>05</b>
	<b>Declaration</b>	<b>06</b>
	<b>Acknowledgement</b>	<b>07</b>
	<b>Abstract</b>	<b>08</b>
	<b>Contribution of individual team members</b>	<b>09</b>
	<b>List of figures</b>	<b>10</b>
	<b>Timeline/ Gantt Chart</b>	<b>11</b>
1.	<b>INTRODUCTION</b>	<b>13</b>
	1.1 Preface 1.2 Problem Definition and Objectives(s) 1.3 Motivations 1.4 Project Overview/ Specification 1.5 Hardware Specification 1.6 Software Specification 1.7 Organization of the report	13 13 14 14 15 15 15
2.	<b>LITERATURE SURVEY</b>	<b>16</b>
	2.1 Preface 2.2 PostgreSQL 2.3 Proposed System	16 17 17

3.	<b>AIM AND SCOPE OF INVESTIGATION</b>	<b>18</b>
	3.1 Aim	18
	3.2 Scope	18
4.	<b>EXPERIMENTAL OR MATERIALS AND METHODS</b>	
	4.1 Random forest	19-24
	4.2 XGboost	25-31
	4.3 Svm	32-33
	4.4 Knn	33-34
	4.5 Decision tree	35-36
	4.6 Bagging	37-38
	4.7 Ada boost	38-39
	4.8 Gradient boosting	40
	4.9 Extra tree	40-41
	4.10 System Architecture	
	4.11 Software environment	42
	4.12 Modules	47
	4.613 Uml Diagram	48

5.	<b>DATA FLOW DIAGRAM</b>	<b>49</b>
	5.1 Data flow diagram 5.2 Use case diagram 5.3 Activity diagram 5.4 <b>INPUT AND OUTPUT DESIGN</b> 5.4.1 Input design 5.4.2 Output design 5.4.3 System testing	49-50 51-52 53-55 57 57 59 61-62
6.	6.1 HTML5 6.2 CSS 6.3 Install & Setup PyCharm	63-67 67-70 70-75
7.	<b>RESULT &amp; DISCUSSIONS</b>	<b>76</b>
	7.1 Screenshot of Our Coding section 7.2 Diagram of Database Connectivity	77-84 85
8.	<b>CONCLUSION</b>	<b>91</b>
9.	<b>TAKE AWAY FROM THE PROJECT</b>	<b>94</b>
	9.1 References / Bibliography	94-95

# **1. INTRODUCTION**

## **1.1 Preface**

In the world of developing technologies, the success of sharing information will help the agriculturists in realizing and developing their potential. The

information sharing is that the valuable and timely information is being shared between agriculturists, either formally or informally. The willingness of information sharing refers to the open attitude among agriculturists. This open attitude determines the degree and scope of information sharing. Using web-technologies like html and CSS we build the web application; we create dataset by gathering data from multiple resources and place them in place which is used to predict the price of the crop and results are subjected to non-linear test later priorities are set and rankings are given to the list of crops. Place information in our application and share that information to agriculturists whose data is collected and stored in the

PostgreSQL server. we software to automatically send the updated information to the agriculturists in the form of text message. So that agriculturists no need to go to nearby towns and cities to know the updated information. We will be machine learning algorithms to predict the price of the crop for the next two months. For prediction purpose we will be using Support vector machine (SVM), Naïve Baye's (NB) and K-Nearest Neighbor (KNN) algorithms to predict the cost of the crop production. Further, a ranking process is applied for decision making in order to select the classifiers results.

## **1.2 Problem Definition and Objective(s):**

- Data set collection from various sources.
- Data parsing and cleansing technique is applied to make the raw data into processing data.
- The data collected is subject to machine learning system along with run

time analysis makes an efficient crop value updating system.

- Usage of Ensemble of classifiers makes the model more robust and efficient.
- Ranking technique used in the project helps us to make efficient decisions.
- Creating a web application for user registrations and collection of data.
- The main objective is to obtain a better variety of crops that can be grown over the season. The proposed system would help to minimize the difficulties faced by farmers in choosing a crop and maximize the yield.
- The model predicts the crop yield by studying factors such as rainfall, temperature, area, season, soil type etc.

### **1.3 Motivation:**

Precision agriculture is in trend nowadays. Precision agriculture is a modern farming technique that uses the data of soil characteristics, soil types, crop yield data, weather conditions and suggests the farmers with the most optimal crop to grow in their farms for maximum yield and profit. This technique can reduce the crop failures and will help the farmers to take informed decision about their farming strategy.

### **1.4 Project Overview/ Specification:**

This project is to showcase of how this application will allow user to:

Technology based crop recommendation system for agriculture helps the farmers to increase the crop yield by recommending a suitable crop for their land with the help of geographic and the climatic parameters. The

proposed hybrid recommender model is found to be effective in recommending a suitable crop.

### **1.5 Hardware Specifications:**

- 8GB RAM
- 512GB Hard disk
- Processor should be greater than i3

### **1.6 Software Specifications:**

- Windows: 8,10,11
- Language: Python
- Database: PostgreSQL

### **1.7 Organization of the project:**

The entire purpose of this project is to design a website that will answer the queries related to a particular soil content asked by the user. Also, the website will give the better result.

The front end of the entire project is designed using HTML & CSS, and then we run it in the browser to get convinced that our frontend works perfectly fine. Then we designed the backend part using PYTHON and flask which is connected to the front-end part and the database so that our data is well stored for each and every user.

We went through the overall application setup and tried to make it as user-friendly as possible so that each and every user can access it well. After completion of the entire project, we set up the report by classifying the entire report into several headings and sub points and then we described the entire project.

## **2. LITERATURE STUDY**

To keep up nutrition levels in the soil in case of deficiency, fertilizers are added to soil. The standard issue existing among the Indian agriculturists choose approximate amount of fertilizers and add them manually. Excess or

deficient extension of fertilizers can harm the plants life and reduce the yield. This paper gives overview of various data mining frameworks used on cultivating soil dataset for fertilizer recommendation.

## PostgreSQL:

PostgreSQL, also known as Postgres, is a free and open-source relational database management system emphasizing extensibility and SQL compliance. It was originally named POSTGRES, referring to its origins as a successor to the Ingres database developed at the University of California, Berkeley. **PostgreSQL** is a powerful, open-source object-relational database system.

### 2.2 Setting up the PostgreSQL:

1. Open the PostgreSQL install file to start the installation wizard.  
Click **next** to continue.
2. Choose an install location for PostgreSQL and click **Next** to proceed.
3. Select which software components you want to install:
  - i. **PostgreSQL Server:** Installs the PostgreSQL database server.
  - ii. **pgAdmin 4:** Provides a graphical interface for managing PostgreSQL databases.
  - iii. **Stack Builder:** Allows you to download and install additional tools to use with PostgreSQL.
  - iv. **Command Line Tools:** Installs the command line tool and client libraries. Required when installing PostgreSQL Server or pgAdmin 4.
2. Once you check the boxes next to the components you want to install, click **Next** to continue. Choose a database directory to store data and click **Next** to continue.
  3. Enter and retype the password for the database super user. Click **Next** to proceed.
  4. Enter the port number (5432) for the PostgreSQL server to listen on and click **Next** to continue.
  5. Choose the locale for the database to use. Selecting the **[Default locale]** option uses the locale settings for your operating system. Once you have chosen a locale, click **Next** to continue.



6. The last step offers a summary of the installation settings. Click **Back** if you want to change any of the settings you made, or click **Next** to proceed.
7. The setup wizard informs you it is ready to start the installation process. Click **Next** to begin installing PostgreSQL.

## 2.3 PROPOSED SYSTEM

In proposed system, the data analysis technology is used to update the crop yield rate change. The concept of this paper is to implement the crop selection method rate change. The concept of this paper is to implement the crop selection method so that this method helps in solving many agriculture and farmers problems. This improves our Indian economy by maximizing the yield rate of crop production.

Different types of land condition. So, the quality of the crops is identified using ranking process. By this process the rate of the low quality and high-quality crop is also intimated. The usage of ensemble of classifiers paves a path way to make a better decision on predictions due to the usage of multiple classifiers. Further, a ranking process is applied for decision making in order to select the classifiers results. This system is used to predict the cost of the fertilizers for further. This project uses Ensemble of classifiers such as Decision tree and Random Forest classifier. In addition, this project uses Ranking technique.

## ADVANTAGES OF PROPOSED SYSTEM

- Useful to people far away from towns/cities.
- Better time efficiency.
- Reduction of repeated work.

### **3. AIM AND SCOPE OF THE PRESENT INVESTIGATION**

#### **3.1 Aim:**

Our Aim from the project is to make a ML model which takes student data trains itself using various Machine Learning techniques and Algorithms (Random Forest, Decision Tree) and predict the yield and best fertilizer that suits for the crops in virtual environment by considering the overall factors that contribute in his overall yield. Secondly, to learn the required tech stacks and use it to make model with a python application and lastly to execute it get output about yield and best fertilizer for the crop.

#### **3.2 Scope:**

This Project can be used to get the student performance with more accuracy than any other model published earlier and we can also make some mobile or web application based on the model.

### **4. EXPERIMENTAL OR MATERIALS AND METHODS;**

#### **ALGORITHMS USED.**

##### **4.1 RANDOM FOREST ALGORITHM:**

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."

Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of over fitting.

Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps:

**Step-1:** Select random K data points from the training set.

**Step-2:** Build the decision trees associated with the selected data points.

**Step-3:** Choose the number N for decision trees that you want to build.

**Step-4:** Repeat Step 1 & 2.

**Step-5:** For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

### **Assumptions for Random Forest**

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random Forest classifier:

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

### **Why use Random Forest?**

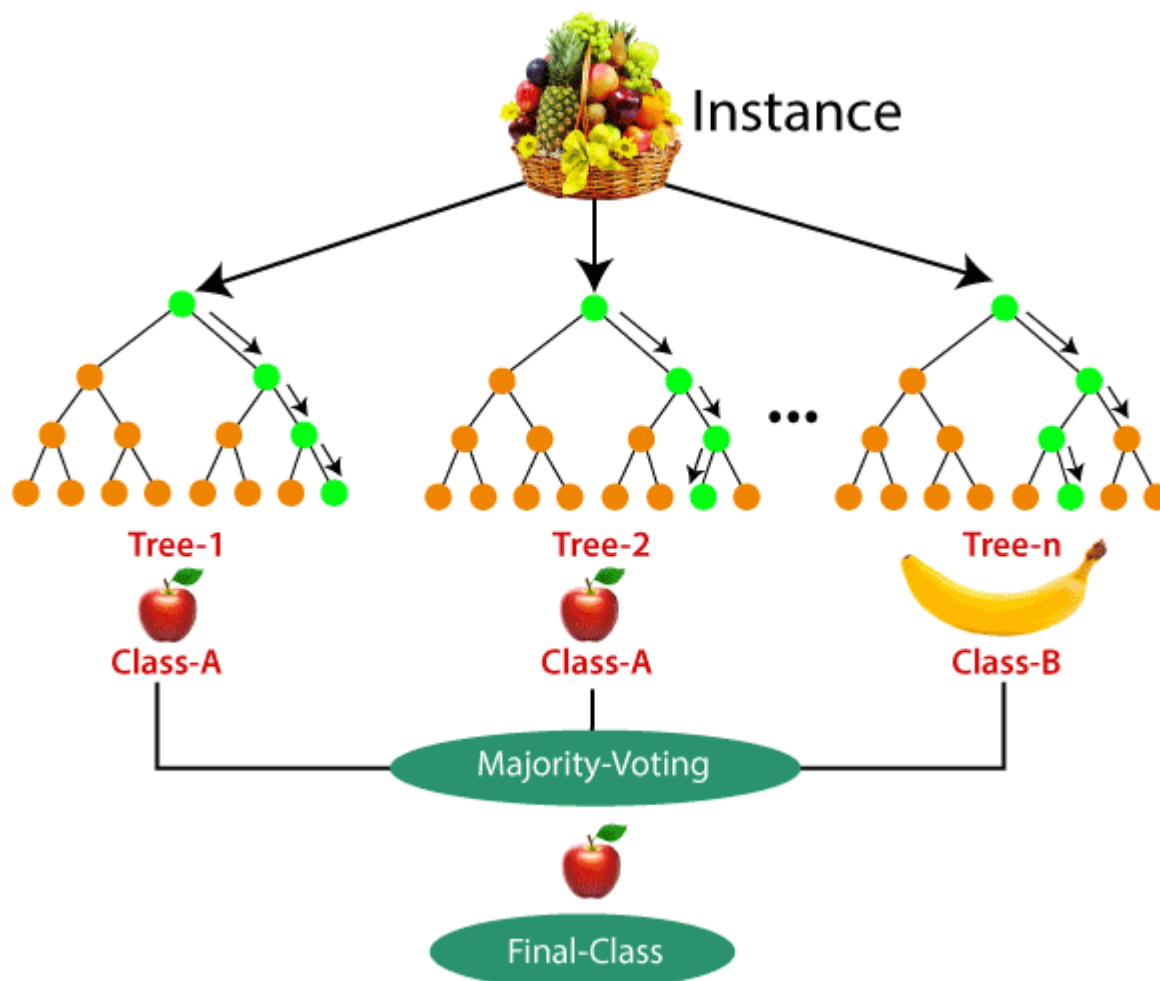
Below are some points that explain why we should use the Random Forest algorithm:

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.

- It can also maintain accuracy when a large proportion of data is missing.

## How does Random Forest algorithm work?

**Example:** Suppose there is a dataset that contains multiple fruit images. So, this dataset is given to the Random Forest classifier. The dataset is divided into subsets and given to each decision tree. During the training phase, each decision tree produces a prediction result, and when a new data point occurs, then based on the majority of results, the Random Forest classifier predicts the final decision. Consider the below image:



## Applications of Random Forest: -

There are mainly four sectors where Random Forest mostly used:

1. **Banking:** Banking sector mostly uses this algorithm for the identification of loan risk.
2. **Medicine:** With the help of this algorithm, disease trends and risks of the disease can be identified.
3. **Land Use:** We can identify the areas of similar land use by this algorithm.
4. **Marketing:** Marketing trends can be identified using this algorithm.

## Advantages of Random Forest

- Random Forest is capable of performing both Classification and Regression tasks.
- It is capable of handling large datasets with high dimensionality.
- It enhances the accuracy of the model and prevents the overfitting issue.

## Disadvantages of Random Forest

- Although random forest can be used for both classification and regression tasks, it is not more suitable for Regression tasks.

## Python Implementation of Random Forest Algorithm

Now we will implement the Random Forest Algorithm tree using Python. For this, we will use the same dataset "user\_data.csv", which we have used in previous classification models. By using the same dataset, we can compare the Random Forest classifier with other classification models.

Implementation Steps are given below:

- Data Pre-processing step
- Fitting the Random Forest algorithm to the Training set

- Predicting the test result
- Test accuracy of the result (Creation of Confusion matrix)
- Visualizing the test set result.

Random Forest: Accuracy: 0.9909

## 4.2 What is XGBoost Algorithm?

XGBoost is a robust machine-learning algorithm that can help you understand your data and make better decisions.

XGBoost is an implementation of gradient-boosting decision trees. It has been used by data scientists and researchers worldwide to optimize their machine-learning models.

### What is XGBoost in Machine Learning?

XGBoost is designed for speed, ease of use, and performance on large datasets. It does not require optimization of the parameters or tuning, which means that it can be used immediately after installation without any further configuration.

#### XGBoost Features:

- XGBoost offers regularization, which allows you to control overfitting by introducing L1/L2 penalties on the weights and biases of each tree. This feature is not available in many other implementations of gradient boosting.
- Another feature of XGBoost is its ability to handle sparse data sets using the weighted quantile sketch algorithm. This algorithm allows us to deal with non-zero entries in the feature matrix while retaining the same computational complexity as other algorithms like stochastic gradient descent.
- XGBoost also has a block structure for parallel learning. It makes it easy to scale up on multicore machines or clusters. It also uses cache awareness,

which helps reduce memory usage when training models with large datasets.

- Finally, XGBoost offers out-of-core computing capabilities using disk-based data structures instead of in-memory ones during the computation phase.

### **XgBoost Formula:**

XGBoost is a gradient boosting algorithm for supervised learning. It's a highly efficient and scalable implementation of the boosting algorithm, with performance comparable to that of other state-of-the-art machine learning algorithms in most cases.

### **XGBoost Benefits and Attributes:**

XGBoost is a highly portable library on OS X, Windows, and Linux platforms. It's also used in production by organizations across various verticals, including finance and retail.

XGBoost is open source, so it's free to use, and it has a large and growing community of data scientists actively contributing to its development. The library was built from the ground up to be efficient, flexible, and portable.

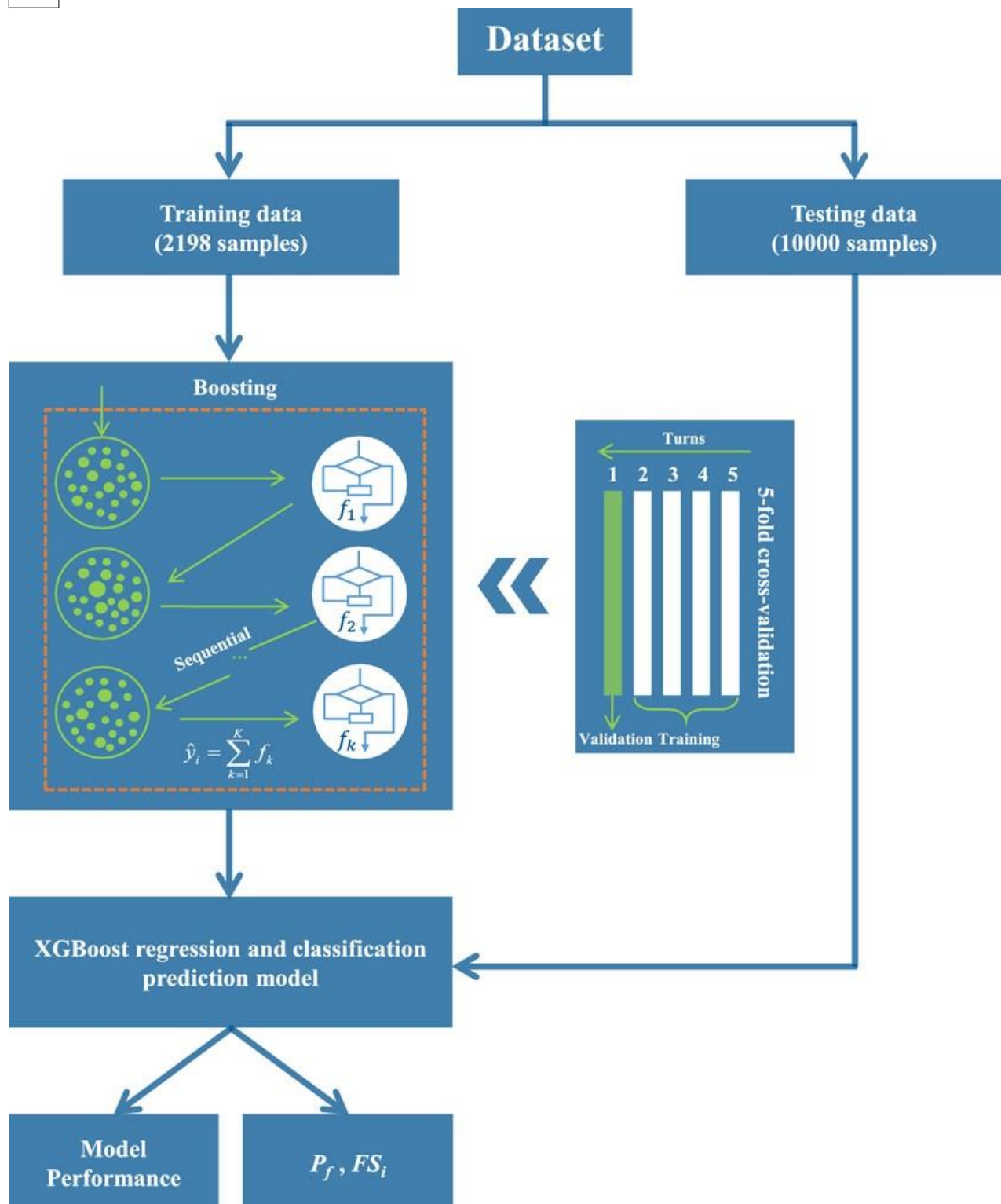
You can use XGBoost for classification, regression, ranking, and even user-defined prediction challenges! You can also use this library with other tools like H2O or Scikit-Learn if you want to get more out of your model-building process.

### **Why does XGBoost perform so well?**

XGBoost and Gradient Boosting Machines (GBMs) are both ensemble tree methods that apply the principle of boosting weak learners (CARTs generally) using the gradient descent architecture.

However, XGBoost improves upon the base GBM framework through systems optimization and algorithmic enhancements.





## System Optimization:

1. **Parallelization:** XGBoost approaches the process of sequential tree building using parallelized implementation. This is possible due to the interchangeable nature of loops used for building base learners; the outer loop that enumerates the leaf nodes of a tree, and the second inner loop that calculates the features. This nesting of loops limits parallelization because without completing the inner loop (more computationally demanding of the two), the outer loop cannot be started. Therefore, to improve run time, the order of loops is interchanged using initialization through a global scan of all instances and sorting using parallel threads. This switch improves algorithmic performance by offsetting any parallelization overheads in computation.
2. **Tree Pruning:** The stopping criterion for tree splitting within GBM framework is greedy in nature and depends on the negative loss criterion at the point of split. XGBoost uses 'max-depth' parameter as specified instead of criterion first, and starts pruning trees backward. This 'depth-first' approach improves computational performance significantly.
3. **Hardware Optimization:** This algorithm has been designed to make efficient use of hardware resources. This is accomplished by cache awareness by allocating internal buffers in each thread to store gradient statistics. Further enhancements such as 'out-of-core' computing optimize available disk space while handling big data-frames that do not fit into memory.

## Algorithmic Enhancements:

1. **Regularization:** It penalizes more complex models through both LASSO (L1) and Ridge (L2) regularization to prevent overfitting.
2. **Sparsity Awareness:** XGBoost naturally admits sparse features for inputs by automatically 'learning' best missing value depending on training loss and handles different types of sparsity patterns in the data more efficiently.

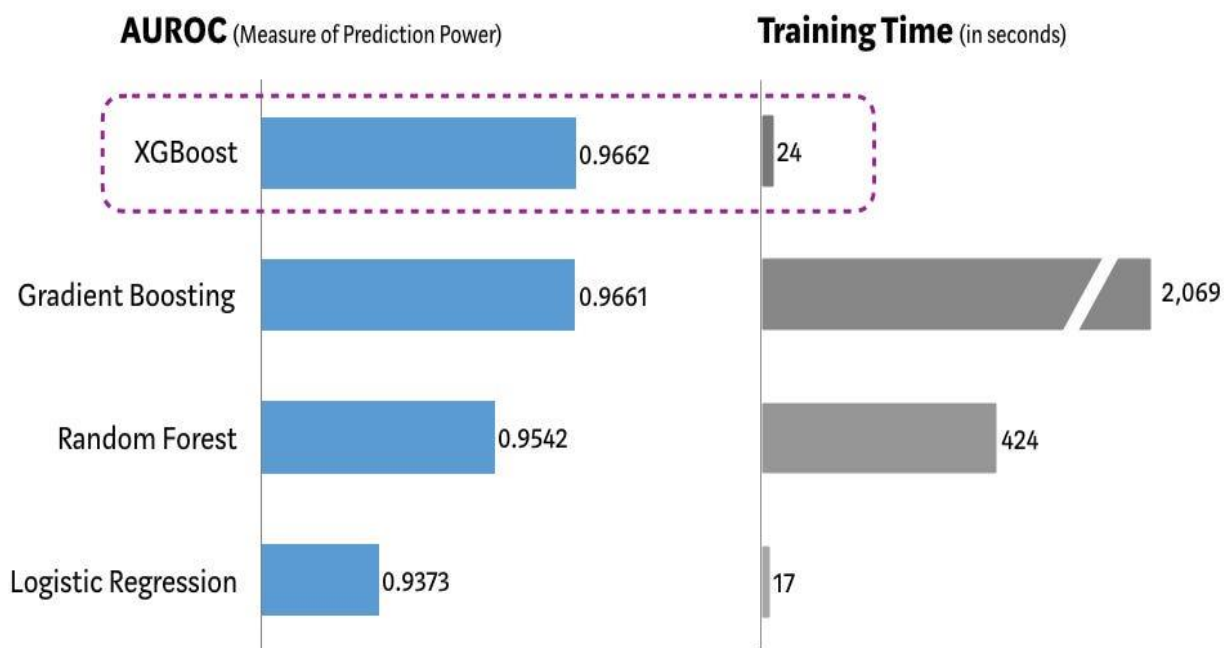
3. **Weighted Quantile Sketch:** XGBoost employs the distributed weighted Quantile Sketch algorithm to effectively find the optimal split points among weighted datasets.
4. **Cross-validation:** The algorithm comes with built-in cross-validation method at each iteration, taking away the need to explicitly program this search and to specify the exact number of boosting iterations required in a single run.

### Where is the proof?

We used Scikit-learn's 'Make Classification' data package to create a random sample of 1 million data points with 20 features (2 informative and 2 redundant). We tested several algorithms such as Logistic Regression, Random Forest, standard Gradient Boosting, and XGBoost.

### Performance Comparison using SKLearn's 'Make\_Classification' Dataset

(5 Fold Cross Validation, 1MM randomly generated data sample, 20 features)



XGBoost vs. Other ML Algorithms using SKLearn's Make Classification Dataset

As demonstrated in the chart above, XGBoost model has the best combination of prediction performance and processing time compared to other algorithms. Other rigorous benchmarking studies have produced

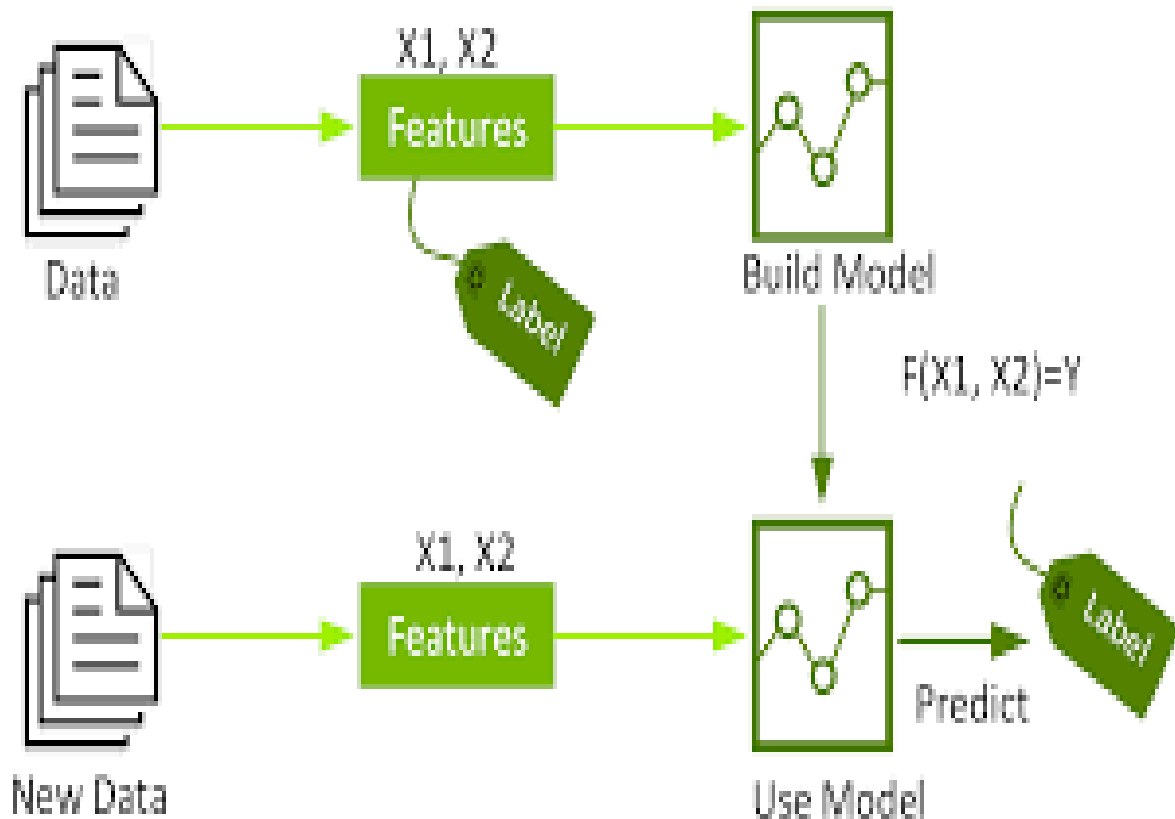
similar results. No wonder XGBoost is widely used in recent Data Science competitions.

So should we use just XGBoost all the time?

When it comes to Machine Learning (or even life for that matter), there is no free lunch. As Data Scientists, we must test all possible algorithms for data at hand to identify the champion algorithm. Besides, picking the right algorithm is not enough. We must also choose the right configuration of the algorithm for a dataset by tuning the hyper-parameters. Furthermore, there are several other considerations for choosing the winning algorithm such as computational complexity, explain ability, and ease of implementation. This is exactly the point where Machine Learning starts drifting away from science towards art, but honestly, that's where the magic happens!

### **What does the future hold?**

Machine Learning is a very active research area and already there are several viable alternatives to XGBoost. Microsoft Research recently released LightGBM framework for gradient boosting that shows great potential. CatBoost developed by Yandex Technology has been delivering impressive bench-marking results. It is a matter of time when we have a better model framework that beats XGBoost in terms of prediction performance, flexibility, explainability, and pragmatism. However, until a time when a strong challenger comes along, XGBoost will continue to reign over the Machine Learning world!



### 4.3 SUPPORT VECTOR MACHINE:

A crop recommendation system based on Support Vector Machines (SVM) is a machine learning algorithm that can predict which crops are best suited to a particular area based on various factors such as soil type, climate, and water availability. SVM is a powerful classification algorithm that can handle high-dimensional datasets with a relatively small number of observations, making it an ideal choice for crop recommendation systems. SVC (Support Vector Classifier) is a type of machine learning algorithm that falls under the category of supervised learning. It is a variant of the Support Vector Machine (SVM) algorithm that is used for classification tasks. SVC works by finding the hyperplane that best separates the different classes in the data. This hyperplane is chosen such that it maximizes the margin between the classes, i.e., the

distance between the hyperplane and the nearest data points of each class. Once the hyperplane is found, SVC can classify new data points by determining which side of the hyperplane they lie on. SVC is a powerful algorithm that can handle both linear and non-linear classification tasks. It can also handle high dimensional data and works well with small to medium-sized datasets. One of the advantages of SVC is that it is less prone to over fitting than other classification algorithms, as it maximizes the margin between classes. In Python, SVC is implemented in the scikit-learn library, which provides a range of tools for machine learning tasks. The SVC class in scikit-learn provides a range of hyper parameters that can be tuned to optimize the algorithm's performance, such as the regularization parameter and the kernel function. Overall, a crop recommendation system based on SVM and SVC can be a valuable tool for farmers to optimize crop yields, reduce waste, and increase profits while contributing to sustainable agriculture practices.

## **EXPLANATION:**

The first line imports the `train_test_split` function from the model selection module of scikit-learn, which is used to split the dataset into training and testing subsets. The function takes in the input data `x` and the target variable `y`, and splits them into random training and testing sets, with a test set size of 30% of the data, and a random state seed of 123. The next line imports the SVC class from the svm module of scikit-learn which is used to create an instance of the SVC classifier. The kernel parameter is set to 'linear', indicating that a linear kernel function will be used. The next line fits the SVC classifier to the training data by calling the `fit` method on the `svcclassifier` object. This trains the classifier on the training data, using the specified kernel function and other hyper parameters. The next line predicts the target variable values for the test data by calling the `predict` method on the `svcclassifier` object, passing in the test input data `x_test`. The predicted values are assigned to the variable `y_pred`. Finally, the last line prints out the predicted target variable values for the test data. This allows the user to evaluate the performance of the SVC classifier on the test data and compare it to the true target variable values `y_test`...

In conclusion, crop recommendation systems based on machine learning techniques, such as Support Vector Machines, can help farmers and agricultural organizations make data-driven decisions about crop selection, maximize yields, and increase profits. These systems rely on a large and diverse dataset containing information on crops, soil types, weather.

Support Vector Machine: Accuracy: 0.9636

#### **4.4 K-NN Algorithm:**

The k-nearest neighbor (k-NN) method is a data mining technique considered to be among the top five techniques for data mining. In this, we consider each of the characteristics in our training set as a different dimension in some space, and take the value an observation has for this characteristic to be its coordinate in that dimension, so getting a set of points in space. We can then consider the similarity of two points to be the distance between them in this space under some appropriate metric. The way in which the algorithm decides which of the points from the training set are similar enough to be considered when choosing the class to predict for a new observation is to pick the k closest data points to the new observation, and to take the most common class among these. This is why it is called the k Nearest Neighbors algorithm.

The implementation of algorithm can be noted as below:

1. Load the data
2. Initialize K to your chosen number of neighbors
3. For each example in the data □ Calculate the distance between the query example and the current example from the data. □ Add the distance and the index of to an ordered collection.
4. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances
5. Pick the first K entries from the sorted collection 6. Get the labels of the selected K entries 7. If regression, return the mean of the K labels 8. If classification, return the mode of the K labels



### Prediction of Crop Yield through KNN:

Here we consider parameters like soil type, PH, Nitrogen, Phosphorous, Potassium etc. We have assigned N, P, K, soil type as input parameters although other parameters may also be considered. The crop yield which is an unknown value can be predicted using the values of the nearest known neighbors. This is possible by calculating Euclidian distance between those points. Thus, we will be able to predict crop yield for the given input parameters.

The calculation of distance between points in a feature space, different distance functions could be used, in which the Euclidean distance function is the most commonly used one. Say p and q are represented as feature vectors. To measure the distance between p and q, the Euclidean metric is generally used by if  $a = (a_1, a_2)$  and  $b = (b_1, b_2)$  then the distance is given by:

$$d(a, b) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}$$

K-Nearest Neighbors: Accuracy: 0.9773

### 4.5 DECISION TREE:

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

In a Decision tree, there are two nodes, which are the **Decision Node** and **Leaf Node**.

Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of



features of the given dataset. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees. The complete process can be better understood using the below algorithm:

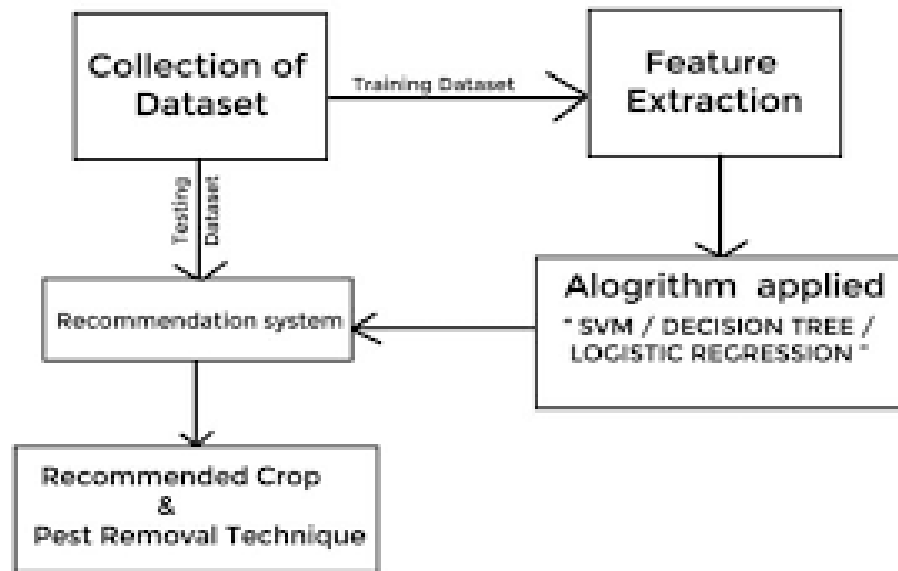
**Step-1:** Begin the tree with the root node, says S, which contains the complete dataset.

**Step-2:** Find the best attribute in the dataset using Attribute Selection Measure (ASM).

**Step-3:** Divide the S into subsets that contains possible values for the best attributes.

**Step-4:** Generate the decision tree node, which contains the best attribute.

**Step-5:** Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.



Decision Tree: Accuracy: 0.9848

#### 4.6 Bagging algorithm:

**Bootstrap Aggregating**, also known as bagging, is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression.

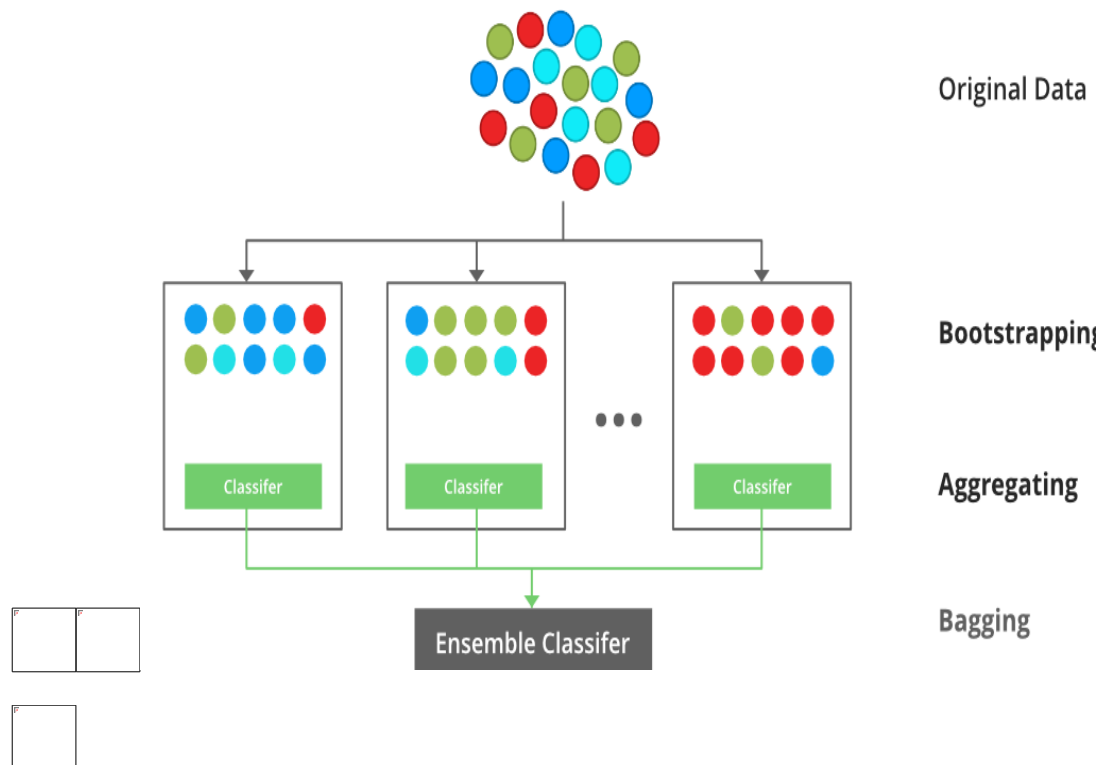
It decreases the variance and helps to avoid overfitting. It is usually applied to decision tree methods. Bagging is a special case of the model averaging approach.

##### Description of the Technique:

Suppose a set  $D$  of  $d$  tuples, at each iteration  $i$ , a training set  $D_i$  of  $d$  tuples is selected via row sampling with a replacement method (i.e., there can be repetitive elements from different  $d$  tuples) from  $D$  (i.e., bootstrap). Then a classifier model  $M_i$  is learned for each training set  $D < I$ . Each classifier  $M_i$  returns its class prediction. The bagged classifier  $M^*$  counts the votes and assigns the class with the most votes to  $X$  (unknown sample).

### Implementation Steps of Bagging:

- **Step 1:** Multiple subsets are created from the original data set with equal tuples, selecting observations with replacement.
- **Step 2:** A base model is created on each of these subsets.
- **Step 3:** Each model is learned in parallel with each training set and independent of each other.
- **Step 4:** The final predictions are determined by combining the predictions from all the models.



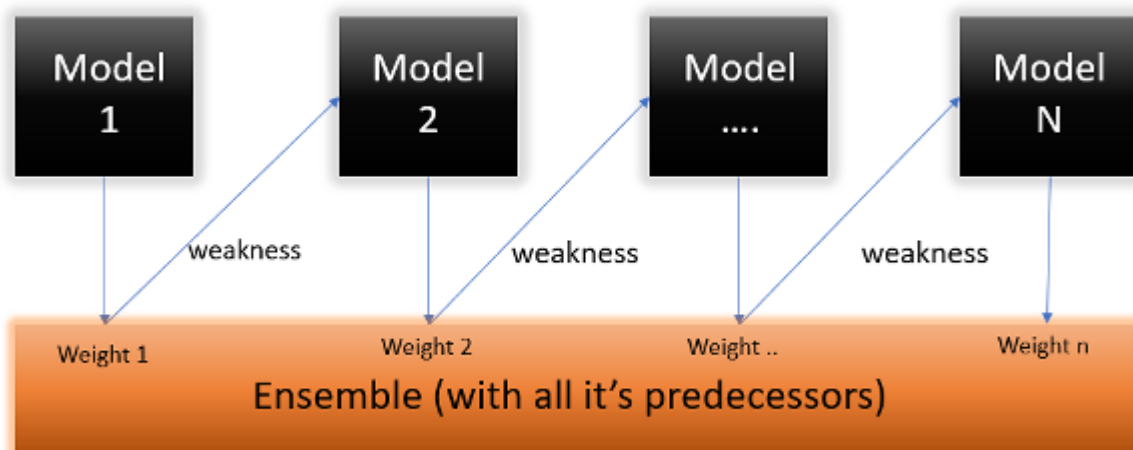
Bagging: Accuracy: 0.9879

## 4.7 What Is the AdaBoost Algorithm?

There are many machine learning algorithms to choose from for your problem statements. One of these algorithms for predictive modeling is called AdaBoost.

AdaBoost, also called Adaptive Boosting, is a technique in Machine Learning used as an Ensemble Method. The most common estimator used with AdaBoost is decision trees with one level which means Decision trees with only 1 split. These trees are also called **Decision Stumps**.

What this algorithm does is that it builds a model and gives equal weights to all the data points. It then assigns higher weights to points that are wrongly classified. Now all the points with higher weights are given more importance in the next model. It will keep training models until and unless a lower error is received.



AdaBoost: Accuracy: 0.2106

## 4.8 Gradient Boosting:

Gradient Boosting is also coming under supervised machine learning technique for solving classification and regression issues. It's a poor prediction model with an ensemble. As with previous boosting approaches, a gradient-boosted trees model is built stage by stage.

Crop Recommendation: Based on the N P K, temperature, humidity, and ph, the model will recommend the optimum crop to grow on the given soil.

Performance Analysis: Performance analysis is a specialized subject that uses systemic objectives to improve performance and decision-making.

Gradient Boosting: Accuracy: 0.9833

## 4.9 Extra Tree:

### Extra Tree Classifier / Regressor

It is a type of ensemble learning technique that aggregates the results of different de-correlated decision trees similar to Random Forest Classifier.

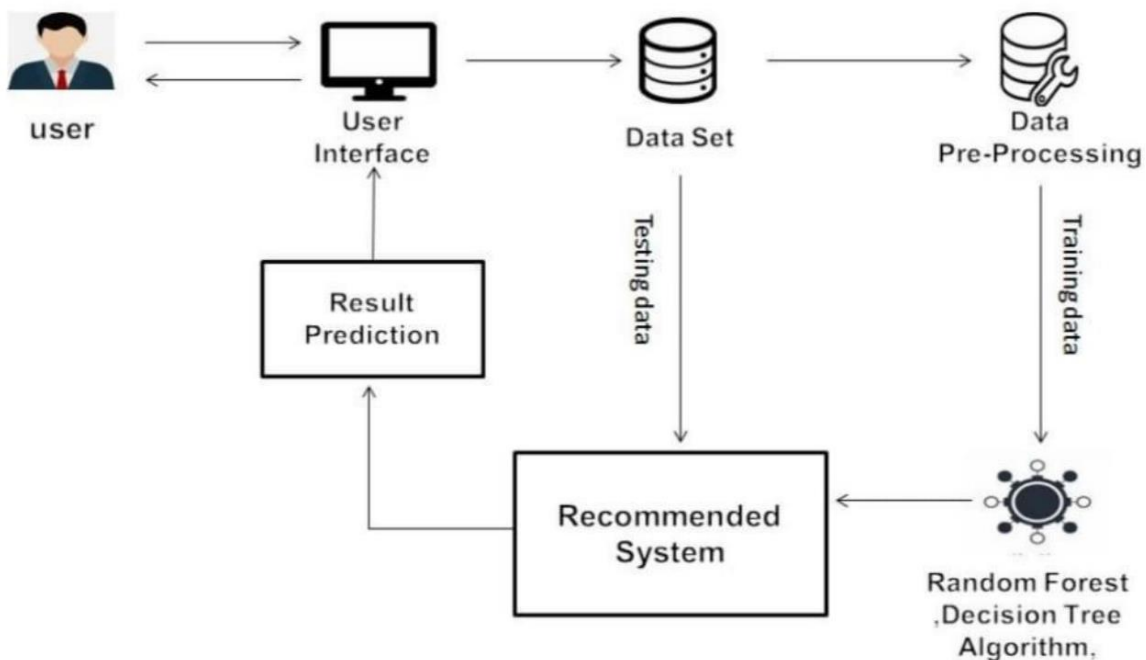
**Extra Tree** can often achieve a good or better performance than the random forest. The key difference between Random Forest and Extra Tree Classifier is,

- Extra Tree Classifier does not perform bootstrap aggregation like in the random forest. In simple words, takes a random subset of data without replacement. Thus, nodes are split on random splits and not on best splits.
- So, in Extra Tree Classifier randomness doesn't come from bootstrap aggregating but comes from the random splits of the data.

Extra Trees: Accuracy: 0.9515 0.9924242424242424

#### 4.10 SYSTEM ARCHITECTURE:

Crop Recommendation This module can be implemented by this method. Datasets can be acquired from Kaggle to train and test the data for tillage. Values are taken by the following site-specific factors are required of users: pH, N, P, and K (all of them in%), temperature (in °C), relative humidity (in%), and rainfall in(mm)



## 4.11 SOFTWARE ENVIRONMENT

### Python:

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

- **Python is Interpreted** – Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP
- **Python is Interactive** – You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- **Python is Object-Oriented** – Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- **Python is a Beginner's Language** – Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games

### History of Python:

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, Smalltalk, and Unix shell and other scripting languages.

Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).

Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

### **Python Features:**

Python's features include –

- **Easy-to-learn** – Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
- **Easy-to-read** – Python code is more clearly defined and visible to the eyes.
- **Easy-to-maintain** – Python's source code is fairly easy-to-maintain.
- **A broad standard library** – Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- **Interactive Mode** – Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- **Portable** – Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- **Extendable** – You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
- **Databases** – Python provides interfaces to all major commercial databases.
- **GUI Programming** – Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
- **Scalable** – Python provides a better structure and support for large programs than shell scripting.



## Features Available in Crop Dataset:

The dataset consists of a single CSV file. This dataset is mainly made concerning Indian climatic conditions. There are, in total, seven input features and one output feature.

The seven input features are as follows:

1. **N** – Ratio of Nitrogen content in the soil.
2. **P** – Ratio of Phosphorous content in the soil.
3. **K** – Ratio of Potassium content in the soil.
4. **Temperature** – The temperature in degrees Celsius.
5. **Humidity** – Relative humidity in %
6. **pH** – ph. value of the soil.
7. **Rainfall** – Rainfall in mm.

Let us perform some exploratory data analysis on our dataset and get to know the data. First, import all the required libraries.

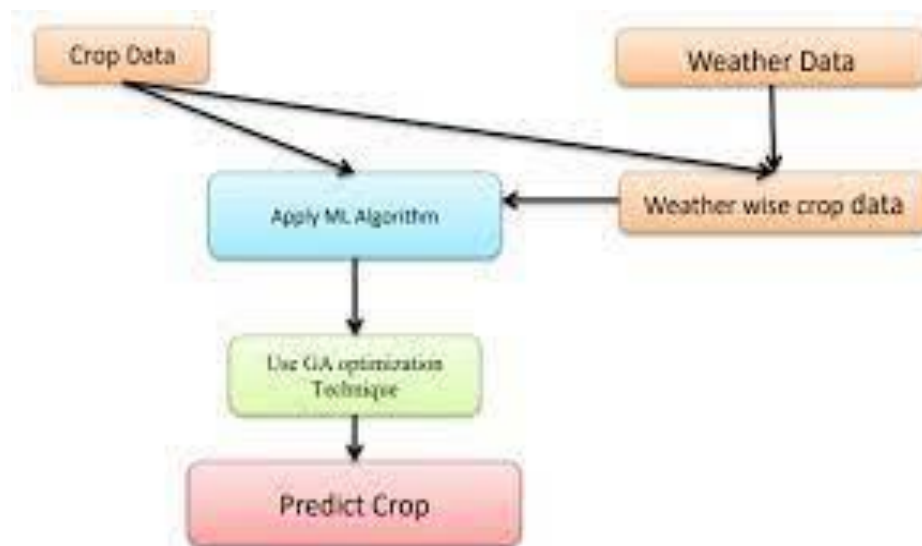
## Importing Libraries

Python libraries make it very easy for us to handle the data and perform typical and complex tasks with a single line of code.

- Pandas – This library helps to load the data frame in a 2D array format and has multiple functions to perform analysis tasks in one go.
- Numpy – Numpy arrays are very fast and can perform large computations in a very short time.
- Matplotlib – This library is used to draw visualizations.
- Sklearn – This module contains multiple libraries having pre-implemented functions to perform tasks from data preprocessing to model development and evaluation.
- TensorFlow – This is an open-source library that is used for Machine Learning and Artificial intelligence and provides a range of functions to achieve complex functionalities with single lines of code.

### **Exploratory Data Analysis (EDA):**

EDA is an approach to analyzing the data using visual techniques. It is used to discover trends, and patterns, or to check assumptions with the help of statistical summaries and graphical representations.



### **4.12 MODULES:**

- Admin Login
- Metadata
- Data Pre-processing
- Crop Prediction Module
- Crop Recommendation Module

## **MODULES DESCRIPTION:**

### **Admin Login:**

This is the first activity, Admin needs to provide a correct contact number and a password, which user enters while registering, in order to login into the webpage.

If information provided by the admin matches with the data in the database table, then user successfully login into the webpage else message of login failed is displayed and user need to re-enter correct information.

### **Metadata:**

All the main data used in the data set are initialized with the number to use in the algorithm it is like initializing all the details. In this metadata, we are going to initialize all the crop names with the numbers. This data makes us use the data easily in the algorithm. Hear the metadata of all the crops is given with a particular number. This number is not duplicated that is one number is given to one crop, the same number is not given to the other crop. This metadata consists of more than a hundred crops that grown all over India.

### **Data Pre-processing:**

Hear the raw data in the crop data is cleaned and the metadata is appending to it by removing the things which are converted to the integer. So, the data is easy to train. Hear all the data. In this pre-processing, we first load the metadata into this and then this metadata will be attached to the data and replace the converted data with metadata. Then this data will be moved further and remove the unwanted data in the list and it will divide the data into the train and the test data.

### **Crop Prediction Module:**

The obtained result will be helpful for the farmers to know the Yield of the crop so, he can go for the better crop which gives high yield and also say them the efficient use of agriculture field. This way we can help the farmers to grow the crop which gives them better yield.

### **Crop Recommendation Module:**

In this module, we have proposed a model that addresses these issues. The novelty of the proposed system is to guide the farmers to maximize the crop yield as well as suggest the most profitable crop for the specific region.

#### **4.13 UML DIAGRAMS:**

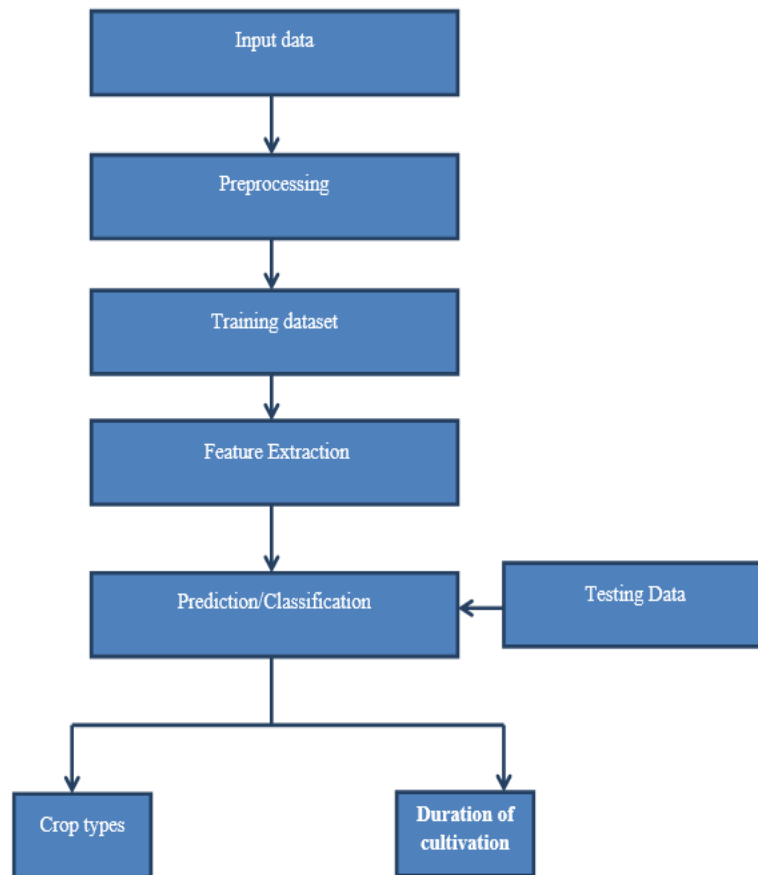
UML is simply another graphical representation of a common semantic model. UML provides a comprehensive notation for the full lifecycle of object-oriented development.

#### **ADVANTAGES**

- To represent complete systems using object-oriented concepts
  - To establish an explicit coupling between concepts and executable code
  - To take into account the scaling factors that are inherent to complex and Critical end.
  - To creating a modelling language usable by both humans and machines
- UML defines several models for representing systems
- The class model captures the static structure
  - The state model expresses the dynamic behavior of objects
  - The use case model describes the requirements of the user
  - The interaction model represents the scenarios and messages flows
  - The implementation model shows the work units
  - The deployment model provides details that pertain to process

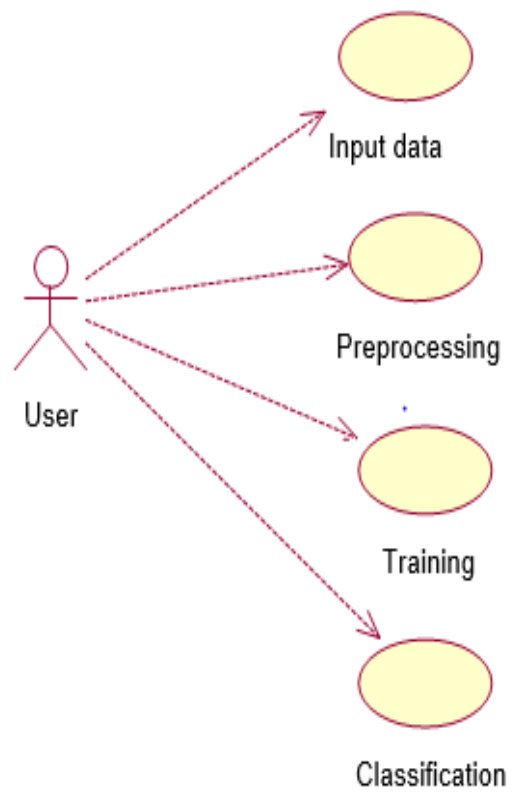
## **5.1 DATA FLOW DIAGRAM:**

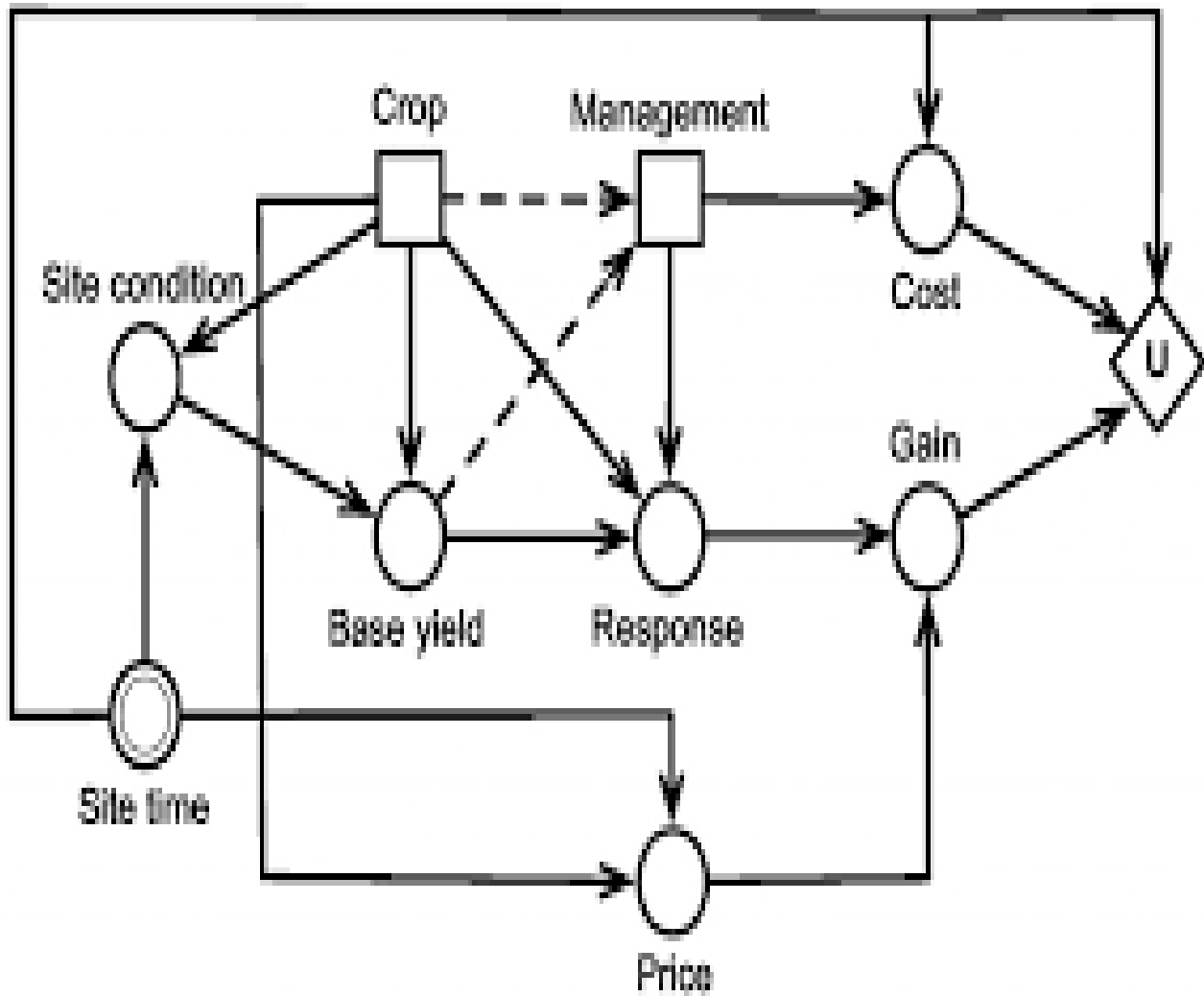
1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.



## 5.2 USECASE DIAGRAM:

Use case diagrams overview the usage requirement for system. They are useful for presentations to management and/or project stakeholders, but for actual development you will find that use cases provide significantly more value because they describe “the meant” of the actual requirements. A use case describes a sequence of action that provides something of measurable value to an action and is drawn as a horizontal ellipse.

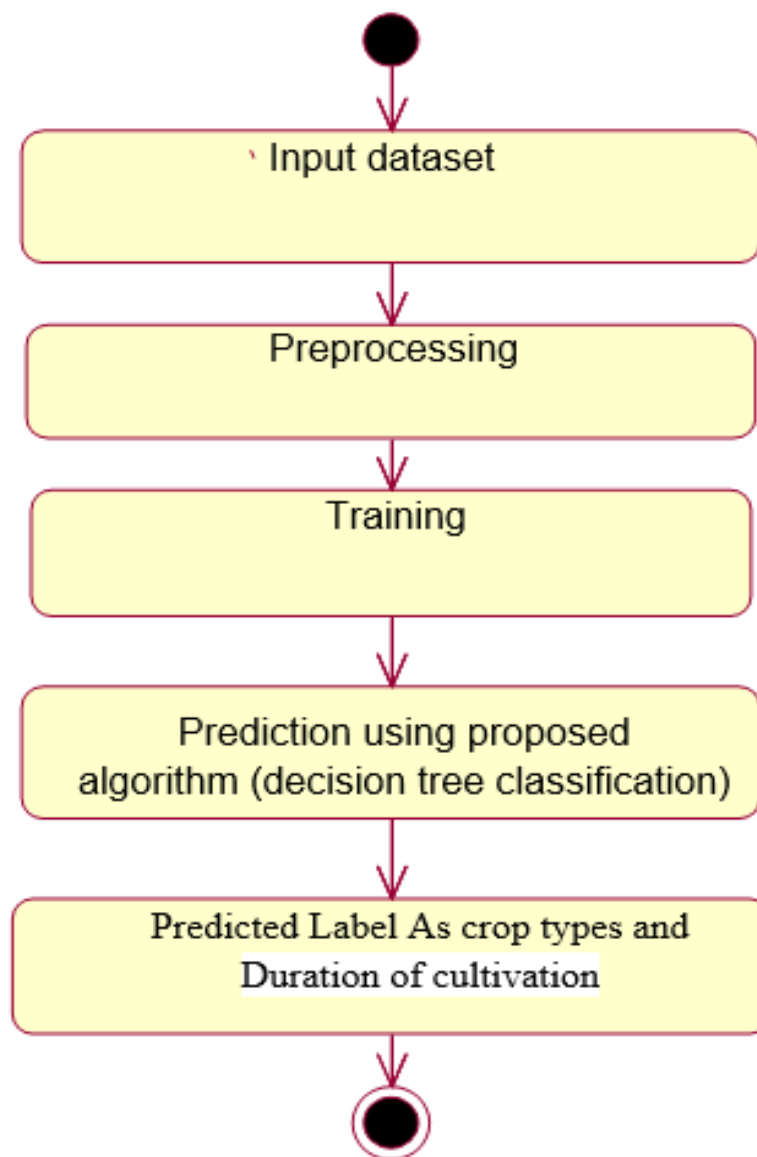


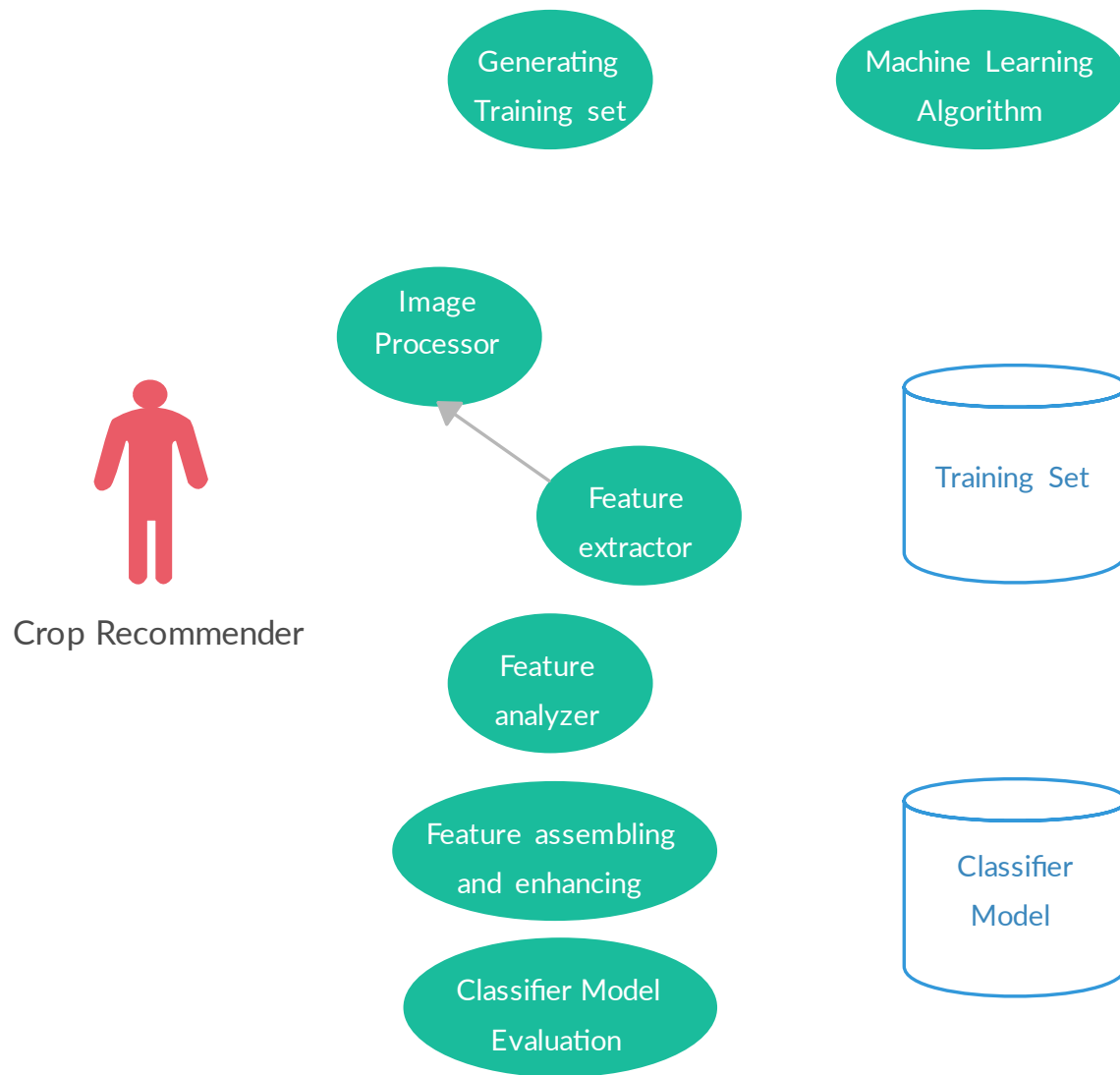




### 5.3 ACTIVITY DIAGRAM:

Activity diagram are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. The activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. Activity diagram consist of Initial node, activity final node and activities in between.



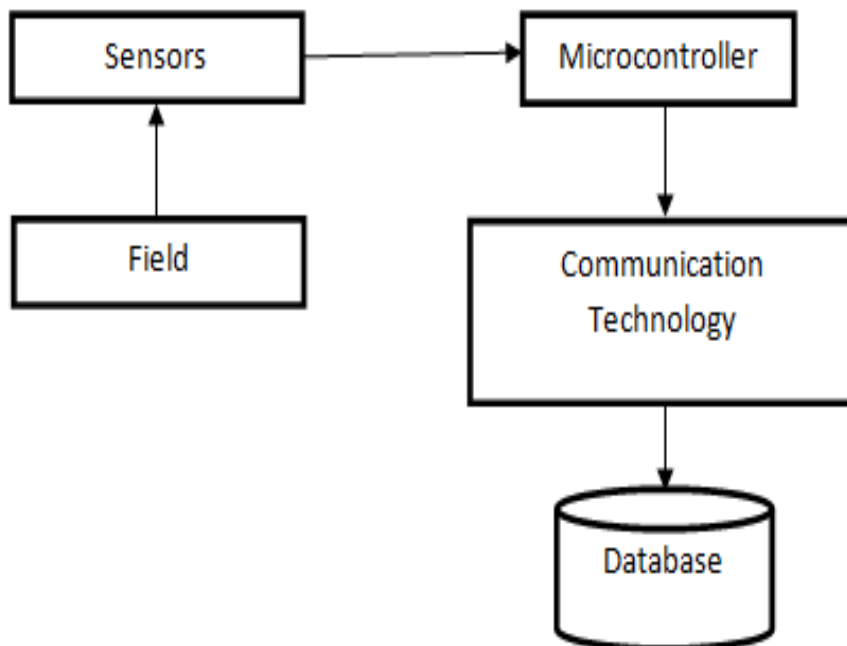


### **lot For Data Collection:**

Firstly, It Involves Collecting Data from The Field Area. The Parameters Such as Moisture, Temperature, Humidity and Ph Are Collected from The Particular Field Area. The Collected Data Is Then Stored and Given

As Input To GUI. The Water Content in The Soil Is Measured Using Soil Moisture Sensor [13][14]. This Is Very Important Parameter Which Is the Basis for Providing Nutrients to The Soil. LM35 Temperature Sensor

Is Used to Measure the Temperature of The Soil. DHT22 Humidity Sensor Is Used to Measure Air Temperature and Moisture. Ph Meter Is Used to Measure the Ph Level of The Soil and It Should Be Constantly Maintained [15][16]. To Collect the Information from The Sensors Arduino Microcontroller Is Used Which Is Responsible for Collecting Information from The Sensors. The Collected Information from The Sensor Is Stored in The Excel Sheet Using Wi-Fi.



## **5.4 INPUT DESIGN AND OUTPUT DESIGN:**

### **5.4.1 INPUT DESIGN:**

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy.

Input Design considered the following things:

What data should be given as input?

How the data should be arranged or coded?

The dialog to guide the operating personnel in providing input.

Methods for preparing input validations and steps to follow when error occur.

### **OBJECTIVES:**

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way

that all the data manipulates can be performed. It also provides record viewing facilities.

3.the user will not be in maize of instant. Thus, the objective of input design is to create an input layout that is easy to follow

### **5.4.2 OUTPUT DESIGN:**

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

3. Create document, report, or other formats that contain information produced by the system. The output form of an information system should accomplish one or more of the following objectives.

- ❖ Convey information about past activities, current status or projections of the Future.
- ❖ Signal important events, opportunities, problems, or warnings.
- ❖ Trigger an action.
- ❖ Confirm an action.

### **5.4.3 SYSTEM STUDY:**

#### **FEASIBILITY STUDY:**

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out.

This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ◆ **ECONOMICAL FEASIBILITY**
- ◆ **TECHNICAL FEASIBILITY**

#### **ECONOMICAL FEASIBILITY:**

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus, the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

#### **TECHNICAL FEASIBILITY:**

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being

placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

## **SYSTEM TESTING:**

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

## **TYPES OF TESTS:**

### **Unit testing:**

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

### **Functional test:**

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

**Valid Input:** identified classes of valid input must be accepted.

**Invalid Input:** identified classes of invalid input must be rejected.

**Functions:** identified functions must be exercised.

**Output:** identified classes of application outputs must be exercised.

**Systems/Procedures:** interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

### **System Test:**

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

## **6.1 What is HTML5?**

HTML stands for Hyper Text Markup Language. It is used to design web pages using a markup language. HTML is an abbreviation of Hypertext and Markup language. Hypertext defines the link between the web pages. The markup language is used to define the text document within the tag which defines the structure of web pages. HTML 5 is the fifth and current version of HTML. It has improved the markup available for documents and has introduced application programming interfaces (API) and Document Object Model (DOM).

### **Features:**

- It has introduced new multimedia features which support audio and video controls by using

<audio> and <video> tags.



- There are new graphics elements including vector graphics and tags.
- Enrich semantic content by including <header> <footer>, <article>, <section> and <figure> are added.
- Drag and Drop- The user can grab an object and drag it further dropping it to a new location.
- Geo-location services- It helps to locate the geographical location of a client.
- Web storage facility which provides web application methods to store data on the web browser.
- Uses SQL database to store data offline.
- Allows drawing various shapes like triangle, rectangle, circle, etc.
- Capable of handling incorrect syntax.
- Easy DOCTYPE declaration i.e., <!doctype html>
- Easy character encoding i.e., <meta charset=" UTF-8">

### **Browser Support:**

The latest versions of Apple Safari, Google Chrome, Mozilla Firefox, and Opera all support many HTML5 features and Internet Explorer 9.0 will also have support for some HTML5 functionality.

The mobile web browsers that come pre-installed on iPhones, iPads, and Android phones all have excellent support for HTML5.

### **New Features:**

HTML5 introduces a number of new elements and attributes that can help you in building modern websites. Here is a set of some of the most prominent features introduced in HTML5.

- **New Semantic Elements** – These are like <header>, <footer>, and <section>.

- **Forms 2.0** – Improvements to HTML web forms where new attributes have been introduced for <input> tag.
- **Persistent Local Storage** – To achieve without resorting to third-party plugins.
- **WebSocket** – A next-generation bidirectional communication technology for web applications.
- **Server-Sent Events** – HTML5 introduces events which flow from web server to the web browsers and they are called Server-Sent Events (SSE).
- **Canvas** – This supports a two-dimensional drawing surface that you can program with JavaScript.
- **Audio & Video** – You can embed audio or video on your webpages without resorting to third-party plugins.
- **Geolocation** – Now visitors can choose to share their physical location with your web application.
- **Microdata** – This lets you create your own vocabularies beyond HTML5 and extend your web pages with custom semantics.
- **Drag and drop** – Drag and drop the items from one location to another location on the same webpage.

## Backward Compatibility

HTML5 is designed, as much as possible, to be backward compatible with existing web browsers. Its new features have been built on existing features and allow you to provide fallback content for older browsers.

It is suggested to detect support for individual HTML5 features using a few lines of JavaScript.

The HTML 5 language has a "custom" HTML syntax that is compatible with HTML 4 and XHTML1 documents published on the Web, but is not compatible with the more esoteric SGML features of HTML 4.

HTML 5 does not have the same syntax rules as XHTML where we needed lower case tag names, quoting our attributes, an attribute had to have a value and to close all empty elements.

HTML5 comes with a lot of flexibility and it supports the following features –

- Uppercase tag names.
- Quotes are optional for attributes.
- Attribute values are optional.
- Closing empty elements are optional.

DOCTYPEs in older versions of HTML were longer because the HTML language was SGML based and therefore required a reference to a DTD.



## 6.2 What is CSS?

**Cascading Style Sheets**, fondly referred to as **CSS**, is a simple design language intended to simplify the process of making web pages presentable. CSS handles the look and feel part of a web page. Using CSS, you can control the colors of the text, the style of fonts, the spacing between paragraphs, how columns are sized and laid out, what background images or colors are used, layout designs, variations in display for different devices and screen sizes as well as a variety of other effects.

CSS is easy to learn and understand but it provides powerful control over the presentation of an HTML document. Most commonly, CSS is combined with the markup languages HTML or XHTML.

### **Advantages of CSS:**

- **CSS saves time** – You can write CSS once and then reuse same sheet in multiple HTML pages. You can define a style for each HTML element and apply it to as many Web pages as you want.
- **Pages load faster** – If you are using CSS, you do not need to write HTML tag attributes every time. Just write one CSS rule of a tag and apply it to all the occurrences of that tag. So, less code means faster download times.
- **Easy maintenance** – To make a global change, simply change the style, and all elements in all the web pages will be updated automatically.
- **Superior styles to HTML** – CSS have a much wider array of attributes than HTML, so you can give a far better look to your HTML page in comparison to HTML attributes.
- **Multiple Device Compatibility** – Style sheets allow content to be optimized for more than one type of device. By using the same HTML document, different versions of a website can be presented for handheld devices such as PDAs and cell phones or for printing.

• **Global web standards** – Now HTML attributes are being deprecated and it is being recommended to use CSS. So it's a good idea to start using CSS in all the HTML pages to make them compatible to future browsers.

### **How to add CSS?**

CSS is added to HTML pages to format the document according to information in the style sheet. There are three ways to insert CSS in HTML documents.

1. Inline CSS
2. Internal CSS
3. External CSS

#### **1) Inline CSS**

Inline CSS is used to apply CSS on a single line or element.

For example:

```
<p style="color: blue">Hello CSS</p>
```

#### **2) Internal CSS**

Internal CSS is used to apply CSS on a single document or page. It can affect all the elements of the page. It is written inside the style tag within head section of html.

For example:

1. `<style>`
2. `p{color: blue}`
3. `</style>`

#### **3) External CSS**

External CSS is used to apply CSS on multiple pages or all pages. Here, we write all the CSS code in a CSS file. Its extension must be .CSS for example style.css.

For example:

1. p {color: blue}

```
<link rel="stylesheet" type="text/css" href="style.css">
```



### **6.3 How to Install and Setup PyCharm for Python in Windows 10 in 2023?**

Are you just starting your journey to become a Pythoneer or Pythonista? If yes, then your first step should be choosing the best IDE for you. There is a number of python IDEs available on the internet, but I would recommend you to kick start your Python journey by installing PyCharm community edition. Why? Because PyCharm community edition comes with a lot of exciting features such as version control, visual debugging, syntax highlighting and many more. In this article, I will guide you to install PyCharm on your computer without any errors.

Follow the steps given below to install and set up PyCharm IDE for Python:

## 7 RESULTS AND DISCUSSIONS:

In this project the data analysis technology is used to update the rate change through notification. Further, a ranking process is applied for decision making.

Similarly, for Random Forest Algorithm and XG Boost Algorithm the value of  $R^2$  is 0.9314 and 0.9391 respectively i.e., the accuracy level will be 93.14% and 93.91% respectively. For  $R^2$  how much the high the value of  $R^2$  that much accurate that algorithm will be. So, from above result the XG Boost Algorithm has the best value so it will be the best algorithm for predicting the crop yield.

### 7.1 SCREENSHOTS OF OUR CODING SECTION:

#### app.py:

```
## Importing all necessary packages and libraries

from flask import Flask, render_template, request

import numpy as np

import pickle

import warnings

import psycopg2

import config

app = Flask(__name__)

#app.static_folder = 'static'

loaded_model = pickle.load(open("model.pkl", 'rb'))

warnings.filterwarnings('ignore')
```

```
con = psycopg2.connect(host=config.host, dbname=config.database,
                        user=config.username, password=config.password,

                        port=config.port)

#
con=psycopg2.connect(host=config.host,dbname=config.database,user=config
g.username,password=config.password,port=config.port)

cur = con.cursor()

con.commit()

cur.close()

con.close()

@app.route('/')

def home():

    return render_template('Home.html')

@app.route('/predict', methods=['POST'])

def predict():

    con = psycopg2.connect(host=config.host, dbname=config.database,
                        user=config.username, password=config.password,

                        port=config.port)

    cur = con.cursor()

    N = int(request.form['Nitrogen'])

    P = int(request.form['Phosphorus'])
```



```
K = int(request.form['Potassium'])

temp = float(request.form['Temperature'])

humidity = float(request.form['Humidity'])

ph = float(request.form['pH'])

rainfall = float(request.form['Rainfall'])

feature_list = [N, P, K, temp, humidity, ph, rainfall]

single_pred = np.array(feature_list).reshape(1, -1)

prediction = loaded_model.predict(single_pred)

crop_dict = {1: "Rice", 2: "Maize", 3: "Jute", 4: "Cotton", 5: "Coconut", 6:
"Papaya", 7: "Orange",

            8: "Apple", 9: "Muskmelon", 10: "Watermelon", 11: "Grapes", 12:
"Mango", 13: "Banana",

            14: "Pomegranate", 15: "Lentil", 16: "Blackgram", 17: "Mungbean",
18: "Mothbeans",

            19: "Pigeonpeas", 20: "Kidneybeans", 21: "Chickpea", 22: "Coffee"}
```

```
if prediction[0] in crop_dict:

    crop = crop_dict[prediction[0]]

    result = "{} is the best crop to be cultivated right there".format(crop)

else:

    result = "Sorry, we could not determine the best crop to be cultivated
with the provided data."
```

```
#insert_script = "INSERT INTO CropPredict (Nitrogen,
Phosphorus,Potassium,Temperature,Humidity,pH,Rainfall) values (" \

                #"%s,%s,%s,%s,%s,%s,%s) "

#value = ('0','Trilochan Sahu', str(feature_list), int(prediction[0]))

cur.execute("INSERT INTO croppredict (Nitrogen, Phosphorus, Potassium,
Temperature, Humidity, pH, Rainfall, predicted_crop) VALUES (%s, %s, %s,
%s, %s, %s, %s, %s)",(N, P, K, temp, humidity, ph, rainfall, crop))

con.commit()

cur.close()

con.close()

return render_template('Home.html', prediction=result)

@app.route('/about')

def about():

    return render_template('about.html')

@app.route('/contact')

def contact():

    return render_template('contact.html')

@app.route('/submit_contact', methods=['POST'])

def submit_contact():

    if request.method == 'POST':

        name = request.form['name']
```

```
email = request.form['email']
```

```
message = request.form['message']
```

# You can add code here to handle the contact form submission, such as sending an email to the team.

```
return "Thank you for your message! We will get back to you shortly."
```

```
if __name__ == '__main__':
```

```
    app.run(debug=True)
```

#### **config.py:**

```
host= "localhost"
```

```
username= "postgres"
```

```
database="postgres"
```

```
password= "lochan123"
```

```
port= 5432
```

#### **\_\_model.py:**

```
import pandas as pd
```

```
import pandas_profiling as pp
```

```
import numpy as np
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

```
crop = pd.read_csv('Crop_recommendation.csv')
```

```
crop.head()
```

```
crop.shape
```

```
crop.isnull().sum()
```

```
crop.duplicated().sum()
```

```
crop.info()
```

```
crop.describe()
```

```
crop_dict = {
```

```
    'rice': 1,
```

```
    'maize': 2,
```

```
    'jute': 3,
```

```
    'cotton': 4,
```

```
    'coconut': 5,
```

```
    'papaya': 6,
```

```
    'orange': 7,
```

```
    'apple': 8,
```

```
    'muskmelon': 9,
```

```
    'watermelon': 10,
```

```
    'grapes': 11,
```

```
'mango': 12,  
'banana': 13,  
'pomegranate': 14,  
'lentil': 15,  
'blackgram': 16,  
'mungbean': 17,  
'mothbeans': 18,  
'pigeonpeas': 19,  
'kidneybeans': 20,  
'chickpea': 21,  
'coffee': 22
```

```
}
```

```
crop['label_num'] = crop['label'].map(crop_dict)
```

```
crop.drop('label', axis=1, inplace=True)
```

```
crop.head()
```

```
# Split the dataset into features and labels
```

```
X = crop.iloc[:, :-1]
```

```
y = crop.iloc[:, -1]
```

```
from sklearn.model_selection import train_test_split
```

```
# Split the data into training and testing sets
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,  
random_state=42)
```

```
from sklearn.preprocessing import MinMaxScaler
```

```
scaler = MinMaxScaler()
```

```
X_train_scaled = scaler.fit_transform(X_train)
```

```
X_test_scaled = scaler.transform(X_test)
```

```
from sklearn.preprocessing import StandardScaler
```

```
scaler = StandardScaler()
```

```
X_train_scaled = scaler.fit_transform(X_train)
```

```
X_test_scaled = scaler.transform(X_test)
```

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
```

```
from sklearn.linear_model import LogisticRegression
```

```
from sklearn.naive_bayes import GaussianNB
```

```
from sklearn.svm import SVC
```

```
from sklearn.neighbors import KNeighborsClassifier
```

```
from sklearn.tree import DecisionTreeClassifier
```

```
from sklearn.tree import ExtraTreeClassifier
```

```
from sklearn.ensemble import RandomForestClassifier
```

```
from sklearn.ensemble import BaggingClassifier

from sklearn.ensemble import GradientBoostingClassifier

from sklearn.ensemble import AdaBoostClassifier


# create instances of all models

models = {

    'Linear Discriminant Analysis': LinearDiscriminantAnalysis(),

    'Logistic Regression': LogisticRegression(),

    'Naive Bayes': GaussianNB(),

    'Support Vector Machine': SVC(),

    'K-Nearest Neighbors': KNeighborsClassifier(),

    'Decision Tree': DecisionTreeClassifier(),

    'Random Forest': RandomForestClassifier(),

    'Bagging': BaggingClassifier(),

    'AdaBoost': AdaBoostClassifier(),

    'Gradient Boosting': GradientBoostingClassifier(),

    'Extra Trees': ExtraTreeClassifier(),

}
```

```
from sklearn.metrics import accuracy_score
```

```
for name, model in models.items():
```

```
    model.fit(X_train, y_train)
```

```
    y_pred = model.predict(X_test)
```

```
    acc = accuracy_score(y_test, y_pred)
```

```
    print(f'{name}:\nAccuracy: {acc:.4f}')
```

```
# Selecting decision tree model:
```

```
rdf = RandomForestClassifier()
```

```
rdf.fit(X_train, y_train)
```

```
y_pred = rdf.predict(X_test)
```

```
print(accuracy_score(y_test, y_pred))
```

```
# Define function to make predictions
```

```
def predict_crop(N, P, K, temperature, humidity, pH, rainfall):
```

```
    # Create a numpy array with the input values
```

```
    input_values = np.array([[N, P, K, temperature, humidity, pH, rainfall]])
```

```
    # Use the model to make a prediction
```



```
prediction = rdf.predict(input_values)
```

```
# Return the predicted crop label
```

```
return prediction[0]
```

```
N = 21
```

```
P = 26
```

```
K = 27
```

```
tem = 27.003155
```

```
humidity = 47.675254
```

```
ph = 5.699587
```

```
rainfall = 95.851183
```

```
pred = predict_crop(N, P, K, tem, humidity, ph, rainfall)
```

```
if pred == 1
```

```
    print("Rice is the best crop to be cultivated right there")
```

```
elif pred == 2:
```

```
    print("Maize is the best crop to be cultivated right there")
```

```
elif pred == 3:
```

```
    print("Jute is the best crop to be cultivated right there")
```

```
elif pred == 4:
```

```
print("Cotton is the best crop to be cultivated right there")

elif pred == 5:

    print("Coconut is the best crop to be cultivated right there")

elif pred == 6:

    print("Papaya is the best crop to be cultivated right there")

elif pred == 7:

    print("Orange is the best crop to be cultivated right there")

elif pred == 8:

    print("Apple is the best crop to be cultivated right there")

elif pred == 9:

    print("Muskmelon is the best crop to be cultivated right there")

elif pred == 10:

    print("Watermelon is the best crop to be cultivated right there")

elif pred == 11:

    print("Grapes is the best crop to be cultivated right there")

elif pred == 12:

    print("Mango is the best crop to be cultivated right there")

elif pred == 13:

    print("Banana is the best crop to be cultivated right there")
```

```
elif pred == 14:
    print("Pomegranate is the best crop to be cultivated right there")
elif pred == 15:
    print("Lentil is the best crop to be cultivated right there")
elif pred == 16:
    print("Blackgram is the best crop to be cultivated right there")
elif pred == 17:
    print("Mungbean is the best crop to be cultivated right there")
elif pred == 18:
    print("Mothbeans is the best crop to be cultivated right there")
elif pred == 19:
    print("Pigeonpeas is the best crop to be cultivated right there")
elif pred == 20:
    print("Kidneybeans is the best crop to be cultivated right there")
elif pred == 21:
    print("Chickpea is the best crop to be cultivated right there")
elif pred == 22:
    print("Coffee is the best crop to be cultivated right there")
else:
```

```
print("Sorry, we could not determine the best crop to be cultivated with the  
provided data.")
```

```
#X_train
```

```
# loading library
```

```
import pickle
```

```
# create an iterator object with write permission - model.pkl
```

```
with open('model.pkl', 'wb') as files:
```

```
    pickle.dump(model, files)
```

### **Diagram in database connectivity:**

pgAdmin 4

File Object Tools Help

Object Explorer

- Foreign Tables
- Functions
- Materialized Views
- Operators
- Procedures
- Sequences
- Tables (2)
  - croppredict
    - Columns (9)
      - id
      - nitrogen
      - phosphorus
      - potassium
      - temperature
      - humidity
      - ph
      - rainfall
      - predicted\_crop
    - Constraints
    - Indexes
    - RLS Policies
    - Rules
    - Triggers
    - Trigger Functions
    - Types
    - Views
  - Subscriptions
  - Login/Group Roles
  - Tablespaces
  - PostgreSQL 15

Dashboard Properties SQL Statistics Dependencies Dependents public.croppredict/postgres/postgres@PostgreSQL 14

public.croppredict/postgres/postgres@PostgreSQL 14

Query Query History

Show queries generated internally by pgAdmin? ☐

Remove Remove All

23/7/2023 1:18:52 17 489 msec

pm Date Rows affected Duration

Copy Copy to Query Editor

SELECT \* FROM public.croppredict  
ORDER BY id ASC

Messages

Successfully run. Total query runtime: 489 msec. 17 rows affected.

Data Output Messages Notifications

	id [PK] integer	nitrogen integer	phosphorus integer	potassium integer	temperature double precision	humidity double precision	ph double precision	rainfall double precision	predicted_crop character varying
1	1	45	56	67	32	23	12	12	Mothbeans
2	2	34	44	89	23	35	23	25	Chickpea
3	3	23	45	56	34	24	13.87	25	Mothbeans
4	4	23	45	56	34	24	13.87	25	Mothbeans
5	5	24	24	24	23	35	13.5	65	Mothbeans
6	6	24	24	24	23	35	13.5	65	Mothbeans
7	7	35	36	47	39	46	23	66	Mothbeans
8	8	76	49	89	39	47	36.9	12	Chickpea
9	9	76	49	89	39	47	36.9	12	Chickpea

Total rows: 18 of 18 Query complete 00:00:00.348 Ln 1, Col 1

81°F Mostly cloudy

Search

10:04 PM 27-07-2023

127.0.0.1:5000/submit\_contact x www.google.com x +

127.0.0.1:5000/submit\_contact

Thank you for your message! We will get back to you shortly.

81°F Mostly cloudy

Search

10:08 PM 27-07-2023

→ 127.0.0.1:5000

Crop Recommendation home Contact About Search

## Crop Recommendation

<b>Nitrogen</b>	<b>Phosphorus</b>	<b>Potassium</b>
<input type="text" value="67"/>	<input type="text" value="78"/>	<input type="text" value="82"/>
<b>Temperature</b>	<b>Humidity</b>	<b>pH</b>
<input type="text" value="23"/>	<input type="text" value="12"/>	<input type="text" value="34"/>
<b>Rainfall</b>		
<input type="text" value="45"/>		

Get Recommendation

Crop recommender system x +

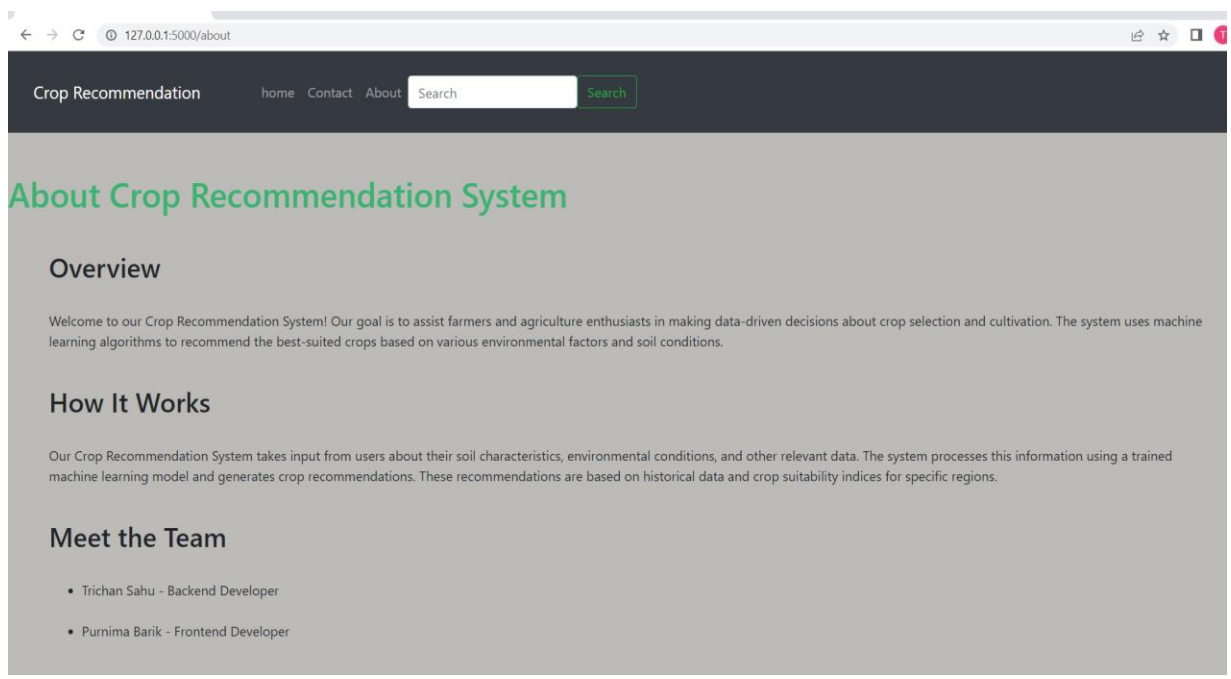
127.0.0.1:5000

Crop Recommendation home Contact About Search Search

## Crop Recommendation

<b>Nitrogen</b>	<b>Phosphorus</b>	<b>Potassium</b>
<input type="text" value="Enter Nitrogen"/>	<input type="text" value="Enter Phosphorus"/>	<input type="text" value="Enter Potassium"/>
<b>Temperature</b>	<b>Humidity</b>	<b>pH</b>
<input type="text" value="Enter Temperature in °C"/>	<input type="text" value="Enter Humidity in %"/>	<input type="text" value="Enter pH value"/>
<b>Rainfall</b>		
<input type="text" value="Enter Rainfall in mm"/>		

[Get Recommendation](#)





← → ↻

127.0.0.1:5000/predict

🔖 ☆ 🏠

## Crop Recommendation

**Nitrogen**

**Phosphorus**


**Potassium**

**Temperature**

**Humidity**

**pH**

**Rainfall**



**Recommend Crop**  
Chickpea is the best crop to be cultivated right there

81 | Page

## **8. CONCLUSION**

### **8.1 Conclusion**

This open attitude determines the degree and scope of information sharing. Big data analysis technology can effectively improve the crop yield production is updation. This project proposes a novel intelligent system for agricultural crop price prediction. The key idea is to use ensemble of classifiers for prediction. The usage of ensemble of classifiers paves a path way to make a better decision on predictions due to the usage of multiple classifiers. Further, a ranking process is applied for decision making in order to select the classifiers results. This system is used to predict the cost of the crop rate for further. The solution will benefit farmers to maximize productivity in agriculture, reduce soil degradation in cultivated fields, and reduce fertilizer use in crop production by recommending the right crop by considering various attributes.

This would provide a comprehensive prediction on the basis of geographical, environmental and economic aspects.

From the above results we can easily find that the accuracy level of XG Boost algorithm is higher in every case either it is R2 or MSE or MAE.

for the year from 2000 to 2014. Area of cultivation, maximum temperature, minimum temperature and rainfall are the input variable and production are the output or dependent variable.

The proposed XG Boost algorithm is compared with Random Forest. XG Boost model gives better performance than all other models. In future, it is planned to optimize the hyper parameters of each traditional algorithms and should be tested with new data sets.

## 8.2 Futuristic Scope:

We have to collect all required data by giving GPS locations of a land and by taking access from Rain forecasting system of by the government, we can predict crops by just giving GPS location. Also, we can develop the model to avoid over and under crisis of the food.

The Crop Recommendation Flask Web App is a web application that recommends the best crop to grow based on soil and climate conditions. The project involves building a machine learning model that can predict the crop yield based on several parameters such as soil pH, temperature, rainfall, humidity, and crop type.

From detecting pests to predicting what crops will deliver the best returns, artificial intelligence can help humanity confront one of its biggest challenges: feeding an additional 2 billion people by 2050<sup>2</sup>, even as climate change disrupts growing seasons, turns arable land into deserts, and floods once-fertile deltas ...

Getting the most from every acre is not an academic problem. The United Nations estimates we will need to increase food production 50 percent by the middle of the century.<sup>3</sup>

Agricultural production tripled between 1960 and 2015 as the world's population grew from 3 billion people to 7 billion.<sup>4</sup>

While technology played a role in the form of pesticides, fertilizers, and machines, much of the gains can be attributed to simply plowing more land—cutting forests and diverting fresh water to fields, orchards, and rice paddies. We will have to be more resourceful this time around.

## **9. TAKE AWAY FROM THE PROJECT**

### **9.1 References and Bibliography**

[1] Manpreet Kaur, Heena Gulati, Harish Kundra, "Data Mining in Agriculture on

Crop Price Prediction: Techniques and Applications", International Journal of Computer Applications, Volume 99– No.12, August 2014.

- 1) <https://www.wikipedia.org/>
- 2) <https://www.python.org/>
- 3) <https://github.com>
- 4) <https://www.kaggle.com>

[5] A. Kaloxylou et al., "Farm management systems and the future Internet era," Compute. Electron. Agricult., vol. 89, pp. 130–144, Nov. 2012.

[6] N. N. Li, T. S. Li, Z. S. Yu, Y. Rui, Y. Y. Miao, and Y. S. Li, "Factors influencing farmers' adoption of new technology based on Logistic-ISM model-a case study of potato planting technology in Dingxi City, Gansu Province," Progress in Geography, vol. 33, no. 4, pp. 542-551, 2014

[7] Y. Wang, "A neural network adaptive control based on rapid learning method and its application," Advances In Modeling and Analysis, Vol. 46(3), pp.27-34,1994.

8. <https://www.engati.com/glossary/natural-language-processing>