PREDICTIVE MODELING OF CROP YIELDS USING MACHINE LEARNING

Minor project-II report submitted
in partial fulfillment of the requirement for award of the degree of

Bachelor of Technology
in
Computer Science & Engineering

By

D.MRUDULA	(21UECS0148)	**(VTU20267)**
MANDADI SINDHU	(21UECS0357)	**(VTU20250)**
D.SAI SANTHOSH REDDY	(21UECS0151)	**(VTU20220)**

Under the guidance of
Dr. M. GURU VIMAL KUMAR, B.Tech., M.E., Ph.D.,
ASSOCIATE PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF COMPUTING

VEL TECH RANGARAJAN DR. SAGUNTHALA R&D INSTITUTE OF SCIENCE & TECHNOLOGY

(Deemed to be University Estd u/s 3 of UGC Act, 1956)
Accredited by NAAC with A++ Grade
CHENNAI 600 062, TAMILNADU, INDIA

May, 2024

PREDICTIVE MODELING OF CROP YIELDS USING MACHINE LEARNING

Minor project-II report submitted
in partial fulfillment of the requirement for award of the degree of

Bachelor of Technology
in
Computer Science & Engineering

By

D.MRUDULA	(21UECS0148)	**(VTU20267)**
MANDADI SINDHU	(21UECS0357)	**(VTU20250)**
D.SAI SANTHOSH REDDY	(21UECS0151)	**(VTU20220)**

Under the guidance of
Dr. M. GURU VIMAL KUMAR, B.Tech., M.E., Ph.D.,
ASSOCIATE PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF COMPUTING

VEL TECH RANGARAJAN DR. SAGUNTHALA R&D INSTITUTE OF SCIENCE & TECHNOLOGY

(Deemed to be University Estd u/s 3 of UGC Act, 1956)
Accredited by NAAC with A++ Grade
CHENNAI 600 062, TAMILNADU, INDIA

May, 2024

CERTIFICATE

It is certified that the work contained in the project report titled ”PREDICTIVE MODELING OF CROP YIELDS USING MACHINE LEARNING” by ”D.MRUDULA (21UECS0148), MANDADI SINDHU (21UECS0357), D.SAI SANTHOSH REDDY (21UECS0151)” has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

Signature of Supervisor
Computer Science & Engineering
School of Computing
Vel Tech Rangarajan Dr. Sagunthala R&D
Institute of Science & Technology
May, 2024

Signature of Professor In-charge
Computer Science & Engineering
School of Computing
Vel Tech Rangarajan Dr. Sagunthala R&D
Institute of Science & Technology
May, 2024

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(D.MRUDULA)
Date: / /

(MANDADI SINDHU)
Date: / /

(D.SAI SANTHOSH REDDY)
Date: / /

APPROVAL SHEET

This project report entitled ”PREDICTIVE MODELING OF CROP YIELDS USING MACHINE LEARNING” by D.MRUDULA (21UECS0148), MANDADI SINDHU (21UECS0357), D.SAI SANTHOSH REDDY(21UECS0151) is approved for the degree of B.Tech in Computer Science & Engineering.

Examiners **Supervisor**

Dr. M. Guru Vimal Kumar, B.Tech., M.E., Ph.D.,

Date: / /

Place:

ACKNOWLEDGEMENT

We express our deepest gratitude to our respected **Founder Chancellor and President Col. Prof. Dr. R. RANGARAJAN B.E. (EEE), B.E. (MECH), M.S (AUTO),D.Sc., Foundress President Dr. R. SAGUNTHALA RANGARAJAN M.B.B.S.** Chairperson Managing Trustee and Vice President.

We are very much grateful to our beloved **Vice Chancellor Prof. S. SALIVAHANAN,** for providing us with an environment to complete our project successfully.

We record indebtedness to our **Professor & Dean, Department of Computer Science & Engineering, School of Computing, Dr. V. SRINIVASA RAO, M.Tech., Ph.D.,** for immense care and encouragement towards us throughout the course of this project.

We are thankful to our **Head, Department of Computer Science & Engineering, Dr.M.S. MURALI DHAR, M.E., Ph.D.,** for providing immense support in all our endeavors.

We also take this opportunity to express a deep sense of gratitude to our **Dr. M. GURU VIMAL KUMAR,B.Tech.,M.E.,Ph.D., for his cordial support, valuable information and guidance, he/she helped us in completing this project through various stages.**

A special thanks to our **Project Coordinators Mr. V. ASHOK KUMAR, M.Tech., Ms. U.HEMAVATHI, M.E., Ms. C. SHYAMALA KUMARI, M.E.,** for their valuable guidance and support throughout the course of the project.

We thank our department faculty, supporting staff and friends for their help and guidance to complete this project.

D.MRUDULA	**(21UECS0148)**
MANDADI SINDHU	**(21UECS0357)**
D.SAI SANTHOSH REDDY	**(21UECS0151)**

ABSTRACT

Agriculture is the back bone of the Indian economic system and more than 50 outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values. 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 overfitting. So, for the upliftment of agriculture sector in our nation should need a technology that Predicts the yield of the crops in advance which helps the farmers to take suitable measures for their crops from the damage and can achieve good yield. Machine learning is an essential Approach for achieving this prediction in advance. basically design an application based on Random Forest Algorithm(RFA) which is a Popular supervised machine learning algorithm that predicts the crop yield based on the data of climate, temperature, etc. in advance and helps the farmers to take a decision whether to Grow the crops or not. This type of application based prediction will create drastic changes in terms of countries Economy and also helps the farmers to save their hard-earned money.

Keywords:

Agriculture , Random Forest Algorithm , Decision Tree, Prediction , Crop Yield, Farmers,.

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
CYP	Crop Yield Prediction
GPU	Graphical Processing Unit
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
K	Potassium
KNN	K-Nearest Neighbour
ML	Machine Learning
N	Nitrogen
P	Phosphorous
RFA	Random Forest Algorithm
TPU	Tensor Processing Unit

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Introduction

The major thing why the agriculture area was getting deprecated is due to climatic changes because the productivity of a crop basically depends upon weather conditions like rainfall, temperature, humidity, soil fertility etc. So we can design a system using Random Forest Algorithm(RFA) which most powerful supervised Machine Learning algorithm (ML) that predicts the crop yield based on parameters like rainfall ,temperature ,humidity etc. in advance the central idea of this crop yield prediction is make the farmers aware of the climatic conditions and also about the types of crops to be cultivated during that climate. This helps the farmers to save their crops and there pockets well.This introduction sets the stage for exploring the intricacies of predictive modeling for crop yields using machine learning. From understanding the underlying principles of machine learning algorithms to implementing practical solutions in real-world agricultural settings, this field offers boundless opportunities to revolutionize global food production and ensure a sustainable future for generations to come.

Predicting yield of crops will surely help the farmer. The farmer can make a decision about crop choice and can contribute more to its profit. Machine learning is found to be a very appealing field that can contribute to the agriculture field. The different models built using machine learning can take different inputs to give some real-time output.In recent years, agriculture has faced unprecedented challenges including climate change, resource constraints, and evolving consumer demands. To address these challenges, farmers and researchers are increasingly turning to data-driven solutions. Predictive modeling leverages historical data, weather patterns, soil characteristics, and other variables to forecast crop yields with remarkable accuracy. By analyzing historical crop yield data alongside environmental factors, machine learning models can predict future yields under different scenarios, guiding farmers in making informed decisions throughout the growing season.

In today's dynamic agricultural landscape, where climate change, population growth, and resource constraints pose significant challenges, the need for accurate predictive modeling of crop yields has never been more pressing. Traditional methods of yield estimation often struggle to account for the complexities inherent in agricultural systems, leading to suboptimal decision-making and resource allocation. However, the advent of machine learning offers a transformative solution by enabling the extraction of valuable insights from vast and diverse datasets. By leveraging machine learning algorithms, researchers and practitioners can analyze complex interactions between environmental factors, agronomic practices, and crop performance to forecast yields with unprecedented precision.

1.2 Aim of the Project

The aim of the project is to construct a crop yield prediction model which helps the farmers for taking appropriate decisions before cultivation of the crop. This project will help the farmers to know the yield of their crop before cultivating in the agricultural field.

1.3 Project Domain

This project comes under the domain of Machine learning. Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes.

Mainly it contains six steps in which the first step concentrates on input data set which is weather and forecasting crop yield dataset. The second phase, Data preprocessing techniques involves in transforming the raw data into understandable format. The entire process should undergone by using random forest algorithm in which verification of data and forecasting should be done to achieve good results.

1.4 Scope of the Project

The scope of the project is to design an application about crop yield prediction based on weather conditions using machine learning technique. If weather-based predictions are precise, then farmers can be alerted in advance so that the major loss can decrease and it would be helpful for economic growth.

The prediction will also aid the farmers to make decisions such as the choice of alternative crops or removing a crop at an early stage in case of critical situations. This designed application is extremely useful for the framers by using this application farmers will get the idea about the crop production based on which they can plan their farming activity and market income expectations.

Chapter 2

LITERATURE REVIEW

[1] Anastasiya Kolesnikova, et.al., 2021, has developed Crop Selection method based on various environmental factors using machine learning This system consists of four levels (Data collection, Preprocessing step, Feature Extraction) to predicts the yield of the crop based on the climate. In this system the algorithm used was naive bayes which is supervised machine learning algorithm for crop prediction .The disadvantage is that the native bayes won't produce accurate results compared to random forest.

[2] Anupama C.G., et.al., 2020, has developed Crop yield estimation using machine learning in agricultural crop production. This system predicts the yield of the crop based on climatic factors.For this system the algorithms used are RFA for predicting the crop yield. it contains six steps in which the first step concentrates on collection of crop dataset. The second step is Data pre-processing techniques involves in converting the raw data into understandable data. After preprocessing step, the next step is dimensional reduction which is used to reduce the number of random variables under consideration by obtaining a set of principal variables. This situation mainly focuses on climate gauging, crop yield expectation and harvest cost anticipating. These elements help the ranchers to develop the best nourishment harvests and raise the correct animals by understanding to natural segments. Similarly, the ranchers can somewhat adapt to changes in the climate by shifting planting times, choosing assortments of different development terms or adjusting harvest pivots. The factual numerical information is identified with horticulture is embraced for the test investigation. However, the grouping based systems and the calculations administered are used to deal with the measurable information gathered. The entire process should undergone by using random forest algorithm in which verification of data and forecasting should be done to achieve good results.

[3] Aruvansh Nigam, et.al., 2021 Crop yield prediction using machine learning algorithms. In this model the algorithms used are Neural network, Machine learning, linear regression, multiple regression. this research describes the development of a different crop yield prediction model with ANN, with three Layer Neural Network. The ANN model develops a formula to ascertain the relationship using a large number of input and output examples, to establish model for yield predictions an Ac-

tivation function: Rectified Linear activation unit is used. The backward and forward propagation techniques are used. The Proposed study was conducted using Python matplotlib and Seaborn which is used for data visualization. Data Pre-processing and Data cleaning processes are performed by Pandas library of python. Basically broad five steps are used for experiment that are Data collection, data Wrangling, data Preprocessing, data Visualization (Different Visualization Library Used), explotary data analysis. The above steps are further explained in detail as follows which are followed for processing and preparing the data for applying the multilayer perceptron technique . The disadvantage of this model was it predicts the yield of the crop with less accuracy.

[4] B.Joesphine, et.al., 2021 crop yield prediction using ANN-Algorithm. In this model the algorithms used are Artificial neural networks, Support Vector Machine. This paper assist user the method that would help them to choose the crop which will maximize the craw takings by taking into retainer all the parameter which affect the growth of crop. The different parameters like environmental, economic and other parameters related to the yield in nature can be analyzed for prediction of accurate resultant role. The economical parameters include demand for crop, market rate etc. whereas environmental parameters include quantity of rainfall, temperature, and type of soil. So, all these factors are considered while predicting the most efficient crop to be cultivated based on season. On the basis of crop selection method described in that, we hereby propose our two methods of crop selection which is an extended work on that. The proposed methods are: i. Crop Selection Method ii. Crop Sequencing Method The price factor is one of the most important factors which play a major role in selecting crop. For example, there are two crops and both produce equal yield but one crop is valued at a lower price than the other. If the price factor is not included in the crop selection method, then system may lead to select a wrong crop to grow. Therefore, price is as important as the factors such as soil type, rainfall, temperature etc . The disadvantage is Exact accuracy is not specified.

[5] Krishna kumar, et.al., 2021 developed A smart agricultural model using K-means and clustering techniques. In this model algorithms used are K-means Algorithm, clustering method and concluded the aim of this paper is to propose and implement a rule-based system this system contains two states Training state and Test state. In training state data was collected and preprocessed. The pre-processed data was clustered using k-means clustering algorithm. The association rule mining process apply on clustered data to find the rules. The training state ends with number of generated rules. In testing state, the yield value is predicted based on generated rules. The disadvantage is Suitable only for using association rule and considered less data.

[6] M. Kalimuthu, et.al., 2020 Food production and prediction is getting depleted due to unnatural

climatic changes, which will adversely affect the economy of farmers by getting a poor yield and also help the farmers to remain less familiar in forecasting the future crops. This research work helps the beginner farmer in such a way to guide them for sowing the reasonable crops by deploying machine learning, one of the advanced technologies in crop prediction. Naive Bayes, a supervised learning algorithm puts forth in the way to achieve it. The seed data of the crops are collected here, with the appropriate parameters like temperature, humidity and moisture content, which helps the crops to achieve a successful growth. In addition as the software, a mobile application for Android is being developed. The users are encouraged to enter parameters like temperature and their location will be taken automatically in this application in order to start the prediction process.

[7] Patrick Helber, et.al., 2020 accurate and reliable crop yield prediction is a complex task. The yield of a crop depends on a variety of factors whose accurate measurement and modeling is challenging. At the same time, reliable yield prediction is highly desirable for farmers to optimize crop production. In this paper, we introduce a modeling based on remote sensing data and Machine Learning models evaluated on a large-scale dataset to address the challenge of an operational crop yield estimation and forecasting on field and subfield level. With our approach, we aim towards a global yield modeling based on Machine Learning models which operates across crop types without the need for crop-specific modeling. We demonstrate that our approach learns to map in-field variability for all studied crop types.

[8] Potnuru Sai Nishant, et.al., 2021 developed a model that predicts the yield of almost all kinds of crops that are planted in India. This script makes novel by the usage of simple parameters like State, district, season, area and the user can predict the yield of the crop in which year he or she wants to. The paper uses advanced regression techniques like Kernel Ridge, Lasso and ENet algorithms to predict the yield and uses the concept of Stacking Regression for enhancing the algorithms to give a better prediction.

[9] Vinita Shah and Prachi Shah, et.al., 2020 stated that Yield prediction is a very important agricultural problem. Any farmer is interested in knowing how much yield he is about to expect. In the past, yield prediction was performed by considering farmer's experience on particular field and crop. Based on previous data, we can predict crop yield using machine-learning technique. Crop yield prediction is an important area of research, which helps in ensuring food security all around the world. We analyzed result of multiple linear Regression, Regression Tree, K-nearest Neighbor and Artificial Neural Network on Groundnut data of previous 8 years. We have done prediction based

on Soil, Environmental and Abiotic attributes. KNN algorithm gives better result compared to other algorithms for Groundnut crop yield prediction.

[10] Jeevan Nagendra Kumar, et.al., 2020 Machine learning (ML) is a crucial perspective for acquiring real-world and operative solution for crop yield issue. From a given set of predictors, ML can predict a target/outcome by using Supervised Learning. To get the desired outputs need to generate a suitable function by set of some variables which will map the input variable to the aim output. Crop yield prediction incorporates forecasting the yield of the crop from past historical data which includes factors such as temperature, humidity, ph, rainfall, crop name. It gives us an idea for the finest predicted crop which will be cultivate in the field weather conditions. These predictions can be done by a machine learning algorithm called Random Forest. It will attain the crop prediction with best accurate value. The algorithm random forest is used to give the best crop yield model by considering least number of models. It is very useful to predict the yield of the crop in agriculture sector.

Chapter 3

PROJECT DESCRIPTION

3.1 Existing System

Artificial Neural Networks (ANNs) are powerful tools for crop yield prediction, leveraging their ability to capture complex, non-linear relationships within data. In the context of crop yield prediction, ANNs typically consist of input, hidden, and output layers. Input nodes represent various features influencing crop yield, such as weather conditions, soil properties, and agricultural practices. The hidden layers, containing interconnected nodes, process this information, learning patterns and relationships. The output layer provides the predicted crop yield.

Training an ANN involves feeding historical data into the network, adjusting the weights between nodes through backpropagation, and minimizing the prediction error. Factors like temperature, precipitation, and fertilizer use contribute to the network's learning process. Regularization techniques may be applied to prevent overfitting.

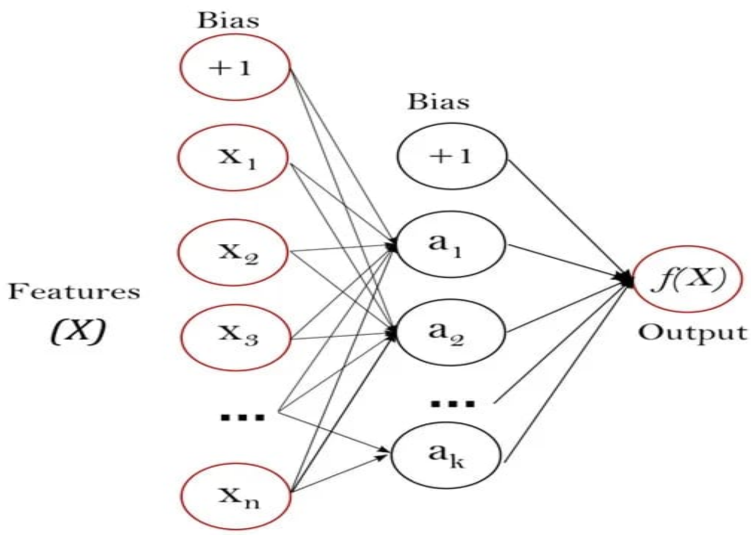


Figure 3.1: **ANN Diagram**

The Fig 3.1 represents the Artificial neural Networks (ANNs) diagram of one neuron is interlinked with all the remaining neurons and output is executed based on the input. Which predicts the error.

3.2 Proposed System

The random forest classifier builds many decision trees after selecting a random slice of data. It then takes the choice to decide on the final categorization of the data after combining the votes from several decision trees as shown in the Fig 3.2.

A single decision tree has a higher chance of making an error, but when several decision trees are used in the classification process, the error decreases and the accuracy rises.

When analysing the influence of each output/decision from any of the decision trees, this method employs the idea of weights. A tree with a high mistake rate receives a low weight, whereas a tree with a low error rate receives a high weight.

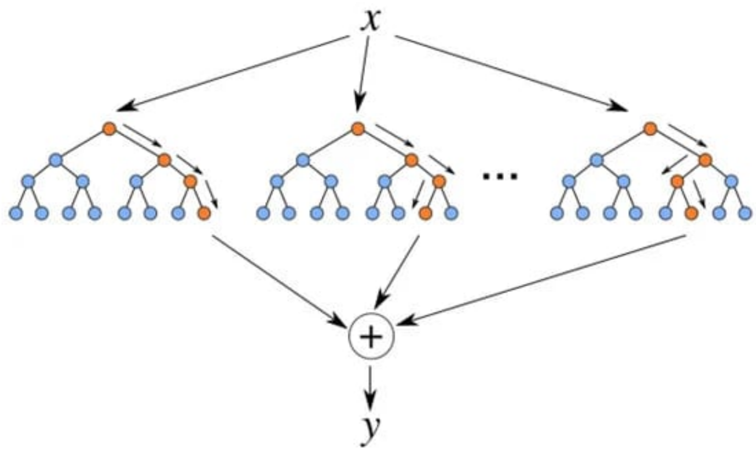


Figure 3.2: **Random Forest Classification**

The Fig 3.2 represents to start with this RFA a data set is needed,then that data is being preprocessed and data clustering will takes place the above process all takes place in training phase . In testing phase input should be given to the algorithm for the output prediction. Here in RFA it wil build decision tress in which each one has its data set attribute value.

Advantages of Random Forest Classifications

1. It reduces overfitting in decision trees and helps to improve the accuracy.

2. It is flexible to both classification and regression problems.

3. This algorithm is also very robust because it uses multiple decision trees to arriveat its result.

4. This algorithm offers you relative feature importance that allows you to select the most contributing features for your classifier easily.

3.3 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.

3.3.1 Economic Feasibility

This study is carried out to check the economic impact that the system will have on the Crop yield. As the developed system was well within the budget and this was achieved because most of the technologies used are freely available.

3.3.2 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 and the prediction of Crop yield. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this Prediction by using Machine learning.

3.3.3 Social Feasibility

Predicting crop yield using Random Forest enhances social feasibility by offering farmers valuable insights for informed decision-making. By leveraging advanced technology, it promotes sustainable agriculture, minimizes resource wastage, and supports food security. Empowering farmers with accurate predictions fosters resilience, mitigates risks, and contributes to economic stability, creating a positive impact on rural communities. This technological integration aligns with societal needs, fostering a collaborative approach towards precision farming and reinforcing the importance of data-driven solutions in agriculture.

3.4 System Specification

3.4.1 Hardware Specification

• System :Intel i3 or Above

• Hard Disk : 40 GB

• Monitor : 14’ Colour Monitor

• Mouse : Optical Mouse

• Ram : 1GB or Above

• Platform : Jupyter Notebook

3.4.2 Software Specification

• System :Intel i3 or Above

• Hard Disk : 40 GB

• Monitor : 14’ Colour Monitor

• Mouse : Optical Mouse

• Ram : 1GB or Above.

3.4.3 Standards and Policies

Google Colaboratory

Colab is a hosted Jupyter Notebook service that requires no setup to use and provides free access to computing resources, including GPUs and TPUs. Colab is especially well suited to machine learning, data science, and education.

Standard Used: ISO/IEC 27001

Jupyter

It’s like an open source web application that allows us to share and create the documents which contains the live code, equations, visualizations and narrative text. It can be used for data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning.

Standard Used: ISO/IEC 27001

Chapter 4

METHODOLOGY

4.1 Architecture for Crop Yield Prediction

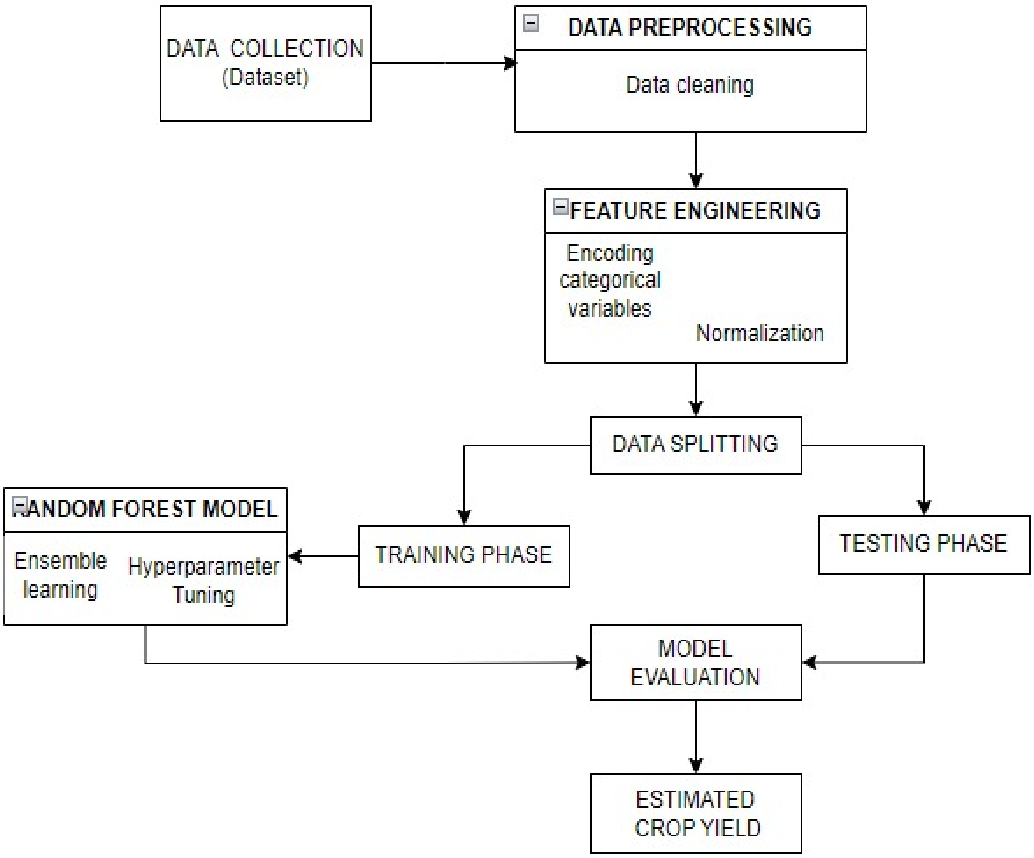


Figure 4.1: **Architecture Diagram**

The Fig 4.1 represents the general architecture of the proposed model which helps us to understand the working of the the project. Continuous monitoring and updating of the model are essential to account for changing environmental conditions, evolving farming practices, and new data inputs.The selected algorithm is trained on the preprocessed data, where it learns patterns and relationships between input features and crop yields.

4.2 Design Phase

4.2.1 Data Flow Diagram

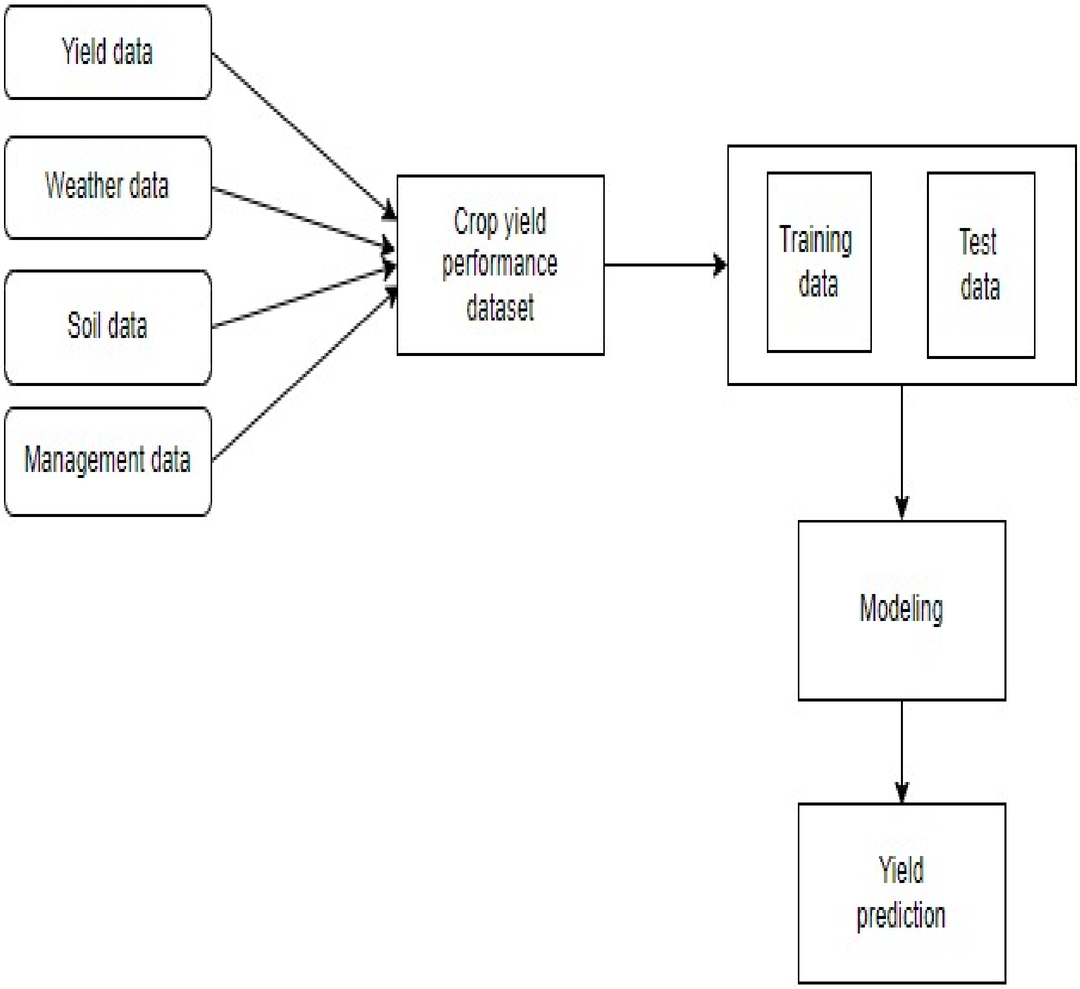


Figure 4.2: **Data Flow Diagram**

The Fig 4.2 represents the data flow diagram of the proposed model which helps us to understand the working of the the project The data set which contains various parameters like temparature, rainfall, previous year crop data e.t.c is preprocessed and a new data is formed for that data Machine learning algorithm is applied and a user input is given so the the predicted yield will be generated .

4.2.2 Use Case Diagram

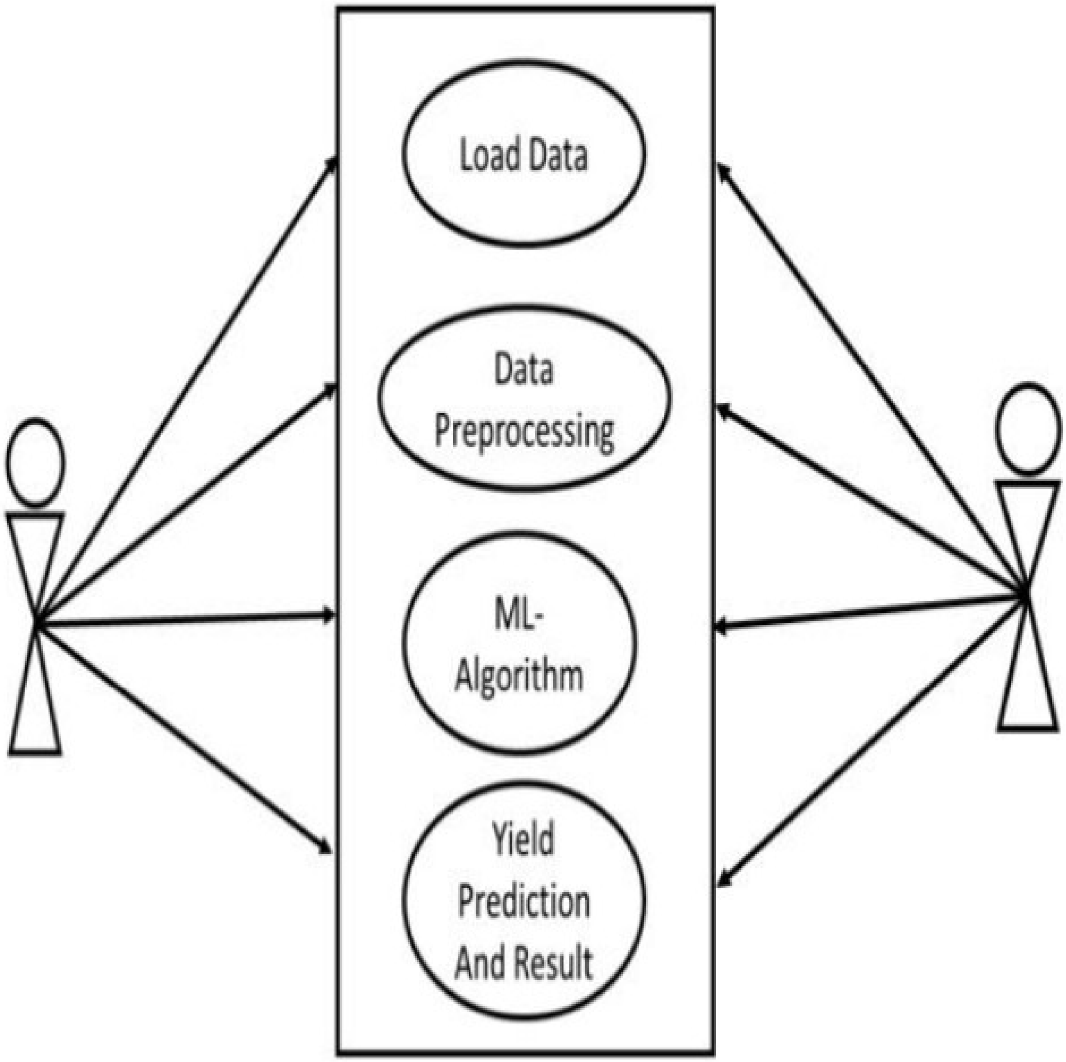


Figure 4.3: **Use Case Diagram**

The Fig 4.3 represents the use case diagram is a collection of situations that describe how a source and a destination connect. The link between actors and use cases is depicted in a use case diagram. Use cases and actors are the two primary components of a use case diagram. This enables them to make informed decisions regarding crop selection, planting schedules, resource allocation, and risk management strategies. For instance, suppose a farmer in a particular region is considering which crops to plant for next growing season.

4.2.3 Class Diagram

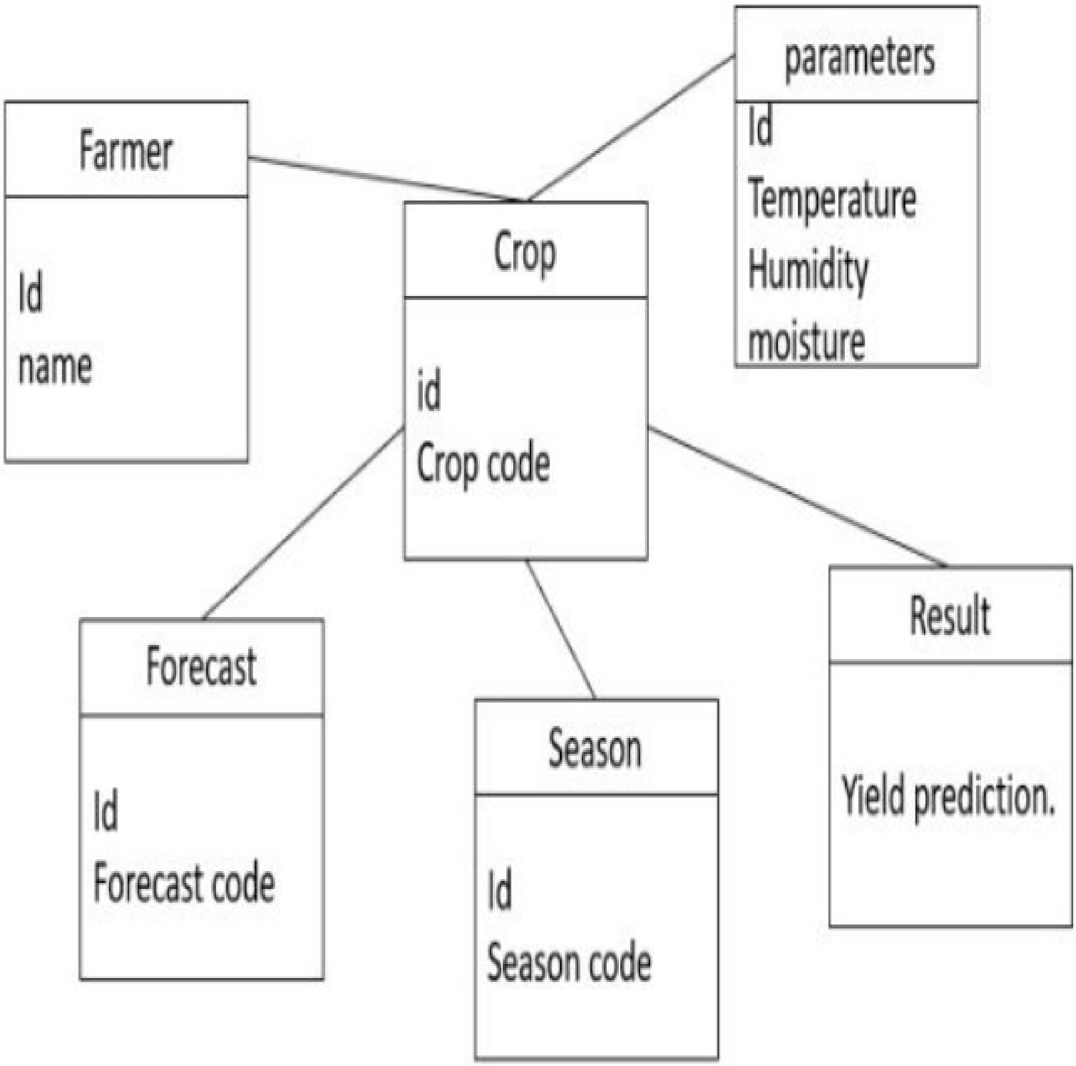


Figure 4.4: **Class Diagram**

The Fig 4.4 represents Class diagram, is the most common diagram type for software documentation. Since most software being created nowadays is still based on the Object- Oriented Programming paradigm, using class diagrams to document the software turns out to be a common-sense solution. This happens because OOP is based on classes and the relations between them.

4.2.4 Sequence Diagram

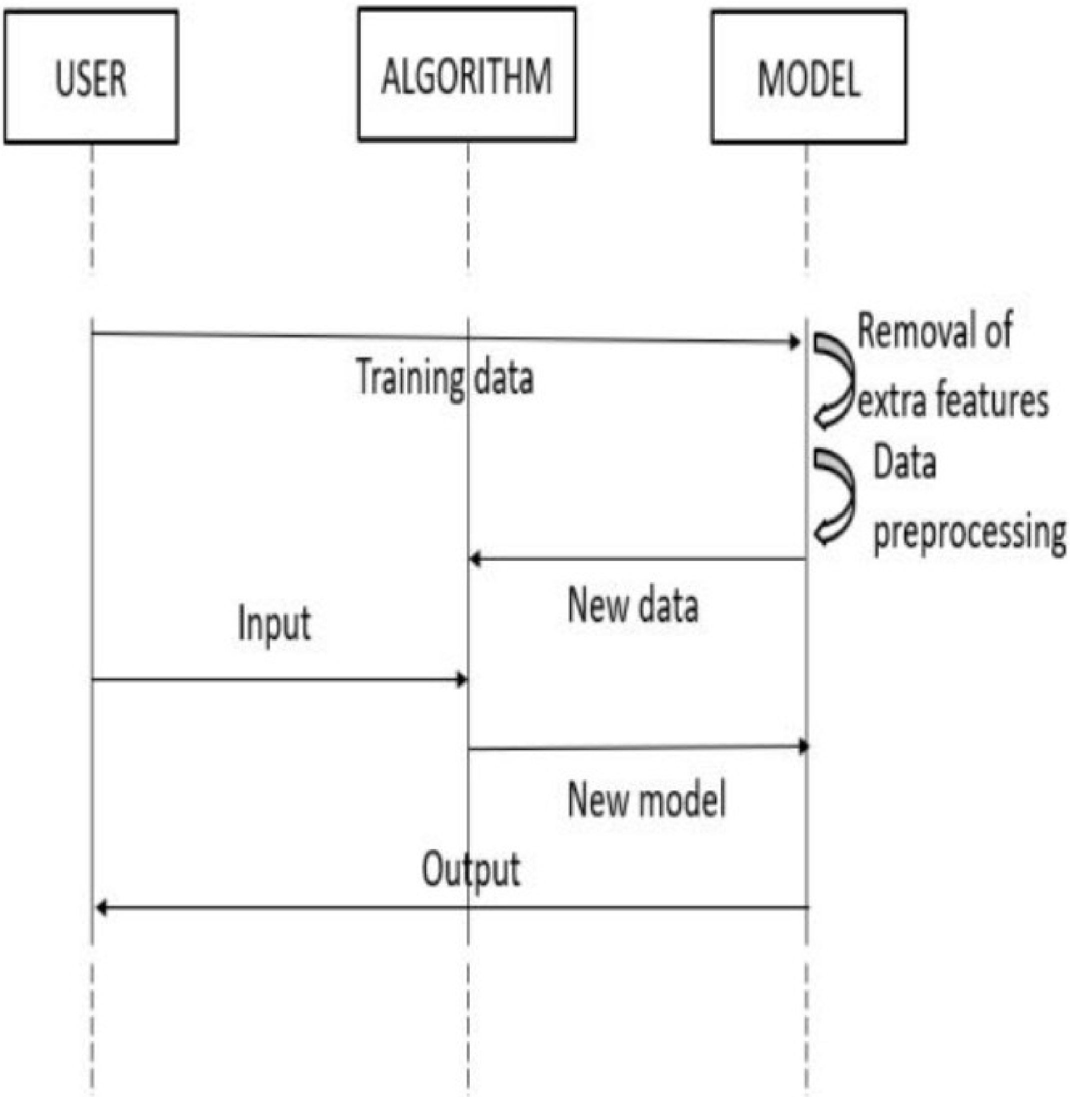


Figure 4.5: **Sequence Diagram**

The Fig 4.5 represents sequence diagram which simply depicts the sequence in which objects link or these activities take place. An event diagram is another name for a sequence diagram. Sequence diagrams demonstrate in what proportion the various components of a system interact. It begins with the collection of diverse agricultural data from sources like weather databases and satellite imagery. This data undergoes preprocessing, including cleaning and feature extraction, to prepare it for analysis.

4.2.5 Collaboration Diagram

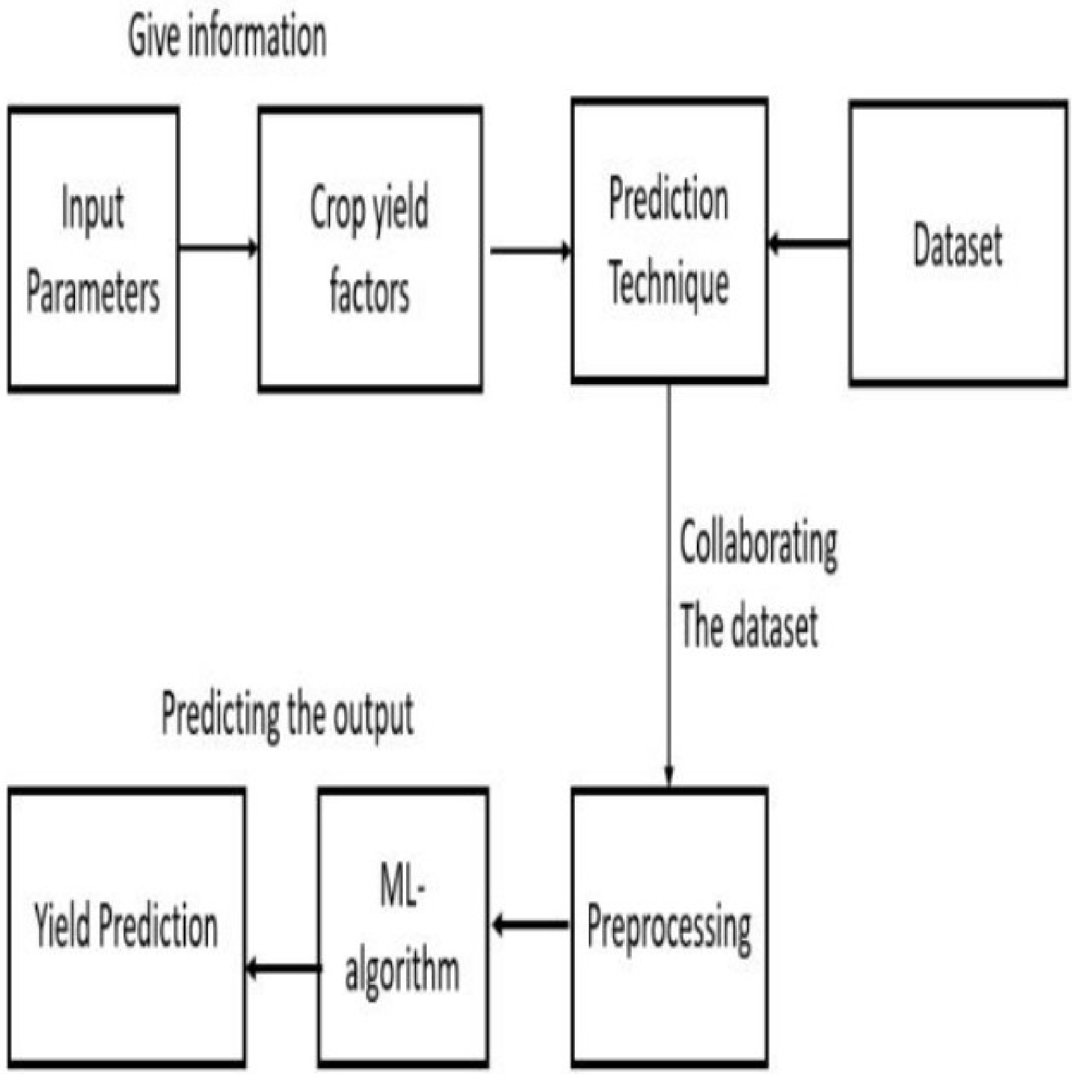


Figure 4.6: **Collaboration Diagram**

The Fig 4.6 represents the collaboration diagram, is also known as a communication diagram, is an illustration of the relation- ships and interactions among software objects in the Unified Modeling Language .The diagrams can be used to portray the dynamic behavior of a particular use case and define the role of each object.

4.2.6 Activity Diagram

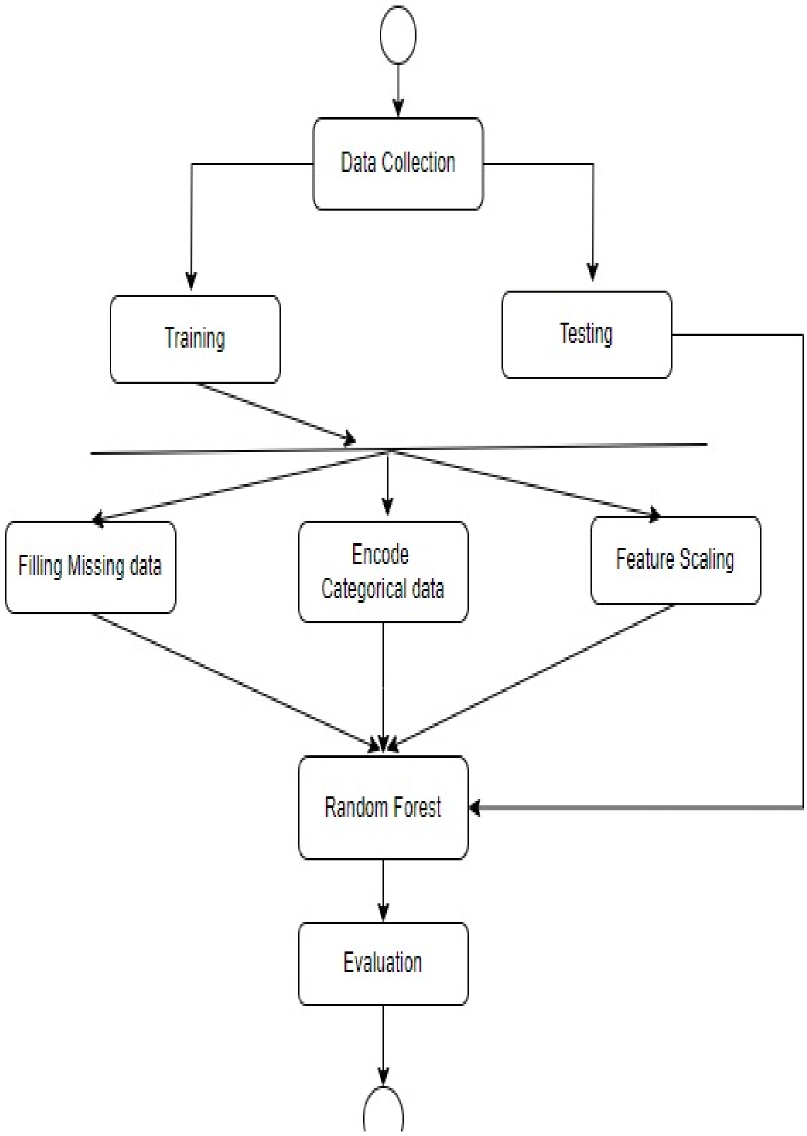


Figure 4.7: **Activity Diagram**

The Fig 4.7 represents the is activity diagram are probably the most important UML diagrams for doing business process modeling. In software development, it is generally used to describe the flow of different activities and actions. These can be both sequential and in parallel. Where it receives new input data, such as current weather conditions and agricultural practices, to generate crop yield predictions.

4.3 Algorithm & Pseudo Code

4.3.1 Enhanced Decision Tree Algoroithm

1. Data Selection:

• A dataset with features and corresponding target variables is required.

2. Bootstrapped Sampling (Bagging):

• Random Forest builds multiple decision trees, and each tree is trained on a bootstrapped sample of the original dataset.

• Bootstrapping involves randomly selecting a subset of data with replacement.

3. Feature Randomization:

• At each node of a decision tree, a random subset of features is considered for splitting.

• This helps to introduce diversity among the trees and reduces the risk of overfitting.

4. Growing Decision Trees:

• Each decision tree is grown by recursively splitting nodes based on the selected features.

• The splitting is done by choosing the feature that provides the best split according to a specified criterion.

5. Voting (Classification):

• For classification, the final prediction is determined by a majority vote from all the trees.

6. Hyperparameter Tuning:

• Random Forest has hyperparameters like the number of trees, the maximum depth of trees, and the minimum number of samples required to split a node.

7. Final Model:

• The final Random Forest model is a collection of decision trees, each trained on a different subset of data and considering a random subset of features at each split.

4.3.2 Pseudo Code

```
# Import necessary libraries
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

# Load your dataset (X features, y labels)
# Assume X is a matrix of features and y is a vector of labels (crop types or categories)

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize the Random Forest Classifier
rf_classifier = RandomForestClassifier(n_estimators=100, random_state=42)

# Train the classifier on the training data
rf_classifier.fit(X_train, y_train)

# Make predictions on the test data
y_pred = rf_classifier.predict(X_test)

# Evaluate the accuracy of the model
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

# Now, you can use the trained model to make predictions on new data
# For example, if you have new features in a variable called 'new_data'
new_predictions = rf_classifier.predict(new_data)
print("Predictions for new data:", new_predictions)
```

Description

1. For each tree:

   a. Randomly select a subset of features.

   b. Build a decision tree using the selected features and bootstrapped data

2. Aggregate predictions from all trees for classification (voting) or regression (averaging).

3. Output the final prediction.

4.4 Module Description

4.4.1 Data Collection

P18 fx

	A	B	C	D	E	F	G	H	I	J
1	N_SOIL	P_SOIL	K_SOIL	TEMPERATU	HUMIDITY	ph	RAINFALL	STATE	CROP_PRIC	CROP
2	90	42	43	20.879744	82.002744	6.5029853	202.93554	Andaman ar	7000	Rice
3	85	58	41	21.770462	80.319644	7.0380964	226.65554	Andaman ar	5000	Rice
4	60	55	44	23.004459	82.320763	7.8402071	263.96425	Andaman ar	7000	Rice
5	74	35	40	26.491096	80.158363	6.9804009	242.86403	Andaman ar	7000	Rice
6	78	42	42	20.130175	81.604873	7.6284729	262.71734	Andaman ar	120000	Rice
7	69	37	42	23.058049	83.370118	7.0734535	251.055	Andaman ar	3500	Rice
8	69	55	38	22.708838	82.639414	5.7008057	271.32486	Andaman ar	7500	Rice
9	94	53	40	20.277744	82.894086	5.7186272	241.97419	Andaman ar	6500	Rice
10	89	54	38	24.515881	83.535216	6.6853464	230.44624	Andaman ar	10000	Rice
11	68	58	38	23.223974	83.033227	6.3362535	221.2092	Andaman ar	11000	Rice
12	91	53	40	26.527235	81.417538	5.3861678	264.61487	Andaman ar	9000	Rice
13	90	46	42	23.978982	81.450616	7.502834	250.08323	Andaman ar	5600	Rice
14	78	58	44	26.800796	80.886848	5.1086818	284.43646	Andaman ar	6000	Rice
15	93	56	36	24.014976	82.056872	6.9843537	185.27734	Andaman ar	3000	Rice
16	94	50	37	25.665852	80.66385	6.9480198	209.58697	Andaman ar	3000	Rice
17	60	48	39	24.282094	80.300256	7.0422991	231.08633	Andhra Prac	620	Rice
18	85	38	41	21.587118	82.788371	6.2490507	276.65525	Andhra Prac	300	Rice
19	91	35	39	23.79392	80.41818	6.9708598	206.26119	Andhra Prac	760	Rice
20	77	38	36	21.865252	80.192301	5.9539333	224.55502	Andhra Prac	4600	Rice
21	88	35	40	23.579436	83.587603	5.8539321	291.29866	Andhra Prac	1900	Rice
22	89	45	36	21.325042	80.474764	6.4424754	185.49747	Andhra Prac	1950	Rice
23	76	40	43	25.157455	83.117135	5.0701757	231.38432	Andhra Prac	1760	Rice
24	67	59	41	21.947667	80.973842	6.0126326	213.35609	Assam	7000	Rice
25	83	41	43	21.052536	82.678395	6.2540285	233.10758	Assam	2400	Rice
26	98	47	37	23.483813	81.332651	7.3754829	224.05812	Assam	2800	Rice
27	66	53	41	25.075635	80.523891	7.7789152	257.00389	Assam	6400	Rice
28	97	59	43	26.359272	84.044036	6.2865002	271.35861	Assam	850	Rice
29	97	50	41	24.529227	80.544986	7.07096	260.2634	Assam	850	Rice
30	60	49	44	20.775761	84.497744	6.2448415	240.08106	Assam	350	Rice

indiancrop_dataset

Figure 4.8: **Crop Dataset**

The Fig 4.8 represents the dataset used in this project is downloaded from Kaggle website. The size of the entire dataset itself is around 1 GB. The data in the train folder consists of samples for temparature, rainfall and previous crop year's data respectively. Soil characteristics such as nutrient levels, pH, texture, and moisture content play a significant role in determining plant health and nutrient availability. These datasets encompass a wide range of variables essential for understanding the complex interactions influencing crop growth and productivity.

4.4.2 Data Preprocessing

Data preprocessing is an essential step in building machine learning models, including Random Forests for crop yield prediction. This process involves several key steps to ensure that the data is clean, consistent, and suitable for use in predictive algorithms.

•Data Cleaning

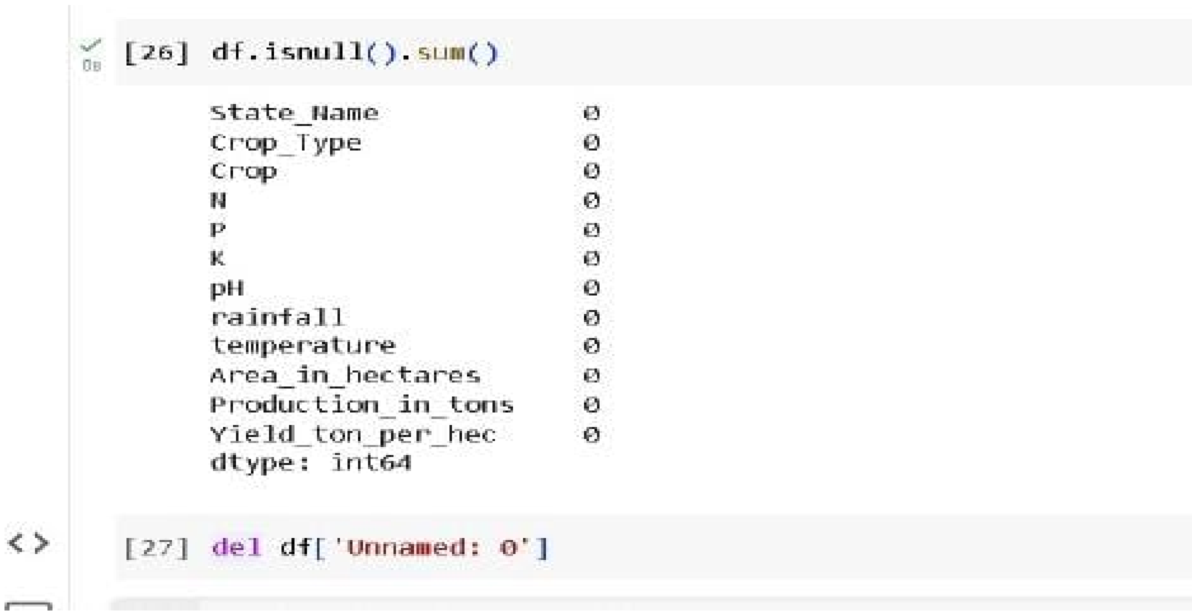


Figure 4.9: **Data Cleaning**

The Fig 4.9 represents the Data cleaning is a critical step in the process of crop yield prediction using machine learning, ensuring that the input data is accurate, consistent, and suitable for analysis. This process involves identifying and rectifying errors.This process involves identifying and rectifying errors, inconsistencies, and missing values to ensure the quality and reliability of the data. One aspect of data cleaning involves addressing outliers, anomalies, and erroneous data points that can distort the predictive patterns and compromise the accuracy of the models.

• Data Transformation

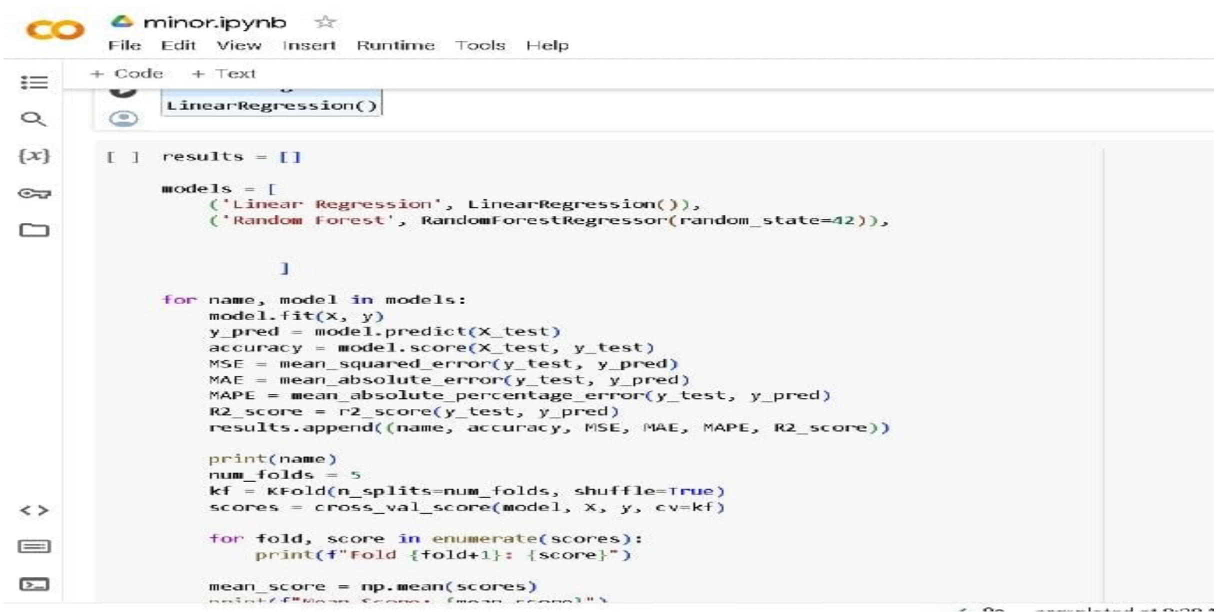


Figure 4.10: **Data Transformation**

The Fig 4.10 represents the Data transformations play a crucial role in enhancing the quality and suitability of agricultural datasets for crop yield prediction using machine learning. These transformations involve converting, scaling, and encoding raw data into formats that are more conducive to analysis and modeling.Furthermore, data cleaning entails standardizing units, formats, and scales across different variables to facilitate meaningful comparisons and analyses. For example, converting temperature measurements to a consistent scale ensures uniformity in the data.providing valuable insights to support sustainable farming practices and food security efforts.

4.4.3 Data Splitting

```
[ ] train, test = data[data['is_train']==True],data[data['is_train']==False]
    print('training data:', len(train))
    print('test data:', len(test))

    training data: 1667
    test data: 533
```

```
features = data.columns[:10]
print(features)

Index(['N_SOIL', 'P_SOIL', 'K_SOIL', 'TEMPERATURE', 'HUMIDITY', 'ph',
       'RAINFALL', 'STATE', 'CROP_PRICE', 'CROP'],
      dtype='object')
```

```
[ ] y= pd.factorize(train['CROP'])[0]
    y

    array([-1, -1, -1, ...,  1,  1,  1])
```

Figure 4.11: **Data Splitting**

The Fig 4.11 represents the split data into training and testing sets to assess the model's performance on unseen data.enhancing the interpretability, efficiency, and predictive performance of crop yield prediction models.The process of data splitting is crucial for developing reliable and effective crop yield prediction models, ensuring that they are trained, validated, and evaluated using representative datasets that capture the variability present in agricultural systems.The process of data splitting is crucial for developing reliable and effective crop yield prediction models.Finally, the testing set, completely separate from the training and validation sets, is reserved for assessing the final performance of the trained model on unseen data.

4.4.4 Random Forest Model

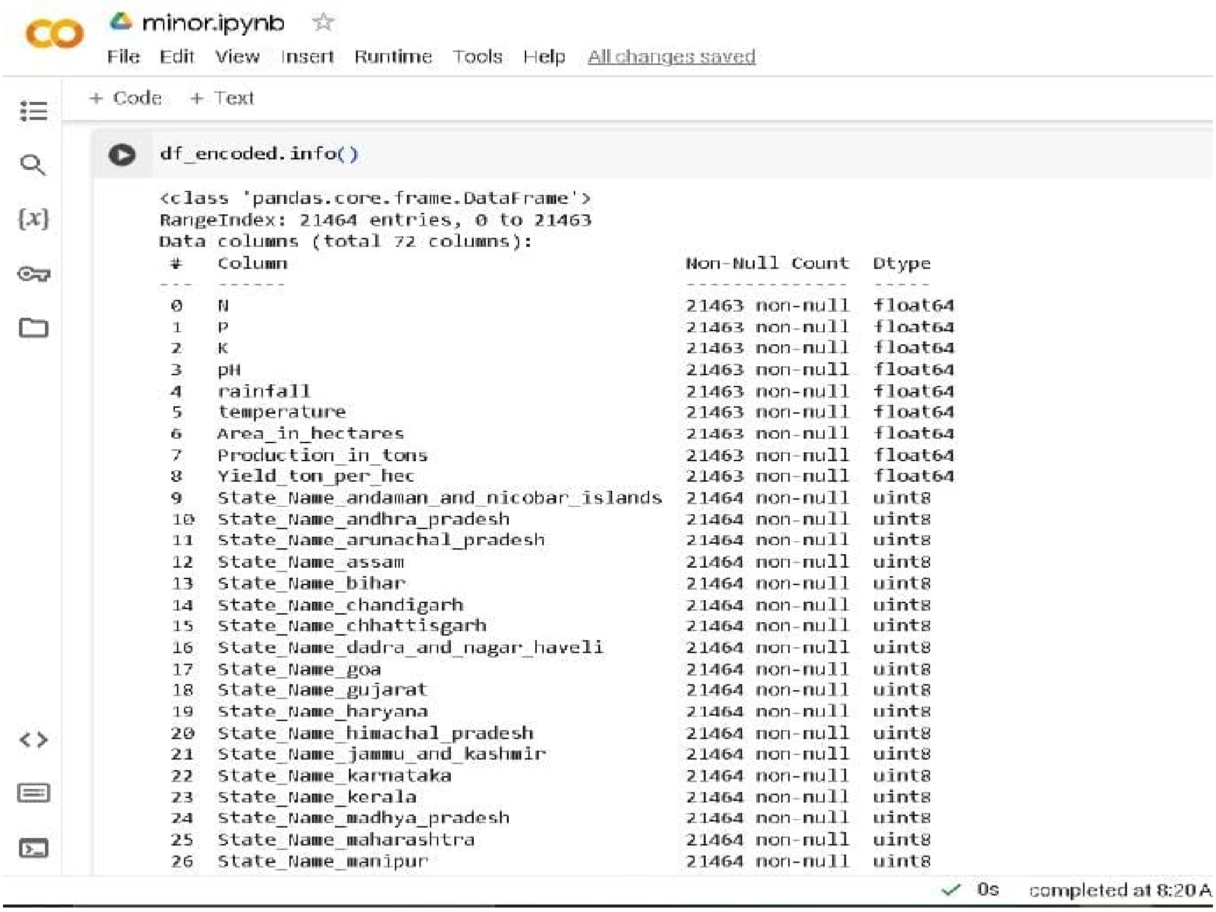


Figure 4.12: **Random Forest Model**

The Fig 4.12 represents the random Forest model lies in its resilience to overfitting, thanks to its inherent randomness in feature selection and bootstrapping. This property helps mitigate the risk of model bias and variance, resulting in more robust and generalizable predictions. For classification tasks, the algorithm typically employs a majority voting scheme, where the most commonly predicted class across all trees is selected. The algorithm averages the predictions of individual trees to produce the final output. This randomness helps to decorrelate the trees and prevents overfitting, leading to improved generalization performance.

4.4.5 Evaluation

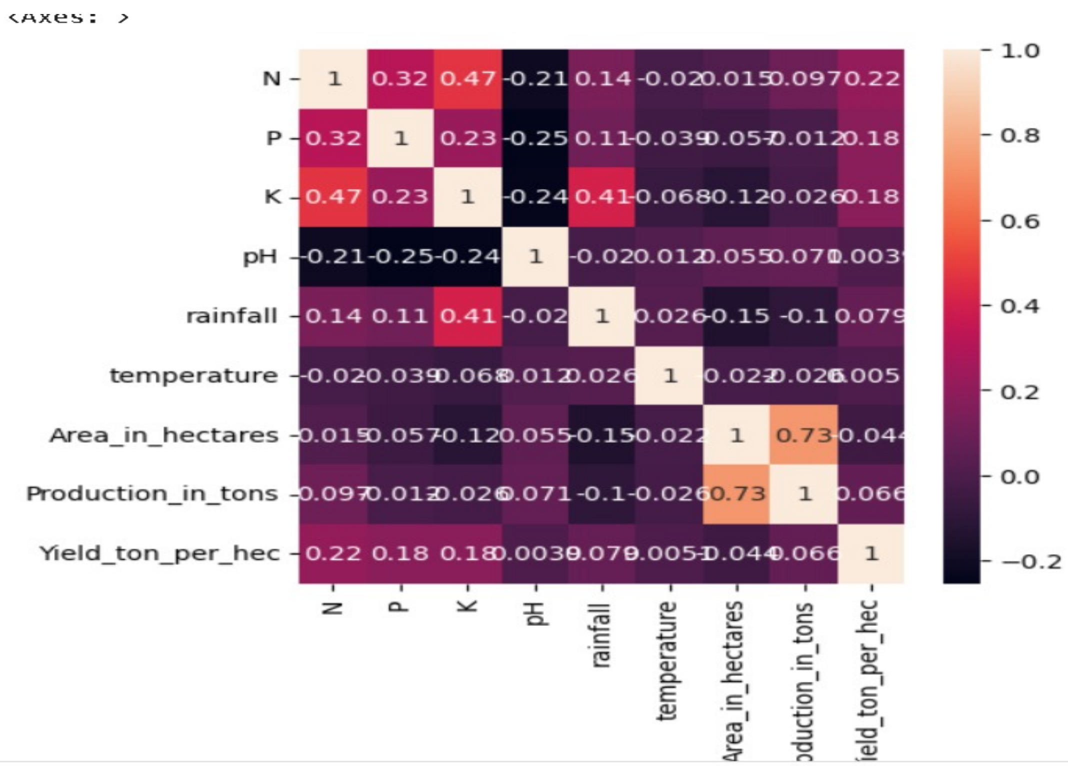


Figure 4.13: **Evaluation**

The Fig 4.13 represents the Evaluation of crop yield prediction models using machine learning is essential to assess their accuracy, reliability, and effectiveness in supporting agricultural decision-making. Visualization techniques, such as scatter plots and residual plots, provide further insights into the relationship between predicted and actual crop yields. Through rigorous evaluation, stakeholders can ensure that crop yield prediction models. Mean Absolute Percentage Error (MAPE) offers a percentage-based measure of prediction accuracy, facilitating interpretation across different crop yield levels.

4.5 Steps to execute/run/implement the project

4.5.1 Data Collection and Preprocessing

1. Gather historical data on crop yields which includes relevant features such as weather conditions, soil quality, crop type, and any other factors that might affect yield.
2. Handle missing values by either removing incomplete entries or imputing values based on statistical methods.
3. Split the dataset into training and testing sets

4.5.2 Model Training using Random Forest

1. Use a programming language such as Python and libraries like scikit-learn for implementing the Random Forest algorithm.

2. Create a Random Forest model using the appropriate parameters and train the model on training dataset.

3. Evaluate the model's performance on the testing set using metrics

4. If the model performance is not satisfactory, consider tuning hyperparameters or trying other algorithms.

4.5.3 Crop Yield Prediction

1. Collect new data for the upcoming crop season, including weather forecasts, soil conditions, etc. This data should have the same features used during training.

2. Apply the same preprocessing steps used for the training data to the new input data.

3. Use the trained Random Forest model to predict crop yield for the upcoming season based on the preprocessed input data.

4. Analyze the model predictions, compare them with the actual yields if available, and assess the model's accuracy and reliability.

Chapter 5

IMPLEMENTATION AND TESTING

5.1 Input and Output

5.1.1 Input Design

P18

	A	B	C	D	E	F	G	H	I	J
1	N_SOIL	P_SOIL	K_SOIL	TEMPERATU	HUMIDITY	ph	RAINFALL	STATE	CROP_PRIC	CROP
2	90	42	43	20.879744	82.002744	6.5029853	202.93554	Andaman ar	7000	Rice
3	85	58	41	21.770462	80.319644	7.0380964	226.65554	Andaman ar	5000	Rice
4	60	55	44	23.004459	82.320763	7.8402071	263.96425	Andaman ar	7000	Rice
5	74	35	40	26.491096	80.158363	6.9804009	242.86403	Andaman ar	7000	Rice
6	78	42	42	20.130175	81.604873	7.6284729	262.71734	Andaman ar	120000	Rice
7	69	37	42	23.058049	83.370118	7.0734535	251.055	Andaman ar	3500	Rice
8	69	55	38	22.708838	82.639414	5.7008057	271.32486	Andaman ar	7500	Rice
9	94	53	40	20.277744	82.894086	5.7186272	241.97419	Andaman ar	6500	Rice
10	89	54	38	24.515881	83.535216	6.6853464	230.44624	Andaman ar	10000	Rice
11	68	58	38	23.223974	83.033227	6.3362535	221.2092	Andaman ar	11000	Rice
12	91	53	40	26.527235	81.417538	5.3861678	264.61487	Andaman ar	9000	Rice
13	90	46	42	23.978982	81.450616	7.502834	250.08323	Andaman ar	5600	Rice
14	78	58	44	26.800796	80.886848	5.1086818	284.43646	Andaman ar	6000	Rice
15	93	56	36	24.014976	82.056872	6.9843537	185.27734	Andaman ar	3000	Rice
16	94	50	37	25.665852	80.66385	6.9480198	209.58697	Andaman ar	3000	Rice
17	60	48	39	24.282094	80.300256	7.0422991	231.08633	Andhra Prac	620	Rice
18	85	38	41	21.587118	82.788371	6.2490507	276.65525	Andhra Prac	300	Rice
19	91	35	39	23.79392	80.41818	6.9708598	206.26119	Andhra Prac	760	Rice
20	77	38	36	21.865252	80.192301	5.9539333	224.55502	Andhra Prac	4600	Rice
21	88	35	40	23.579436	83.587603	5.8539321	291.29866	Andhra Prac	1900	Rice
22	89	45	36	21.325042	80.474764	6.4424754	185.49747	Andhra Prac	1950	Rice
23	76	40	43	25.157455	83.117135	5.0701757	231.38432	Andhra Prac	1760	Rice
24	67	59	41	21.947667	80.973842	6.0126326	213.35609	Assam	7000	Rice
25	83	41	43	21.052536	82.678395	6.2540285	233.10758	Assam	2400	Rice
26	98	47	37	23.483813	81.332651	7.3754829	224.05812	Assam	2800	Rice
27	66	53	41	25.075635	80.523891	7.7789152	257.00389	Assam	6400	Rice
28	97	59	43	26.359272	84.044036	6.2865002	271.35861	Assam	850	Rice
29	97	50	41	24.529227	80.544986	7.07096	260.2634	Assam	850	Rice
30	60	49	44	20.775761	84.497744	6.2448415	240.08106	Assam	350	Rice

indiancrop_dataset

Figure 5.1: **Data Set for Crop Yield**

The Fig 5.1 represents the input design contains the dataset of Crop prediction with numerous parameters.These inputs typically include historical and real-time data sources, such as weather patterns, soil characteristics, crop types, planting dates, and agricultural management practices. Weather data encompasses factors like temperature, rainfall, humidity, and solar radiation, which directly influence crop growth and development.

5.1.2 Output Design

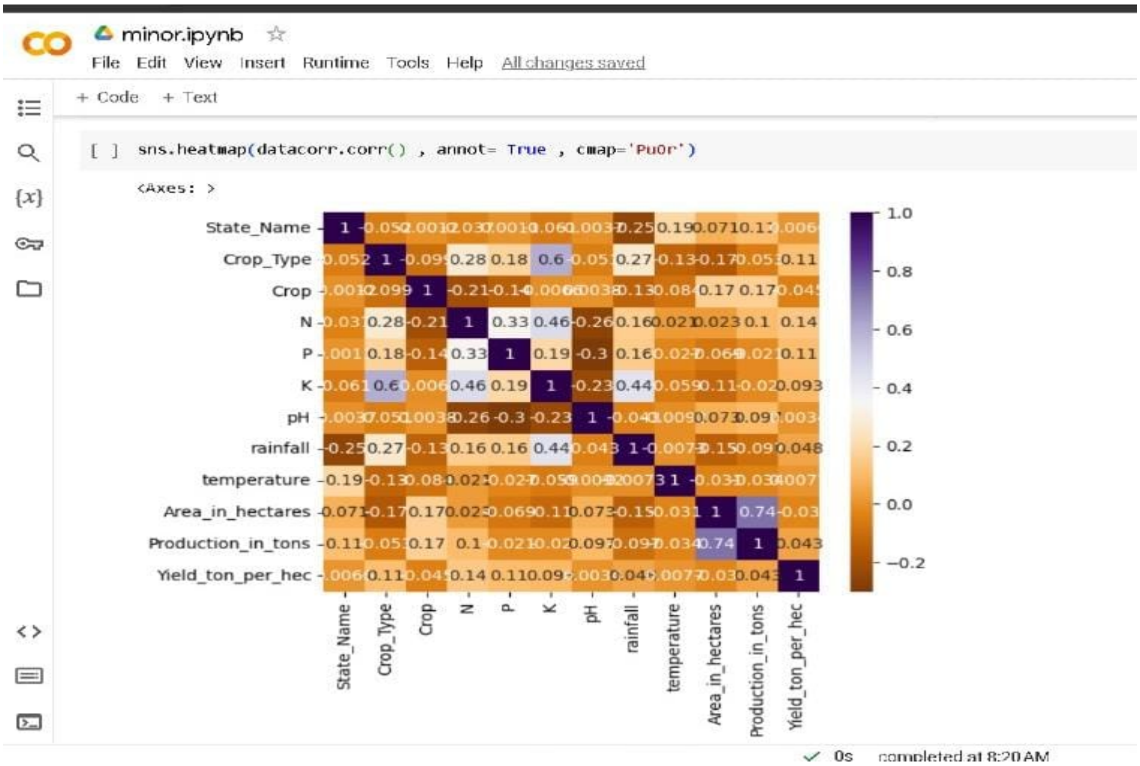


Figure 5.2: **Prediction of Crop Yield**

The Fig 5.2 represents the integrating these multifaceted inputs, crop yield prediction models can effectively forecast future harvests, empowering farmers with actionable insights to optimize agricultural practices, mitigate risks, and ensure sustainable production.

5.2 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 test. Each test type addresses a specific testing requirement.

5.3 Types of Testing

5.3.1 Unit Testing

Unit testing is a software testing technique where individual units or components of a software application are tested in isolation to ensure they work as expected. In the context of a machine learning model, a unit test might involve testing individual functions or methods responsible for tasks such as data preprocessing, feature engineering, model training, and prediction.

```
df['State_Name'].value_counts()

for i,j in enumerate(df['State_Name']):
  df.at[i,'State_Name']=str(j).replace(' ','_')

plt.figure(figsize=(5,5))
sns.heatmap(df.corr(numeric_only=True),annot=True)

df_encoded=pd.get_dummies(df,columns=['State_Name','Crop','Crop_Type'])

df_encoded.head(20)
```

5.3.2 Integration Testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent.

```
df=pd.read_csv("/content/Crop_production.csv")
df.info()

df.isnull().sum()

del df['Unnamed: 0']

df['Crop'].value_counts()

df['Area_in_hectares'].max()
df['Area_in_hectares'].min()
df['Yield_ton_per_hec'].max()
df.head()

from sklearn.model_selection import train_test_split
```

```
datacorr=df.copy()

from sklearn.preprocessing import LabelEncoder
categorical_columns = datacorr.select_dtypes(include=['object']).columns.tolist()
label_encoder = LabelEncoder()
for column in categorical_columns:
    datacorr[column] = label_encoder.fit_transform(datacorr[column])
```

5.3.3 System Testing

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

```
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import KFold
from sklearn.preprocessing import LabelEncoder
categorical_columns = datacorr.select_dtypes(include=['object']).columns.tolist()
label_encoder = LabelEncoder()
for column in categorical_columns:
    datacorr[column] = label_encoder.fit_transform(datacorr[column])

sns.heatmap(datacorr.corr() , annot= True , cmap='PuOr')

sns.set(palette='BrBG')
df.hist(figsize=(5,10));

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

5.3.4 Test Result

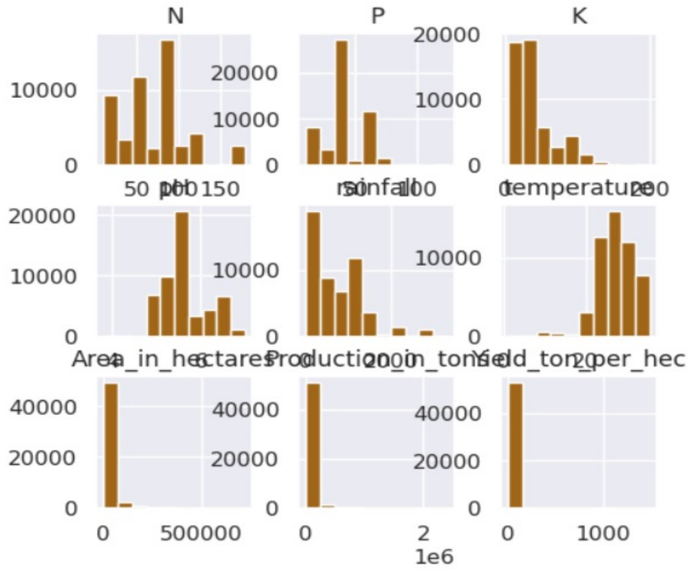


Figure 5.3: **Area in Production**

The Fig 5.3 represents the production across various regions and cropping areas. By harnessing advanced analytics and predictive modeling techniques, farmers and agricultural stakeholders can gain valuable insights into anticipated harvests, enabling them to make informed decisions.

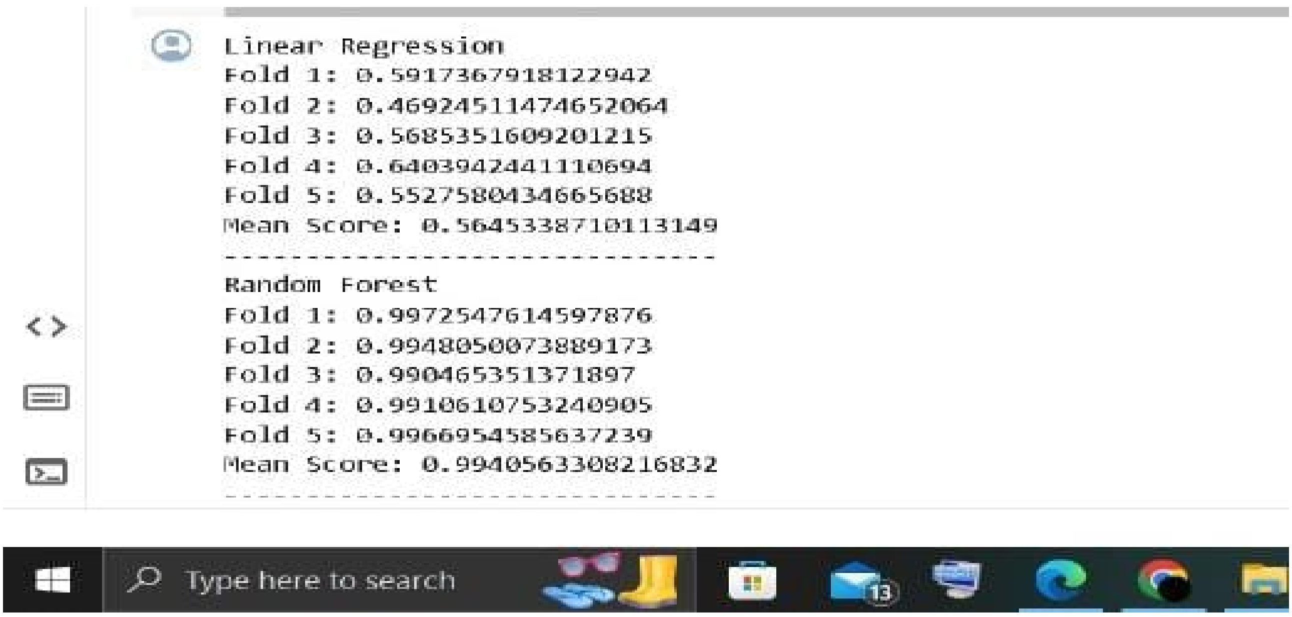


Figure 5.4: **Cross-Validation**

The Fig 5.4 represents the cross-validation test case provides valuable insights into the performance and generalization capabilities of the crop yield prediction model. By systematically evaluating the model's performance across multiple validation folds.

5.4 Efficiency of the Proposed System

The proposed system is based on the Random forest Algorithm that creates many decision trees. Accuracy of proposed system is done by using random forest gives the ouput approximately 98 percent. Random forest implements many decision trees and also gives the most accurate output when compared to the decision tree. Random Forest algorithm is used in the two phases. Firstly, the RF algorithm extracts subsamples from the original samples by using the bootstrap resampling method and creates the decision trees for each testing sample and then the algorithm classifies the decision trees and implements a vote with the help of the largest vote of the classification as a final result of the classification. The random Forest algorithm always includes some of the steps as follows: Selecting the training dataset:Using the bootstrap random sampling method we can derive the K training sets from the original dataset properties using the size of all training set the same as that of original training dataset. Building the random forest algorithm: Creating a classification regression tree each of the bootstrap training set will generate the K decision trees to form a random forest model, uses the trees that are not pruned. Looking at the growth of the tree, 31 this approach is not chosen the best feature as the internal nodes for the branches but rather the branching process is a random selection of all the trees gives the best features.

5.5 Comparison of Existing and Proposed System

Existing system:(Artificial Neural Network)

ANNs, inspired by the human brain's neural structure, can capture intricate relationships within data. In the context of crop yield prediction, ANNs can learn complex patterns from various input features like weather conditions, soil attributes, and historical yield data. They excel at handling non-linear relationships, but they come with challenges. ANNs often require a large amount of data for effective training, and overfitting can occur if not properly regularized. Moreover, ANNs are considered "black-box" models, making it challenging to interpret the reasoning behind their predictions.

Proposed system:(Random forest algorithm)

Random Forests, a popular ensemble learning method, offer several advantages for crop yield prediction. RF models consist of multiple decision trees, each trained on a random subset of the data. They are robust to overfitting, require less hyperparameter tuning, and handle both numerical and categorical features well. Additionally, RF provides a feature importance measure, aiding in the interpretation of the model's predictions.

Chapter 6

CONCLUSION AND FUTURE ENHANCEMENTS

6.1 Conclusion

Agriculture plays an important role for the economic growth of our country. There are many techniques to develop agriculture in different ways. Implementation of an algorithm so called Random Forest Algorithm using machine learning to improve the yield rate of the crops.

Since, the number of farmer suicides has been increasing day by day this system can be of great help in predicting crop sequences as well as maximizing yield rates and monetary benefits to the farmers. Also, successfully integrating machine learning with agriculture in predicting crop diseases, different irrigation patterns, studying crop simulations etc. This project will help the farmers to know the yield of their crop before cultivating onto the agricultural field and help them to take the appropriate decisions.0.8 is the accuracy value of crop yeild prediction using machine learning

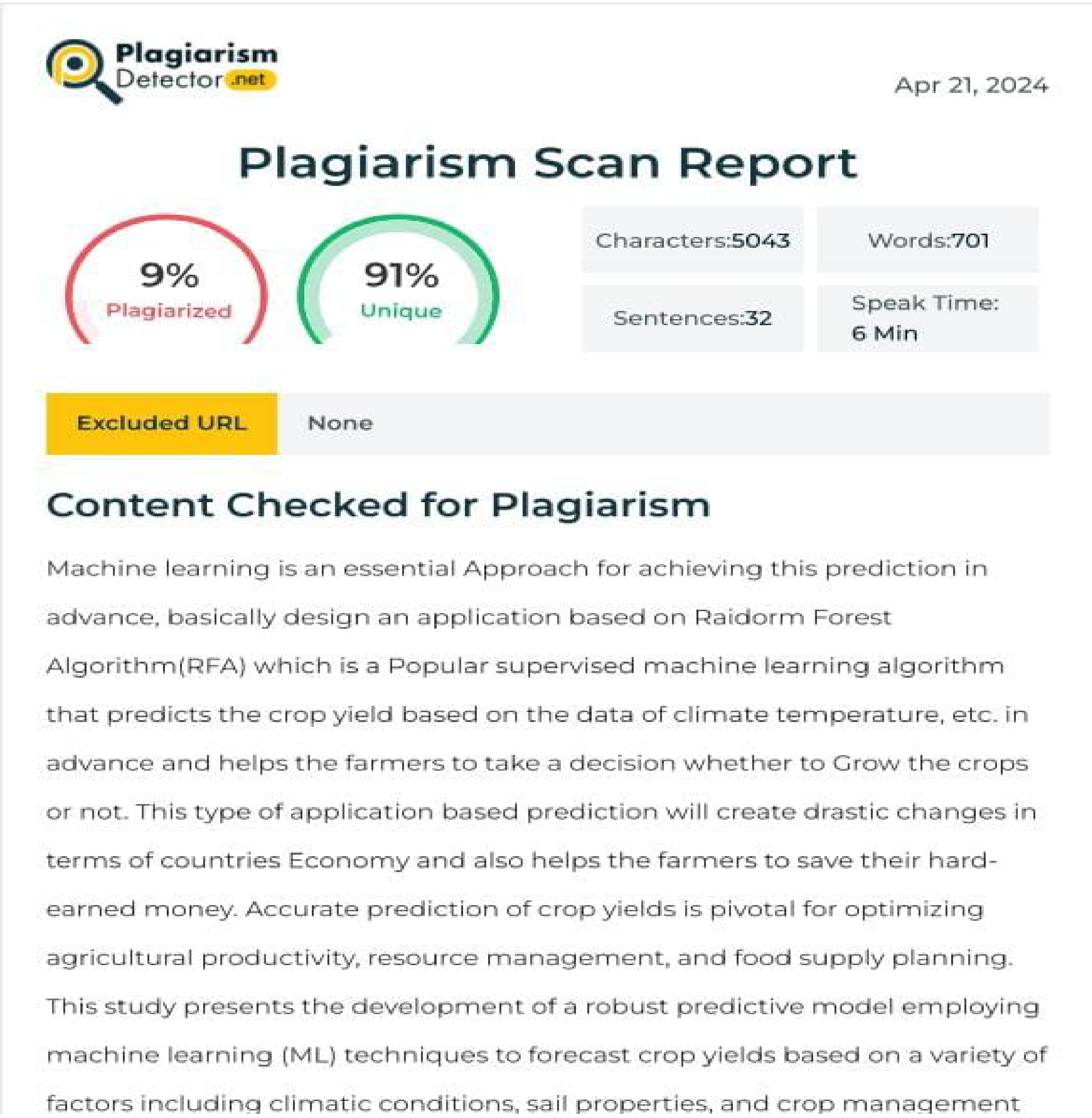
6.2 Future Enhancements

The application used here for crop yield prediction is just an application, but in future using upcoming software's and latest technologies this can increase the efficiency of the model and develop the model as a application in which framers can use it as a app through there smartphones by converting the whole system into their regional language.

Along with this system in future some features like crop price prediction, fertilizer recomder e.t.c can be added to the system. So that it might be usefull for the farmers.

Chapter 7

PLAGIARISM REPORT



Plagiarism Detector .net

Apr 21, 2024

Plagiarism Scan Report

9% Plagiarized

91% Unique

Characters:5043	Words:701
Sentences:32	Speak Time: 6 Min

Excluded URL	None

Content Checked for Plagiarism

Machine learning is an essential Approach for achieving this prediction in advance, basically design an application based on Raidorm Forest Algorithm(RFA) which is a Popular supervised machine learning algorithm that predicts the crop yield based on the data of climate temperature, etc. in advance and helps the farmers to take a decision whether to Grow the crops or not. This type of application based prediction will create drastic changes in terms of countries Economy and also helps the farmers to save their hard-earned money. Accurate prediction of crop yields is pivotal for optimizing agricultural productivity, resource management, and food supply planning. This study presents the development of a robust predictive model employing machine learning (ML) techniques to forecast crop yields based on a variety of factors including climatic conditions, sail properties, and crop management

Figure 7.1: **Plagiarism Report**

Chapter 8

SOURCE CODE & POSTER PRESENTATION

8.1 Source Code

```python

from flask import Flask, render_template
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor, BaggingRegressor
from xgboost import XGBRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error,
    mean_absolute_percentage_error
from sklearn.model_selection import train_test_split, cross_val_score, KFold
from sklearn.preprocessing import LabelEncoder
import numpy as np

app = Flask(_name_)

# Load the dataset
df = pd.read_csv("Crop_production.csv")
del df['Unnamed: 0']

# Preprocess the data
df['State_Name'] = df['State_Name'].apply(lambda x: str(x).replace(' ', '_'))
df_encoded = pd.get_dummies(df, columns=['State_Name', 'Crop', 'Crop_Type'])

# Feature selection
datacorr = df.copy()
categorical_columns = datacorr.select_dtypes(include=['object']).columns.tolist()
label_encoder = LabelEncoder()
for column in categorical_columns:
    datacorr[column] = label_encoder.fit_transform(datacorr[column])

# Split the data
X, y = datacorr.drop(labels='Production_in_tons', axis=1), datacorr['Production_in_tons']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

```

```python
# Models
models = [
    ('Linear Regression', LinearRegression()),
    ('Random Forest', RandomForestRegressor(random_state=42)),
    ('Gradient Boost', GradientBoostingRegressor(n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42)),
    ('XGBoost', XGBRegressor(random_state=42)),
    ('KNN', KNeighborsRegressor(n_neighbors=5)),
    ('Decision Tree', DecisionTreeRegressor(random_state=42)),
    ('Bagging Regressor', BaggingRegressor(n_estimators=150, random_state=42))
]


@app.route('/')
def home():
    results = []
    print("Home")
    for name, model in models:
        print(name,model)
        model.fit(X, y)
        y_pred = model.predict(X_test)
        print(y_pred)
        accuracy = model.score(X_test, y_test)
        print(accuracy)
        MSE = mean_squared_error(y_test, y_pred)
        print(MSE)
        MAE = mean_absolute_error(y_test, y_pred)
        print(MAE)
        MAPE = mean_absolute_percentage_error(y_test, y_pred)
        print(MAPE)
        R2_score = r2_score(y_test, y_pred)
        print(R2_score)
        results.append((name, accuracy, MSE, MAE, MAPE, R2_score))

        num_folds = 5
        kf = KFold(n_splits=num_folds, shuffle=True)
        scores = cross_val_score(model, X, y, cv=kf)
        mean_score = np.mean(scores)

    df_results = pd.DataFrame(results, columns=['Model', 'Accuracy', 'MSE', 'MAE', 'MAPE', 'R2_score'])
    df_styled_best = df_results.style.highlight_max(subset=['Accuracy', 'R2_score'], color='lightblue').highlight_min(
        subset=['MSE', 'MAE', 'MAPE'], color='lightblue').highlight_max(subset=['MSE', 'MAE', 'MAPE'], color='red').highlight_min(
        subset=['Accuracy', 'R2_score'], color='red')

    # Generate a simple Plotly graph for demonstration purposes
    plot_data = {'x': [1, 2, 3, 4, 5], 'y': [10, 11, 8, 14, 9]}

```

```python
    # Prepare data for predictions section
    predictions = []
    for name, model in models:
        model.fit(X, y)
        y_pred = model.predict(X_test)
        accuracy = model.score(X_test, y_test)
        MSE = mean_squared_error(y_test, y_pred)
        MAE = mean_absolute_error(y_test, y_pred)
        MAPE = mean_absolute_percentage_error(y_test, y_pred)
        R2_score = r2_score(y_test, y_pred)
        print({'name': name, 'accuracy': accuracy, 'mse': MSE, 'mae': MAE, 'mape': MAPE, 'r2_score': R2_score})
        predictions.append({'name': name, 'accuracy': accuracy, 'mse': MSE, 'mae': MAE, 'mape': MAPE, 'r2_score': R2_score})

    return render_template('index.html', table=df_styled_best.render(), plot_data=plot_data, predictions=predictions)


if _name_ == '_main_':
    app.run(debug=True)
```

8.2 Poster Presentation

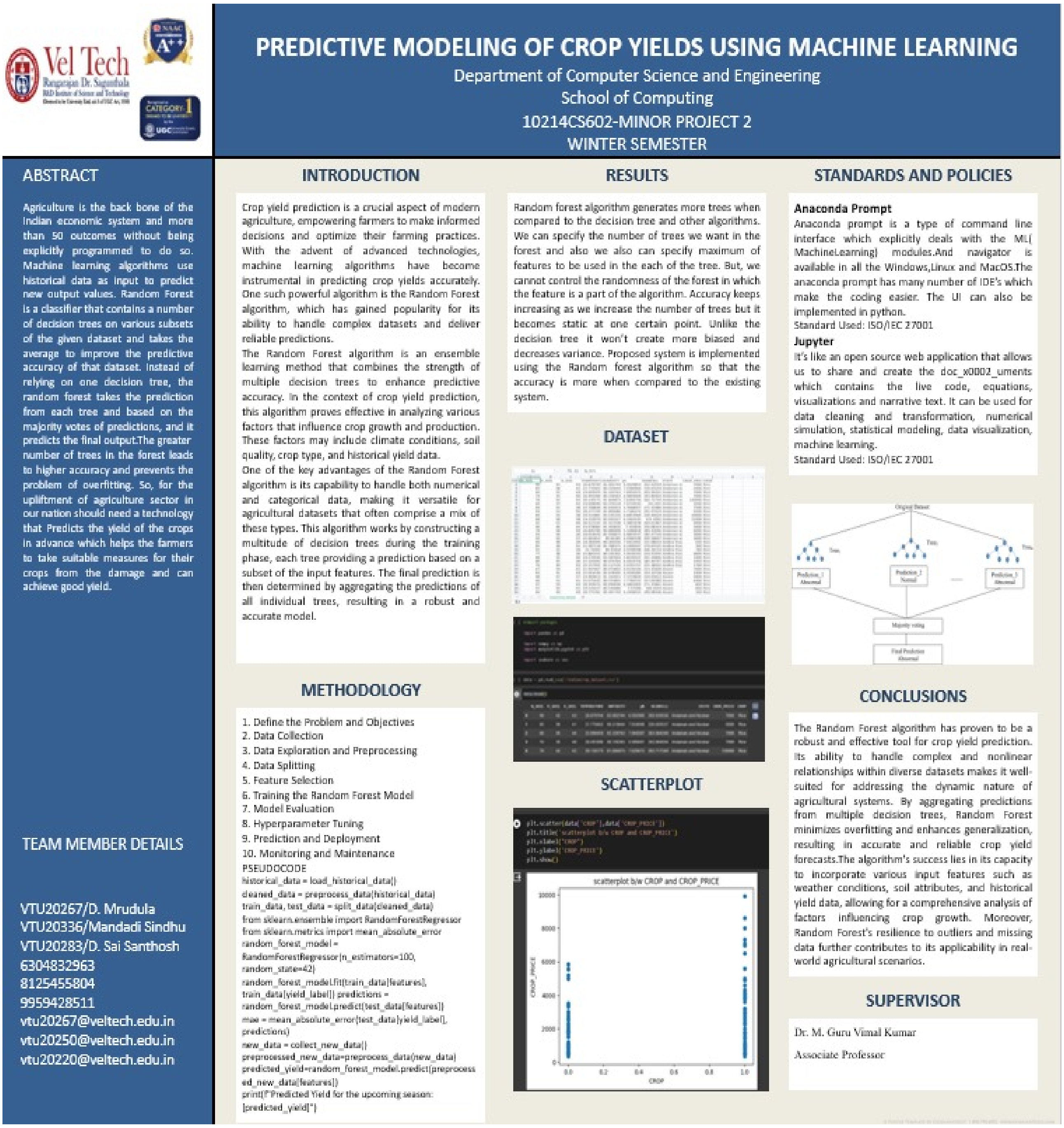


Figure 8.1: **Poster**

References

[1] Anastasiya Kolesnikova, Chi-Hwa Song, Won Don Lee “Crop selection method based on various environmental factors using machine learning,” International research journal of engineering and technology, vol. 04,issue 02,2021.

[2] Anupama C.G., Lakshmi C, “Crop yield estimation using machine learning in agricultural crop production,” Next Generation Computing Technologies (NGCT), 1st International Conference, 2020.

[3] Aruvansh Nigam; Saksham Garg; Archit Agrawal; Parul Agrawal, “Crop Yield Prediction Using Machine Learning Algorithms”, IEEE Fifth International Conference on Image Information Processing (ICIIP), 2021

[4] B.joesphine and K.prabha, “Crop yield prediction using ANN-Algorithm,” International Journal of Scientific and Technology research, vol. 9, issue 02, 2021

[5] Krishna Kumar, K. Rupa Kumar, R. G. Ashrit, “A Smart agricultural model using K-means and clustering techniques,” Indian journal of science and technology, vol. 9(38), 2021.

[6] L.Snehal and S.Rupa kumar, “A model for prediction of crop yield,”International Journal of Computational Intelligence and Informatics, Vol. 6, No. 4, 2021.

[7] M. Kalimuthu, P. Vaishnavi, M. Kishore, ”Crop Prediction using Machine Learning,” Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2021.

[8] N.Sagar Srivatsav, R.Agarwal “A Machine learning approach to predict crop yield and success rate,” IEEE pune , vol. 1-5, 2021.

[9] Patrick Helber, Benjamin Bischke, Peter Habelitz, Cristhian Sanchez, Deepak Pathak, ”Crop Yield Prediction: An Operational Approach to Crop Yield Modeling on Field and Subfield Level with Machine Learning Models,” IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, 2023.

[10] Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, and B. Jabber, “Crop Yield Prediction based on Indian Agriculture using Machine Learning,” International Conference for Emerging Technology (INCET), 2021.