Precision Agriculture Using Machine Learning

A PROJECT REPORT

Submitted by

Ishan Khare (23BCE10317)
Amogh Agarwal (23BCE10220)
Shreedhar Daingade (23BCE10321)
Shubrat Mishra (23BCE10340)
Aayush Singh (23BCE10279)

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

VIT BHOPAL UNIVERSITY

KOTHRIKALAN, SEHORE

MADHYA PRADESH - 466114

DECEMBER 2024

VIT BHOPAL UNIVERSITY, KOTHRIKALAN, SEHORE
MADHYA PRADESH – 466114

BONAFIDE CERTIFICATE

Certified that this project report titled **"Precision Agriculture Using Machine Learning "** is the bonafide work of **Ishan Khare (23BCE10317)** who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported at this time does not form part of any other project/research work based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROGRAM CHAIR	**PROJECT GUIDE**
Dr. Vikas Panthi	Dr. Vijendra Singh Bramhe
School of Computer Science and Engineering	School of Computer Science and Engineering
VIT BHOPAL UNIVERSITY	VIT BHOPAL UNIVERSITY

The Project Exhibition I Examination is held on 19 December 2024

VIT BHOPAL UNIVERSITY, KOTHRIKALAN, SEHORE
MADHYA PRADESH – 466114

BONAFIDE CERTIFICATE

Certified that this project report titled **"Precision Agriculture Using Machine Learning "**is the bonafide work of **Amogh Agarwal** (23BCE10220) who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported at this time does not form part of any other project/research work based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROGRAM CHAIR	**PROJECT GUIDE**
Dr. Vikas Panthi	Dr. Vijendra Singh Bramhe
School of Computer Science and Engineering	School of Computer Science and Engineering
VIT BHOPAL UNIVERSITY	VIT BHOPAL UNIVERSITY

The Project Exhibition I Examination is held on 19 December 2024

VIT BHOPAL UNIVERSITY, KOTHRIKALAN, SEHORE MADHYA PRADESH – 466114

BONAFIDE CERTIFICATE

Certified that this project report titled **"Precision Agriculture Using Machine Learning "** is the bonafide work of **Shreedhar Daingade** (23BCE10321) who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported at this time does not form part of any other project/research work based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROGRAM CHAIR

Dr. Vikas Panthi

School of Computer Science and Engineering

VIT BHOPAL UNIVERSITY

PROJECT GUIDE

Dr. Vijendra Singh Bramhe

School of Computer Science and Engineering

VIT BHOPAL UNIVERSITY

The Project Exhibition I Examination is held on 19 December 2024

VIT BHOPAL UNIVERSITY, KOTHRIKALAN, SEHORE
MADHYA PRADESH – 466114

BONAFIDE CERTIFICATE

Certified that this project report titled **"Precision Agriculture Using Machine Learning "** is the bonafide work of **Shubrat Mishra** (23BCE10340)who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported at this time does not form part of any other project/research work based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROGRAM CHAIR	**PROJECT GUIDE**
Dr. Vikas Panthi	Dr. Vijendra Singh Bramhe
School of Computer Science and Engineering	School of Computer Science and Engineering
VIT BHOPAL UNIVERSITY	VIT BHOPAL UNIVERSITY

The Project Exhibition I Examination is held on 19 December 2024

VIT BHOPAL UNIVERSITY, KOTHRIKALAN, SEHORE MADHYA PRADESH – 466114

BONAFIDE CERTIFICATE

Certified that this project report titled **"Precision Agriculture Using Machine Learning "** is the bonafide work of **Aayush Singh** (23BCE10279) who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported at this time does not form part of any other project/research work based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROGRAM CHAIR	**PROJECT GUIDE**
Dr. Vikas Panthi	Dr. Vijendra Singh Bramhe
School of Computer Science and Engineering	School of Computer Science and Engineering
VIT BHOPAL UNIVERSITY	VIT BHOPAL UNIVERSITY

The Project Exhibition I Examination is held on 19 December 2024

ACKNOWLEDGEMENT

First and foremost I would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

I wish to express my heartfelt gratitude to Dr. **Pushpinder Singh Patheja,** Head of the Department, School of Computer Science and Engineering for much of his valuable support and encouragement in carrying out this work.

I would like to thank My Programme Chair **Dr. Vikas Panthi**, for giving us this valuable opportunity to present this project.

I would like to thank my internal guide **Dr. Vijendra Singh Bramhe**, for continually guiding and actively participating in my project, giving valuable suggestions to complete the project work.

I would like to thank all the technical and teaching staff of the School of Computer Science and Engineering, who extended directly or indirectly all support.

Last, but not least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

LIST OF ABBREVIATIONS

Sr. no.	Abbreviation	Full Form
1.	AI	Artificial Intelligence
2.	ML	ML - Machine Learning
3.	GPU	GPU - Graphics Processing Unit
4.	HTML	HTML - HyperText Markup Language
5.	CSS	CSS - Cascading Style Sheets
6.	JS	JS - JavaScript

LIST OF FIGURES AND GRAPHS

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER-1:
PROJECT DESCRIPTION AND OUTLINE

1.1 Introduction

Agriculture serves as the backbone of the Indian economy, supporting over 50% of the population.According to the 2011 Census, 54.6% of India's population is engaged in agriculture and allied activities.Additionally, the Food and Agriculture Organization (FAO) reports that 70% of India's rural households depend primarily on agriculture for their livelihood.

Despite its significance, it faces numerous challenges, including poor crop choices, outdated methods, and unsustainable practices. Issues such as climate change and unpredictable weather result in low productivity and financial instability for farmers. One critical problem is the lack of scientific guidance on crop selection based on soil, climate, and geography, leading to reduced yields and economic hardships.

To address this issue, the proposed project introduces a crop recommendation system leveraging Machine Learning (ML) algorithms. This system will analyze soil properties, climate conditions, and historical productivity data to recommend optimal crop choices. By adopting data-driven approaches, the project aims to enable sustainable farming practices, improve farmer incomes, and enhance agricultural productivity.

1.2 Motivation for the Work

Indian farming still relies on old methods and poor crop choices, causing low productivity. This project aims to use technology to guide farmers with smart crop suggestions, solving key problems, improving yields, and helping farmers earn better.

1.3 Techniques and Approach

This project employs advanced ML techniques to analyze soil, climate, and geographical data. The methodology integrates:

- Predictive modeling to identify the amount of yield.
- Predictive modeling to identify suitable crops.
- Historical data analysis to ensure optimal decision-making.

The acquired approach ensures robust, sustainable, and data-driven crop recommendations tailored to individual farms.

1.4 Problem Statement

Indian agriculture faces severe challenges due to poor crop selection, uninformed decision-making, and reliance on outdated practices. Current systems either focus solely on weather or soil parameters, leading to inaccurate or incomplete recommendations. A comprehensive system considering both soil and climatic conditions is crucial to mitigate these issues.

1.5 Objectives of the Work

The **Intelligent Crop Recommendation System** aims to:

1. Develop a predictive ML model analyzing soil nutrients, climate, and geography to recommend suitable crops.
2. Improve farm productivity by enabling informed crop selection.
3. Reduce farmers' financial risks by minimizing monetary losses due to poor crop decisions.
4. Empower farmers with data-driven insights for better decision-making regarding crops, fertilizers, and irrigation.

1.6 Organization of the Project

The project is structured into the following sections:

The project is organized into seven chapters:

- **Project Description and Outline**: Introduces the problem, motivation, and objectives.
- **Related Work Investigation**: Summarizes existing literature and diagnostic methods.
- **Requirement Artifacts**: Details the hardware, software, and data requirements.
- **Design Methodology and Novelty**: Explains the methods and unique aspects of the project.
- **Technical Implementation and Analysis**: Describes the development and evaluation of the model.
- **Project Outcome and Applicability**: Highlights the results and real-world applications.
- **Conclusions and Recommendations**: Provides a summary of findings and suggestions for future work.

1.7 Summary

This chapter outlined the motivation, techniques, problem statement, objectives, and organization of the project. The Intelligent Crop Recommendation System leverages ML to provide data-driven crop recommendations, addressing critical issues in Indian agriculture. By offering sustainable solutions, it seeks to transform farming practices and improve farmer livelihoods.

Chapter 2: Related Work Investigation

2.1 Introduction

This chapter reviews the existing approaches, techniques, and research work relevant to the crop recommendation and yield prediction system based on algorithms of ML

2.2 Core Area of the Project

The core area of this project is Develop a predictive ML model that analyzes soil, climate, and geographic data to make recommendations for what crop should be cultivated in a region.
Empower the farmer by using data and science for better decision-making around crops, fertilizers, and irrigation.

2.3 Existing Approaches/Methods

In Order to understand the current trends and existing Methods /Approaches , We have studied many research articles related to crop prediction /recommendation and yield prediction.
Lets understand the methods through the given below table which we have prepared after analyzing different case studies related to the same .

2.3 Crop Recommendation

S.NO	STUDY	OBJECTIVE	RESULTS	TECHNIQUES USED	DATASETS
1	Soil Classification, 2018	Classify Soil types	92% accuracy	Random Forest, Support Vector Machine (SVM) and KNN	Soil analysis datasets
2	Intelligent Crop Recommendation System using Machine Learning, 2021	Suggest suitable crops	89.88% accuracy	Random Forest, KNN	Agricultural soil and crop yield datasets
3	Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique	Improving crop selection	99.91% accuracy	Ensemble Learning, Random Forest, Naive Bayes, Linear SVM, Majority Voting	Soil Type, pH value, NPK content (Nitrogen, Phosphorus, Potassium), Average rainfall, Surface temperature

Table 2.3.1

Lets see the above case studies in details for clear understanding about the methods / Approaches

Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series, 2018:
Author:Sk Al Zaminur Rahman Kaushik Chandra Mitra S.M. Mohidul Islam

Objective:This study highlights how machine learning can help farmers identify soil types and recommend the best crops for their land. By analyzing factors like soil texture, pH level, organic matter, and nutrient content, the model predicts soil types and suggests crops that are most likely to thrive. This approach aims to improve crop yields by aligning farming decisions with the specific characteristics of the soil.

The research tested several machine learning algorithms, including Random Forest, Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN). Among these, the Random Forest algorithm stood out, achieving 92% accuracy in classifying soil types. The analysis found that soil texture and pH level were the most critical factors for accurate predictions, making the model highly reliable for practical use.

Integrating soil analysis with crop recommendations can revolutionize farming by helping farmers make smarter, more sustainable decisions. By tailoring crop selection to soil conditions, this approach maximizes yields, reduces resource waste, and promotes efficient agricultural practices. This tool is a step toward improving productivity and supporting farmers in managing their land effectively.

Intelligent Crop Recommendation System using Machine Learning, 2021:

Authors: Priyadharshini A, Swapneel Chakraborty, Aayush Kumar, Omen Rajendra Pooniwala

Objectives: This article shows how machine learning can be used to help farmers choose the best crops. The aim is to make farming better and lessen money problems from bad crop choices. It aims to provide data for the farmer to understand the soil, seasons, and the land he or she is working on much better.

The article talks about a smart system encouraging farmers to cultivate the proper crops through machine learning. It is, in fact, an important part of the Indian economy because it comprises nearly 50% of the people. In case, inappropriate crops are selected, then low yields and financial issues may occur. These issues get worse due to ignorance of the kind of soil, weather, and geography. This literature explains how researchers would collect heavy data, including crop yield, soil nutrient, cultivation costs, weather condition, and crop prices, gathered from credible sources such as government websites and Kaggle. In managing data inconsistencies, they substituted missing values with -1 in order to have smoother predictions. To develop their system, they used multiple machine learning models, including **linear regression**, **neural networks**, **K-Nearest Neighbors (KNN)**, and **decision trees**. Their system can be categorized into three sections: profit analysis, crop recommendations, and sustainability predictions.

The neural network got an accuracy of **89.88%**, linear regression was at **88.26%**, and KNN with cross-validation was at **88%**. The system is planned to have real uses, like a mobile app and a web interface, to give farmers personalized crop advice, helping millions. The writing demonstrates how technology, especially machine learning, can greatly enhance farming practices. It helps farmers make better choices and promotes progress in sustainable farming.

Yield Prediction

S.NO	STUDY	OBJECTIVE	RESULTS	TECHNIQUES USED	DATASETS
1	Yield Prediction, 2020	Predict crop yield	87.5% accuracy	KNN, Random Forest, Naive Bayes	Regional yield data
2	Crop Yield Prediction Using Machine Learning: A Pragmatic Approach	Predict crop yield	97.01%	Random Forest (RF) ,AdaBoost Decision Tree:,Support Vector Machine (SVM)	Soil pH ,Rainfall ,Temperature ,Humidity

Table 2.3.2

Crop Prediction Method to Maximize Crop Yield Using Machine Learning, 2020

Authors:Neelam Singh1, Deeksha Pant2, Devesh Pratap Singh, Bhasker Pant:

Objective: This study shows how machine learning can help farmers choose crops that suit their region, improve yields, and address agricultural challenges across India. For example, water-saving crops like wheat and mustard are ideal for Punjab and Haryana, while rice and pulses work well in monsoon-dependent states like Tamil Nadu. Drought-resistant crops like cotton and sorghum are recommended for Maharashtra and Gujarat.

Different machine learning models tackle specific challenges. K-Nearest Neighbors (KNN) predicts yields in states with varied climates, while Naïve Bayes helps regions with unpredictable rainfall like Madhya Pradesh. Random Forest handles complex factors like drought in Maharashtra, and K-Star works well in tough terrains like Himachal Pradesh. Future advancements could include real-time weather and soil monitoring, supported by mobile apps to provide farmers with instant advice. These innovations can transform Indian agriculture, boosting productivity and food security.

Crop Yield Prediction Using Machine Learning: A Pragmatic Approach

Authors:Rajswee Surana,Ritu Khandelwal

Objective: to develop a machine learning model for accurate crop yield prediction to help farmers improve agricultural productivity. By evaluating algorithms like Random Forest, Adaboost, Gradient Boost, and SVM, the research aims to identify the key environmental factors influencing crop yields and determine the most accurate prediction method. The goal is to provide farmers with reliable insights for optimal crop selection, enhancing yield and resource utilization in a changing agricultural landscape.

This research uses machine learning to help farmers predict crop yields more accurately, improving productivity and reducing risks. It evaluates algorithms like Random Forest, Adaboost, Gradient Boost, and SVM, focusing on factors like soil pH, temperature, rainfall, and humidity. A dataset of 2,201 samples was split for training (80%) and testing (20%). Random Forest proved most accurate. The study aims to provide farmers with practical insights for better crop choices and efficient resource use, boosting yields and supporting smarter farming in a changing agricultural landscape.

2.4 Pros and Cons of the Stated Approaches/Methods

One shortcoming we identified in all these notable published works was that the authors of each paper focused on only one parameter (either weather or soil) for predicting the suitability of crop growth. However, in our opinion, both these factors should be taken together into consideration concomitantly for the most effective and most accurate prediction. This is often because a specific soil type could also be used for supporting one variety of crop, but if the climatic conditions of the regions don't seem to be suitable for that crop type, then the yield will suffer.

2.5. Summary

Machine learning provides a promising solution to the challenges in Agriculture such as Crop Recommendation and Yield Prediction , emphasizing the need for further research and development.

Chapter 3: Requirement Artifacts

3.1 Introduction

This chapter outlines the hardware, software, and specific requirements necessary for the successful implementation of the project.

3.2 Hardware and Software Requirements

- **Hardware**: Standard computing systems with GPUs for training models.
- **Software:** Python, HTML, CSS and JavaScript

3.3 Specific Project Requirements

3.3.1 Data Requirement

Yield Prediction Model

- Detailed agricultural datasets containing crop types, crop years, and growing seasons.
- Regional farming records with information about states and cultivated areas.
- Production data showing the total output of various crops.
- Climatic records including annual rainfall measurements.
- Fertilizer and pesticide usage reports to analyze their impact on crop yield.
- Historical yield data to train the model for accurate forecasting.

Crop Prediction Model

- Soil nutrient datasets detailing levels of nitrogen, phosphorus, and potassium.
- Environmental condition records including temperature, humidity, and rainfall data.
- Soil pH measurements to assess soil acidity or alkalinity.
- Crop mapping data to correlate soil and environmental conditions with suitable crops.

3.3.2 Look and Feel Requirements

- User-friendly interface for Farmers and other users

Figure 3.1

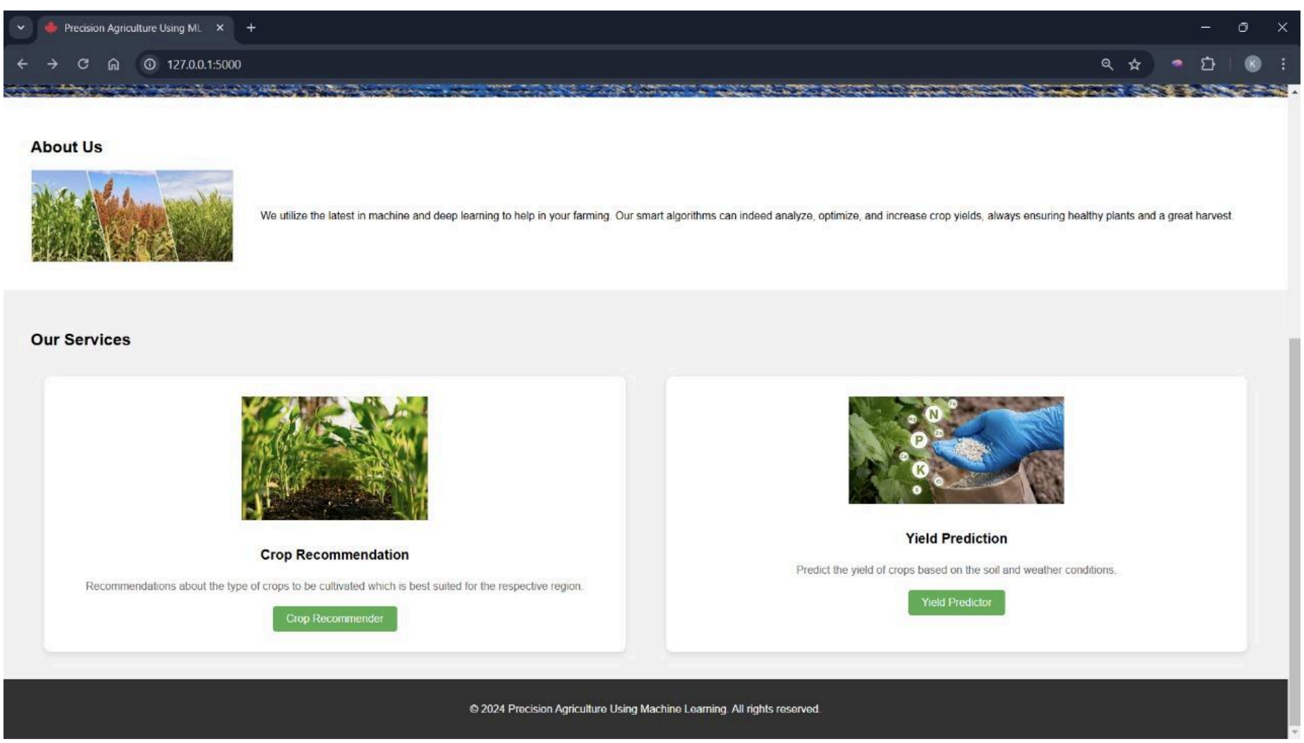


Figure 3.2

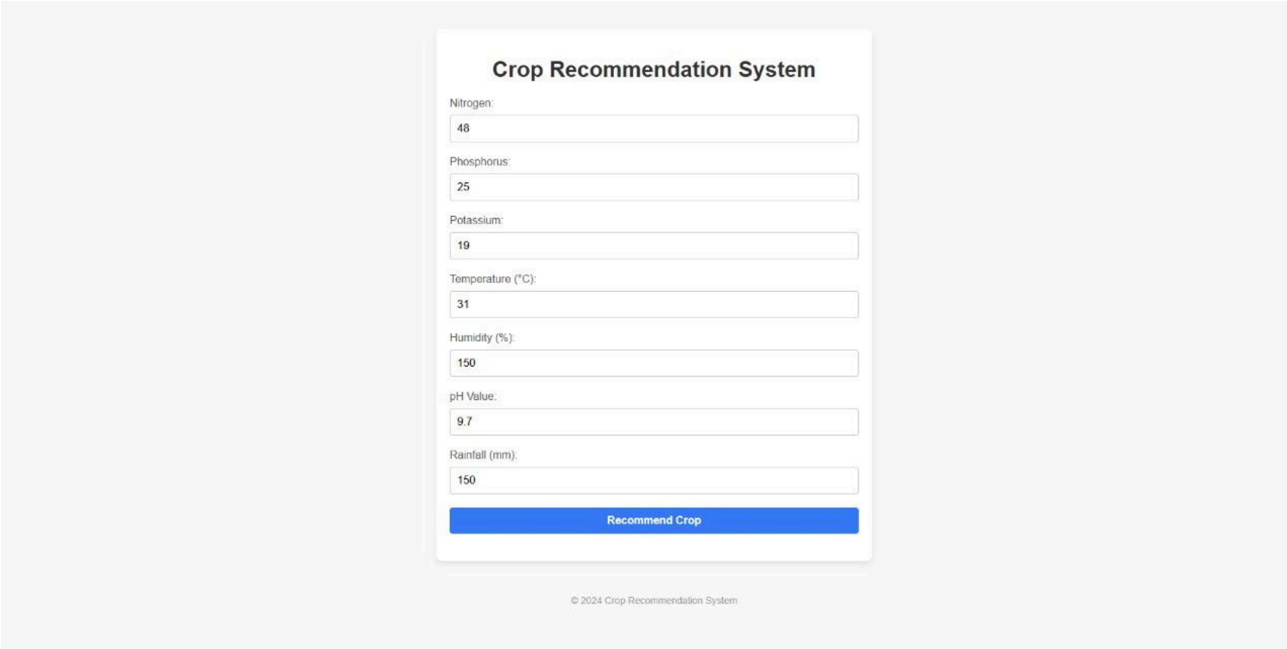


Figure 3.3

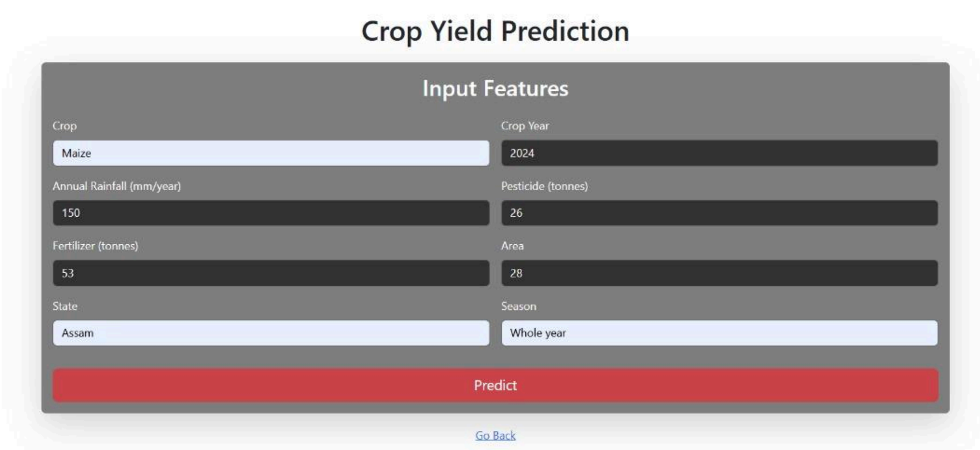


Figure 3.4

3.4 Summary

This chapter detailed the necessary requirements to achieve the project objectives. The next chapter will focus on the design methodology and the novelty of the proposed solution.

Chapter 4: Design Methodology and Its Novelty

4.1 Methodology and Goal

The goal of this project is to help farmers make informed decisions on which crops to plant and predict the potential yield based on soil and environmental factors. The project involves two key processes: predicting the right crop to grow based on soil conditions (Crop Prediction) and forecasting the crop yield (Yield Prediction).

4.2 Functional Modules Design and Analysis

The system has two main parts:

- **Crop Prediction**: Based on factors like soil nutrients and weather, it recommends the most suitable crop.
- **Yield Prediction**: It predicts the amount of crop yield based on factors like rainfall, temperature, and the use of pesticides.

4.3 Methodology

A: Crop Prediction

1. **Input Data**
   To make a recommendation, the model uses seven pieces of information:
    - **Nitrogen**: Amount of nitrogen in the soil .
    - **Phosphorus**: Amount of phosphorus in the soil .
    - **Potassium**: Amount of potassium in the soil .
    - **Temperature**: Weather temperature (°C).
    - **Humidity**: Amount of moisture in the air .
    - **pH Value**: The acidity or alkalinity of the soil .
    - **Rainfall**: Total yearly rainfall (in millimeters).
2. **Target Output**:
   The model recommends a **crop**, such as "Wheat," "Rice," etc., based on these inputs.

Flow of Operation

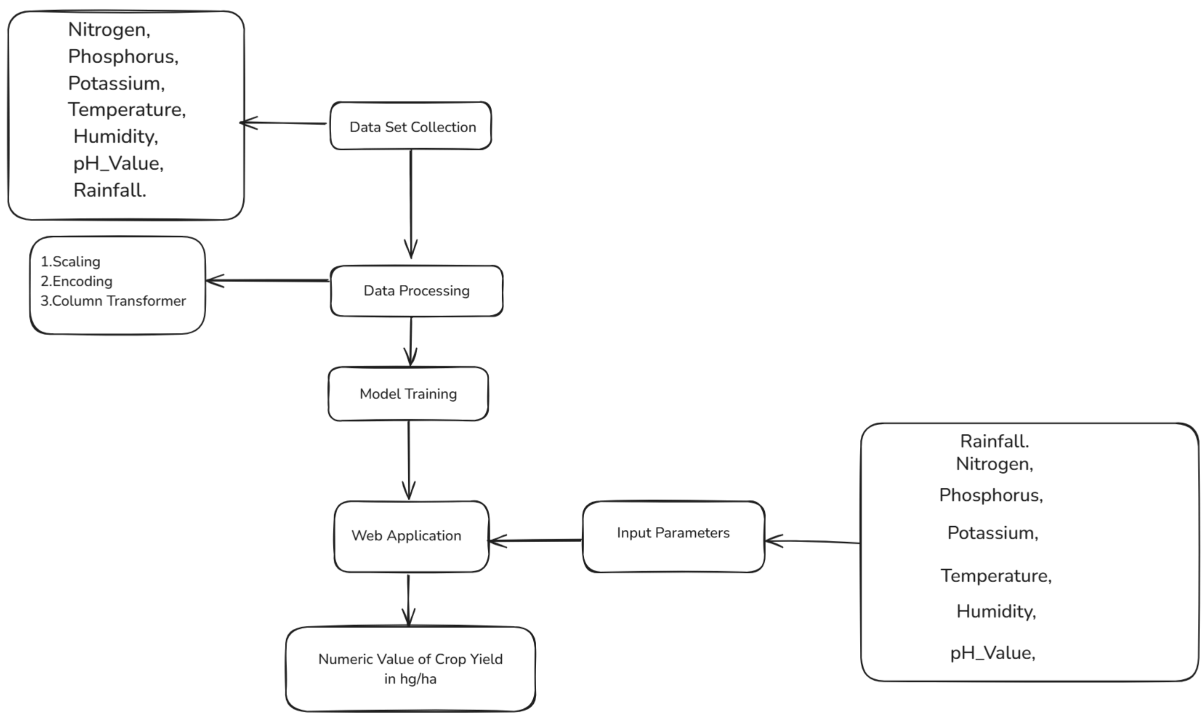


Figure 4.1

3. **Data Preprocessing**

   The data undergoes a few simple steps to ensure it's ready for the model:

   - **Cleaning Column Names**: Removes extra spaces to avoid errors in the column names.
   - **Verify Dataset**: Ensures that all the expected data is available (like Nitrogen, Phosphorus, etc.). If not, it raises an error.
   - **Label Encoding for Crop**: Converts crop names (like "Wheat") into numbers that the model can understand, which makes the target variable machine-readable.

4. **Splitting the Dataset**

   The dataset is split into two parts:

   - **Training Data (80%)**: Used to teach the model how to make predictions.
   - **Testing Data (20%)**: Used to test how well the model works on new data.

5. **Algorithm Used**: **Random Forest Classifier**

   This method uses several decision trees to predict the crop. Each tree makes its prediction, and the final recommendation is decided by the majority vote. The more trees, the more accurate the result. The algorithm uses:

   - **n_estimators=100**: 100 decision trees are created for better prediction.
   - **random_state=42**: Ensures the results are consistent across different runs.

6. **Model Training**

   The model learns by using the training data and is adjusted to make accurate predictions about which crop to recommend.

7. **Model Evaluation**

   After training, the model is tested on new data to check its performance. The model's accuracy is calculated by comparing its predictions to the actual values in the test dataset.

   Model Accuracy: 99.32%

8. **Model Saving**

   Once trained, the model is saved in a file (e.g., **crop_model.pkl,** pkl - pickle file) so it can be used later for predictions. The **LabelEncoder** is also saved as **label_encoder.pkl** for consistent use of crop labels.

B: Yield Prediction

1. **Inputs**:
   - **Year**: The year the data was collected.
   - **Average Rainfall (mm/year)**: The average amount of rainfall in millimeters over a year.
   - **Pesticides (tonnes)**: The amount of pesticides used in tonnes.
   - **Average Temperature (°C)**: The average temperature in degrees Celsius during the growing season.
   - **Area**: The geographical region or area where the crop is grown (categorical, e.g., “Assam”).
   - **Item**: The type of crop or agricultural product being grown (categorical, e.g., "Wheat," "Rice").

2. **Output**:

   The model predicts the **crop yield** in hectograms per hectare.

Flow of Operation

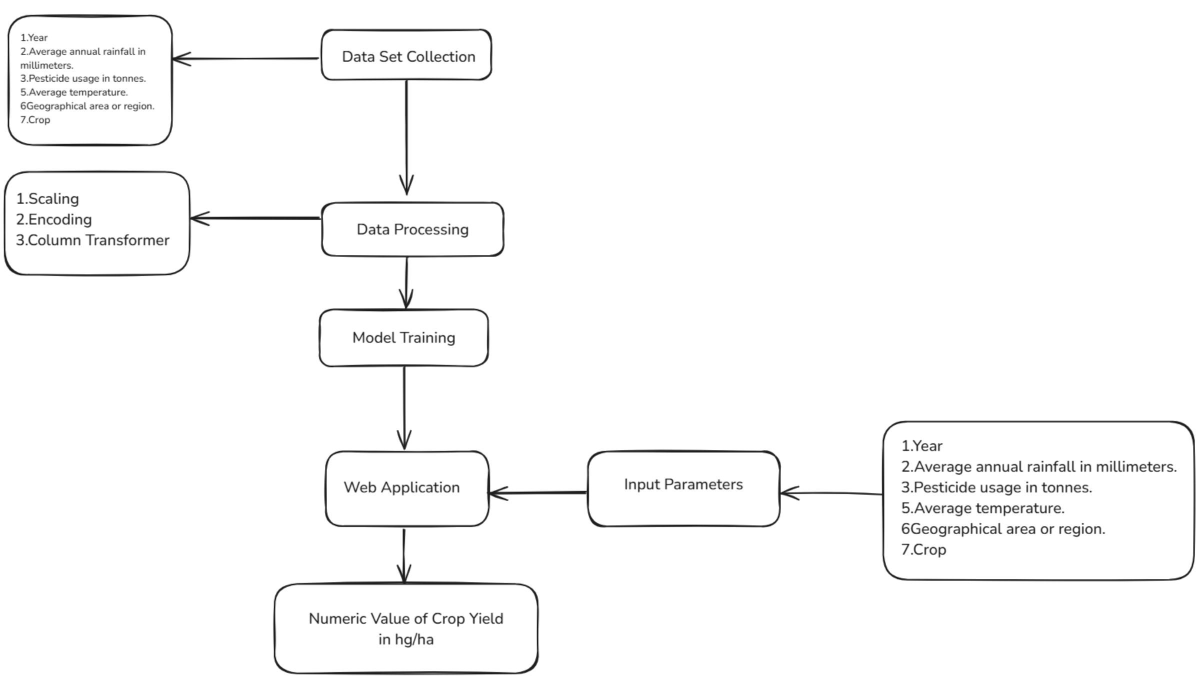


Figure 4.2

3. **Data Splitting**

   The dataset is split into:

   - **Training Set (80%)**: For training the model.
   - **Testing Set (20%)**: For testing the model's performance.

4. **Data Preprocessing**

   - **Scaling**: Numerical data (e.g., Year, Rainfall, Temperature) is scaled to ensure all values are within a similar range, using **StandardScaler**.
   - **Encoding**: Categorical data (e.g., Area, Item) is converted into numerical data using **OneHotEncoder**, so the model can process it.

5. **ColumnTransformer** is used to combine scaling and encoding in a single pipeline for simplicity.

6. **Model Training**: **Decision Tree Regressor (DTR)**

   This model splits the data based on the features to predict the crop yield. The decision tree works well for understanding how different factors affect the yield and provides interpretable predictions.

7. **Model Evaluation**

   After training, the model is tested on new data to assess its performance. The model's accuracy is measured by comparing its predicted crop yield to the actual values in the test dataset. **Model Accuracy: 96.95%**

8. **Model Saving**

   The trained model (e.g., **dtr.pkl**) and preprocessing steps (e.g., **preprocessor.pkl**) are saved, so they can be reused for future predictions.

4.4 User Interface Designs

The web app is simple to use. For both crop and yield predictions, users input relevant data (like rainfall, temperature, etc.), and the system displays the results on a dashboard.

- **Input Forms**: Easy-to-use forms for entering data.

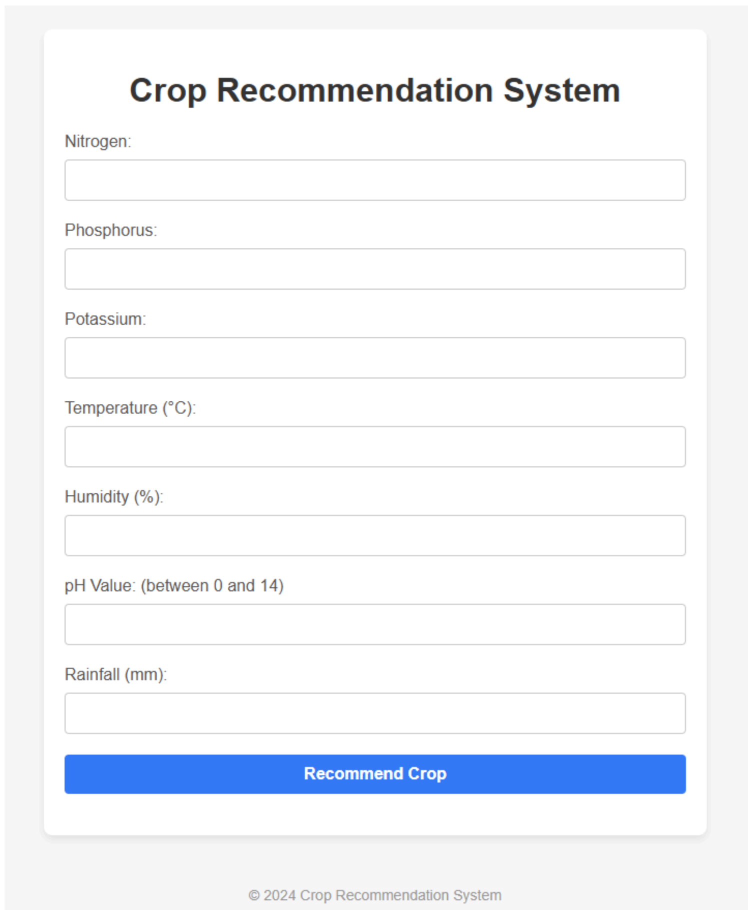


Figure 4.3

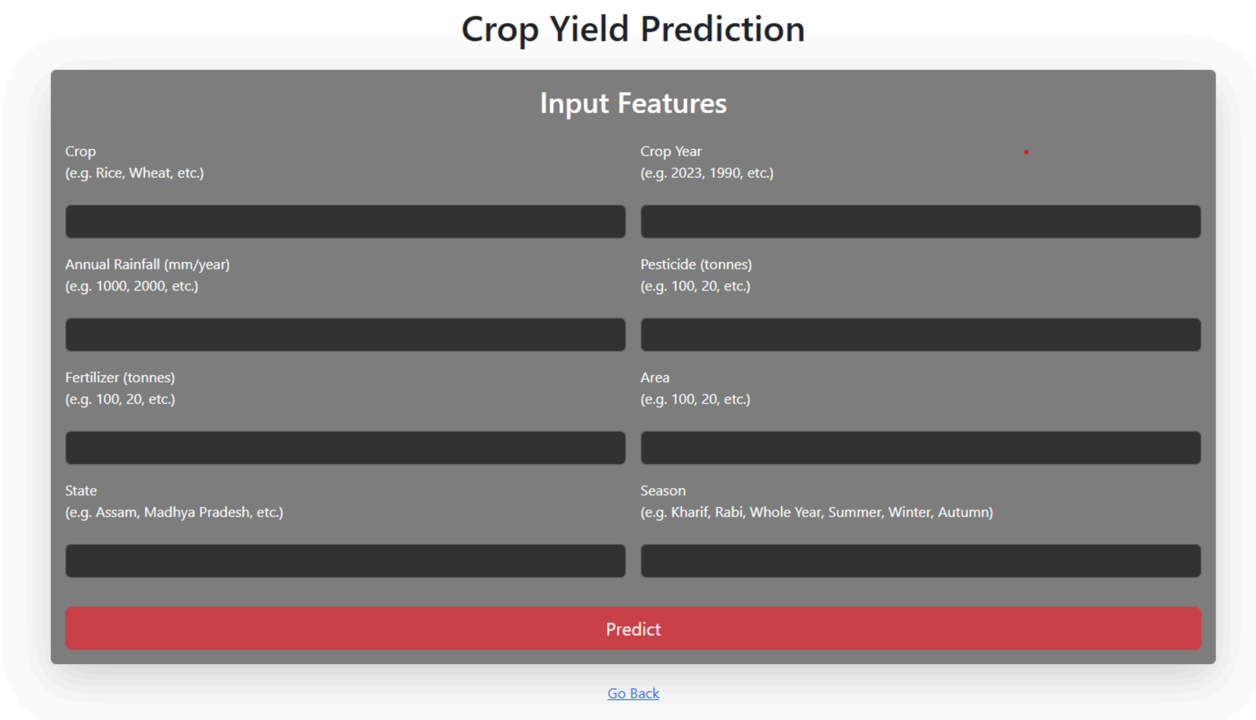


Figure 4.4

- **Output Dashboard**: Shows the crop recommendation or yield prediction.

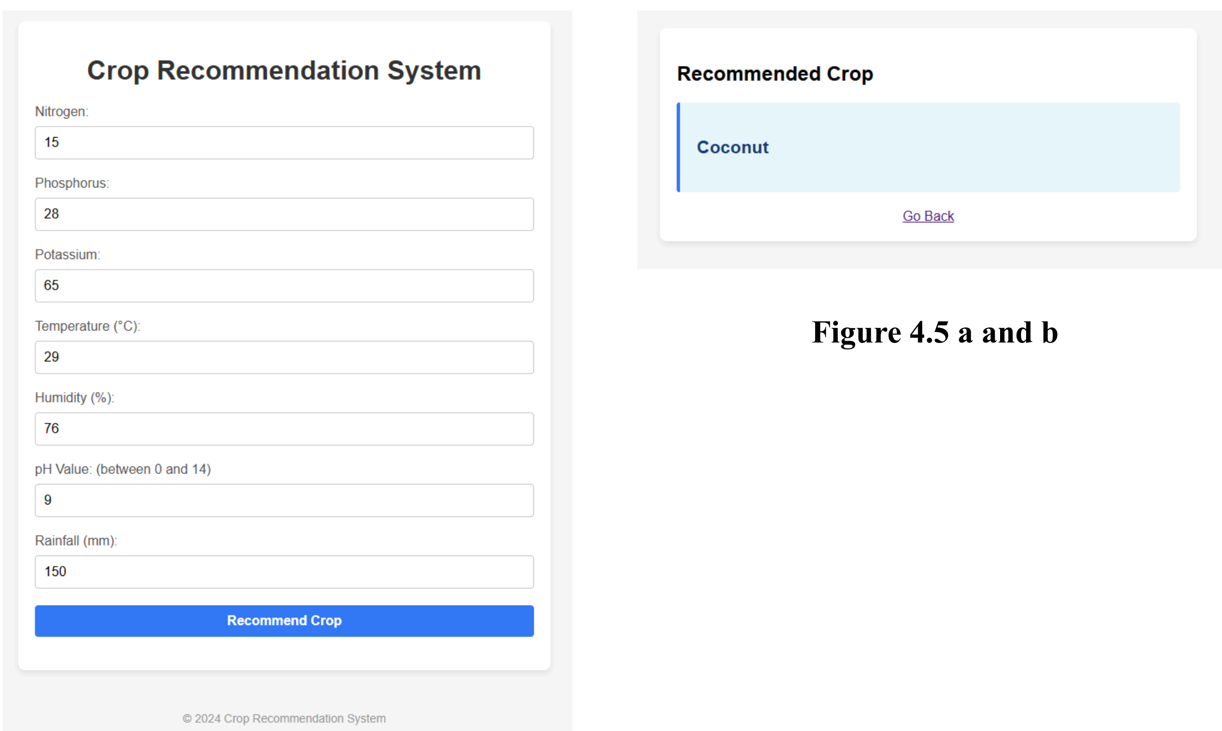


Figure 4.5 a and b

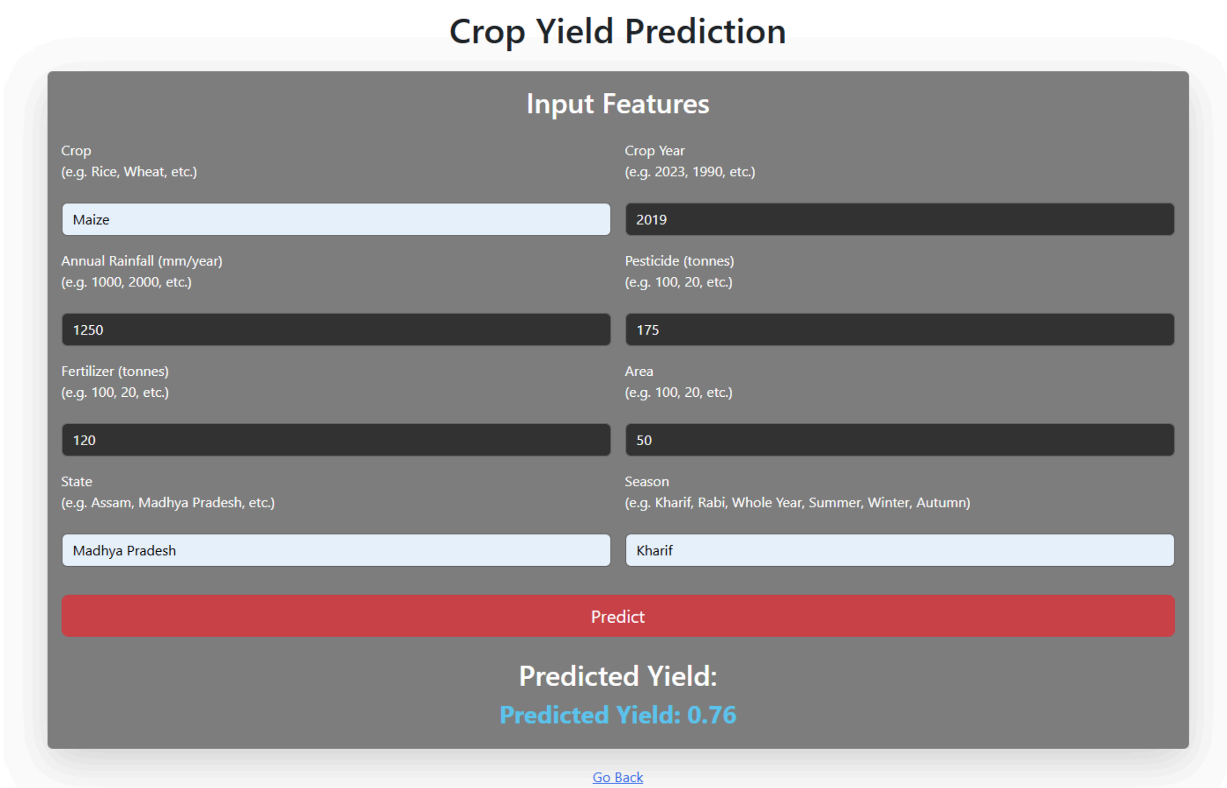


Figure 4.6

4.5 Summary

This chapter has explained the methodology behind the crop prediction and yield prediction models, including the steps taken for data processing, model training, and evaluation. The next chapter will discuss the technical details of the system's implementation and performance evaluation

Chapter 5: Technical Implementation and Analysis

5.1 Outline

This chapter details the technical aspects of the project, including coding solutions, system design, and performance analysis

5.2 Technical Coding And Code Solutions

The Implementation Involves The Following Steps :-

Step 1: Data Preparation and Preprocessing

- **Load Dataset:** Import and clean the dataset (remove duplicates, unnecessary columns, and validate fields).
- **Preprocessing:**
  - Scale numerical columns using StandardScaler.
  - Encode categorical columns using OneHotEncoder or Label Encoding.
  - Split the data into training (80%) and testing (20%) subsets.

Step 2: Model Training and Evaluation

- **Train Model:**
  - Use appropriate machine learning algorithms (e.g., Random Forest Classifier for classification tasks or Decision Tree Regressor for regression tasks).
- **Evaluate Model:**
  - Use metrics like precision, recall, and F1-score for classification tasks.
  - Use Mean Absolute Error (MAE) and R-squared (R^2) for regression tasks.

Step 3: Web Application Deployment

- **Save Models:** Store trained models and preprocessing pipelines as `.pkl` files.
- **Deploy Web App:** Build a Flask web app to:
  - Accept user inputs, preprocess them, generate predictions, and display results

5.3 Working Layout of Forms

1. Crop Recommendation Form

Purpose: Collects soil and environmental data (e.g., Nitrogen, Phosphorus, etc.) to predict the best crop.
Action: User submits numeric inputs; system processes and suggests a crop.

2. Crop Yield Prediction Form

Purpose: Gathers climatic, regional, and crop-related data to estimate crop yield.
Action: User provides inputs (numeric and categorical); system predicts yield.

3. Result Display Page

For Recommendation: Displays the suggested crop based on user inputs.
For Yield Prediction: Shows the estimated crop yield in hectograms per hectare.

5.4 Prototype

A prototype was developed with:

Frontend: Developed using **HTML**, **CSS**, and **JavaScript** to create responsive and interactive user interfaces for input forms and result display.

Backend: Built using **Python Flask** to handle input processing and machine learning model predictions.

Sample Data: Simulated dataset for validation.

5.5 Test And Validation

The system was tested using real-world and simulated datasets. Metrics used for validation include:

- **Crop Recommendation:**
  - **Classification Report:** Outputs metrics like precision, recall, F1-score, and accuracy for each crop class.
- **Crop Yield Prediction:**
  - **Mean Absolute Error (MAE):** Measures average prediction error.**(8.26)**
  - **R-squared (R^2):** Indicates how well the model explains variance in the target variable.**(0.97)**

Precision:

Precision measures the proportion of true positive predictions (correct crop recommendations) out of all positive predictions made by the model (true positives + false positives). High precision means that when the model predicts a crop, it is likely to be correct.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall:

Recall measures the proportion of actual positive cases (true crops) that the model successfully predicted. It is important when we want to ensure the model correctly identifies as many positive samples as possible.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

F1-Score:

The F1-Score is the harmonic mean of precision and recall. It is particularly useful when there is an imbalance in the classes , as it provides a balance between the two metrics.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

1

Mean Absolute Error (MAE):

MAE measures the average of the absolute errors between predicted crop yield values and actual values. It gives an idea of how close the predictions are to the true values on average. It is a straightforward metric that does not penalize larger errors more than smaller ones.

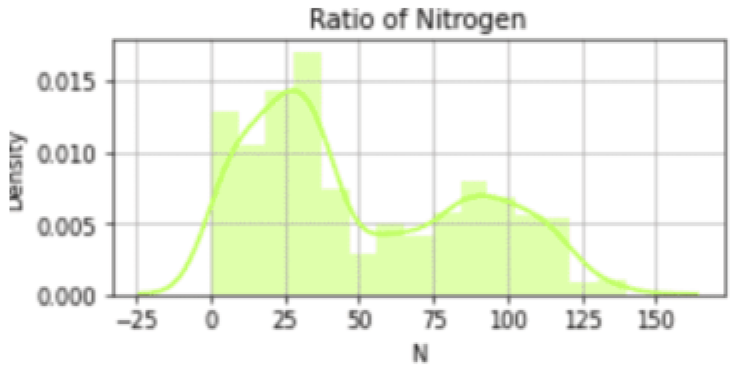
$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\text{Predicted}_i - \text{Actual}_i|$$

R-squared (R^2):

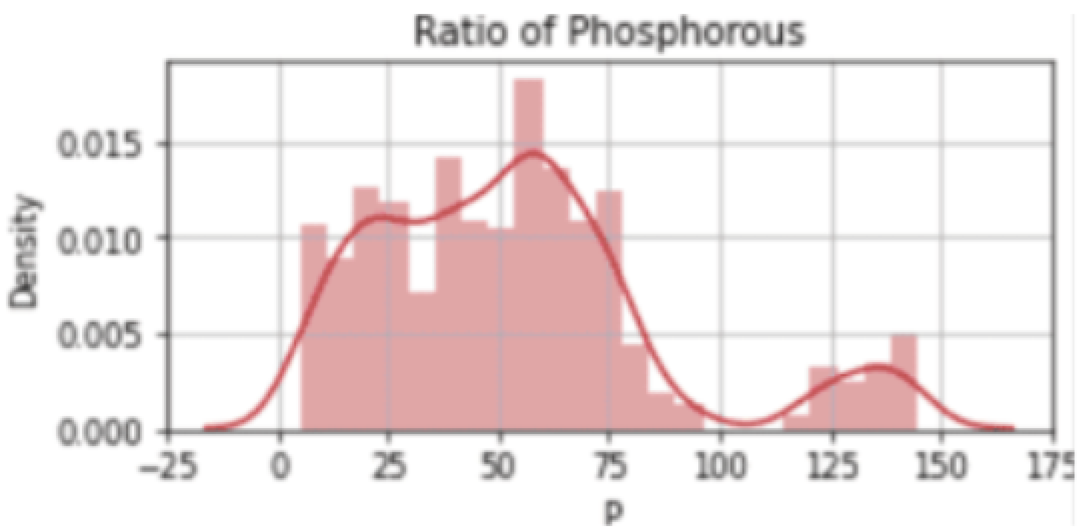
R-squared, also known as the coefficient of determination, indicates how well the model's predictions explain the variance in the observed data. An R^2 value of 1 means perfect predictions, while 0 means that the model does not explain any of the variance. Negative values can occur if the model performs worse than a simple mean-based model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\text{Actual}_i - \text{Predicted}_i)^2}{\sum_{i=1}^{n} (\text{Actual}_i - \overline{\text{Actual}})^2}$$

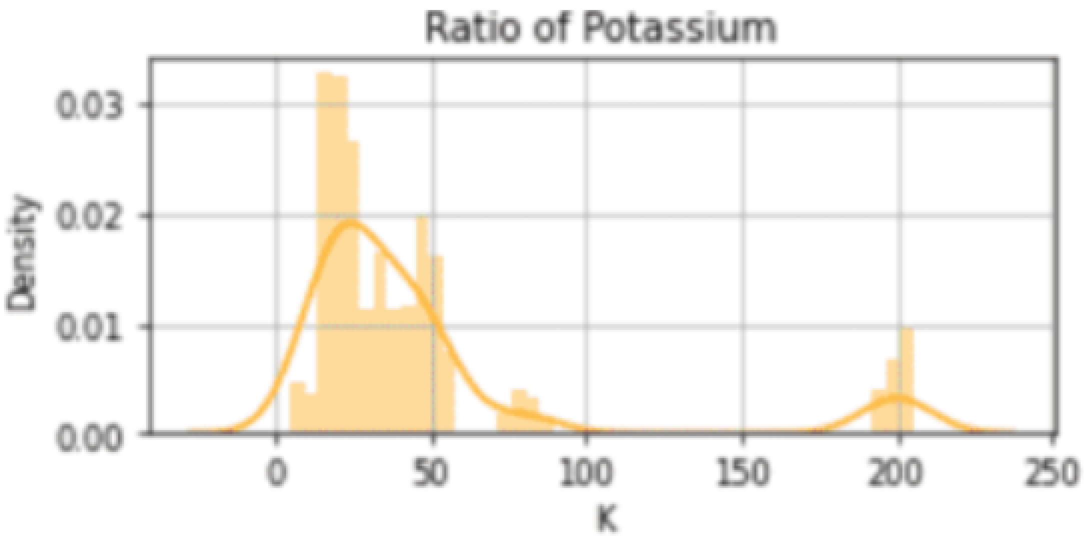
5.6 Performance Analysis (Graphs)



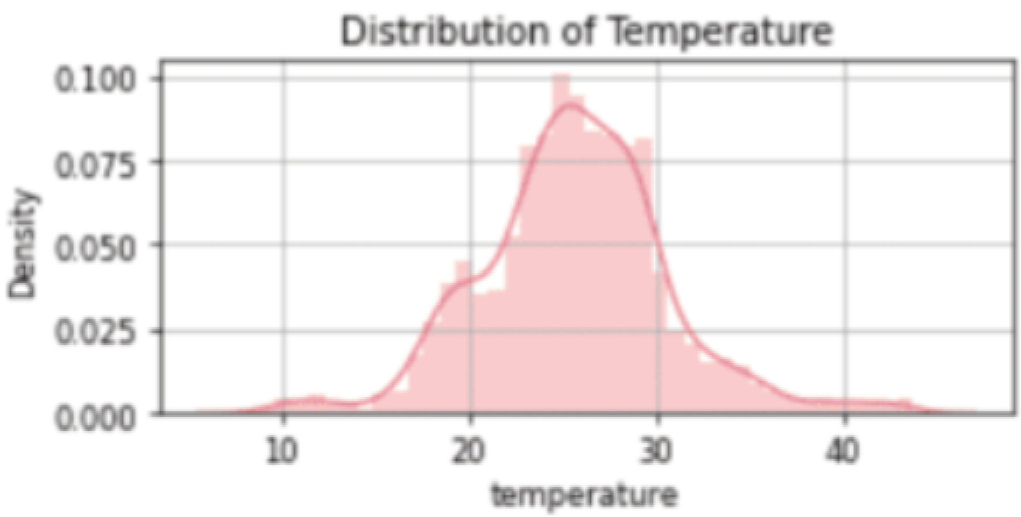
GRAPH - 1



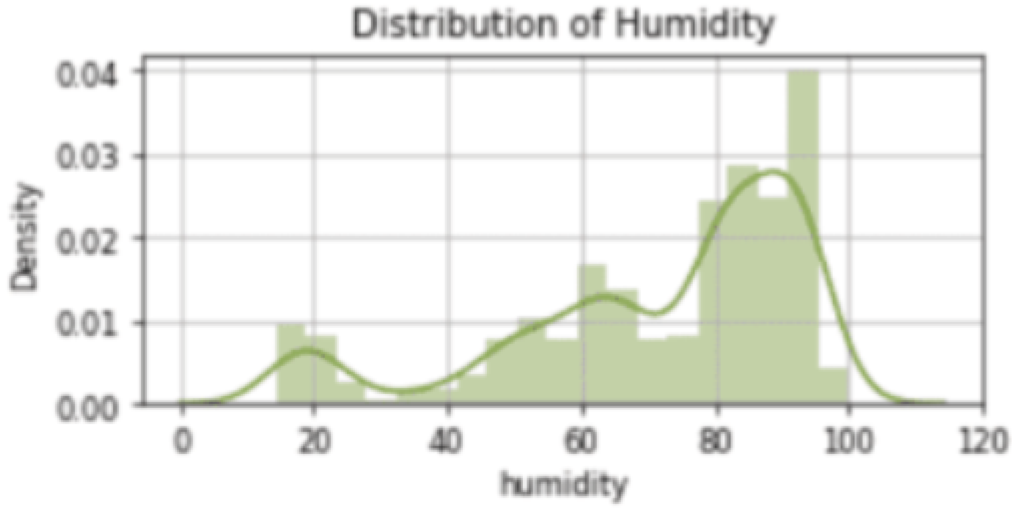
GRAPH - 2



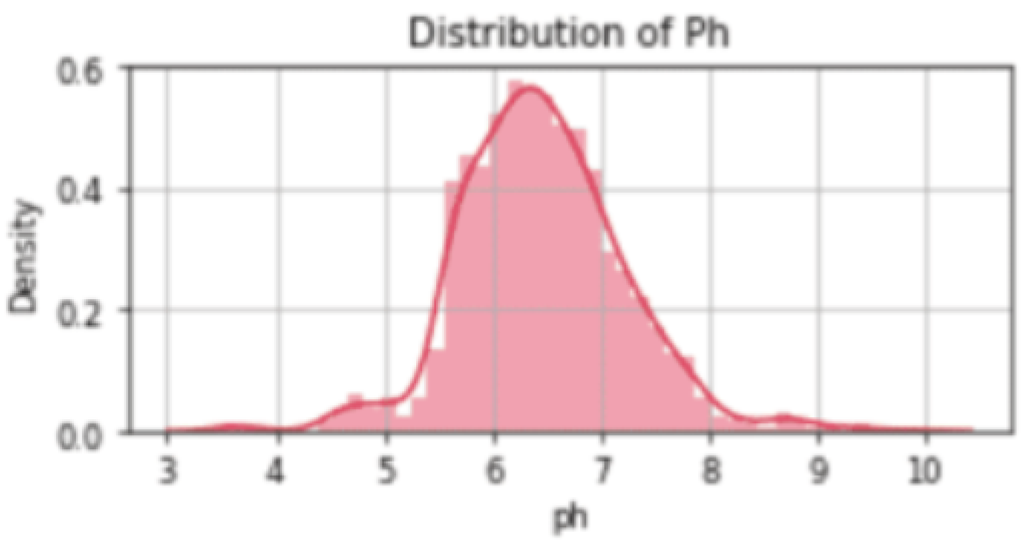
GRAPH - 3



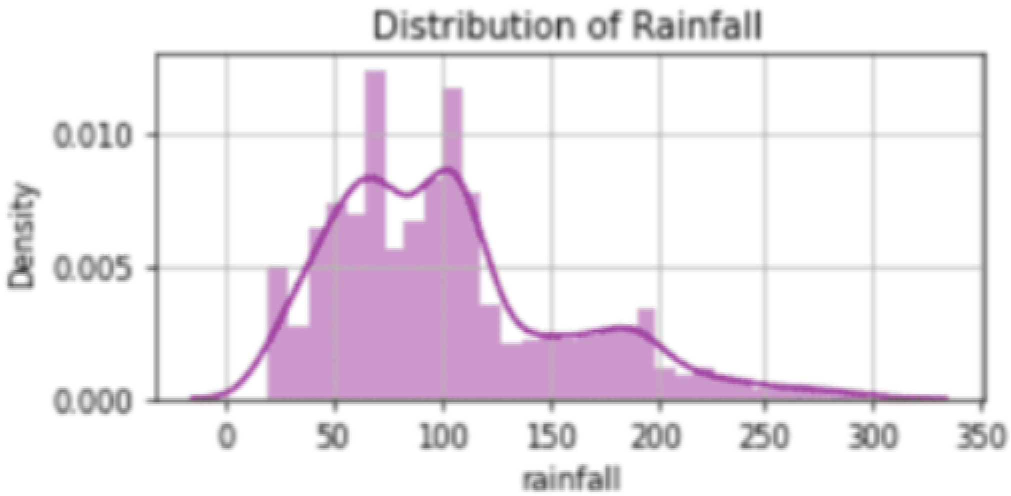
GRAPH - 4



GRAPH - 5



GRAPH - 6



Distribution of Rainfall
Density
0.010
0.005
0.000
0
50
100
150
200
250
300
350
rainfall

GRAPH - 7

5.6.1 Importance of Features During Evaluation

The data from the seven graphs reveals various distributions across different environmental factors. The distribution of nitrogen ratios is roughly normal, with a peak around 25-30 and a slight right skew, indicating a concentration of values in this range and a few higher outliers. Phosphorus levels show a multimodal distribution with peaks around 50 and 125, suggesting variability across samples or groups. Potassium levels exhibit a bimodal distribution, implying two distinct subpopulations with different potassium concentrations. Temperature data is approximately bell-shaped, indicating most values fall within a certain range, with fewer observations at the extremes. Humidity levels are also normally distributed, with a peak around 80-90%, suggesting this range is most common in the dataset. pH values follow a bell-shaped curve centered around pH 6, indicating a concentration of measurements in this range. Finally, rainfall distribution is skewed to the right, with most rainfall events occurring between 50 and 100 millimeters, peaking at around 75 millimeters, and gradually decreasing as rainfall amounts increase, reflecting that heavier rainfall events are less frequent.

5.7 Summary

This chapter described the implementation process, including coding solutions, testing, and performance evaluation . The next chapter will focus on project outcomes and applicability.

Chapter 6: Project Outcome and Applicability

6.1 Outline

This chapter outlines the results, key implementations, and real-world applicability of the proposed intelligent crop recommendation and yield prediction system using machine learning to optimize crop selection, improve productivity, reduce financial risks, and enhance decision-making for farmers based on soil, climate, and environmental conditions.

6.2 Key Implementation Outlines of System

1. **Data Collection & Preparation:**
    - **Inputs:** Collect soil and environmental conditions (e.g., Nitrogen, Phosphorus, Temperature, etc.).
    - **Data Preprocessing:** Clean column names, verify dataset integrity, and use LabelEncoder for converting crop names into numeric labels for machine learning.
    -
2. **Model Development & Training:**
    - **Algorithm:** Use a Random Forest Classifier for crop prediction, training on 80% of the dataset.
    - **Training: T**rain the model with key hyperparameters and evaluate using a Classification Report.
    -
3. **Model Evaluation & Saving:**
    - **Evaluation:** Assess model accuracy using precision, recall, and F1-score on the test set.
    - **Saving:** Serialize the trained model and LabelEncoder for future use.

4. **Web Application for Crop Recommendation:**

   - **User Interface:** Build a web app where users input soil/environmental data to receive a crop recommendation.
   - **Backend Processing:** Format user inputs, predict crop using the trained model, and display the result on a recommendation page.

5. **Yield Prediction System:**

   - **Objective:** Predict crop yield (hg/ha) using environmental factors (e.g., rainfall, temperature).
   - **Model:** Use a Decision Tree Regressor (DTR) to predict yield and deploy it via Flask.
   - **Preprocessing:** Scale numerical data with StandardScaler and encode categorical data with OneHotEncoder. Save the trained model and preprocessing pipeline .

6.3 Significant Project Outcomes

1. Improved Crop Recommendations:

- The system accurately recommends suitable crops for a given region based on soil and environmental conditions (e.g., Nitrogen, Phosphorus, Temperature, Humidity, etc.). This helps farmers make informed decisions and improve their crop productivity**.**

2. Reduced Financial Risk for Farmers:

- By providing accurate crop recommendations, the system minimizes the risk of poor crop choices, which reduces monetary losses due to crop failure or suboptimal yield.

3. Enhanced Farm Productivity:

- By tailoring crop suggestions to specific land and environmental conditions, farmers can grow crops that are more suited to their land, leading to increased agricultural output and better resource utilization.

4. Empowered Farmers Through Data-Driven Decisions:

- The system uses Machine Learning to give farmers access to scientifically-backed crop suggestions, helping them make better decisions regarding crop selection, fertilizer use, and irrigation, thus fostering more sustainable farming practices.

5. Optimized Crop Yield Prediction:

- The system also includes a crop yield prediction model that uses factors like rainfall, temperature, pesticides usage, and geographical area to predict crop yield (hg/ha). This provides farmers with valuable insights into potential yields, allowing for better planning and resource allocation.

6.4 Project Adaptability On Real World Applications

1. Precision Agriculture: The system can be integrated with precision agriculture tools like satellite imagery and weather data to create even more accurate yield predictions.

2. Futures Market Trading: Yield predictions can inform commodity traders in futures markets, allowing them to make more strategic bets on crop prices.

3. Supply Chain Management: Food processors can use yield predictions to plan their production schedules and secure the necessary raw materials in advance.

4. Market Forecasting: By analyzing regional yield predictions, agricultural stakeholders like commodity traders and food processors can make informed decisions about buying and selling crops, potentially stabilizing market prices.

5. Agricultural Extension Services: Government agencies and NGOs can leverage the system to provide targeted advice to farmers in different regions, taking into account both suitable crops and predicted yields.

Chapter 7: Conclusions and Recommendations

7.1 Outline

This chapter summarizes the project's findings, highlights its limitations, and outlines future improvements for expanding the system's capabilities.

7.2 Limitations/Constraints of the System

1. Data Availability and Quality:

   The system's performance depends on the availability of accurate and comprehensive datasets on soil properties, climatic conditions, and crop yields. Missing or incorrect data can reduce prediction accuracy.

2. Regional Specificity:

   Models trained on data from specific regions may not generalize well to other areas with different soil types, climates, or farming practices.

3. Dynamic Environmental Changes:

   Climate conditions and soil properties are not static. The system may require frequent updates to adapt to changing environmental factors.

4. Limited Farmer Access to Technology:

   Small-scale and remote farmers may lack access to the internet, smartphones, or other tools required to utilize the system effectively.

5. Language and Usability:

   The system may not be user-friendly for non-technical farmers or those unfamiliar with the language used in the interface.

7.3 Future Enhancements

1. Incorporating Real-Time Data:

    Integrate IoT devices, weather APIs, and satellite imagery to provide real-time updates on soil moisture, weather, and pest infestations for dynamic crop recommendations.

2. Localization and Language Support:

    Add support for regional languages and tailor the system to local farming practices to improve accessibility and usability.

3. Pest and Disease Management:

    Include modules to predict and recommend control measures for crop diseases and pest outbreaks based on weather and historical trends.

4. Expanded Data Sources:

    Collaborate with agricultural agencies, universities, and governments to collect more robust and diverse datasets for improved accuracy and broader applicability.

5. Cloud and Mobile Accessibility:

    Develop mobile applications and cloud-based solutions to ensure larger access and usability, even in remote areas.

6. Personalized Fertilizer and Irrigation Plans:

    Extend the system to provide specific fertilizer and irrigation recommendations based on individual crop and soil conditions.

7. Machine Learning Model Improvements:

    Continuously update and refine ML models with new algorithms and data to enhance prediction accuracy and scalability.

7.4 Inference

The crop recommendation system powered by Machine Learning can greatly improve farming by helping farmers choose the right crops based on soil, climate, and location. This boosts productivity, reduces financial risks, and supports better decision-making.

However, it faces challenges like needing accurate data, adapting to different regions, and being accessible to small farmers. Adding features like real-time updates, regional languages, and smarter tools can make it more useful. By overcoming these limits, the system can help farmers earn more, grow better crops, and promote sustainable farming.

References

1. Intelligent Crop Recommendation System using Machine Learning. IEEE Xplore, 2021. https://ieeexplore.ieee.org/abstract/document/9418375.

2. Rahman, M.,Hasan, K., Karim, R. et al., "Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series." In: Proceedings of the 2018 International Conference of Computer and Information Technology (ICCIT), 2018. 10.1109/ICCITECHN.2018.8631943.

3. Keys, T. G., "Predicting USCS Soil Classification from Soil Property Variables Using Random Forest." Journal of Soil Mechanics, 2016. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0022489816300040.

4. Singh, N., Garg, R. K., "Crop Prediction Method to Maximize Crop Yield Using Machine Learning: A Case Study for Uttarakhand Region." ResearchGate, 2020. https://www.researchgate.net/publication/344198702.

5. Kamilaris, M., Prenafeta-Boldú, F. X., "Crop yield prediction using machine learning: A systematic literature review." Computers and Electronics in Agriculture, 2020. 10.1016/j.compag.2020.105731.

6. Singh, N., et al., "Soil Classification Using Deep Learning and Computer Vision Techniques." Multimedia Tools and Applications, vol. 80, no. 10, 2021. 10.1007/s11042-021-10544-5.

7. Ghosh, A., et al., "Agricultural Recommendation System for Crop Protection." Computers and Electronics in Agriculture, vol. 159, 2018. 10.1016/j.compag.2018.04.006.

8. Lin, X., "Use of Machine Learning Techniques in Soil Classification." Sustainability, vol. 15, no. 3, 2023. 10.3390/su15030237.