AI Driven Crop Recommendation and Decision Support System

Master of Science in Artificial Intelligence and Machine Learning

E.Karthic
(MSA23006)

DEPARTMENT OF INFORMATION TECHNOLOGY
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, LUCKNOW
2023 - 2025

AI Driven Crop Recommendation and Decision Support System

Capstone project submitted as part of the fulfillment of the course curriculum for the third semester of the

Master of Science

in

Artificial Intelligence and Machine Learning

by

E.Karthic

(MSA23006)

under the guidance of

Dr. Sushil Kumar Tiwari

विद्या ददाति विनयं विनयाद् याति पात्रताम्

Indian Institute of Information Technology, Lucknow

2023-25

Declaration by Candidate

I hereby declare that the project titled "**AI Driven Crop Recommendation and Decision Support System**" is the result of my own work, conducted with dedication and under the valuable guidance of **Dr. Sushil Kumar Tiwari** at the Indian Institute of Information Technology, Lucknow. This project has not been submitted, either wholly or in part, for any other degree or academic credit at this or any other institution. All sources of information used have been acknowledged.

E.Karthic
Roll No: MSA23006

Certificate by Supervisor

This is to certify that the capstone project titled "**AI Driven Crop Recommendation and Decision Support System**" submitted by **E.Karthic (Roll No.: MSA23006)** has been carried out under my guidance and supervision. This project has been conducted as part of the fulfillment of the course curriculum for the third semester of the Master of Science degree in Artificial Intelligence and Machine Learning. The work presented in this report is, to the best of my knowledge, a result of the student's independent efforts and it meets the academic standards required for this degree program. I hereby endorse the project and recommend it for evaluation.

Dr. Sushil Kumar Tiwari
(Department of Information Technology)
Indian Institute of Information Technology, Lucknow

Date:

Signature:

Contents

Chapter 1

Introduction

1.1 Introduction

Agriculture is one of the most important sector for the country's economy in a world wide,especially in the countries like India.It is also important for ensuring food security for the rapid growing population. But, Since from the last decades we are facing enormous problems in this sector, including frequent change in weather conditions, decreasing soil fertility by the overuse of chemical fertilizers, and improper way of crop selection. These problems not only affecting the crop production but also leads to significant economic losses for agricultural farmers and also leads to depletion of country's economy worldwide. In such a case, leveraging the most recent technologies nothing but Artificial Intelligence (AI) in agriculture to support farmers in making decisions regarding crop production and cultivation leads to an innovative perfect solution to the problems whatever we are facing right now.
This project aims to overcome some of the problems facing in this sector by building a system that provides data-driven insights and recommendations to the farmers. This system is designed in a way that will analyze various types of agricultural factors such as soil composition, nutrient levels in the soil, frequent change in weather patterns, and climatic conditions to predict the most suitable crops for cultivation in a given region. The primary goal is to assist the agricultural farmers in optimizing their crop yields while maintaining the sustainability and improving their overall income.

One of the major motivations behind this project is the realization of that many farmers still now depends on the traditional knowledge and appropriate guess, when it comes to choosing for the most suitable crops for cultivation in a specific agricultural land. While this traditional way had been worked efficiently over the centuries, it is often inadequate in today's world, where environmental conditions are rapidly changing, and the world is switching to the precision agriculture. With the recent advancements in AI and data science, we forcefully want to use these technologies in this sector to give the each and every insights directly to the knowledge of farmers.

In this project, we majorly focus on developing the recommendation system that integrates machine learning models with the real-world agricultural data. The system takes inputs such as soil test results, historical weather data, and crop requirements to give the recommendations based on the specific farming conditions. Additionally, the system includes a decision support that surely helps the farmers to take the necessary measures and fertilizer recommendations based on the Weather Stress Index that had been calculated by the given environmental conditions.

The significance of this project lies in its potential to bridge the gap between modern agricultural research and practical way of agricultural farming. By simplifying the complexity way of data analysis and presenting it in a user-friendly way, this system can help farmers with valuable insights. Moreover, the use of AI in agriculture aligns with the broader visionary mind of smart farming, which is necessary for addressing the recent global challenges in agriculture like food security, rapid climatic change, and sustainable development in this sector.

This system has the enormous potential to transform the traditional way of agricultural farming to a AI driven agricultural practices by making them more data-driven and efficient method. Ultimately, this project is not only just about applying AI in agriculture; it is about contributing to a cause or problem that affects millions and millions of lives and ensuring a brighter future for the farming communities.

1.2 Problem Statement

Agricultural sector faces numerous problems, including reliance on traditional practices, lack of soil and crop-specific knowledge, and rapid change in weather patterns, leading to poor crop selection, resource inefficiency, and economic losses. Existing support systems facing a lots of struggle to provide precise recommendations and personalized guidance to farmers. Farmers are also vulnerable to risks like improper crop selection and no proper guidance with limited tools to mitigate these challenges. This project aims to address these issues by leveraging AI and machine learning to provide the most optimal crop recommendations, risk management insights, and sustainable agricultural farming practices. It aims to help farmers by giving them easy-to-use tools based on data. This will make agricultural farming more productive, sustainable, and profitable.

Chapter 2

Literature Survey

Predicting most suitable crop to cultivate based on a given environmental conditions is a complex task due to numerous factors and vast amounts of data. This literature review explores recent research on this area to address the challenges facing, specifically focusing on their application in crop recommendation and decision support systems.
In 2018,a paper [1] describes the development of a crop recommendation system dedicated to enhance agricultural farming practices and decision-making by utilizing machine learning techniques.

- The system utilizes algorithms like Decision Trees, Support Vector Machines, and Naive Bayes to analyze soil characteristics, weather conditions, and crop types to suggest the most suitable crops to farmers.
- It also highlights the importance of region-specific agricultural practices, including variations in their soil type and climatic patterns, to offer valuable guidance for crop selection.
- Key limitations include the need for region-specific algorithms to improve precision and to address the computational problem in real-time decision-making.

In 2021,another paper [2] provides an study of various algorithms, specially aimed at improving agricultural farming practices and decision-making processes.

- Explores various machine learning classifiers, including Support Vector

Machines (SVM), Random Forest (RF), Naive Bayes (NB), and Decision Trees (DT), to predict the best crop to grow in a particular area.

- Compares the accuracy and performance metrics of each and every algorithm to determine the most optimal model for Crop recommendation based on the specific environmental conditions.
- Emphasis the need for a structured and data-driven approach in agriculture farming practices, especially in getting the optimal crop choices that ensures high crop yield and the most efficient way of resource utilization.
- Addresses the problems in data collection and the integration of IoT sensors for real-time data, which could enhance the accuracy of crop prediction in dynamic environments.

In 2023,a paper [3] discusses about the combined crop and fertilizer recommendation model that utilizes machine learning algorithms and rule-based approaches to optimize agricultural farming practices, specifically in Rwanda, a country in East Africa.

- The system utilizes the models trained on soil properties like nitrogen, phosphorus, potassium, and pH levels, achieving an accuracy of 97%.
- Highlights the importance of how AI can enhance resource efficiency and sustainability in agriculture farming practices, especially through precision agricultural farming that gives recommendations to specific crop and soil needs.
- A limitation of this approach is that soil conditions, such as nutrient levels and acidity, can differ significantly between regions and countries. This might affect the accuracy of the crop recommendations.

Recently in 2024,a paper [4] briefly discusses about an AI-based model that suggests the best crops to grow.

- Employs machine learning techniques in agricultural sector to analyze large number of datasets from agricultural fields and adapt to tailored recommendations dynamically, supporting sustainable farming practices by optimizing the resource usage.

- Integrating the IoT devices and sensor data for real-time monitoring. This type of approach enhances the precision agriculture of crop recommendations and improves crop yield predictions.
- Highlights the benefits of utilizing deep learning and IoT for improved agricultural outcomes.
- Discusses the limitations such as high computational requirements and problems faced in deploying real-time systems.

Recently in 2024,another paper [5] discusses how AI can enhance the most optimal crop selection and offer guidance to farmers.

- Explores the potential of implementing the machine and deep learning methods in smart agricultural farming, highlighting the role of AI in optimizing agricultural farming practices.
- Incorporating timely satellite data from sensors, the system also uses machine learning algorithms to analyze specific field environmental conditions and recommend the most optimal crop for precise agricultural farming.
- Deploying the machine learning models like Decision Trees and SVM, utilizing IoT for timely data collection, and applying the techniques of cloud storage to handle large datasets.

Chapter 3

Methodology

This project follows a systematic approach to develop and evaluate a machine learning model for crop recommendation and decision making, leveraging agricultural dataset.The methodology can be implemented as follows:

3.1 Data Collection

- The collection of the agricultural dataset for crop recommendation and decision making was sourced from the open-source service **Kaggle**.
- The dataset contains 2200 entries with total 7 important features that are necessarily required for crop prediction.
- The Crop labels belongs to the multi class labels of about 22 labels.
- The dataset is a balanced dataset as it contains 100 entries for each labels, totally accounts for an 2200 entries with 7 features as mentioned above.

3.2 Data Preprocessing

- Since the dataset contains zero null values, so there is no need for deletion or imputation of the missing values.
- Since the labels are in categorical,we convert them in the numerical format using the dictionary mapping method to make it suitable for model training.

- In this dataset,we consider a diverse set of features with different range of values, to make it best suitable for model training, we can scale down into the same range of each and every features by using MinMaxScaler method from Scikit-learn library, that will increase the accuracy better.

3.3 Data Visualization and Segmentation

- We can explore the dataset and gain valuable insights from the dataset both statistically and mathematically by plotting graphs and various types of interactive visualizations using matplotlib and plotly library of python.
- Furthermore, we divide the dataset into training and testing sets for model training and validation.

3.4 Model Training

In this step, various machine learning models are trained on the agricultural dataset for crop recommendation to choose most suitable cultivable crop in the agricultural land based on given environmental specifications.The various machine learning models considered for training the dataset and tuned their hyperparameters to increase their accuracy are as follows:

1. Logistic Regression

Description: Logistic Regression is a statistical method used to predict the probability of an event happening.

Tuned Hyperparameters:

- `C=1.0`: The C parameter controls how much the model is penalized for complexity.
- `solver='lbfgs'`: The lbfgs optimizer is used to find the best settings for the model.This approach extends binary logistic regression to handle multiple classes by modeling the probability distribution across all classes simultaneously, using the softmax function.

- max_iter=500: The model is trained for a maximum of 500 iterations.

2. Decision Tree

Description: A Decision Tree is a simple yet powerful machine learning algorithm that makes decisions like a flowchart. It starts with a root node and branches out into decision nodes and leaf nodes.

Tuned Hyperparameters:

- `max_depth=10`: The tree can grow up to 10 levels deep to prevent it from becoming too complex.
- `min_samples_split=5`: A node can only split if it has at least 5 data points.
- `min_samples_leaf=4`: Each leaf node must have at least 4 data points.

3. Random Forest

Description: Random Forest is a machine learning technique that combines many decision trees to make more accurate predictions. It works by randomly selecting different subsets of data and features to train each tree, reducing overfitting.

Tuned Hyperparameters:

- `n_estimators=300`: The model creates 300 decision trees.
- `max_depth=10`: Each tree has a maximum depth of 10 levels.
- `min_samples_split=5`: A node in a tree is split only if it has at least 5 data points.
- `min_samples_leaf=3`: Each leaf node must have at least 3 data points.
- `class_weight='balanced'`: The model automatically adjusts the importance of different classes, especially if they are imbalanced.
- `max_features='sqrt'`: For each split in a tree, the model considers the square root of the total number of features.

4. Support Vector Machine (SVM)

Description: SVM is a technique that finds the best line (or hyperplane) to divide data points into different categories.

Tuned Hyperparameters:

- `C=1.2`: Regularization parameter; larger values emphasize correct classification.
- `kernel='rbf'`: Radial basis function kernel for non-linear classification.
- `gamma='scale'`: Kernel coefficient, automatically scaled based on the data.

5. Gaussian Naive Bayes (GaussianNB)

Description: Gaussian Naive Bayes assumes that features are normally distributed and uses Bayes' theorem to classify data points.

Tuned Hyperparameters:

- `var_smoothing=1e-12`: Smoothing parameter to avoid division by zero in calculations.

6. Multilayer Perceptron (MLP)

Description: MLP is a type of neural network that learns non-linear mappings from input to output using multiple layers of neurons.

Tuned Hyperparameters:

- `hidden_layer_sizes=(100, 50)`: Two hidden layers with 100 and 50 neurons, respectively.
- `activation='relu'`: Rectified Linear Unit (ReLU) activation function for neurons.
- `solver='adam'`: Stochastic gradient-based optimizer.
- `learning_rate_init=0.001`: Initial learning rate for weight updates.
- `max_iter=1000`: Maximum number of iterations during training.

7. K-Nearest Neighbors (KNN)

Description: K-Nearest Neighbors (KNN) is a straightforward machine learning algorithm that classifies data points by looking at the labels of the data points closest to it.

Tuned Hyperparameters:

- `n_neighbors=10`: The model will consider the 10 closest data points to make a prediction.
- `weights='uniform'`: All 10 neighbors will have equal importance in the final decision.
- `metric='minkowski'` and `p=2`: The model will use Euclidean distance to measure the distance between data points.

8. Bagging Classifier

Description: Bagging is a technique that creates multiple models by randomly selecting a portion of the data and training a model on each subset. These models are then combined to make a more accurate prediction.

Tuned Hyperparameters:

- `n_estimators=150`: The model will create 150 different decision trees.
- `max_samples=0.6`: Each tree will be trained on 60% of the data, randomly selected.
- `max_features=0.6`: Only 60% of the features will be used to train each tree.
- `estimator=DecisionTreeClassifier(max_depth=8)`: Each individual model will be a decision tree with a maximum depth of 8.

9. Gradient Boosting Classifier

Description: Gradient Boosting is a machine learning technique that creates a series of decision trees. Each tree learns from the mistakes of the previous one, gradually improving the model's accuracy.

Tuned Hyperparameters:

- `n_estimators=250`: This controls the number of decision trees used in the model.
- `learning_rate=0.01`: This determines how much each tree contributes to the final prediction. .
- `max_depth=2`: This limits the depth of each tree, preventing overfitting.
- `min_samples_split=15`: A node in the tree won't split unless it has at least 15 data points.
- `min_samples_leaf=10`: A leaf node must have at least 10 data points.
- `subsample=0.8`: Only 80% of the data is used to train each tree, reducing overfitting.

3.5 Model Evaluation

- Each model is evaluated based on its accuracy and robustness for the selection of most suitable crop to be cultivated in a given environmental conditons.
- **Selection of Best Model:** The model with the highest performance metrics, including accuracy, precision, recall and F1-score is chosen. The Corresponding model will serve as the primary multiclass classifier model in the recommendation of most suitable crop to cultivate.

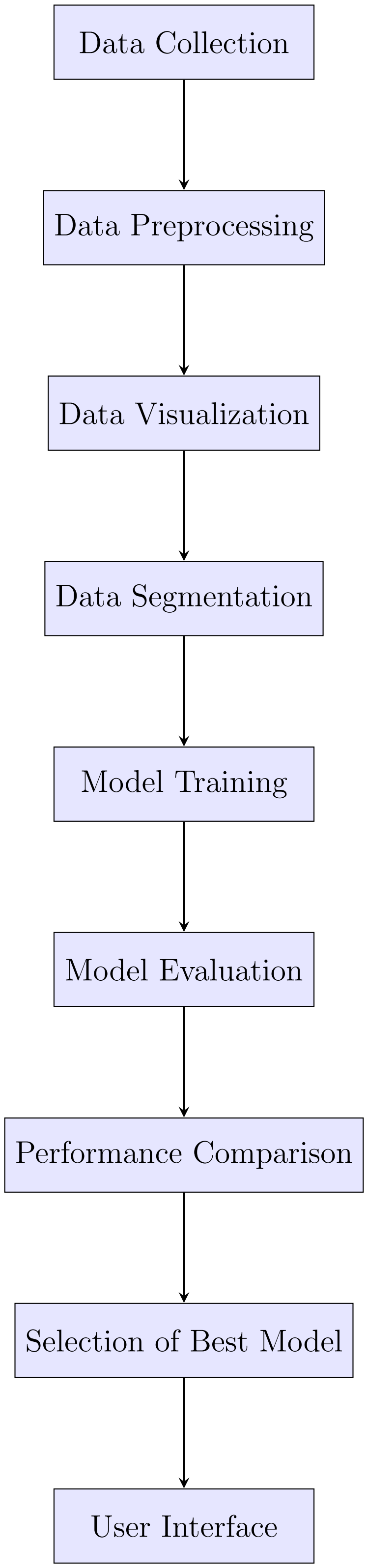
3.6 User Interface

- As mentioned in the above section,the selection of best model for recommendation system for choosing the most suitable crop to be cultivated in a given environmental conditions is solely decided by the evaluation metrics.

- The best model is appended into the system using pickle file that runs in the backend of the system.
- The system will give the informed decisions to the farmers regarding the necessary actions that they want to take immediately is solely based on WSI calculated using the features of our dataset.
- The Weather Stress Index (WSI) is a measure used primarily in agriculture to assess the impact of weather-related stresses on crops. It provides an indication of how environmental factors might negatively affect crop health, growth, and yield.
- The user interface will get the inputs of the environmental conditons from the user and recommends the most suitable crop to cultivate and also give the suggestions to the user based on the WSI calculated in the backend.
- The user interface that made for AI Driven Crop Recommendation and Decision Support System is user friendly to the farmers that can give suggestions and decisions in multiple languages for their easy understanding purpose.

Flowchart

The flowchart below visually represents the process of the methodology.



Chapter 4

Results and Discussions

The graph below compares the various machine learning models based on multiple performance metrics:

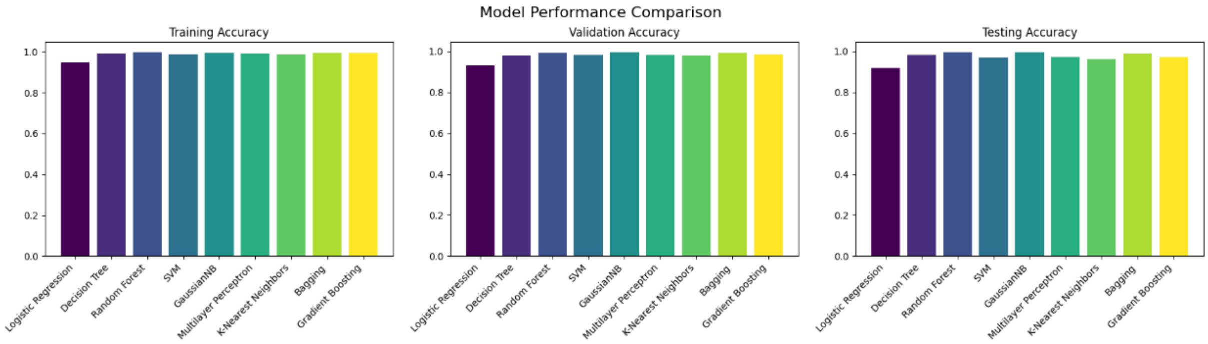


Figure 4.1: Comparison of Training,Validation and Testing Accurcies of different models

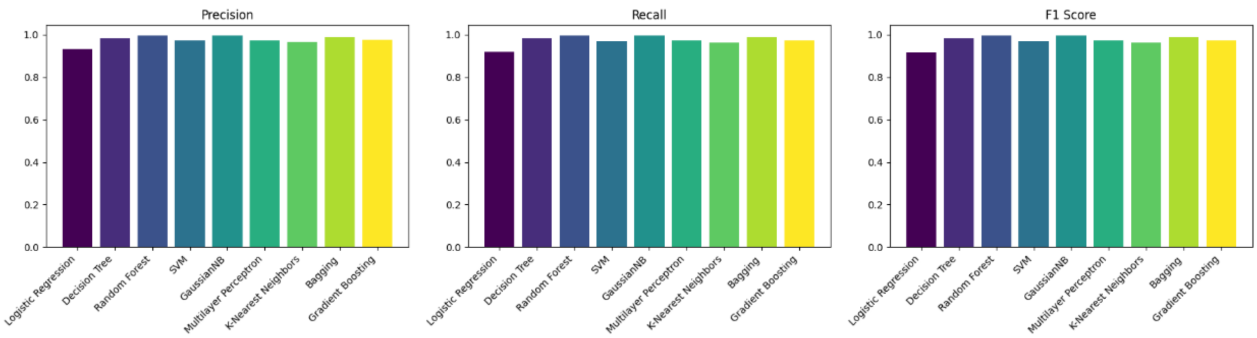


Figure 4.2: Comparison of Precision,Recall and F1 score of different models

The result of the analysis of performance of the various models based on the above two figures are interpreted as follows :

- **Logistic Regression:** Decent performance but slightly lower than other models.
- **Decision Tree:** High accuracy, but Decision Trees can sometimes overfit. This model generalizes well here, possibly due to hyperparameter tuning.
- **Random Forest:** Exceptional performance across all metrics, indicating robust handling of both training and unseen data. Likely due to its ensemble nature and hyperparameters tuning.
- **Support Vector Machine (SVM):** Performs well but slightly behind ensemble models like Random Forest and Gradient Boosting.
- **Gaussian Naive Bayes:** Surprisingly strong performance, likely due to well-separated feature distributions aligning with the Naive Bayes assumption.
- **Multilayer Perceptron (MLP):** Performs well but falls short of ensemble models. This could be due to limited capacity to handle this dataset compared to ensembles.
- **K-Nearest Neighbors (KNN):** Performs slightly worse than ensemble models, possibly due to its sensitivity to data distribution and feature scaling.
- **Bagging:** Excellent performance, showing strong generalization. It benefits from combining multiple Decision Trees for stability.
- **Gradient Boosting:** Consistent across all metrics, though slightly lower than Random Forest.

From the result of the analysis of performance of the various models, it can be found that:

- **Best Overall Model:** Random Forest stands out with the highest accuracy and F1 score, showing it is the most robust for this dataset.

- **Notable Alternatives:** Gradient Boosting and GaussianNB also perform exceptionally well and could be considered depending on computational constraints or dataset characteristics.

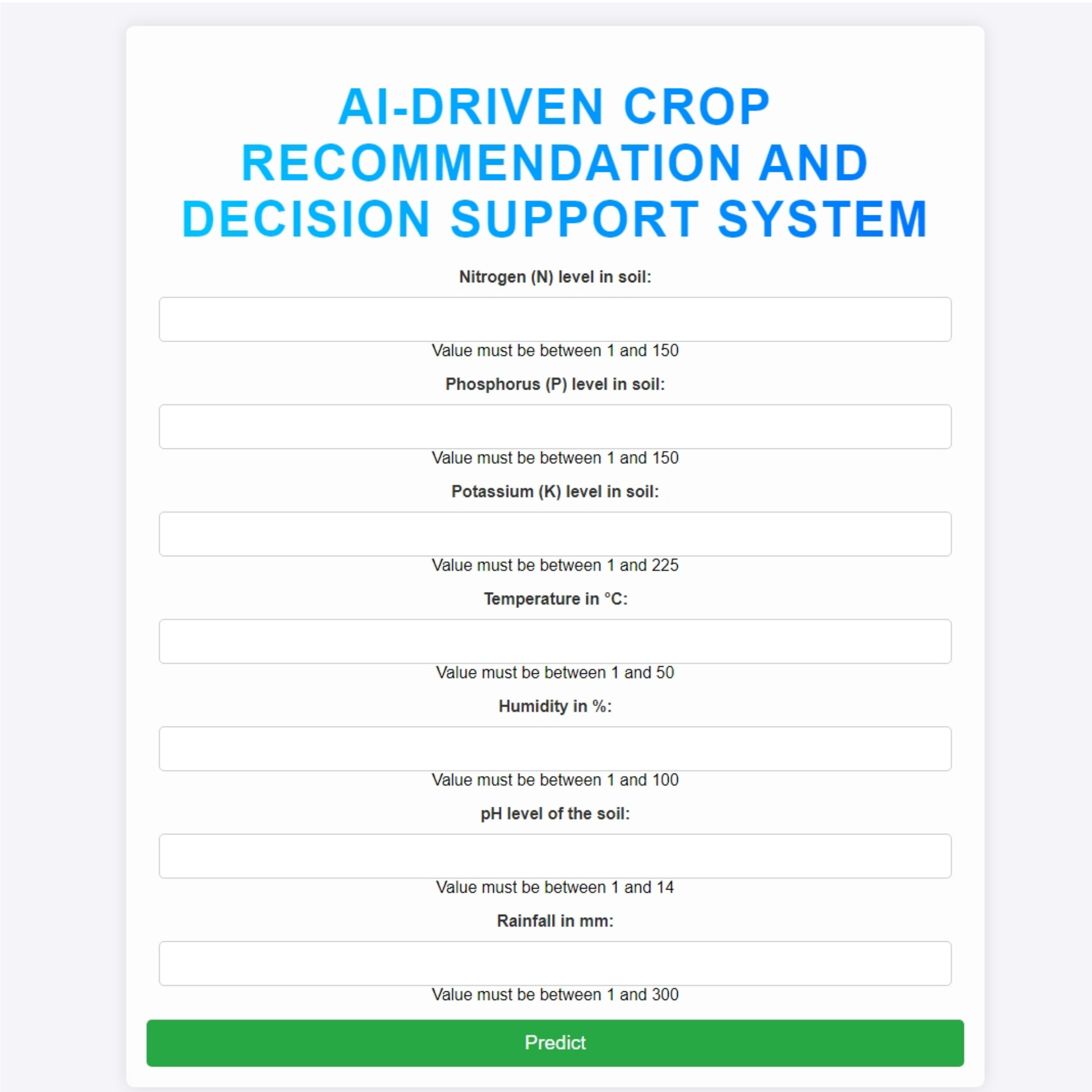


Figure 4.3: User Interface of AI Driven Crop Recommendation and Decision Making System

Chapter 5

Conclusion and Future Work

Agriculture being the major sector in the worldwide, supporting the millions of lives of agricultural farmers and ensuring the food security among the rapid growing population. However, the problem faced by the agricultural sector are huge. This project demonstrates how AI can be effectively utilized in this sector to address these problems and brings the transformation from the traditional way to modern AI driven agricultural practices.
Through this project,designed a system that is capable of providing the personalized and data-driven crop recommendations, the system also more helpful to farmers to make more informed decisions regarding the farming practices, ensuring the better crop yields and sustainable farming practices. One of the key insights gained from this project is that AI bridges the gap between scientific advancements in farming and real-world farming practices.
This project provided an opportunity to apply theoretical knowledge of machine learning and AI to solve a real-world practical problem.
While this project designed well good for an AI-driven recommendation system, there is still required further enhancements. Future improvements might involve expanding the dataset with more localized information, integrating real-time satellite data for more accurate recommendations, and enabling multilingual support to make the system broader accessible to a farmers. Continuous feedback from farmers and agricultural experts will be considered for refining its accuracy, and effectiveness in addressing the problems faced by modern AI driven agriculture.

References

[1] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, "Agroconsultant: Intelligent crop recommendation system using machine learning algorithms," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pp. 1–6, 2018.

[2] G. Chauhan and A. Chaudhary, "Crop recommendation system using machine learning algorithms," in *2021 10th International Conference on System Modeling Advancement in Research Trends (SMART)*, pp. 109–112, 2021.

[3] C. Musanase, A. Vodacek, D. Hanyurwimfura, A. Uwitonze, and I. Kabandana, "Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices," *Agriculture*, vol. 13, no. 11, 2023.

[4] T. Dey, S. Bera, L. P. Latua, M. Parua, A. Mukherjee, and D. De, "icrop: An intelligent crop recommendation system for agriculture 5.0," *IEEE Transactions on AgriFood Electronics*, vol. 2, no. 2, pp. 587–595, 2024.

[5] H. Verma, A. Singh, S. Avasthi, and T. Sanwal, "Ai-based agriculture application for crop recommendation and guidance system for farmers," in *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*, pp. 401–406, 2024.