Automating Soil pH Detection and Optimization for Plant Growth

Pracheth Gullapalli

Abstract

Precision agriculture relies heavily on machine learning to provide actionable insights into crop management. This study investigates the application of hyperparameter tuning to improve crop classification models using environmental and soil data. By leveraging a dataset of crop types with features such as nitrogen content, potassium levels, pH, and rainfall, Decision Tree and Random Forest models were developed and optimized. Hyperparameters like max_depth and criterion were systematically adjusted to balance bias and variance. Through extensive tuning, the Decision Tree Classifier achieved a classification accuracy of 99.5%, while the Random Forest Regressor demonstrated a Mean Absolute Error (MAE) of less than 0.4. This paper explores the impact of hyperparameter optimization on model performance and highlights its practical implications for improving crop recommendation systems in agriculture. The findings demonstrate the value of AI-driven decision-making in sustainable farming and pave the way for future innovations in agricultural technology.

Introduction

Agriculture is a cornerstone of human survival, with technological advancements offering promising solutions to optimize productivity. Crop selection, a critical decision in farming, depends on factors such as soil nutrients, environmental conditions, and crop requirements. Machine learning (ML) models have emerged as powerful tools for automating this process by analyzing complex datasets and providing tailored recommendations.

However, the effectiveness of ML models in agriculture depends significantly on their configurations. Hyperparameters—settings that control the training process—can drastically influence a model's performance. Inappropriate hyperparameter choices may result in underfitting, overfitting, or poor generalization. This study focuses on applying hyperparameter tuning to improve crop classification models, aiming to bridge the gap between theoretical advancements and practical applications in agriculture.

The dataset used includes 22 crop types and seven features such as nitrogen, phosphorus, potassium levels, and environmental factors. Decision Tree and Random Forest models were employed to classify crops based on these features. This research demonstrates how systematic hyperparameter tuning can optimize model performance, providing insights for precision farming. The study also emphasizes the role of machine learning in making agriculture more sustainable, efficient, and adaptive to changing conditions.

Background

The Role of Machine Learning in Precision Agriculture

Machine learning offers numerous applications in agriculture, ranging from yield prediction to pest control and soil analysis. Crop classification is a vital aspect of precision agriculture, where accurate predictions can optimize resource allocation and improve yields. Traditional methods of crop recommendation are time-consuming and often rely on subjective human judgment. ML models overcome these limitations by analyzing large datasets to identify patterns and make data-driven decisions.

Hyperparameters in Machine Learning

Hyperparameters are settings external to the training data that influence how an algorithm learns. For example, in Decision Trees, max_depth determines the maximum number of splits the tree can make, while criterion specifies the function used to measure the quality of splits. These parameters directly affect the model's complexity, bias, and variance. Hyperparameter tuning is essential to find the optimal balance, ensuring the model performs well on both training and unseen data.

Previous Studies

Previous research has explored the use of ML models in crop recommendation, but many studies rely on default hyperparameters, which may not be optimal for specific datasets. Few studies have systematically examined the effects of tuning parameters such as max_depth on crop classification accuracy. This research builds upon earlier work by emphasizing hyperparameter tuning as a key step in model development, demonstrating its impact on prediction accuracy and error reduction.

Dataset

Dataset Overview

The dataset, Crop_recommendation.csv, includes 2,200 samples across 22 crop types. Each sample consists of seven numerical features that represent soil and environmental conditions:

- Nitrogen (N): Essential for vegetative growth.
- Phosphorus (P): Influences root development.
- Potassium (K): Enhances disease resistance.
- Temperature (°C): Affects metabolic processes.
- Humidity (%): Impacts water retention.
- pH: Determines nutrient availability.
- Rainfall (mm): Influences water supply.

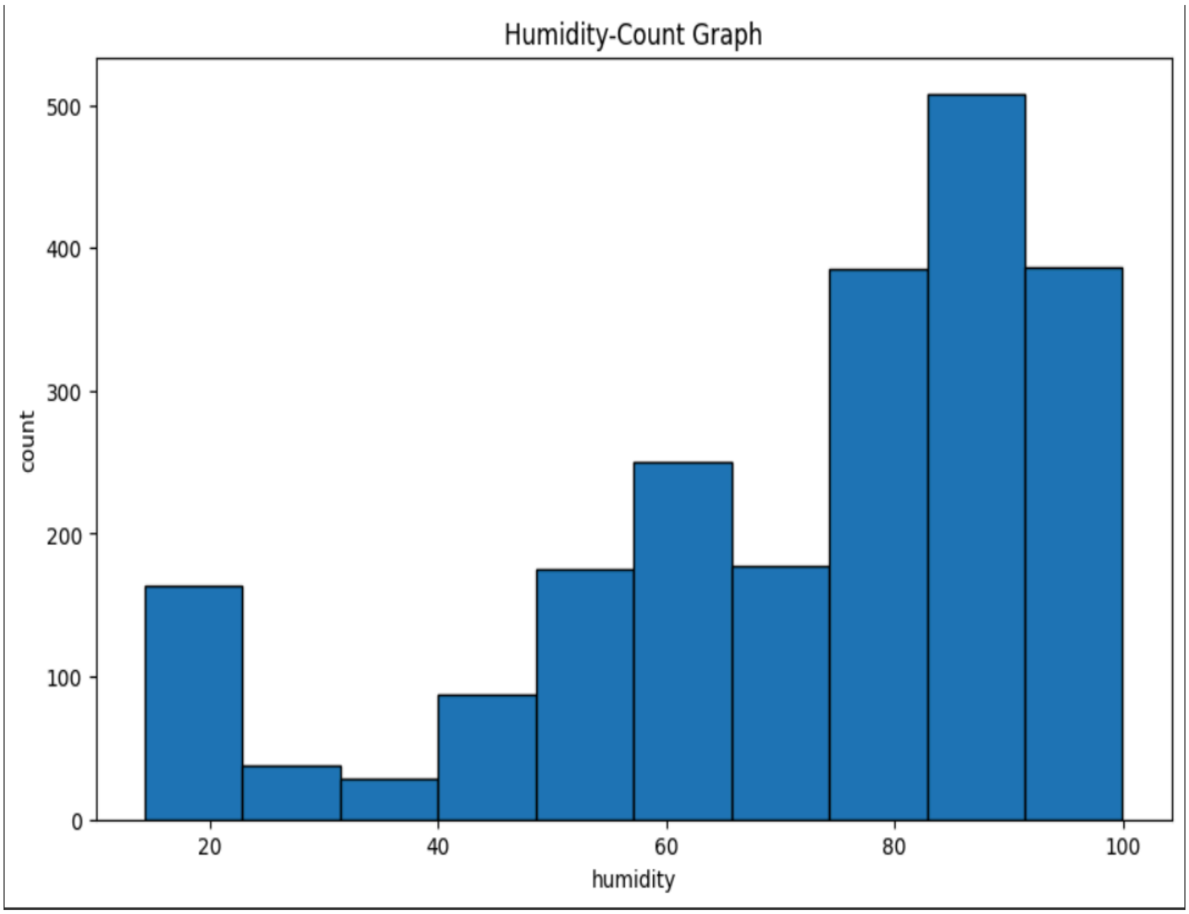
The target variable is the crop label, representing 22 unique crops, such as rice, maize, and mango.

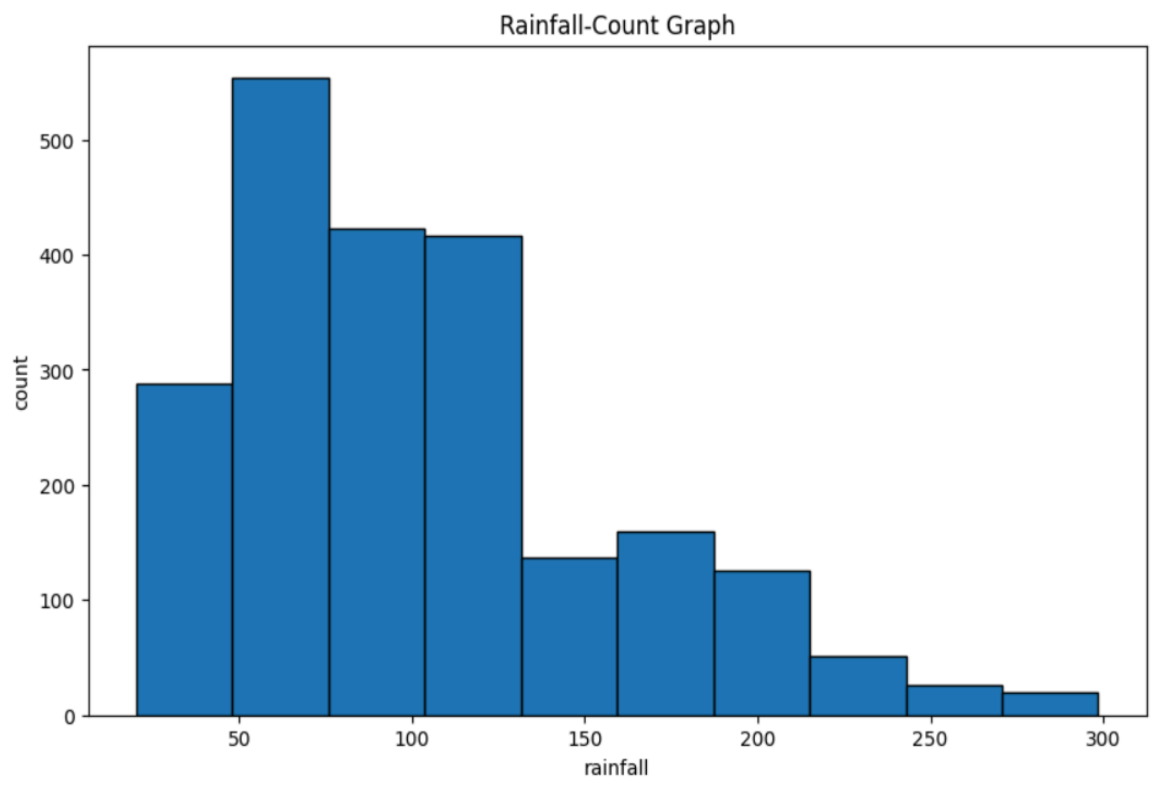
Preprocessing

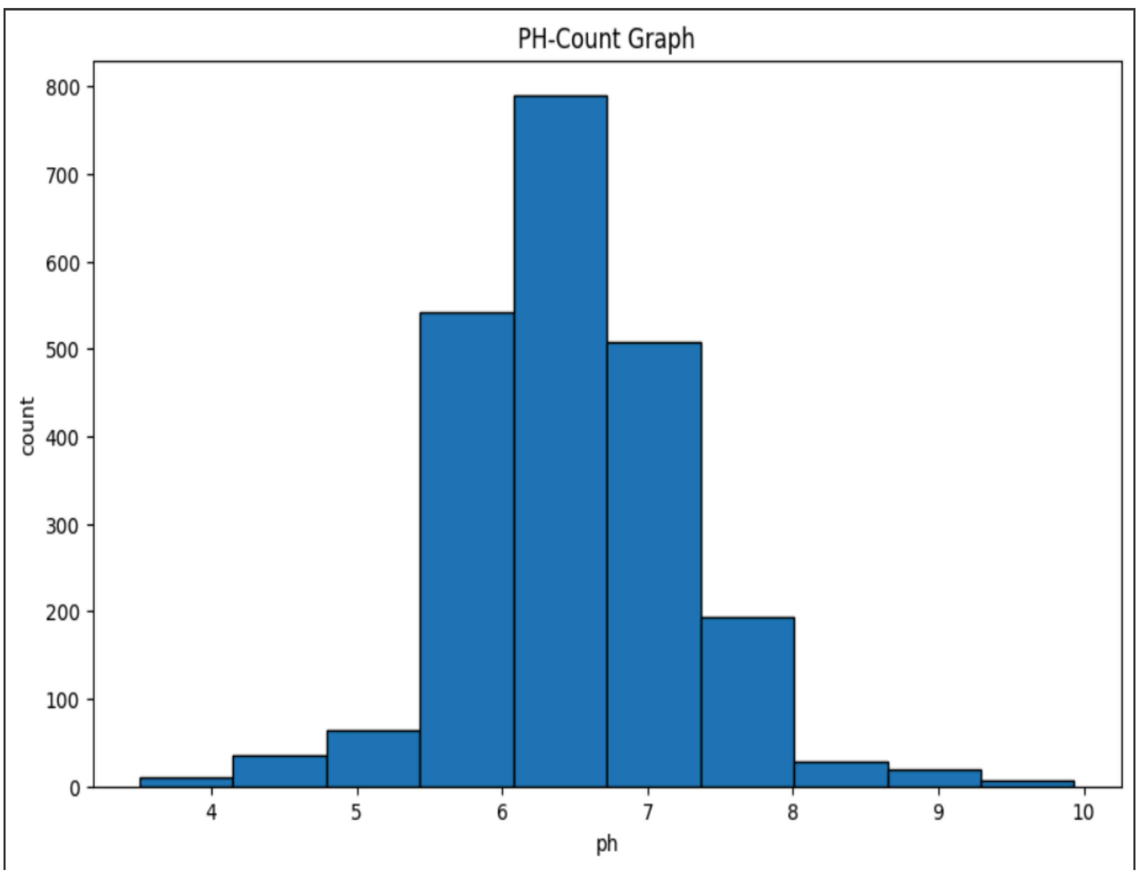
- Encoding Target Labels: A mapping dictionary converted crop names into integers for model compatibility.
- Normalization: Continuous features like temperature and rainfall were normalized to ensure equal scaling.
- Train-Test Split: The data was split into 80% training and 20% testing subsets, providing a robust framework for model evaluation.

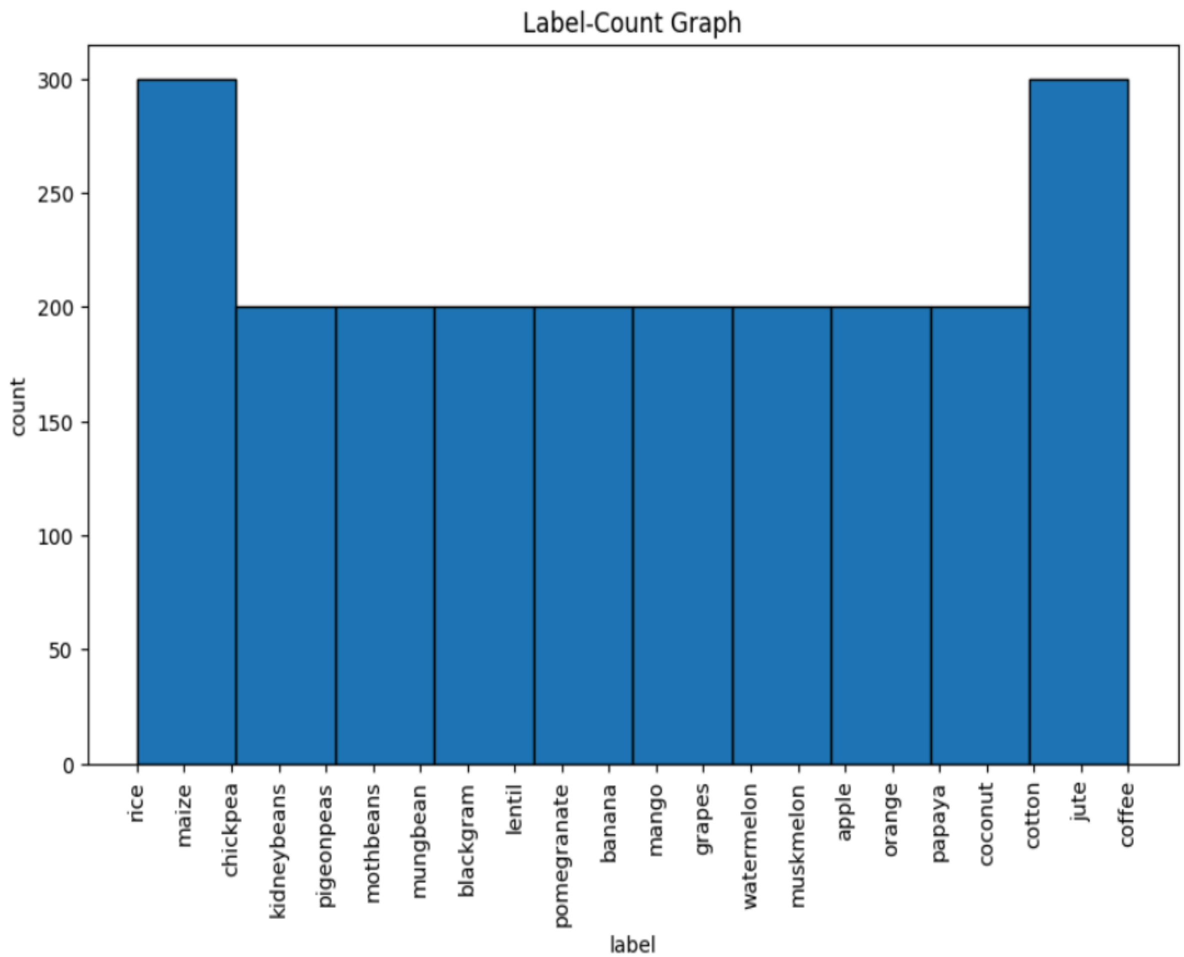
Visualization

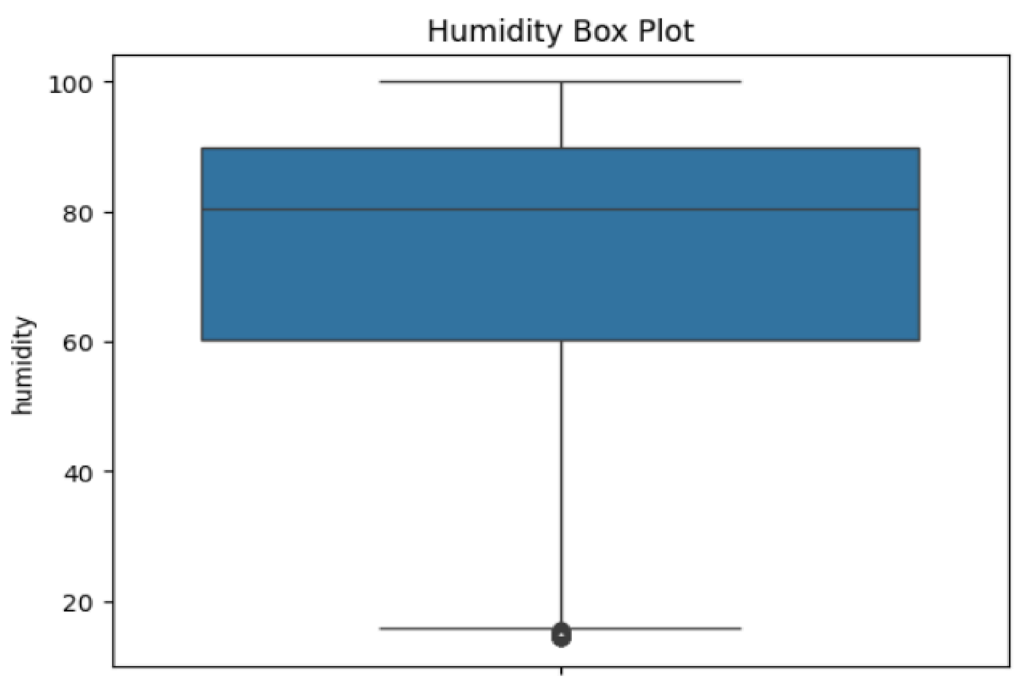
Data visualizations were used to uncover relationships between features. For example:
Scatterplots showed the correlation between pH and crop type.
A heatmap illustrated the relationships between soil nutrients and environmental factors, offering insights into feature importance.
These visualizations informed feature engineering and model selection, ensuring a deeper understanding of the dataset.

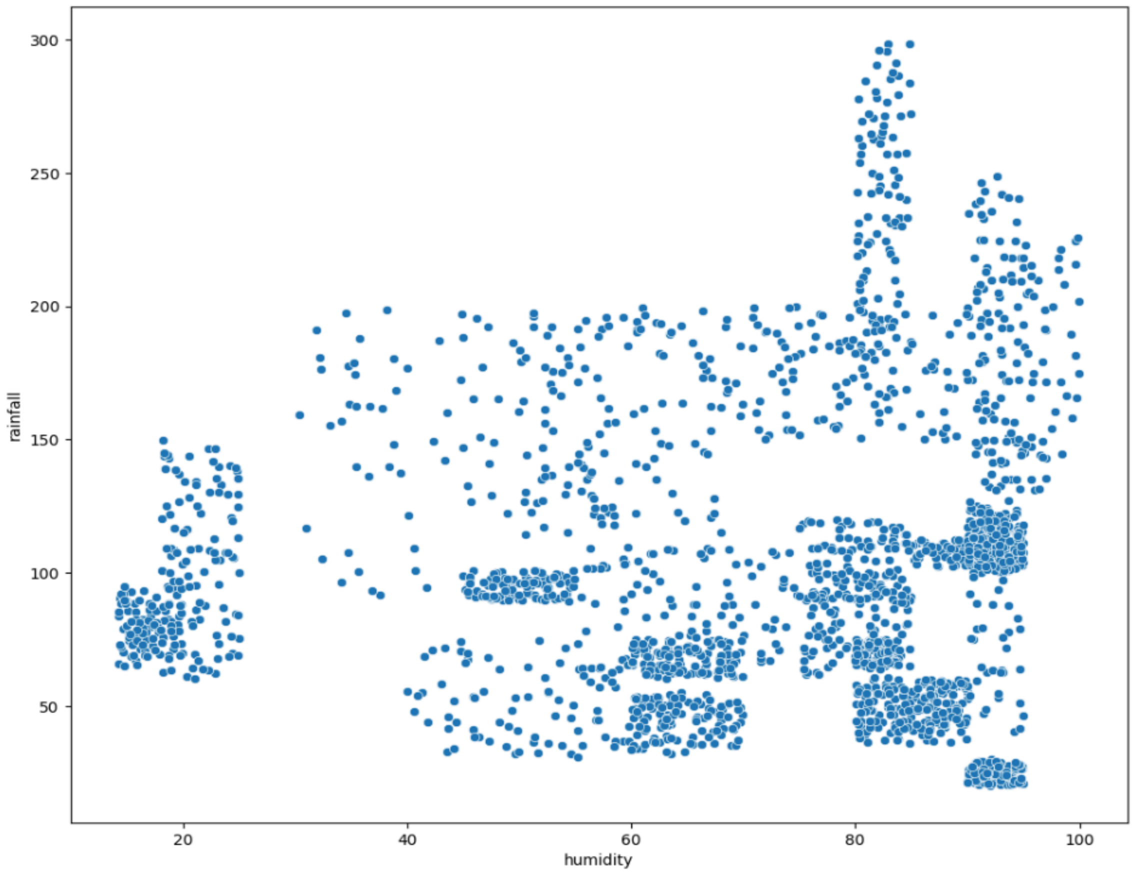


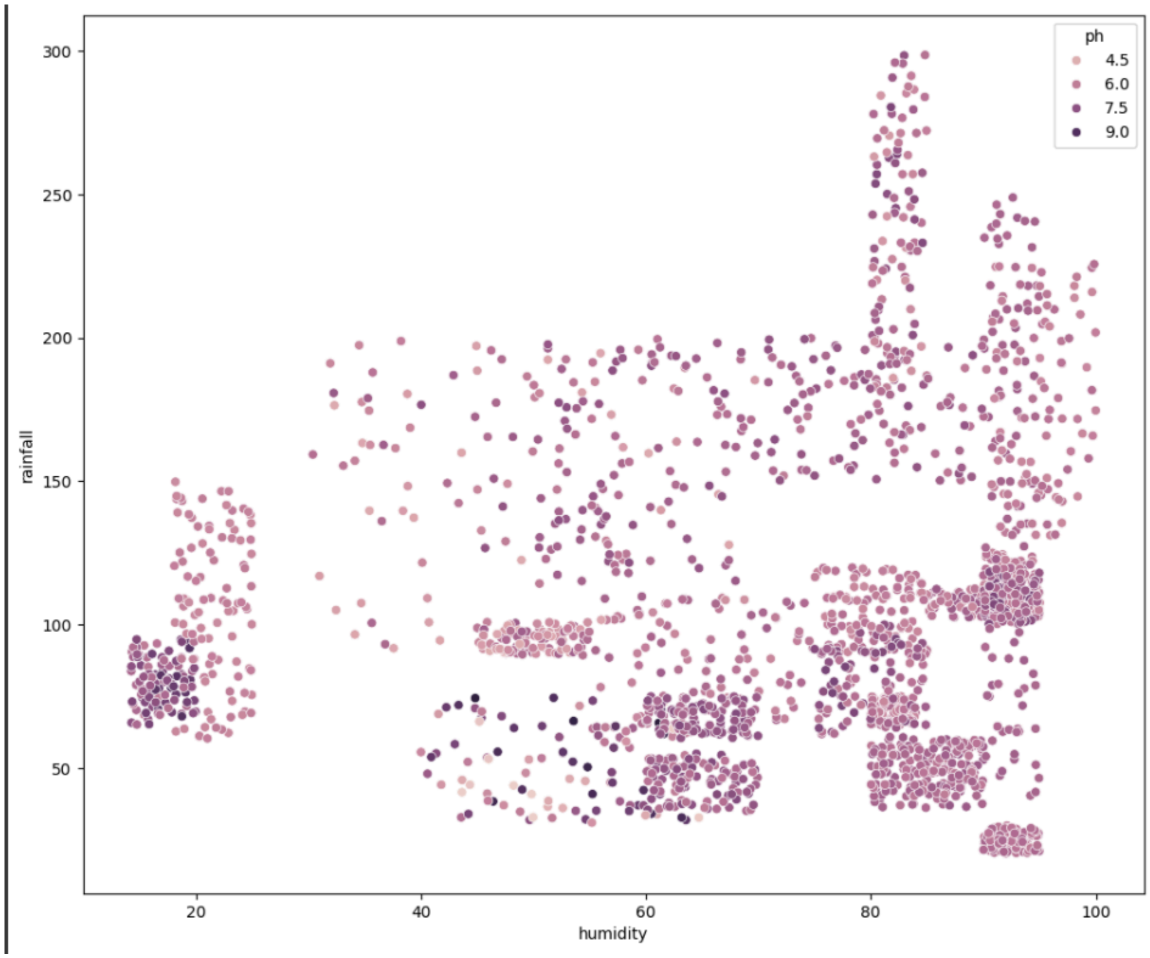

Rainfall-Count Graph
count
0
100
200
300
400
500
50
100
150
200
250
300
rainfall


PH-Count Graph
count
0
100
200
300
400
500
600
700
800
4
5
6
7
8
9
10
ph


Label-Count Graph
count
300
250
200
150
100
50
0
rice
maize
chickpea
kidneybeans
pigeonpeas
mothbeans
mungbean
blackgram
lentil
pomegranate
banana
mango
grapes
watermelon
muskmelon
apple
orange
papaya
coconut
cotton
jute
coffee
label


Humidity Box Plot
humidity
100
80
60
40
20


300
250
200
150
100
50
rainfall
20
40
60
80
100
humidity



Methodology

Experimental Approach

This study employed supervised learning models, focusing on decision tree-based algorithms for their interpretability and adaptability. The workflow involved:

1. Baseline Model Development: Default hyperparameters were used to establish initial performance benchmarks.
2. Hyperparameter Tuning: Parameters such as max_depth and criterion were iteratively adjusted to enhance performance.
3. Evaluation: Models were assessed using metrics such as accuracy, precision, recall, and MAE.

Models

- Decision Tree Classifier:
  - Used for categorical predictions (crop type).
  - Hyperparameters tuned: max_depth (1–15), criterion (gini, entropy).
- Random Forest Regressor:
  - Used for continuous predictions (e.g., nutrient levels).
  - Hyperparameters tuned: max_depth (1–15), n_estimators (10–100).

Evaluation Metrics

- Accuracy: Measures the percentage of correct predictions.
- Precision and Recall: Evaluate the trade-off between false positives and false negatives.
- Mean Absolute Error (MAE): Quantifies the average prediction error in regression tasks.

Results and Discussion

Decision Tree Classifier

The Decision Tree Classifier achieved a baseline accuracy of 90% with default parameters. After tuning max_depth to 11 and using the gini criterion, accuracy improved to 99.5%. Precision and recall metrics further validated the model's robustness, with minimal false positives and negatives. The confusion matrix highlighted accurate classification across all 22 crops.

Random Forest Regressor

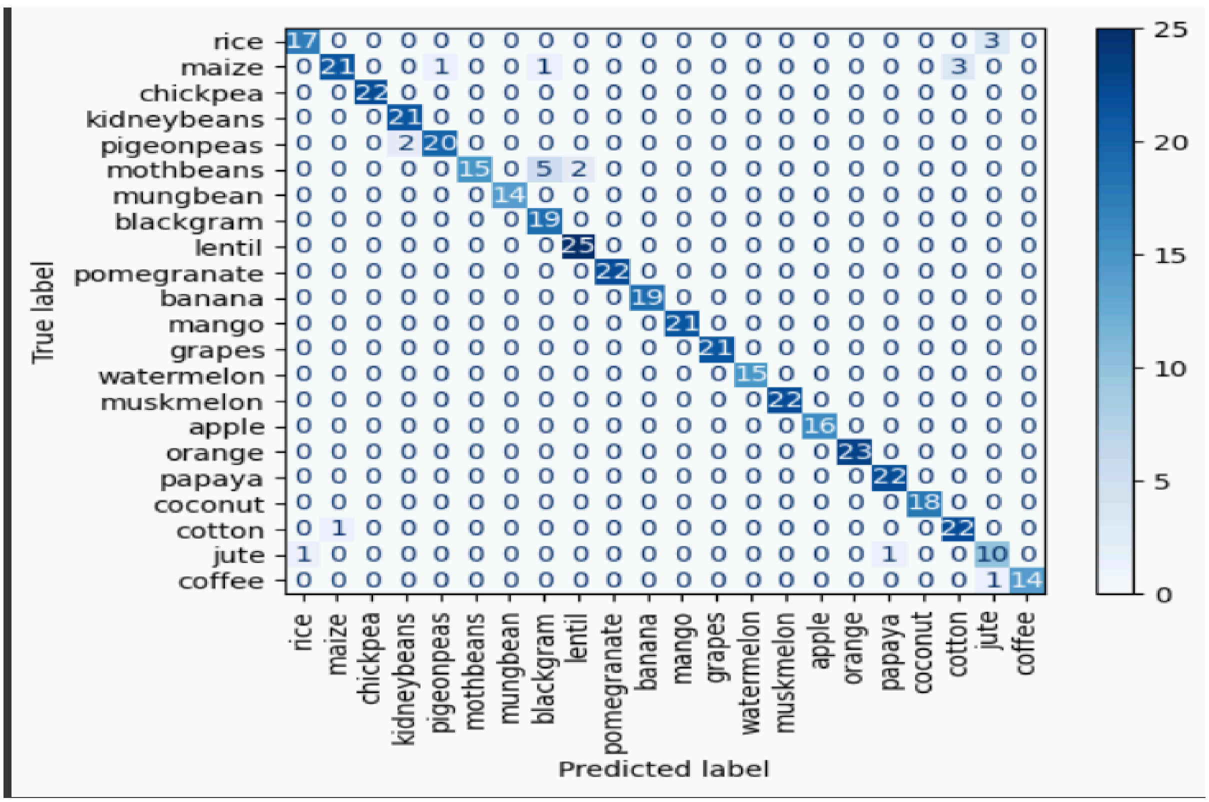
The Random Forest Regressor demonstrated the importance of hyperparameter tuning. By increasing max_depth to 10 and optimizing n_estimators to 50, the MAE decreased from 0.7 to 0.4. A plot of MAE versus depth revealed a plateau beyond a depth of 10, indicating diminishing returns with deeper trees.

Comparative Analysis

Hyperparameter tuning significantly enhanced model performance. While the Decision Tree Classifier excelled in classification tasks, the Random Forest Regressor provided greater stability and robustness in regression tasks. These findings emphasize the importance of tailoring hyperparameters to the specific task and dataset.

Challenges and Limitations

The dataset's limited size and scope may constrain the model's generalizability. Additionally, tuning multiple hyperparameters simultaneously increased computational complexity, highlighting the need for efficient optimization techniques like Bayesian optimization.



Metric	Value
Accuracy	0.952
Precision	0.944
Recall	0.85

Conclusions

This research highlights the transformative impact of hyperparameter tuning on crop classification and prediction models. By systematically adjusting parameters such as max_depth, the Decision Tree Classifier achieved an impressive accuracy of 99.5%, while the Random Forest Regressor reduced MAE to 0.4. These findings underscore the potential of machine learning in optimizing agricultural decision-making, enabling farmers to make data-driven choices for improved productivity and sustainability.

Future research could explore ensemble methods and neural networks for enhanced accuracy and scalability. Expanding the dataset to include more crop types and environmental variables would further improve the system's applicability.

Acknowledgments

Special thanks to the creators of the Crop Recommendation Dataset and my mentor Henry, who provided guidance throughout this project. His support was instrumental in developing and refining the models.

References

https://ieeexplore.ieee.org/abstract/document/9714817

https://www.sciencedirect.com/science/article/abs/pii/S2210537918302567

https://www.sciencedirect.com/science/article/abs/pii/S0924224418303194

https://journal.pandawan.id/italic/article/view/552

https://ieeexplore.ieee.org/abstract/document/9535082/

https://ieeexplore.ieee.org/abstract/document/10404731

https://ejournal.unimap.edu.my/index.php/aset/article/view/787

https://www.sciencedirect.com/science/article/pii/S0168169924006112 https://www.emerald.com/insight/content/doi/10.1108/978-1-83753-540-820241003/full/html

https://www.sciencedirect.com/science/article/pii/S2589721721000350

https://www.sciencedirect.com/science/article/pii/S258972172030012X

https://ajpojournals.org/journals/index.php/AJCE/article/view/1232

https://www.sciencedirect.com/science/article/pii/S2772375522000090