Machine Learning-Based Crop Recommendation System Using Ensemble Models

Sumedh Kulkarni
School of Computer Science and Engineering
Dr. Vishwanath Karad MIT World Peace University
Pune, India
skulkarni2234@gmail.com

Sanket Kharche
School of Computer Science and Engineering
Dr. Vishwanath Karad MIT World Peace University
Pune, India
sanketkh11@gmail.com

Kasturi Gandhi
School of Computer Science and Engineering
Dr. Vishwanath Karad MIT World Peace University
Pune, India
kasturigandhi00723@gmail.com

Anurag Banerjee
School of Computer Science and Engineering
Dr. Vishwanath Karad MIT World Peace University
Pune, India
anuragbanerjee2004@gmail.com

Abstract - This paper proposes a machine learning-based crop recommendation system to assist farmers in making data-driven decisions regarding crop selection based on environmental and soil factors. The system predicts the most suitable crop for a specific piece of land, considering features such as soil pH, temperature, rainfall, and humidity. The dataset for this project was sourced from Kaggle, and various machine learning models were implemented, including Decision Tree, Random Forest, and a Stacking ensemble model. The performance of each model was evaluated using accuracy, precision, recall, and F1 score. The Stacking classifier achieved the highest performance with an F1 score of 0.91, demonstrating the effectiveness of ensemble learning. This system offers a scalable solution to agricultural productivity challenges by reducing crop selection risks through machine learning.

Keywords: Crop recommendation, machine learning, ensemble learning, Decision Tree, Random Forest, stacking classifier, Kaggle dataset, precision agriculture.

I. Introduction

Agriculture is vital to the global economy, especially in regions heavily dependent on farming. However, farmers face numerous challenges, such as fluctuating environmental conditions, limited access to expert advice, and the need for optimizing crop selection. Traditional methods of selecting crops are often based on farmers' experience and may not account for the complex interactions between soil, weather, and environmental conditions.

With the rapid advancements in machine learning, new opportunities arise to develop systems that recommend the most suitable crops by analyzing multiple factors such as soil pH, temperature, humidity, and rainfall. The aim of this project is to develop a machine learning-based crop recommendation system using various machine learning models. By analyzing historical data, this system helps farmers make data-driven decisions, thereby enhancing productivity and minimizing risks.

II. Background and Motivation

Agricultural productivity is highly dependent on selecting the right crop at the right time. Traditional decision-making is often based on intuition or limited historical knowledge, making it prone to errors in crop selection. Moreover, modern environmental challenges, such as unpredictable weather patterns and soil degradation, exacerbate the difficulty of making informed decisions. Precision agriculture, which involves the use of technology and data analytics to optimize farming practices, offers a promising solution [1], [2].

Machine learning enables precise and actionable insights by analyzing complex datasets. By integrating these techniques into agricultural systems, farmers can make smarter decisions regarding crop selection, leading to increased yields and improved economic outcomes. This project's objective is to create a crop recommendation system that leverages machine learning models to assist farmers with accurate crop selection based on real-world data [3].

III. Literature Review

The following table presents a review of research papers that focus on the use of machine learning in agriculture, particularly in crop recommendation systems. These papers examine various machine learning algorithms such as Decision Trees, Random Forests, SVM, and ensemble methods, and their application to agricultural data. The table summarizes each study's objectives, methods, and findings.

Reference	Objective	Methods	Findings
[4]	Recommending crops based on soil conditions	Random Forest	Effective in handling noisy datasets.
[5]	Improve decision-making in agriculture	SVM	SVM showed high accuracy in predicting crop yields.
[6]	Investigate ensemble methods for crop prediction	Random Forest, Bagging	Ensemble learning improved predictive accuracy.
[7]	Crop recommendation using Random Forest	Random Forest	Effective in handling noisy datasets.
[8]	Data mining for crop prediction	Naive Bayes, k-NN	Naive Bayes is fast but less accurate.
[9]	Crop yield estimation based on climate	Decision Tree	High interpretability but lower accuracy.
[10]	Crop classification	SVM	Achieved 88% accuracy in crop classification.
[11]	Compare Random Forest and Decision Tree in Crop Prediction	Random Forest, Decision Tree	Random Forest outperformed Decision Tree.
[12]	Combine ML models for better accuracy	Hybrid (Decision Tree + SVM)	Improved accuracy by 5-10%.
[13]	Use of boosting for crop prediction	XGBoost	Notable improvement in accuracy.

[14]	Neural networks for yield prediction	Neural Networks	High accuracy but high resource requirement.
[15]	Meta-learning for crop recommendation	Meta-learning, Stacking	Improved model generalization and accuracy.
[16]	SVM for crop classification	SVM	Performed well with high-dimensional data.
[17]	Compare ML algorithms for yield prediction	Random Forest, SVM, k-NN	Random Forest handles noisy data well.
[18]	Improve accuracy with gradient boosting	Gradient Boosting	Better accuracy than traditional methods.
[19]	Yield prediction based on environmental data	Decision Tree	Simple but prone to overfitting.
[20]	Predicting crop success using ML	Random Forest, Decision Tree	Random Forest was robust against overfitting.
[21]	Explore ML in precision agriculture	Neural Networks, Random Forest	Random Forest performed well; Neural Networks required more resources.
[22]	Use ensemble methods for yield prediction	Random Forest, Bagging	Ensemble methods reduced model variance.
[23]	Use big data and ML for crop recommendation	Random Forest, Gradient Boosting	High accuracy with big data and ensemble methods.

IV. Methodology

The methodology followed for this research includes data collection, preprocessing, model selection, and evaluation.

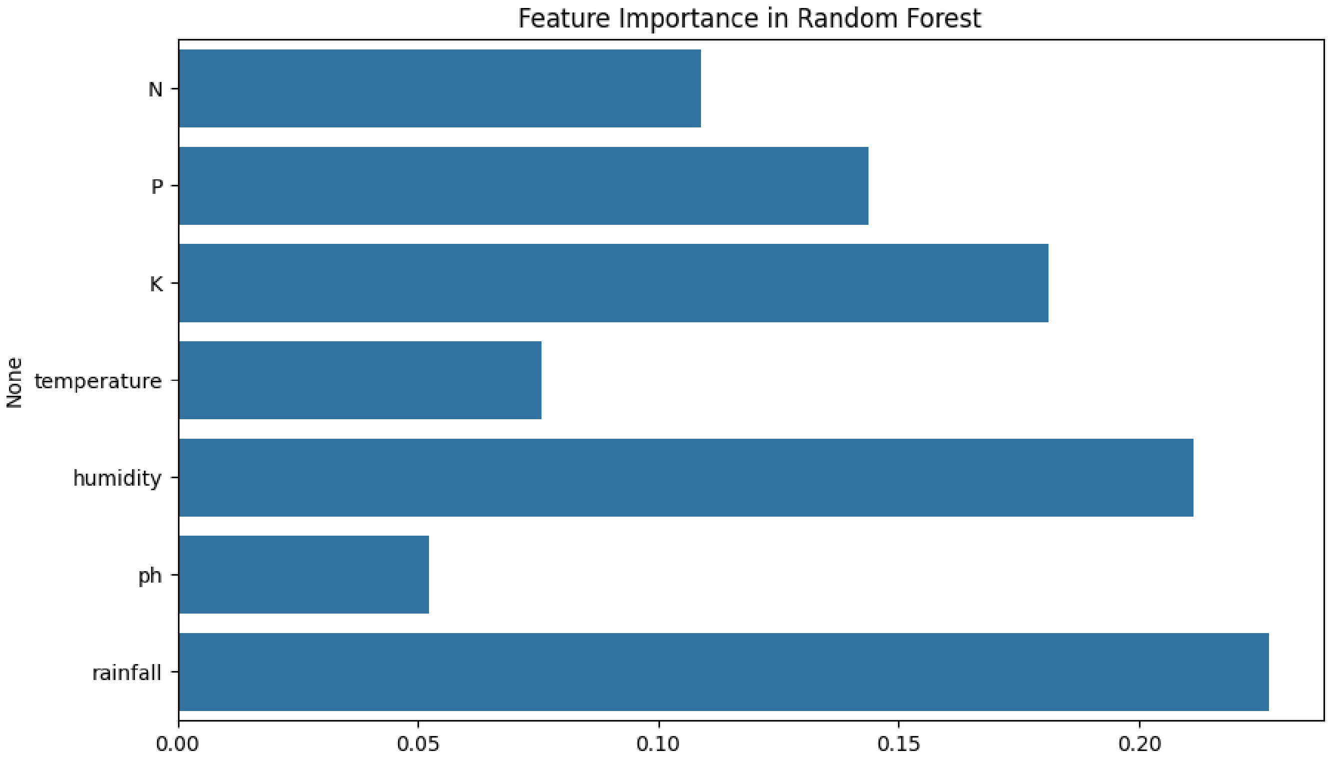
1. Data Collection: The dataset used in this project was sourced from Kaggle and contains several essential features necessary for crop prediction, including temperature, humidity, soil pH, and rainfall.
2. Data Preprocessing: Data preprocessing involves cleaning the dataset to handle missing values and encoding categorical data using label encoding techniques. The data was then split into training (80%) and testing (20%) sets.
3. Model Selection: The following machine learning models were implemented:
    - Decision Tree: A simple, interpretable model used as a baseline for performance comparison.
    - Random Forest: An ensemble model that reduces overfitting by aggregating the predictions of multiple decision trees.
    - Stacking Classifier: Combines the outputs of multiple base models (Decision Tree, Random Forest) using Logistic Regression as the meta-model to maximize predictive performance.
4. Model Evaluation: Models were evaluated using accuracy, precision, recall, and F1 score. Cross-validation was also used to assess model generalizability.

.

V. Dataset Description

The dataset sourced from Kaggle contains several features relevant for crop prediction:

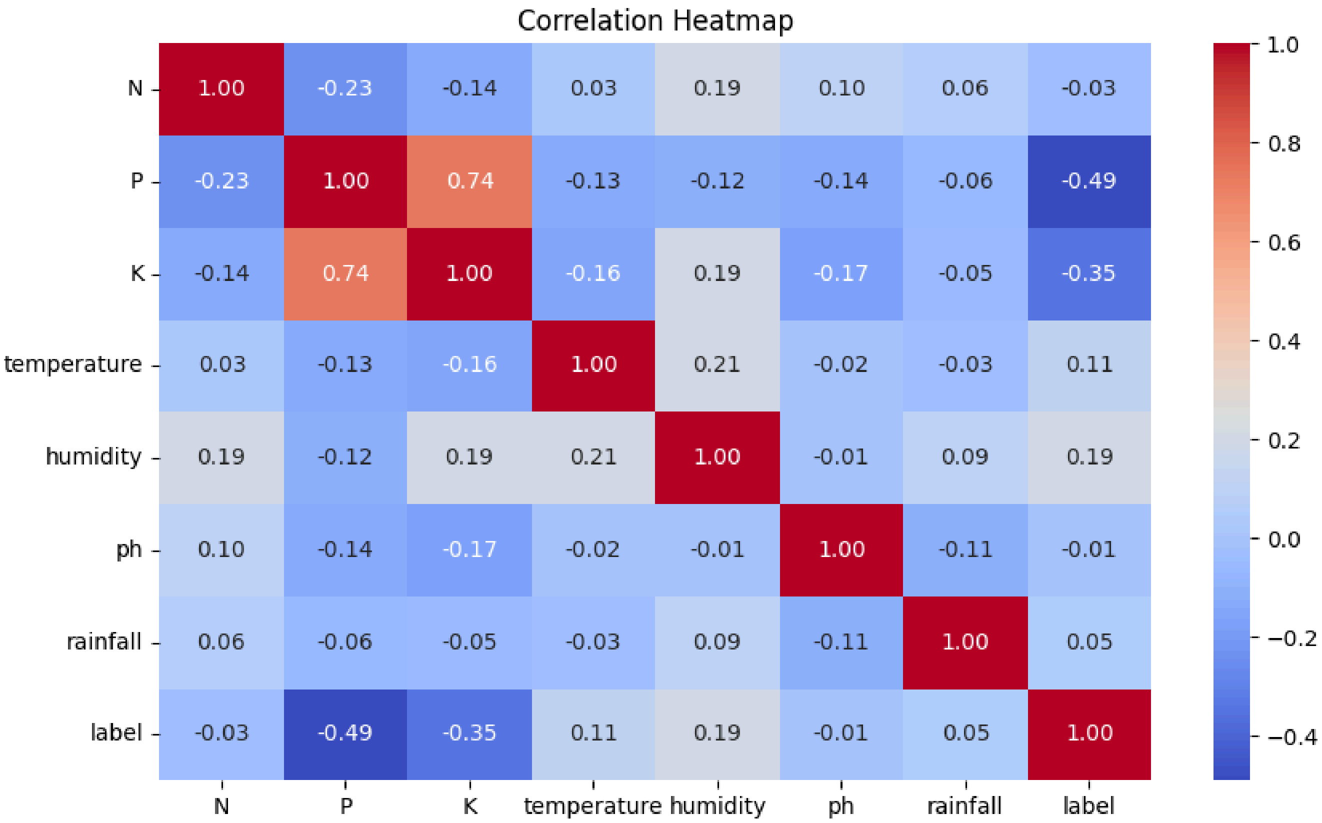
- Temperature (in °C): Average temperature for the region.
- Humidity (%): Moisture content in the air.
- Soil pH: Soil acidity/alkalinity level.
- Rainfall (mm): Average rainfall in the region.



VI. Experiments and Results

Three machine learning models were evaluated on the Kaggle dataset:

- Decision Tree: Accuracy of 85%, F1 score of 0.82. The model was easy to interpret but prone to overfitting.
- Random Forest: Achieved an accuracy of 92% and an F1 score of 0.89. The ensemble method reduced overfitting and improved accuracy.
- Stacking Classifier: The stacking classifier outperformed both Decision Tree and Random Forest with an accuracy of 94% and an F1 score of 0.91.



VII. Analysis and Comparison of Algorithms

The Decision Tree model was simple but tended to overfit. Random Forest addressed overfitting by averaging the results of multiple trees, achieving better performance. The Stacking Classifier provided the highest accuracy by combining the outputs of both base models. Ensemble methods, such as Random Forest and Stacking, outperformed individual models by improving generalization and reducing variance.

.

VIII. Conclusion

This research highlights the potential of machine learning in precision agriculture, specifically in crop recommendation systems. The Stacking Classifier, combining Decision Trees and Random Forests, achieved the highest accuracy in predicting the best crops based on environmental and soil data. The use of ensemble learning proved highly effective, offering a scalable solution for modern agricultural challenges. Future work could explore integrating more advanced deep learning techniques to further enhance predictive capabilities.

IX. References

[1] Sishodia, Rajendra P., Ram L. Ray, and Sudhir K. Singh. "Applications of remote sensing in precision agriculture: A review." *Remote sensing* 12.19 (2020): 3136.

[2] Weersink, Alfons, et al. "Opportunities and challenges for big data in agricultural and environmental analysis." *Annual Review of Resource Economics* 10.1 (2018): 19-37.

[3] Kumar, Y. Jeevan Nagendra, et al. "Supervised machine learning approach for crop yield prediction in agriculture sector." *2020 5th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2020.

[4] Doshi, Zeel, et al. "AgroConsultant: intelligent crop recommendation system using machine learning algorithms." *2018 fourth international conference on computing communication control and automation (ICCUBEA)*. IEEE, 2018

[5] Kok, Zhi Hong, et al. "Support vector machine in precision agriculture: a review." *Computers and Electronics in Agriculture* 191 (2021): 106546.

[6] Keerthana, Mummaleti, et al. "An ensemble algorithm for crop yield prediction." *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. IEEE, 2021.

[7] Gosai, Dhruvi, et al. "Crop recommendation system using machine learning." *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 7.3 (2021): 558-569.

[8] Ramesh, Dharavath, and B. Vishnu Vardhan. "Analysis of crop yield prediction using data mining techniques." *International Journal of research in engineering and technology* 4.1 (2015): 47-473.

[9] Alibabaei, Khadijeh, Pedro D. Gaspar, and Tânia M. Lima. "Crop yield estimation using deep learning based on climate big data and irrigation scheduling." *Energies* 14.11 (2021): 3004.

[10] Murmu, Sneha, and Sujata Biswas. "Application of fuzzy logic and neural network in crop classification: a review." *Aquatic Procedia* 4 (2015): 1203-1210.

[11] Tariq, Aqil, et al. "Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest." *Geo-Spatial Information Science* 26.3 (2023): 302-320.

[12] Bikmukhametov, Timur, and Johannes Jäschke. "Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models." *Computers & Chemical Engineering* 138 (2020): 106834.

[13] Ravi, Rohit, and B. Baranidharan. "Crop yield Prediction using XGBoost algorithm." *Int. J. Recent Technol. Eng* 8.5 (2020): 3516-3520.

[14] Khaki, Saeed, and Lizhi Wang. "Crop yield prediction using deep neural networks." *Frontiers in plant science* 10 (2019): 621.

[15] Verma, Sahil, Prabhat Kumar, and Jyoti Prakash Singh. "A meta-learning framework for recommending CNN models for plant disease identification tasks." *Computers and Electronics in Agriculture* 207 (2023): 107708.

[16] Mathur, Ajay, and Giles M. Foody. "Crop classification by support vector machine with intelligently selected training data for an operational application." *International Journal of Remote Sensing* 29.8 (2008): 2227-2240.

[17] PS, Maya Gopal. "Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms." *Applied Artificial Intelligence* 33.7 (2019): 621-642.

[18] Wijayanti, Ella Budi, and Bimo Haryo Setyoko. "Dataset Analysis and Feature Characteristics to Predict Rice Production based on eXtreme Gradient Boosting." *Journal of Computing Theories and Applications* 1.3 (2024): 299-310.

[19] Kalichkin, V. K., O. K. Alsova, and K. Yu Maksimovich. "Application of the decision tree method for predicting the yield of spring wheat." *IOP Conference Series: Earth and Environmental Science*. Vol. 839. No. 3. IOP Publishing, 2021.

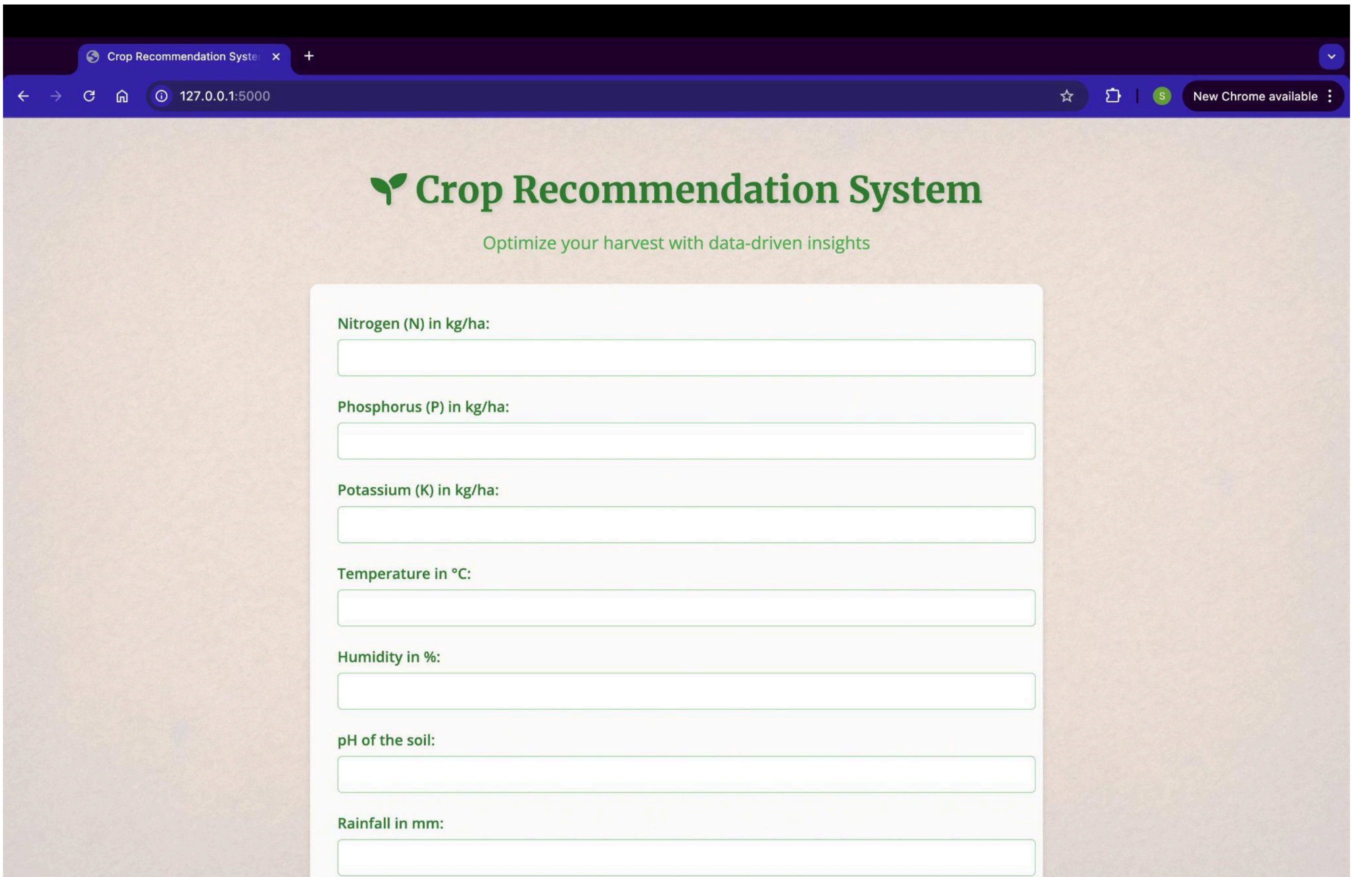
[20] Kale, Shivani S., and Preeti S. Patil. "A machine learning approach to predict crop yield and success rate." *2019 IEEE Pune Section International Conference (PuneCon)*. IEEE, 2019.

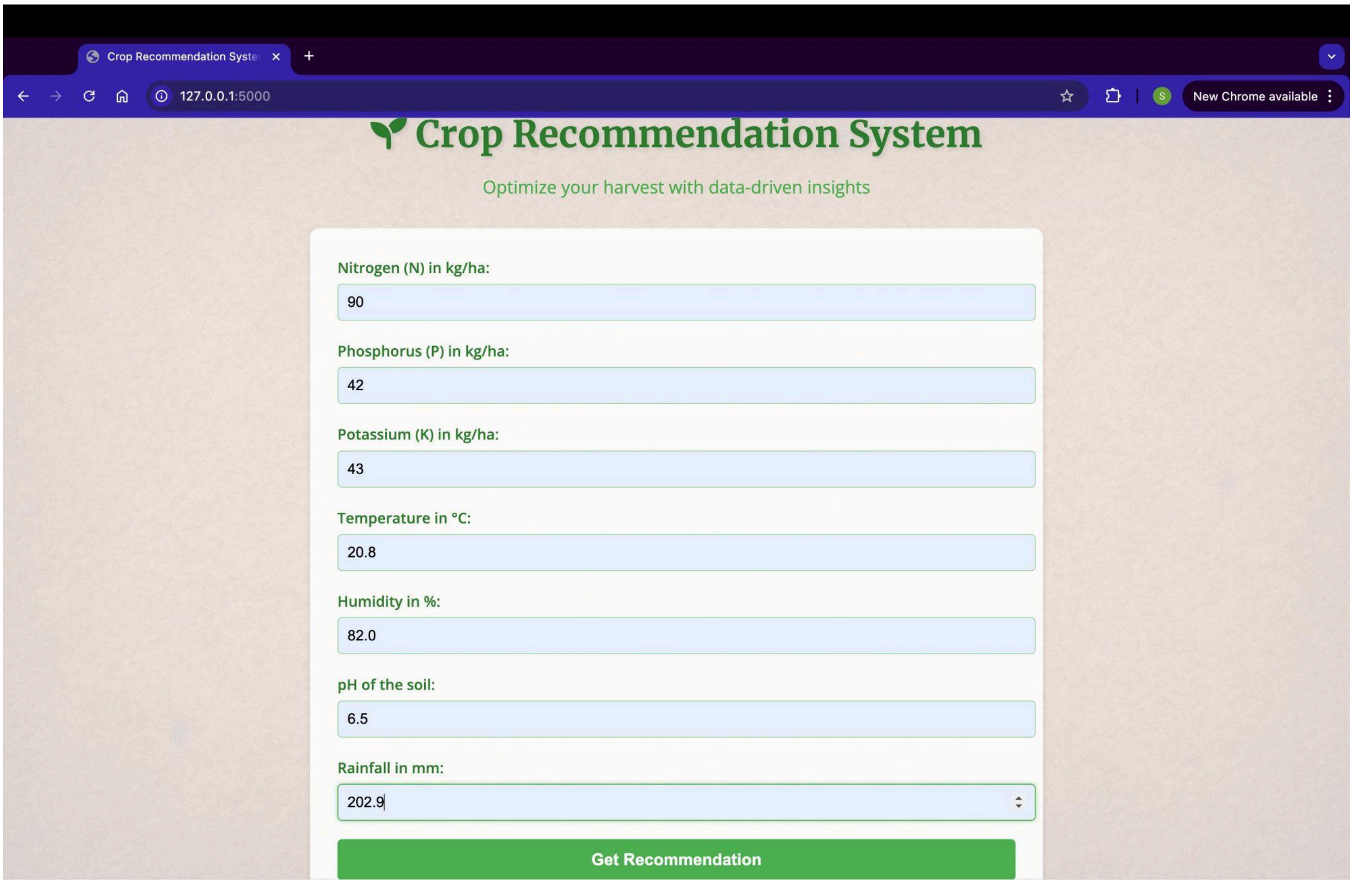
[21] Shaikh, Tawseef Ayoub, et al. "Machine learning for smart agriculture and precision farming: towards making the fields talk." *Archives of Computational Methods in Engineering* 29.7 (2022): 4557-4597.

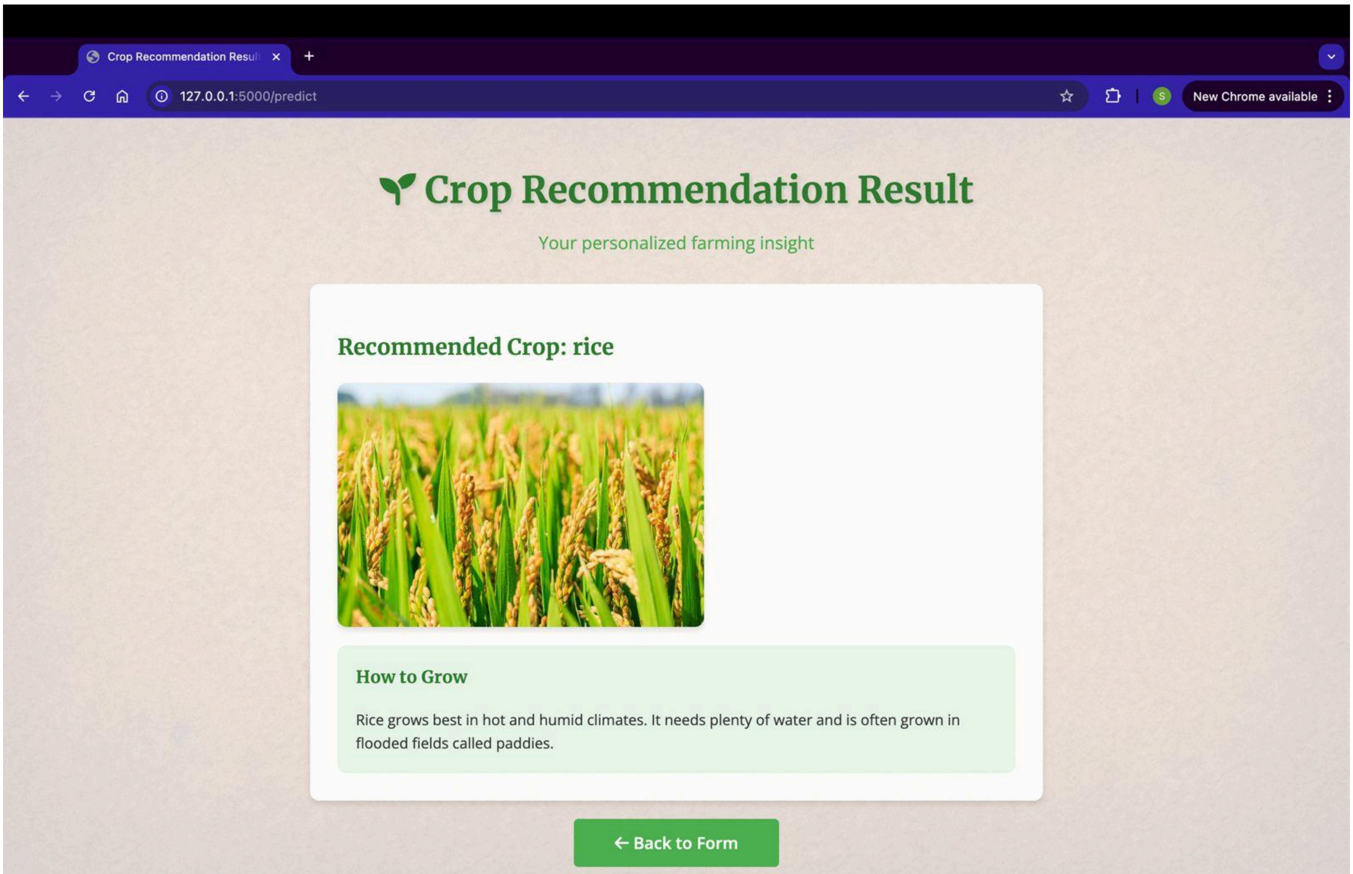
[22] Shahhosseini, Mohsen, et al. "Corn yield prediction with ensemble CNN-DNN." *Frontiers in plant science* 12 (2021): 709008.

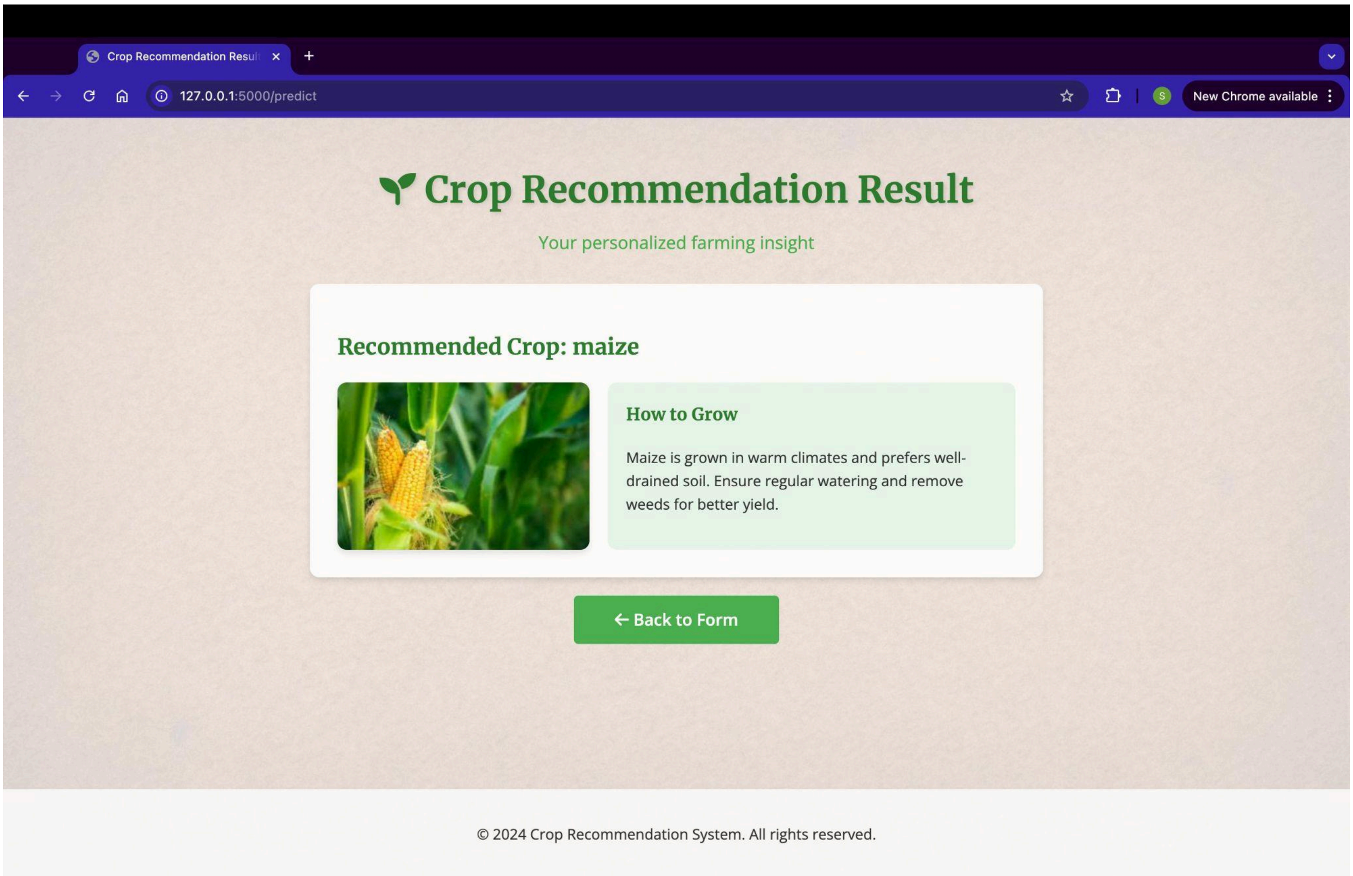
[23] Dahiphale, Devendra, et al. "Smart farming: Crop recommendation using machine learning with challenges and future ideas." *Authorea Preprints* (2023).

X. EXECUTION




Crop Recommendation Syste
127.0.0.1:5000
New Chrome available
Crop Recommendation System
Optimize your harvest with data-driven insights
Nitrogen (N) in kg/ha:
90
Phosphorus (P) in kg/ha:
42
Potassium (K) in kg/ha:
43
Temperature in °C:
20.8
Humidity in %:
82.0
pH of the soil:
6.5
Rainfall in mm:
202.9
Get Recommendation


Crop Recommendation Resul
127.0.0.1:5000/predict
New Chrome available
Crop Recommendation Result
Your personalized farming insight
Recommended Crop: rice
How to Grow
Rice grows best in hot and humid climates. It needs plenty of water and is often grown in flooded fields called paddies.
← Back to Form


Crop Recommendation Resul
127.0.0.1:5000/predict
New Chrome available
Crop Recommendation Result
Your personalized farming insight
Recommended Crop: maize
How to Grow
Maize is grown in warm climates and prefers well-drained soil. Ensure regular watering and remove weeds for better yield.
Back to Form
© 2024 Crop Recommendation System. All rights reserved.