

Multi Crop Recommendation Accuracy and Predictive Analysis Based on Machine Learning Algorithms

¹Wasee Ahsan

*Department of Science and Technology
American International University-Bangladesh
Dhaka, Bangladesh
waseeahsan39@gmail.com*

³Jannat Ara Tasnim

*Department of Science and Technology
American International University-Bangladesh
Dhaka, Bangladesh
jannatara2399@gmail.com*

²Nabiha Tahsin

*Department of Science and Technology
American International University-Bangladesh
Dhaka, Bangladesh
tahsin.n611@gmail.com*

⁴Yeameen Aziz Abdullah

*Department of Science and Technology
American International University-Bangladesh
Dhaka, Bangladesh
yeameenaziz1@gmail.com*

Abstract—Crop recommendation in agriculture is the process of advising farmers on the most suitable crops to grow in specific regions, considering some natural and meteorological factors like soil type, climate etc. to maximize productivity and sustainability. This paper aims to find the suitable crops to recommend to farmers with multiple choices for yielding crops without risks in-season and off-season through machine learning techniques. This study covers the "Crop recommendation" dataset and applies multiple Machine Learning algorithms- Logistic Regression, XGBoost, CATboost, Random Forest. After training with these models, XGBoost classifier gives a much satisfactory results than others with a proper ROC curve as a result. According to this finding, an app system is deployed which displays the best crop choices with XGBoost Classifier highlighting the advanced determining technique and advantage of machine learning. This research contributes valuable insights into the data-driven agricultural practices, helping farmers optimize crop yield and decision-making with multiple recommendation choices.

Keywords: Multi-Crop Recommendation, Agriculture, Sustainability, XGBoost, System, Machine Learning

I. INTRODUCTION

Agriculture is the foundation of global food supply, nutrition security and economic stability, yet farmers frequently face considerable hurdles when selecting the best crops for production. Traditional crop selection approaches fail to account for the complicated interaction of variables such as changing environmental conditions, degrading soil health day by day resulting in poor yields and resource inefficiency. [1] Farmers often follow conventional agricultural procedures when determining which crops to produce in their fields. Crop yield is heavily influenced by meteorological conditions and soil qualities, yet farmers often overlook this. [1] As

natural factors and soil qualities tend of geographical area tend to change by seasons, the crop suitable for production efficiency may be switched to alternate crop choices too. In recent years, machine learning (ML) has emerged as a transformational technology in agriculture, allowing for data-driven crop suggestions based on local circumstances. [2] Crop selection relies on input factors such as temperature, acreage, and irrigation techniques. Precision in crop selection is achieved by adopting appropriate inputs and models without harming or changing the agricultural production system or nature in a geographical state. [3] This study uses the Crop Recommendation Dataset to investigate the possibilities of machine learning for multi-crop recommendation systems. Unlike traditional single-crop frameworks, this method promotes sustainable agriculture practices by maximizing resource utilization and multiple choices for suitable crop. A variety of machine learning methods, including Random Forest, Logistic Regression, XGBoost Classifier, and CATBoost Classifier are examined for their accuracy and predictive potential in crop recommendation. By including different environmental and soil characteristics, the study hopes to create strong, scalable, and flexible models that can meet the diverse demands of modern farming and help in increasing production. This research contributes to society and the economy by developing machine learning-based models for predicting various crop choices. This research demonstrates how comparative and alternative numerous crops may be generated for increased food supply and balanced economic strength. It investigates the prediction of machine learning models, the outcomes and discussion of findings, and finishes with recommendations, limits, and future research prospects.

II. LITERATURE REVIEW

Crop recommendation using machine learning techniques have already been here for some time gradually making progress in agricultural industry for betterment. E. Prem Kumar and Priyadarshini R et al. [1] have proposed a machine learning-based recommendation system for multi-crop farming to enhance precision agriculture. Their method integrates real-time data, soil characteristics, and crop compatibility in order to recommend multiple crops for cultivation. By focusing on nutrient requirements and environmental factors, the proposed system optimizes land use and aids in crop diversification. It evidences the capabilities of advanced machine learning algorithms in maximizing yield and guaranteeing sustainability. Kalimuthu et al. [2] intended to present a crop prediction model that uses the random forest and Naive Bayes methods of supervised learning algorithms. In this work, the study looks at how far soil and climatic conditions are always variable and attempts to give practical solutions in optimizing crop selection. It shows the flexibility of these algorithms in dynamic agricultural datasets and gives a framework for accurate and data-driven recommendations. Priyadarshini et al. [3] This work focuses on intelligent crop recommendations using machine learning algorithms like Decision Trees and Support Vector Machines. The system enhances crop yield prediction by analyzing soil and climate data. The research outlines how environmental factors integrated into decision-making processes can enhance farming outcomes; hence, it is an indispensable tool for single-crop farming systems. Vijayabaskar et al. [4] mentioned in the study that utilizes predictive analytics to improve crop productivity by showing how historical and seasonal data are vital in predicting the best crops. It highlights how data analysis can reduce uncertainties in traditional agriculture, hence serving as an effective approach to sustainable farming by focusing on soil health and weather patterns. Pudumalar et al. [5] authors have designed a precision agriculture system which, based on integration of data from soil quality and environmental parameters, suggests the right crop. This system uses real-time data in dynamically suggesting crops to farmers, hence showing the importance of timely information in enhancing agricultural productivity. Onoriode et al. [6] This work focuses on optimization techniques for machine learning algorithms in crop recommendation systems through model parameter tuning and ensemble techniques that combine several models. The resultant framework ensures that the proposed system will be scalable and can be used in many regions under diverse farming conditions. Aimufua et al. [7] introduces a stacking ensemble-based approach that combines predictions from multiple base algorithms, resulting in improved accuracy in crop recommendations. This technique is especially effective in addressing the challenges of diverse and heterogeneous agricultural environments. Senapaty et al. [8] A decision support system leveraging machine learning classifiers, such as K-Nearest Neighbors and Logistic Regression, is proposed in this study. By integrating soil, climate, and market data, the

system provides comprehensive recommendations tailored to the farmer's needs. The model's practical application in real-world farming makes it a significant contribution to the field.

The review of the related studies together refers to the potential of machine learning algorithms in the recommendation system for crops. Although previous research achieved certain milestones for single-crop prediction and precision agriculture, latest developments such as multi-crop recommendation systems presented holistic solutions for diverse farming. Real-time environmental data integration, nutrient analysis, and ensemble techniques have increased the adaptability and accuracy of these systems. While working on dynamic datasets, scalability improvement is one of the future works, handling factors related to the usability of farmers who are nontechnical still remains very interesting. These results thus provide a foundation from various studies in realizing powerful and user-friendly multi-crop preferring models responsive to the changing applications of modern farming.

III. METHODOLOGY

This research work is focused on developing a multiple and robust recommendation system for crops by using some of the advanced models of machine learning. Instead of just one crop, this study wants to have an overview whether an alternate crop selection choice is perceived. The methodology of this study includes data preprocessing, model implementation, and evaluation. Each model is briefly explained with its working mechanism to provide insight into their roles in the system. In developing the "Multi-Crop Recommendation System" based on machine learning algorithms, a number of tools and libraries were used. Python was the primary language of implementation because it is powerful and flexible. For data manipulation, pandas and numpy were used, while for visualization, matplotlib and seaborn were used, thus allowing comprehensive data analysis and insights. Machine learning models, including Logistic Regression, Random Forest, CatBoost, and XGBoost, were implemented using scikit-learn and specialized libraries. The entire development process was carried out on Google Colab, leveraging its computational efficiency and GPU support to optimize performance.

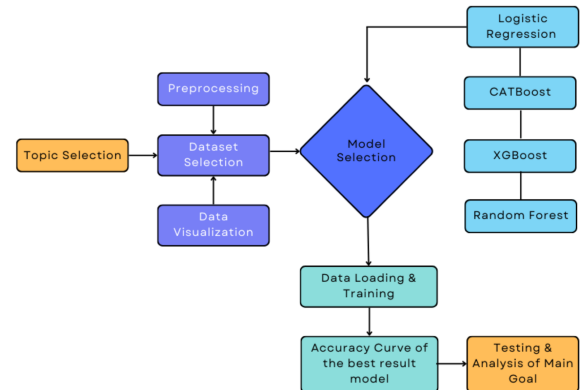


Fig. 1. Fig. 1. Block Diagram of workflow

The first step in this research is finding the dataset related to the study goals. The dataset used for this study was retrieved from Kaggle's "Crop Recommendation.csv," containing 2,201 records. The dataset used for this study was obtained from Kaggle's "Crop Recommendation Dataset." It contains 2,200 records with seven key features: Nitrogen (N), Phosphorus (P), Potassium (K), temperature (°C), humidity (pH), and rainfall (mm). This is the "label" for crop type, which represents the target variable for given environmental parameters. The encoding was done to make the data compatible with model training. This sourced dataset was at a preprocessed stage; it was checked for missing values in the dataset, if any. The target variable is the recommended crop for specific environmental conditions.

After the preprocessing step, a full EDA step was done to understand the dataset. It included the generation of different visualizations, such as histograms and boxplots, studying the distribution of features and detecting potential outliers. Furthermore, a correlation heatmap was created to study the relations among features and get useful insights on how each may influence the target variable. Histograms were used to analyze the distribution of features, boxplots to detect outliers, and correlation heatmaps to identify relationships among features. These tools provided insights into data patterns, variability, and feature interactions, ensuring effective preprocessing and improved model performance.

For the model selection phase, four machine learning models were chosen to evaluate: Logistic Regression (LR), Random Forest (RF), XGBoost Classifier (XGB), and CatBoost Classifier (CB). These models represent diverse algorithmic approaches, ranging from simple linear models to advanced ensemble methods, ensuring a comprehensive comparison:

A. Logistic Regression

Logistic Regression is a statistical model that uses the logistic function to estimate probabilities for a binary or multi-class target variable. It calculates the probability of each class and chooses the class with the highest probability. Formula for Logistic Regression-

$$P(y = 1 | x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}} \quad (1)$$

B. CATBoost Classifier

CATBoost is a type of gradient boosting that can deal with categorical and numerical data. Though it does not require any feature encoding approaches, such as Label Encoder, it transforms category characteristics to numerical ones. It also manages missing values in the dataset automatically so that overfitting can be avoided and the overall performance of the dataset can be improved. [9] CATboost is chosen because it works efficiently with both types of data and reduces the complexity of preprocessing. For loss minimization, the formula is-

$$L = \sum_{i=1}^N l(y_i, \hat{y}_i) + \lambda \sum_{j=1}^M \omega_j^2 \quad (2)$$

where $l(y_i, \hat{y}_i)$ is the loss function, ω_j represents the weights of the characteristics, and λ is the regularization parameter.

C. XGBoost Classifier

XGBoost is an ensemble learning that uses decision trees as base learners for improving the performance of a model. Boosting corrects the errors of the previously made trees. It has inbuilt support for parallel processing, which enables faster training on larger datasets, and it allows customizations to optimize any particular problem at hand. It starts with a base learner, calculates the error, trains the next tree, and this process goes on until the stopping criterion is reached, and finally, the prediction is the sum of the predictions by all trees. [10] This model was chosen because it has very high predictive power in performance and strength for nonlinear data handling. Formula for XGBoost Classifier -

$$o_{bj} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^k \Omega(f_k) \quad (3)$$

D. Random Forrest

A popular decision tree-based ensemble learning algorithm that works by combining predictions from multiple trees: this improves the overall accuracy and reduces overfitting. Because it can handle high-dimensional data with ease and give importance to features, it is chosen. Formula for random forest-

$$\hat{y} = \frac{1}{n} \sum_{i=1}^n T_i(x) \quad (4)$$

The selected dataset was divided into two subsets: one for training and the other for testing. To ensure proper validation, the training dataset was further split into two parts, with 80% allocated for training and 20% for validation. The testing was performed using the test dataset. The evaluation metrics included the AUC value and the ROC curve, which were used to assess the performance of the model. The goal was to identify the best-performing model for recommending the most suitable crop based on the given environmental conditions.

IV. ANALYSIS

In this study, we assessed the performance of several machine learning models, focusing on how their capacity to manage nonlinearity affects their outcomes. It is critical to first understand the characteristics of the dataset used to predict emotional well-being based on social media activity.

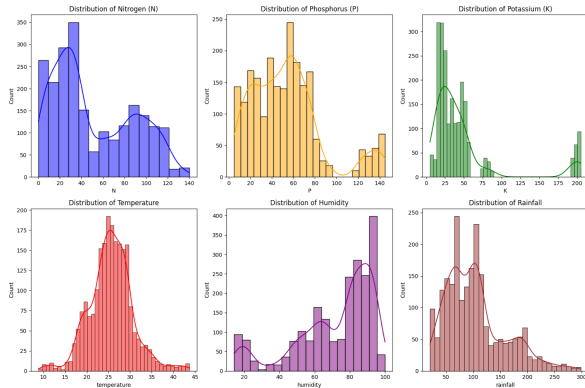


Fig. 2. Variable Distribution in Histogram

These variable distributions provide insight into the variability and common ranges of each parameter, crucial for predicting crop recommendations. This histogram with density plot displays the distribution of six agricultural and environmental variables. Nitrogen (N) and Potassium (K) are positively skewed, the majority of the values falling within the lower range; for N between 20-60, K values fall between 0-50. Phosphorus is multimodal, with double peaks around 30 and 60. Temperature is close to normal around 25 degrees Centigrade, ranging between 10 and 45 degrees Centigrade. The humidity is bimodally distributed around 60% and 90%, showing changes in surroundings. Rainfall is a little skewed and peaks between 50-100 mm, although it further extends to 300 mm. Overall, this dataset is pretty diverse, and a number of the variables suggest regional or environmental variability.

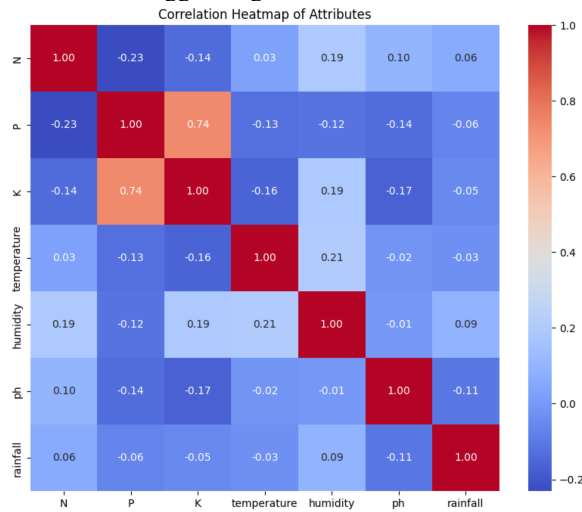


Fig. 3. Heatmap to show correlation of variables

The heatmap represents the various agricultural and environmental attribute correlation coefficients that lie between -1 and 1. A very strong positive correlation of 0.74 exists between P and K, which indicates that both these nutrients mostly increase together. Simultaneously, a weak correlation can be observed for variables such as N and pH or Rainfall and K

close to 0, which indicates negligible or no linear relationship between the two continuous variables. The negative values of some of these correlations, like Nitrogen and Phosphorus, correspond to a value of -0.23, hence showing a weak negative correlation. The diagonal is perfect self-correlation (value = 1), as would be expected. Overall, this heatmap shows how the variables are related and highlights the need to handle overlapping information and how each feature is related to understanding the element's importance in the study's determination.

N	P	K	temperature	humidity	pH	rainfall	real	1st_recommend	Probability_1	2nd_recommend	Probability_2
94	50	37	25.6658205	80.66385045	6.94801983	206.5869708	rice	rice	52.11%	jute	44.41%
91	35	39	23.70381957	80.41817957	6.970859754	206.2611855	rice	rice	56.71%	jute	37.37%
95	55	47	26.7852306	82.1480873	5.950600566	103.3473887	rice	jute	47.18%	rice	42.75%
91	35	38	24.8928223	80.5258888	8.15402921	103.8782027	rice	jute	49.58%	rice	48.38%
74	54	38	25.6553461	83.41821081	7.128272972	217.3786353	rice	rice	49.98%	jute	43.83%
71	46	40	20.2861837	82.1235421	7.236705436	191.9535738	rice	jute	51.08%	rice	46.31%
70	36	42	21.84106875	80.72880384	6.9482098810000015	202.3630319	rice	rice	56.97%	jute	40.25%
61	32	41	24.97888518	83.891885	6.888431223	204.8881847	rice	jute	52.31%	rice	34.74%
78	42	42	21.32262327	83.0020469	7.283748017000001	192.3197536	rice	jute	56.77%	rice	40.33%
77	36	37	26.88444818	81.4803732	6.138131869	194.5786559	rice	jute	50.65%	rice	38.20%
91	41	38	22.67846116	83.72874389	7.5240800760000015	206.9133156	rice	jute	56.19%	rice	41.46%
88	46	42	22.68319059	83.48368271	6.604903475	194.2651719	rice	rice	56.30%	jute	41.27%
98	44	21	25.77125155	74.0849114	6.524478072	167.4031917	mango	cotton	34.16%	mango	33.27%
99	58	17	24.19858297	73.1312281	6.234330358	71.07542236	mango	mango	48.63%	cotton	41.92%
90	52	25	25.97482359	68.36385721	6.822588548	103.2234212	mango	mango	38.05%	cotton	20.42%
65	60	22	25.36783634	72.52054555	6.6088480800000015	167.9124111	mango	mango	37.10%	blackgram	22.73%
18	74	15	24.8035819	22.27512704	5.70830803	148.4727237	kidneybeans	pigeonpeas	54.51%	kidneybeans	45.20%
35	38	28	28.38588982	82.47742011	5.781702819	86.85428653	pigeonpeas	pigeonpeas	41.88%	blackgram	35.78%
40	62	19	27.32186828	34.13731227	4.697726704	98.51524628	pigeonpeas	kidneybeans	40.12%	pigeonpeas	38.18%

Fig. 4. Confidence output of Logistic Regression

Here in logistic regression, the first real or target variable is rice. The result for the first recommendation is also rice with a probability of 52.11%, whereas the second recommendation has a probability of 44.41%. This is not a good result for the second crop choice, as it carries a risk of loss and uncertainty. Among the linear models, the best performance is achieved by logistic regression in binary classification. However, these models may fail to perform well since the classification in this case is not binary, and the data exhibits obvious non-linearity with overlapping class correlations. This suboptimal performance is a result of these models' inability to capture the intricate interactions between features. [11]

Sample 1:
Predicted Primary Crop: muskmelon Confidence: 99.85%
2nd Best Crop: watermelon Confidence: 0.10%
Sample 2:
Predicted Primary Crop: watermelon Confidence: 99.72%
2nd Best Crop: muskmelon Confidence: 0.22%
Sample 3:
Predicted Primary Crop: papaya Confidence: 99.93%
2nd Best Crop: apple Confidence: 0.02%
Sample 4:
Predicted Primary Crop: papaya Confidence: 99.90%
2nd Best Crop: apple Confidence: 0.04%
Sample 5:
Predicted Primary Crop: apple Confidence: 99.88%
2nd Best Crop: papaya Confidence: 0.04%
Primary Crop Recommendation Accuracy: 98.86%

Fig. 5. Confidence analysis of CATBoost Regression

CATBoost predicts the probability of an event or class based on input features. But here in this study, CATBoost couldn't

give the best confidence for our second crop choice as it only focused on the primary crop and measured 2nd one including the first one too when building it.

Sample	Predicted Crop	1st Best Crop	1st Best Confidence	2nd Best Crop	2nd Best Confidence
1	muskmelon	muskmelon	99.51%	watermelon	96.53%
2	watermelon	watermelon	99.63%	mungbean	25.75%
3	papaya	papaya	99.73%	blackgram	32.56%
4	papaya	papaya	99.73%	blackgram	21.39%
5	apple	apple	99.57%	grapes	32.59%
6	mango	mango	99.69%	pigeonpeas	96.59%
7	apple	apple	99.50%	papaya	58.56%
8	mothbeans	mothbeans	99.65%	mungbean	17.19%
9	mungbean	mungbean	99.54%	mothbeans	17.59%
10	lentil	lentil	99.74%	blackgram	35.31%
11	blackgram	blackgram	99.79%	maize	21.23%
12	coconut	coconut	99.66%	coffee	42.60%
13	pomegranate	pomegranate	99.65%	watermelon	14.01%
14	jute	jute	99.40%	coffee	9.08%
15	coconut	coconut	99.21%	coffee	40.63%
16	pomegranate	pomegranate	98.71%	papaya	10.31%
17	apple	apple	99.52%	papaya	58.56%
18	maize	maize	99.70%	coffee	9.40%
19	papaya	papaya	99.01%	coconut	56.57%
20	muskmelon	muskmelon	99.52%	watermelon	96.62%
21	coffee	coffee	99.73%	mango	36.38%
22	papaya	papaya	99.72%	blackgram	12.87%

Fig. 6. Confidence analysis of XGBoost Regression

In that sense, the XGBoost comparatively gave a lot better result for both crop recommendation as we can see. Some 2nd best confidence is above 50-80 % which refers better recommendations technique for example sample 1, 6, 7 etc.

Model Accuracy: 99.32%	
Top 2 Confidence Scores for the first 5 predictions:	
Prediction: muskmelon Confidence: 100.00%	
2nd Best Crop: watermelon Confidence: 0.00%	
Prediction: watermelon Confidence: 100.00%	
2nd Best Crop: rice Confidence: 0.00%	
Prediction: papaya Confidence: 99.00%	
2nd Best Crop: orange Confidence: 1.00%	
Prediction: papaya Confidence: 97.00%	
2nd Best Crop: orange Confidence: 2.00%	
Prediction: apple Confidence: 100.00%	
2nd Best Crop: rice Confidence: 0.00%	

Fig. 7. Confidence analysis of Random Forrest

Random forest classifier which has the least confidence for second crop as random forest uses a random sample of the data and a random subset of features at each split, it's more like a voting system voting for only the best one crop excluding any other. [12] Among the linear models, the best performance comes from logistic regression in binary classification. But as the data here are non-linear so, these models may fail to perform well since the classification here is not binary and the data shows obvious non-linearity with overlapping class correlations. This suboptimal performance is a result of these models' inability to grasp the intricate interactions between features. In that sense, XGBoost Classifier gave the nearest result or best two crop confidences to choose from.

V. RESULT & DISCUSSION

Significant differences among the performances of the models have resulted from this investigation. Of all, the XGBoost Classifier seems to perform highest in recommendation performance. Simpler models, for instance, Random Forest, CATBoost Classifier, and Logistic Regression, on the

other hand, did badly and achieved lesser accuracies. Real-world confidence recommendations for crop selection were less feasible with these models. These models' comparatively poor performance highlights their shortcomings in handling the dataset's non-linear connections and feature dependencies. Unlike the other models., XGBoost provides the crop options with much higher probability which can actually be perceived by farmers without risks and loss.

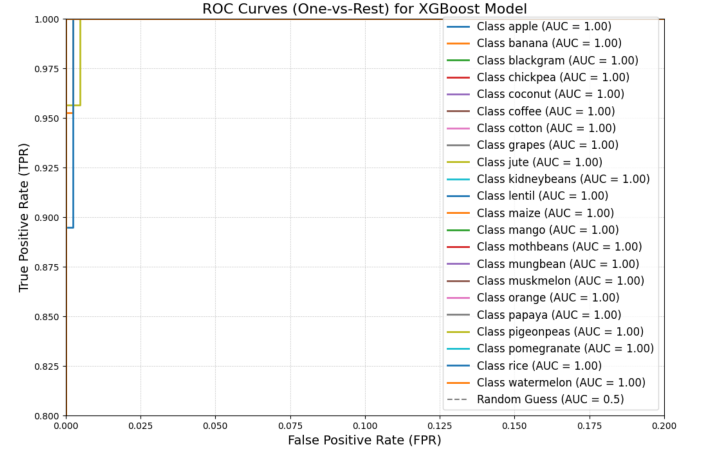


Fig. 8. ROC Curve for XGBoost Classifier

This ROC curve here visualizes the performance of the multi-class classification model using the one-vs-rest approach. Each curve represents the classification performance of the model for one crop class (e.g., apple, banana, rice, etc.), plotted as True Positive Rate (TPR) vs. False Positive Rate (FPR). For all classes (e.g., apple, banana, etc.), the AUC is 1.00, indicating that the model perfectly distinguishes each class from the others. An AUC of 1.00 means the model has no classification errors for these crops in the tested dataset. The curves are clustered at the top-left corner of the graph, suggesting near-perfect performance, with a high TPR and a very low FPR for all classes. The dashed line (AUC = 0.5) represents a random classifier, and all curves are far above it, confirming the model's reliability. The one-vs-rest method distinguishes one class (e.g., apple) against all other classes collectively, repeating this for each crop class. In summary, this ROC curve indicates that the XGBoost model is performing exceptionally well for predicting crop classes, as evidenced by the perfect AUC scores. It suggests the model can reliably differentiate between crops based on the input features.

Using numpy, pandas, sklearn under Django, a system interface was built for multiple crop recommendation which also shows the two best crop analysis according to this model. From this analysis, more models can be implemented this way for better recommendation testing. [13]

As for the limitation, the performance of models depends on the quality and completeness of data. It all depends basically on the data quality and completeness, which are inconsistent in some parts of the world; their accuracies therefore vary in

Crop Recommendation System

Nitrogen (N):

Phosphorus (P):

Potassium (K):

Temperature (°C):

Humidity (%):

pH Level:

Rainfall (mm):

Recommended Crops:

1st Best Crop: muskmelon (99.66%)

2nd Best Crop: watermelon (95.83%)

Fig. 9. App Deployment Interface for Multiple Recommendation

their prediction. Efficiency can hence be brought to a minimal since models could have been produced for specific regions with specific areas of climates and types of soils, uses differing highly from place to place in agriculture. For future works, more machine learning algorithms can be trained and tested for this purpose of this study. More precisely, deep learning-for instance, using CNNs or LSTMs-for feature extraction and prediction would model the complex relationships in these data that the more traditional machine learning models cannot grasp. IoT sensors and real-time weather information might be integrated in further research, thus being able to provide location-based dynamic recommendations of crops with greater precision and adaptability.

VI. CONCLUSION

The proposed model prioritizes important features like N (Nitrogen), P (Phosphorus), K (Potassium), temperature, humidity, pH, and rainfall that are crucial for predicting the soil and climate conditions of a particular region. Planting various crops considering these variables will be highly helpful for improving the soil conditions and increasing the economic revenue of a nation. The model is highly confident of the recommendations it has made to the farmers in terms of Predicted Crop, 1st Best Crop with Confidence, and 2nd Best Crop with Confidence for reliable insights. In this way, it is ensured that agricultural practice is data-driven and hence sustainable, helping farmers optimize crop yield and decision-making with multiple recommendation choices.

ACKNOWLEDGMENT

Gratitude to the resources and tools used in this work: providing insight from ChatGPT, data management by Quiltbot, and language accuracy by Grammarly. GeeksforGeeks, JavaT-point, and IBM provided technical information in developing

this. These tools contributed toward quality and precision in the work.

[14] [15] [16] [17] [18] [19]

REFERENCES

- [1] E. P. Kumar and R. Priyadarshini, "Enhanced precision multi-crop farming using machine learning recommendations," in *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, pp. 1–8, IEEE, 2024.
- [2] M. Kalimuthu, P. Vaishnavi, and M. Kishore, "Crop prediction using machine learning," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 926–932, 2020.
- [3] P. A. S. Chakraborty, A. Kumar, and O. R. Pooniwalla, "Intelligent crop recommendation system using machine learning," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 843–848, 2021.
- [4] P. S. Vijayabaskar, R. Sreemathi, and E. Keertanaa, "Crop prediction using predictive analytics," in *2017 International Conference on Computation of Power, Energy Information and Communication (ICCPEIC)*, pp. 370–373, 2017.
- [5] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavaya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," in *2016 Eighth International Conference on Advanced Computing (ICoAC)*, pp. 32–36, 2017.
- [6] E. Onoriode, N. Nwobodo-Nzeribe, and K. Chibueze, "Optimizing crop recommendation systems using machine learning algorithms," in *Zenodo*, 06 2024.
- [7] G. I. Aimufua, O. Morufu, M. U. Abdullahi, and U. F. Usman, "Stacking ensemble-based predictive system for crop recommendation," *FUDMA Journal of Sciences*, vol. 8, no. 6, pp. 72–83, 2024.
- [8] M. K. Senapaty, A. Ray, and N. Padhy, "A decision support system for crop recommendation using machine learning classification algorithms," *Agriculture*, vol. 14, no. 8, p. 1256, 2024.
- [9] Javatpoint, "Catboost in machine learning," <https://www.javatpoint.com/catboost-in-machine-learning>, 2025. Accessed: Jan. 23, 2025.
- [10] Geeksforgeeks, "Xgboost," <https://www.geeksforgeeks.org/xgboost/>, 2025. Accessed: Jan. 23, 2025.
- [11] "Differences between linear and logistic regression," <https://www.geeksforgeeks.org/understanding-logistic-regression/differences-between-linear-and-logistic-regression>, 2025. Accessed: Jan. 23, 2025.
- [12] IBM, "What is random forest?," <https://www.ibm.com/topics/random-forest>: :text=Random Accessed: Jan. 24, 2025.
- [13] W. Ahsan, "Multi-crop-recommendation-live-showing-confidence," 2024. [Online]. Available: <https://github.com/WaseeAhsan/Multi-Crop-Recommendation-Live-Showing-Confidence-.git>. [Accessed: Jan. 25, 2025].
- [14] P. S. Bharathi, V. Amudha, G. Ramkumar, T. Nagalakshmi, N. Nalini, and P. Jagadeesh, "An experimental analysis of crop yield prediction using modified deep learning strategy," in *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, pp. 1–6, IEEE, 2022.
- [15] S. Ramadasan, M. Ramadasan, K. Vijayakumar, and G. Balaji, "Classification of rice-plant images into healthy/disease class using resnetv2 variants," in *2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICES)*, pp. 1–4, IEEE, 2023.
- [16] B. Kashyap and R. Kumar, "Sensing methodologies in agriculture for soil moisture and nutrient monitoring," *IEEE Access*, vol. 9, pp. 14095–14121, 2021.
- [17] M. Ahmed, R. Hayat, M. Ahmad, M. Ul-Hassan, A. M. Kheir, F. Ul-Hassan, M. H. Ur-Rehman, F. A. Shaheen, M. A. Raza, and S. Ahmad, "Impact of climate change on dryland agricultural systems: A review of current status, potentials, and further work needed," *International Journal of Plant Production*, vol. 16, no. 3, pp. 341–363, 2022.
- [18] A. Shehadeh, O. Alshboul, R. E. Al Mamlook, and O. Hamedat, "Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, lightgbm, and xgboost regression," *Automation in Construction*, vol. 129, p. 103827, 2021.

[19] GeeksforGeeks, “Gradient boosting in ml.”
<https://www.geeksforgeeks.org/ml-gradient-boosting/>, Sept. 2024.
Accessed: Sep. 29, 2024.