International Journal of Life Science and Agriculture Research

ISSN (Print): 2833-2091, ISSN (Online): 2833-2105

Volume 04 Issue 11 November 2025

DOI: https://doi.org/10.55677/ijlsar/V04I11Y2025-11

Impact Factor: 7.88 , Page No : 707-714

Crop Recommendation Model Using Machine Learning

Dr. Emmanuel N.[1], Gabriel N.[1], Gad R.[1], Bwiza D.[1], Manzi J. K.[1], Kennet C.[2]

[1]Carnegie Mellon University - Africa

[2]Kumva Insights Ltd

ABSTRACT

Rwanda's agriculture sector plays a critical role in the country's economy, but farmers often face challenges in selecting the most suitable crops for their fields due to recommendation practices that focus primarily on soil nutrient analysis and neglect important factors such as climate variability and geographic altitude. This study presents the development of a data-driven crop recommendation system designed to address these limitations by using various data sources including soil test results, historical and forecast weather patterns, and crop-specific requirements, with machine learning algorithms applied to generate customized crop recommendations. While XGBoost achieved the highest accuracy at 98.25%, the Random Forest model, which achieved an accuracy of 96.7%, was selected for deployment because its better balance between precision and recall helps minimize the risk of resource inefficiencies from incorrect recommendations, and its more evenly distributed feature importance enhances generalizability across diverse farming conditions. The final model was integrated into a scalable analytics platform to ensure usability and impact, demonstrating how combining advanced data analytics with machine learning can enhance decision-making in agriculture, support sustainable farming practices, and improve yields in Rwanda.

Published Online:
November 27, 2025

Corresponding Author:
Dr. Emmanuel N.

KEYWORDS: Nutrients, predictive modeling, seasonal data, crop yield.

INTRODUCTION

Agriculture remains a cornerstone of Rwanda's economy, employing approximately 64.5% of the population and contributing about a quarter of the national GDP (Qu & Hao, 2018). Smallholder farmers account for more than 80% of total agricultural production, underscoring their central role in ensuring food security and driving sustainable rural development (Musabanganji et al., 2019). Despite this importance, the sector continues to face persistent challenges, including limited access to modern technologies, fragmented agricultural data, and increasing climate variability. These issues constrain productivity and threaten the long-term resilience of farming systems, making it essential to develop innovative approaches that support informed agricultural decision-making (Cantore, 2012).

Globally, agriculture employed an estimated 892 million people over a quarter of the world's workforce in 2022, and agrifood systems supported more than 1.3 billion jobs in 2021 (FAO, 2024). Yet global agriculture is also burdened by challenges such as climate change, unpredictable weather patterns, and resource scarcity, which disproportionately impact smallholder farmers (Cui et al., 2018). Limited access to critical resources and modern technologies often exacerbates farmers' vulnerability to crop failures, reduced yields, and economic instability.

Advancements in data-driven technologies, especially machine learning (ML), present transformative opportunities for addressing these constraints. ML enables the integration and analysis of complex datasets including soil composition, environmental conditions, and multisource climate information to generate actionable insights that enhance decision-making in agriculture (Modi et al., 2021). ML-powered systems have proven capable of recommending appropriate crops, predicting yields, and improving resource efficiency through data-driven guidance (Tuyizere et al., 2024). However, many existing crop recommendation systems rely primarily on basic soil test inputs while overlooking essential factors such as climate variability, geographic altitude, and localized environmental dynamics (Modi et al., 2021). This narrow focus limits the accuracy and relevance of recommendations, particularly in regions characterized by diverse agroecological conditions like Rwanda.

To address these gaps, this research develops a machine learning–based crop recommendation system tailored to Rwanda's unique environmental landscape. The system integrates multiple datasets, including soil characteristics, altitude-based information, and both historical and forecast weather patterns, together with crop-specific requirements. By leveraging advanced ML algorithms within a scalable analytics platform, the system aims to generate localized and reliable crop recommendations that empower farmers, enhance productivity, and strengthen resilience to climate-related risks. Beyond immediate crop selection support, the platform aligns with broader efforts to modernize agriculture through data-driven approaches that promote sustainability, food security, and efficient resource use (Kabirigi et al., 2017; Kumar et al., 2021; Cyemezo et al., 2019).

MATERIALS AND METHODS

Modeling Approach

An Agile development methodology was adopted to support iterative implementation, continuous testing, and adaptive refinement of the crop recommendation system. This approach allowed feedback-driven improvements at each stage of the development cycle, ensuring that system performance aligned with user requirements and project objectives.

System Requirements

Python was used as the primary programming language due to its extensive ecosystem for data analysis and machine learning. Jupyter Notebooks facilitated exploratory analysis and model prototyping, while FastAPI enabled deployment and integration of the final model into the Kumva Insights analytics platform. GitHub was used for version control and collaborative code management. A summary of key libraries and packages is presented in Table 1.

TABLE 1: Libraries/Packages

Category	Library / Module	Purpose
Data Manipulation and Analysis	pandas, NumPy	Data manipulation and numerical operations
Visualization	seaborn, matplotlib.pyplot	Visualization of data and model performance
Preprocessing	LabelEncoder, StandardScaler, MinMaxScaler, SimpleImputer	Encoding, scaling, and imputing missing data
Statistical Analysis	scipy.stats	Performing statistical tests and analyses
Model Training and Evaluation	train_test_split, StratifiedShuffleSplit, StratifiedKFold, cross_val_score, accuracy_score, classification_report	Splitting data, cross-validation, and evaluating model performance
Machine Learning Models	RandomForestClassifier, DecisionTreeClassifier, SVC, XGBClassifier, GaussianNB	Machine learning algorithms for classification
Utility	datetime, joblib, json	Saving models, handling dates, and managing JSON data

Data Collection

To ensure a comprehensive and reliable data set, this project used both internal and external data sources:

1) **Kumva Insights Dataset:** The internal dataset included detailed soil, climate, and geographic data specific to Rwanda.
2) **External Sources:** Additional data, such as weather patterns and geographic features, was sourced from reliable platforms like Google API.

The data collected focused on two primary aspects:

1) **Crop Suitability Data:** Comprehensive information on features and conditions required for optimal crop growth.
2) **Regional Feature Values:** Data representing specific soil, climate, and environmental characteristics across Rwanda's sectors.

The data set incorporated 22 crop types, selected for their prominence in Rwandan agriculture, ensuring relevance and utility for local farmers.

Exploratory Data Analysis

Exploratory analysis was conducted to understand relationships among climatic variables, soil characteristics, and crop requirements. Temperature tolerance ranges were examined using box plots to differentiate heat-sensitive and heat-tolerant crops, which supports climate-based recommendation logic.

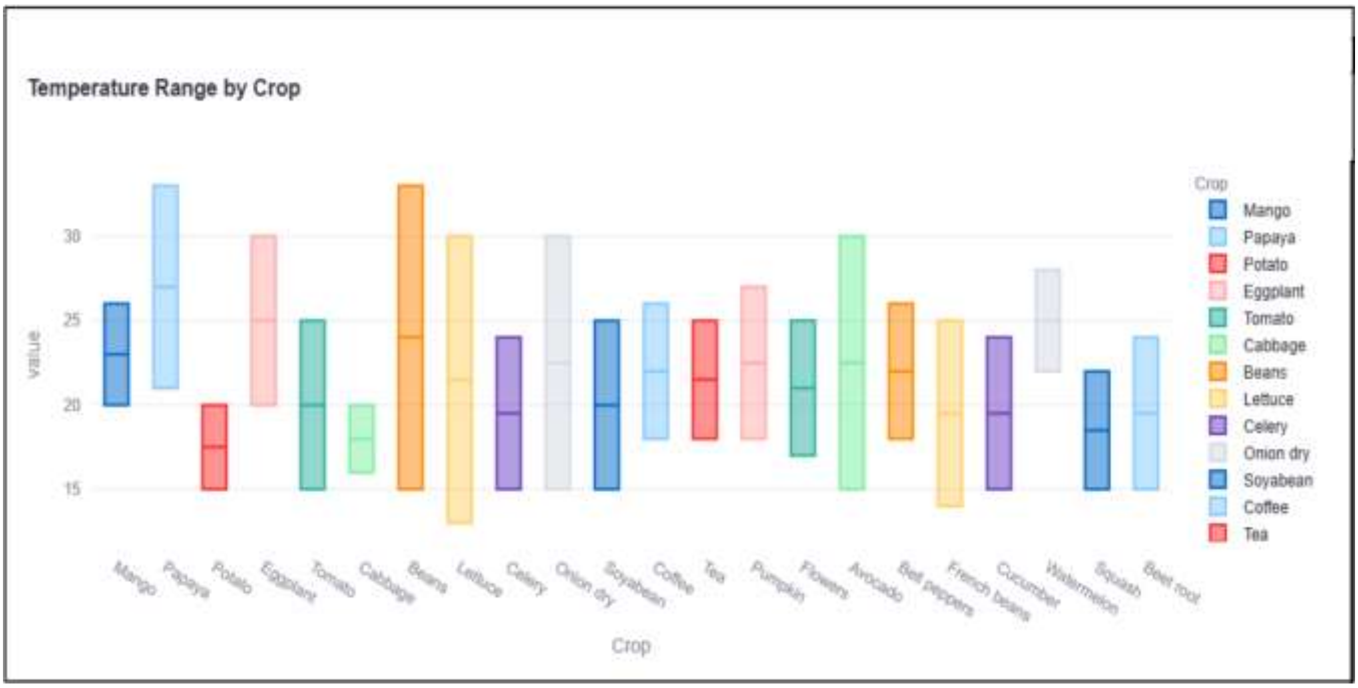


Fig. 1(Temperature range by crop)

A correlation matrix was also generated to assess interactions among key soil and climate variables (Figure 2). Additionally, a radar chart visualized nutrient and pH preferences across crops, helping identify edaphic similarities and clusters useful for classification (Figure 3).

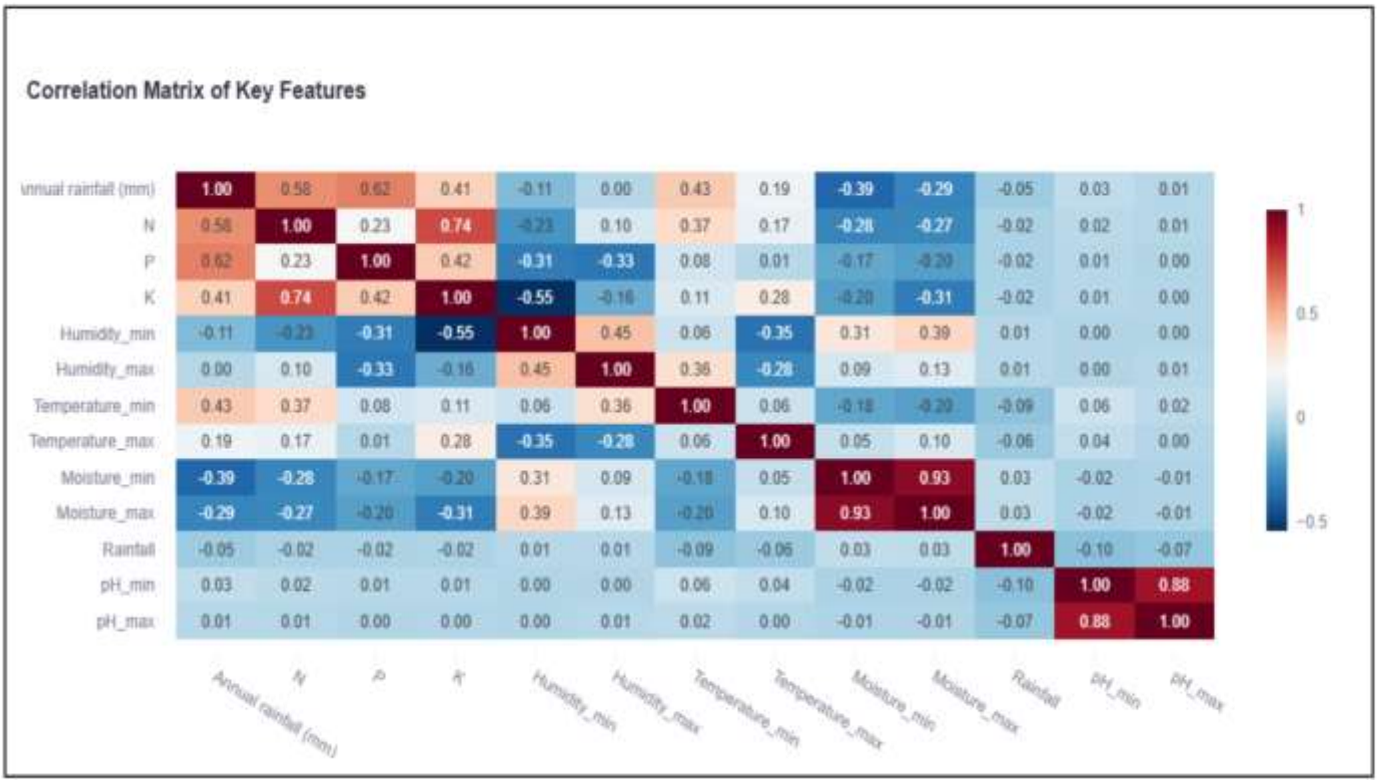


Fig. 2(Correlation matrix of key features)

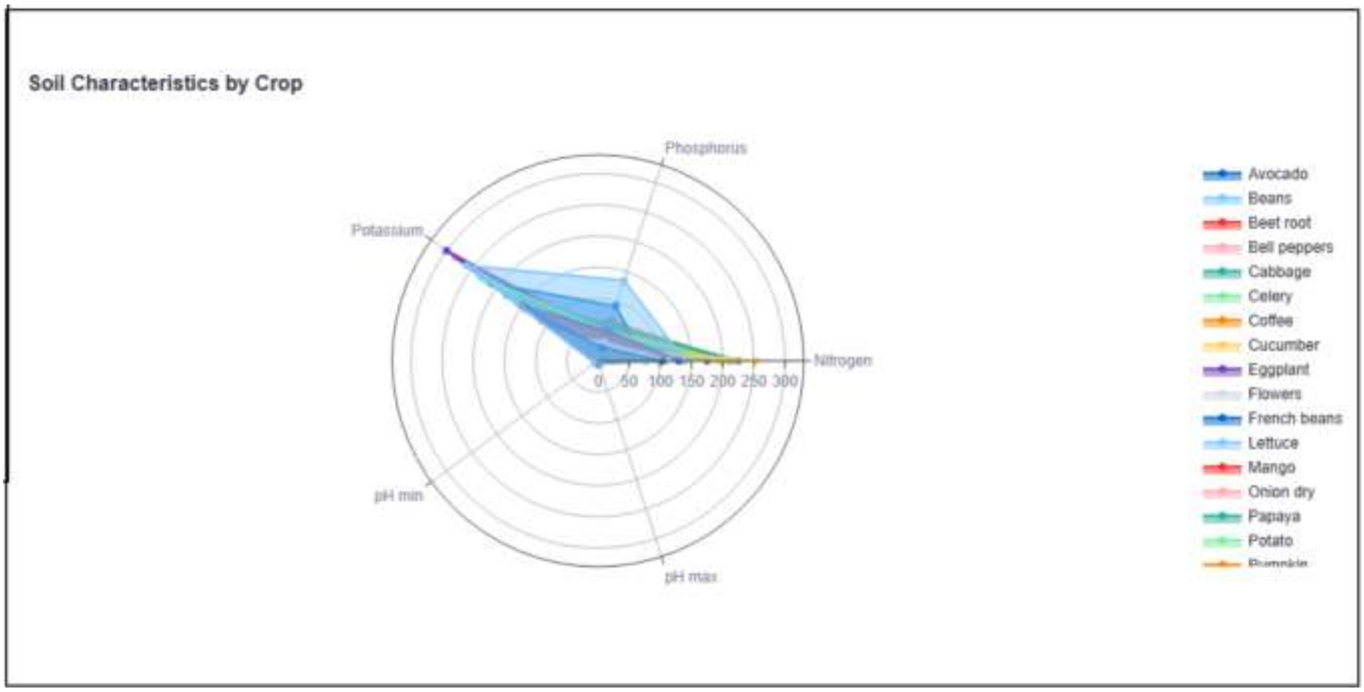


Fig. 3(Soil characteristics by crop)

Feature Selection and Importance Analysis

Features influencing crop suitability were grouped into four major categories:

1) **Crop-Specific Factors:** Crop type, growth duration, and water requirements.
2) **Soil and Environmental Factors:** Temperature, altitude, pH, humidity, and soil type.
3) **Soil Nutrients:** Concentrations of nitrogen (N), phosphorus (P), and potassium (K).
4) **Seasonal Variables:** Planting and harvesting months for Rwanda's Seasons A and B.

Feature importance was assessed using the Random Forest algorithm. The top ten influential features included growing period, nutrient levels (N and P), crop water needs, and temperature-related variables (Figure 4).

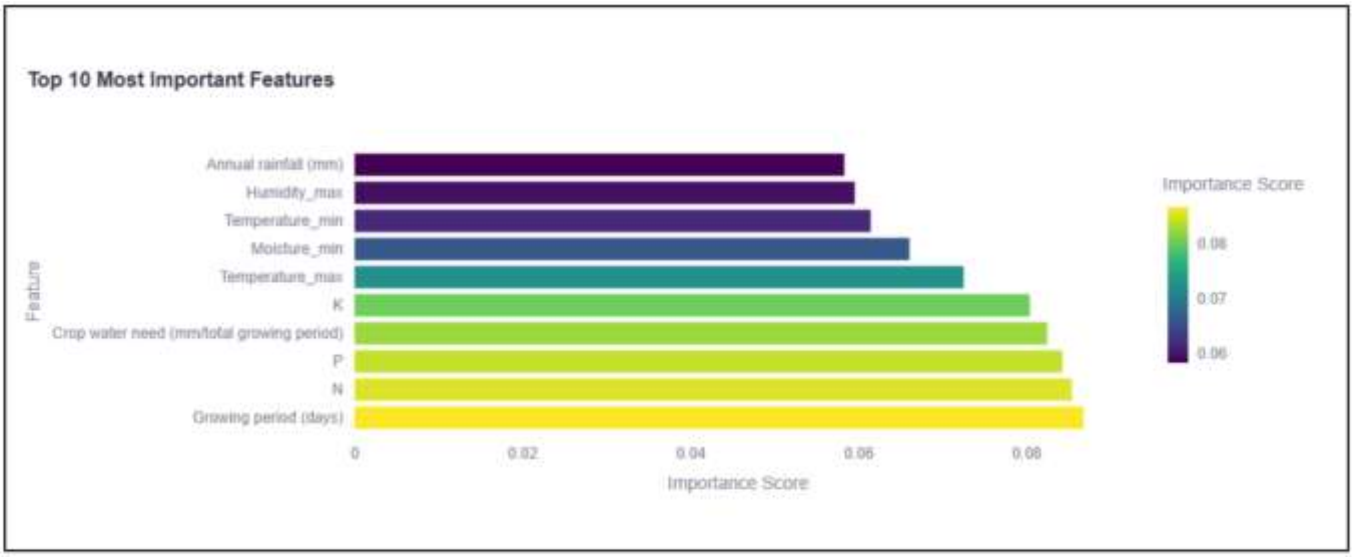


Fig. 4(Top 10 most important features)

Data Preprocessing

Raw data underwent several preprocessing steps to prepare them for modeling:

1) **Data Cleaning:** Removal of duplicates, correction of inconsistencies, and imputation of missing values.
2) **Encoding**: Conversion of categorical variables (e.g., soil type) into numerical formats.
3) **Scaling**: Standardization of numerical values such as temperature and nutrient levels.
4) **Feature Selection:** Retention of the most relevant variables based on importance scores and agronomic relevance.

Feature Engineering

Additional domain-specific features were engineered to improve model performance:

1) **Seasonal Timing Variables:** Optimal planting and harvesting windows.
2) **Altitude-Adjusted Climate Metrics:** Temperature and humidity adjusted for elevation differences.
3) **Soil Fertility Index:** Composite indicator derived from N, P, and K levels.

These engineered features improved the model's generalizability across Rwanda's diverse agroecological zones.

Model Architecture

The model architecture followed a structured pipeline that included data preprocessing, feature engineering, algorithm training, performance evaluation, and final deployment. A simplified architecture is shown in Figure 5.

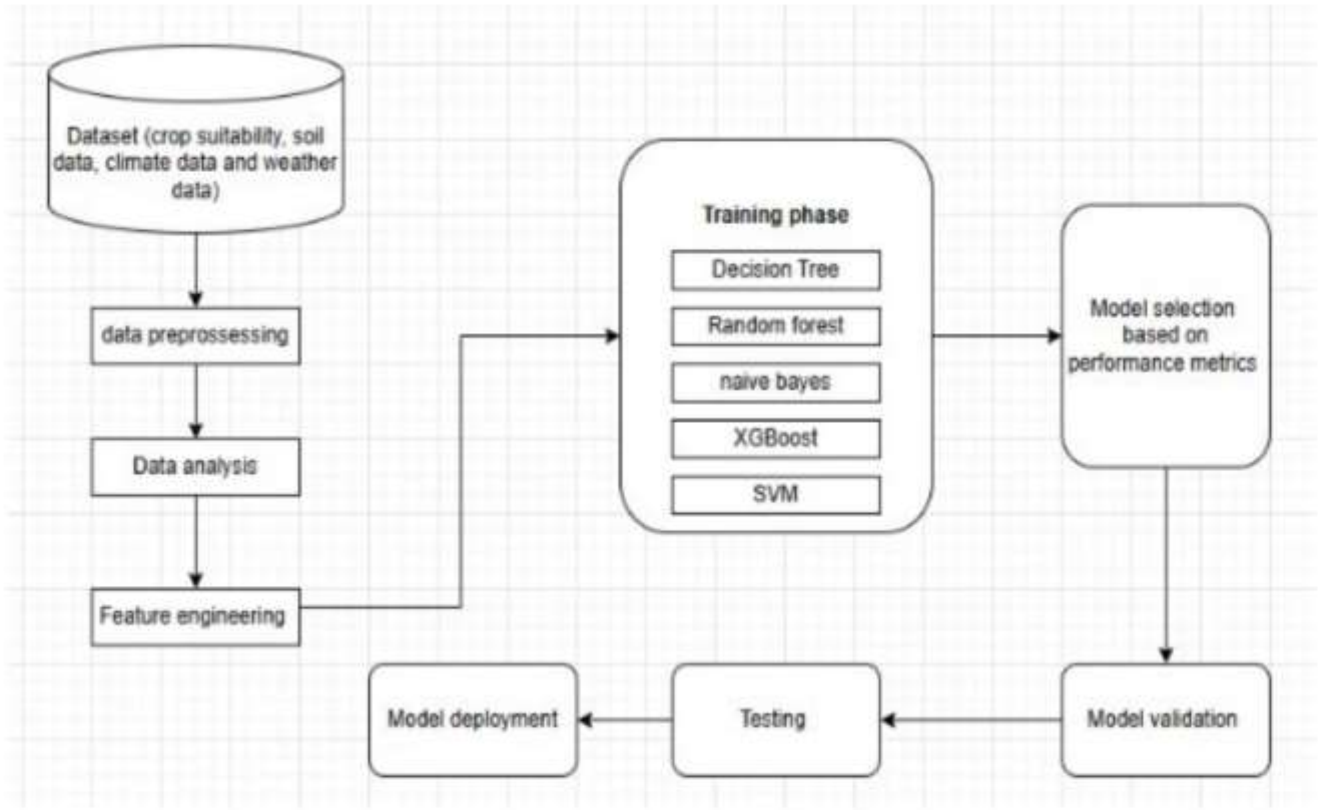


Fig. 5(Model Architecture of the Crop Recommendation System)

Model Training and Evaluation

Five machine learning algorithms were evaluated:

1) Decision Tree
2) Random Forest
3) Naïve Bayes
4) XGBoost
5) Support Vector Machine (SVM)

RESULTS

The crop recommendation model was evaluated using accuracy, precision, recall, and F1-score across several machine learning algorithms. Model performance was assessed on a dataset containing soil, climate, and environmental attributes relevant to crop selection in Rwanda.

Model Performance Comparison

The effectiveness of each algorithm was analyzed to identify the best-performing model for crop suitability prediction. Table 2 summarizes the accuracy and defining characteristics of all tested models.

Table 2. Model Performance Comparison

Model	*Characteristics*	*Accuracy (%)*
Naïve Bayes	Assumes feature independence; struggles with correlated attributes	85.3
XGBoost	Captures complex patterns; handles nonlinear relationships effectively	98.25
Random Forest	Robust ensemble of decision trees; reduces overfitting	96.7
Decision Tree	Simple and interpretable; prone to overfitting without pruning	91.2
SVM	Finds optimal decision boundary; sensitive to outliers	94.8

XGBoost achieved the highest accuracy (98.25%), reflecting its strong ability to model complex dependencies among environmental variables. Random Forest followed closely with 96.7%, benefiting from ensemble learning that stabilizes predictions. In contrast, simpler models such as Naïve Bayes and Decision Tree demonstrated lower accuracy, primarily due to feature correlation issues and susceptibility to overfitting.

Precision and Recall Analysis

Precision and recall comparisons reinforced the superior predictive strength of XGBoost. Its ability to model nonlinear interactions and identify subtle patterns contributed to high recall and reliable identification of suitable crops. However, XGBoost's sensitivity to noise occasionally resulted in borderline misclassifications, which may affect practical deployment.

Random Forest demonstrated a more balanced trade-off between precision and recall, making it a reliable option for real-world use where incorrect recommendations can incur significant resource losses. Moreover, Random Forest produced more evenly distributed feature importance scores, reducing dependence on any single variable and improving generalizability across Rwanda's diverse agroecological zones.

Naïve Bayes performed less effectively due to its assumption of independent features, which is unsuitable for highly correlated agronomic data.

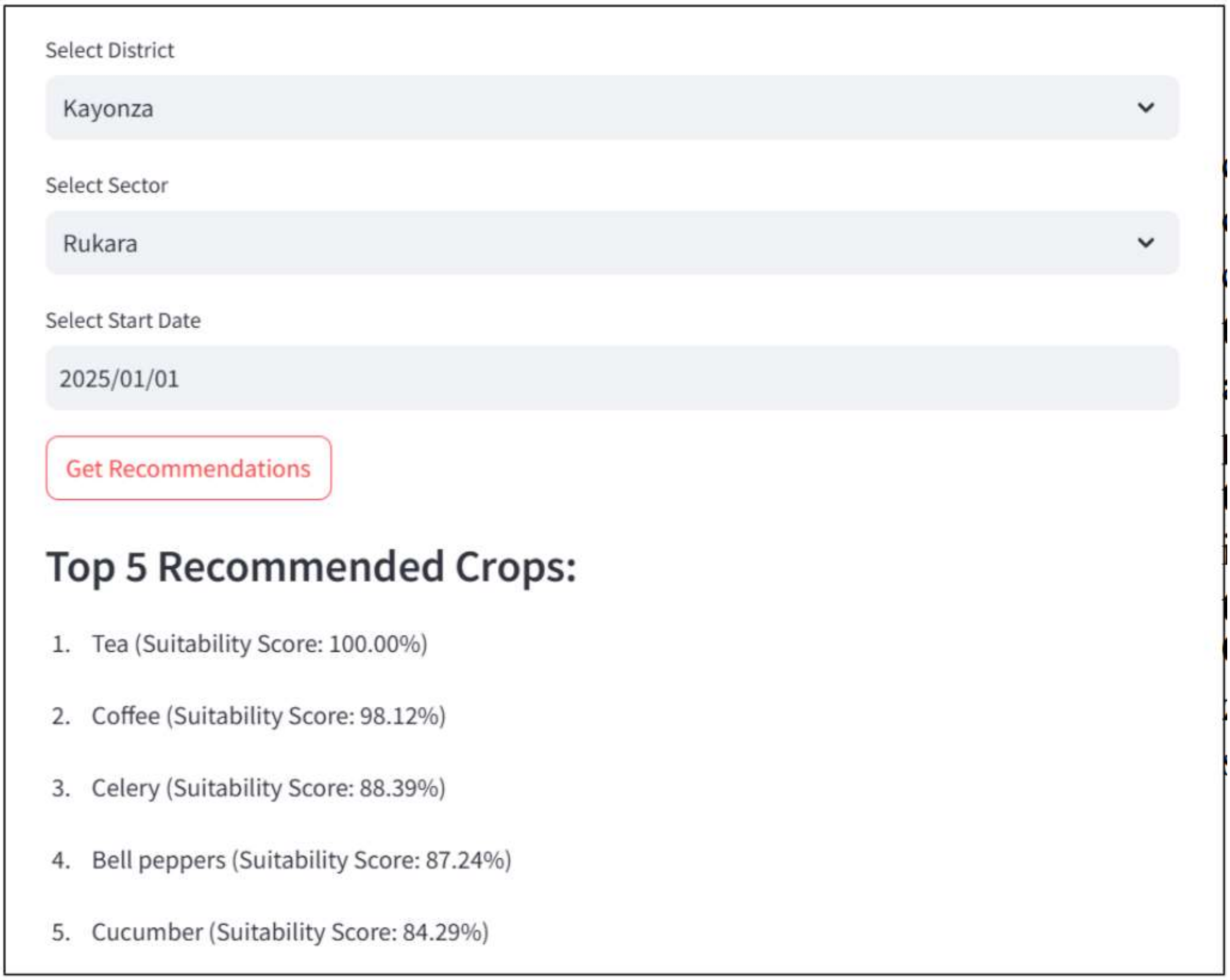
Random Forest's balance of interpretability, computational efficiency, and stability across datasets ultimately contributed to its selection for deployment.

User Interface

A user-friendly web-based interface was developed to demonstrate and test crop recommendations for farmers and agricultural stakeholders. Figure 6 illustrates the system interface and its prediction functionality.

The interface allows users to select their location through a hierarchical menu of Rwanda's administrative divisions (district and sector) and specify a planting date. This streamlined process enables the model to incorporate location-specific soil and climate information without requiring farmers to provide technical measurements. For example, when tested using Kayonza District, Mukarange Sector, with a planting date of January 1, 2025, the system generated suitability scores for multiple crops. Tea emerged as the top recommendation with a suitability score of 100%, followed by Coffee (98.35%) and Celery (88.60%). These outputs align with established agricultural patterns in Rwanda's Eastern Province, where tea and coffee are widely cultivated. The identification of celery as a suitable alternative highlights the system's capacity to recommend diversification options that may not be immediately evident to farmers.

Overall, the interface combines simplicity with analytical rigor, supporting practical decision-making even among users with limited technical expertise.



Select District: Kayonza

Select Sector: Rukara

Select Start Date: 2025/01/01

Top 5 Recommended Crops:

1. Tea (Suitability Score: 100.00%)
2. Coffee (Suitability Score: 98.12%)
3. Celery (Suitability Score: 88.39%)
4. Bell peppers (Suitability Score: 87.24%)
5. Cucumber (Suitability Score: 84.29%)

Fig. 6(User Interface)

DISCUSSION

Random Forest was selected for deployment due to its robustness, balanced precision-recall performance, and superior generalizability across varied environmental conditions. Although XGBoost achieved the highest accuracy, Random Forest offered more efficient performance for real-time, scalable deployment. XGBoost demonstrated the strongest predictive capability (98.25% accuracy) but was more computationally intensive and sensitive to noisy inputs, decreasing its suitability for operational use in resource-constrained environments. Naïve Bayes recorded the lowest accuracy (85.3%) because its assumption of feature independence does not hold for agronomic datasets, which contain correlated soil and climate variables. SVM and Decision Tree performed moderately well, but were outperformed by ensemble models. Decision Tree was particularly prone to overfitting, while SVM showed sensitivity to outliers. Overall, the results indicate that while advanced models such as XGBoost excel in accuracy, Random Forest presents the optimal balance of performance, interpretability, speed, and deployment feasibility for Rwanda's crop recommendation needs.

CONCLUSION

This study demonstrates the effectiveness of machine learning in optimizing crop recommendations for Rwanda. By integrating environmental, soil, and climatic factors, the system provides data-driven insights for smallholder farmers. Random Forest was selected for deployment due to its strong performance and interpretability, while simpler models struggled with correlated features. The model enables realtime, location-specific recommendations, improving accessibility for farmers. This research supports climate-resilient farming, enhances yield optimization, and strengthens food security, highlighting AI's potential in addressing agricultural challenges.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the staff and researchers at Kumva Insights Ltd. for providing access to critical datasets and technical support during the development of the crop recommendation system. Special thanks are extended to

Carnegie Mellon University Africa (CMU-Africa) for facilitating the collaboration and providing guidance throughout the project. The authors also appreciate the constructive feedback and mentorship from colleagues and professors, which greatly enhanced the study design, model evaluation, and overall quality of the work. Finally, heartfelt thanks go to the local farmers and agricultural advisors in Rwanda who shared their practical insights into regional crop practices, enabling the system to deliver accurate and contextually relevant recommendations.

DISCLOSURE

It is reported that no conflicts of interest exist in this work. No financial support or personal relationships influenced the study design, data collection, analysis, interpretation, or the reporting of results.

REFERENCES

1. C. Qu and X. Hao, "Agriculture drought and food security monitoring over the horn of africa (hoa) from space," 2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics), 08 2018.
2. "Employment indicators 2000–2022 (october 2024 update)," Statistics, 10 2024. [Online]. Available: https://www.fao.org/statistics/highlights-archive/highlights-detail/ employment-indicators-2000-2022-%28september-2024-update%29/en
3. Z. Cui and et al., "Pursuing sustainable productivity with millions of smallholder farmers," Nature, vol. 555, pp. 363–366, 2018.
4. D. Modi, A. V. Sutagundar, V. Yalavigi, and A. Aravatagimath, "Crop recommendation using machine learning algorithm," in 2021 5th International Conference on Information Systems and Computer Networks (ISCON). IEEE, 2021, pp. 1–5.
5. D. TuYizere, V. Uwase, M. Niyonkuru, G. Ndanyunzwe, M. Kabutware, P. Singadi, R. Uwera, and G. Okeyo, "Ai-driven precision farming: A holistic approach to enhance food and nutrition security in africa," pp. 1–6, 07 2024. [Online]. Available: https://ieeexplore.ieee.org/stamp/ stamp.jsp?tp=&arnumber=10622109
6. E. Musabanganji and Others, "Improving agricultural productivity in rwanda," Agroforestry Systems, pp. 112–125, 2019.
7. R. Kumar and Others, "Hybrid models for crop recommendation," Agricultural Computing, pp. 78–92, 2021.
8. A. Cyemezo and Others, "Time-series forecasting for crop yield prediction using multilayer perceptrons," International Journal of Scientific Research and Publications, vol. 9, no. 7, pp. 22–35, July 2019.
9. N. Cantore, "The crop intensification program in rwanda: a sustainability analysis." [Online]. Available: https://rema.gov.rw/rema doc/pei/Report ODI AM.pdf
10. M. Kabirigi, S. Mugambi, B. S. Musana, G. T. Ngoga, J. C. Muhutu, J. Rutebuka, V. M. Ruganzu, I. Nzeyimana, and N. L. Nabahungu, "Estimation of soil erosion risk, its valuation and economic implications for agricultural production in western part of rwanda," Journal of Experimental Biology and Agricultural Sciences, vol. 5, pp. 525–536, 09 2017.
11. H. Amit, "Xgboost vs linear regression: A practical guide." [Online]. Available: https://medium.com/@heyamit10/ xgboost-vs-linear-regression-a-practical-guide-aa09a68af12b
12. Sruthi, "Understanding random forest algorithm with examples." [Online]. Available: https://www.analyticsvidhya.com/blog/2021/06/ understanding-random-forest/