A FRAMEWORK FOR A CROP YIELD PREDICTION MODEL BASED ON DECISION TREE AND MIN-MAX SCALING

S.A. Audu, A.A. Rabiu, G.I.O. Aimufua, M. Samaila, A.M. Umar, and M.A. Yau
[1,2,3,5,6]Department of Computer Science, Nasarawa State University, Keffi, Nigeria
[4]Department of Computer Science, Federal University Gusau, Zamfara, Nigeria
e-mail: samailaagp@gmail.com; +234 8065670727

ABSTRACT

The prediction of crop yield has become critical for enhancing global food security. This study presents a framework for agricultural yield prediction, employing a decision tree-based mathematical model integrated with Min-Max scaling. A dataset comprising 28,241 entries was collected from the Food and Agriculture Organisation (FAO) and the World Data Bank (WDB). Variables include nation, crop type, year, rainfall, pesticide usage, and temperature. The model achieved an accuracy of 97.8%, demonstrating that crop variety significantly influences agricultural output, while temperature and pesticide usage impact yield more than rainfall. The findings provide actionable insights into optimizing agricultural productivity by identifying key influencing factors. This research offers a robust decision-making tool for stakeholders aiming to enhance food security and agricultural efficiency.

Keywords: Crop yield, decision tree, prediction framework, Min-Max scaling, agricultural productivity.

INTRODUCTION

The agricultural sector faces numerous challenges, including climate variability, environmental degradation, labor costs, resource limitations, and conflicts such as herdsmen-farmer clashes and banditry. Additionally, the need to sustain agricultural practices while preserving the environment complicates efforts to meet food demands for growing populations. These challenges underscore the importance of technological innovations for improving crop productivity without compromising quality (Mittal et al., 2020).

Crop yield forecasting is vital for achieving food security. It supports decision-making for farmers, industries, and governments by estimating production levels based on various factors, such as soil properties, fertilizer use, irrigation management, and climatic variables like temperature and rainfall (Khaki et al., 2020). Optimizing these factors, coupled with effective policies, can significantly enhance agricultural productivity.

This study focuses on yield management, which integrates all agricultural processes into the final productivity outcome. By leveraging predictive modeling, stakeholders can improve agricultural efficiency, meet growing food demands, and inform government decisions regarding food imports (Juvanna et al., 2021). This research aims to develop a framework for crop yield prediction, emphasizing the role of climatic and agronomic variables. Beyond a shadow of a doubt, predicting crop yield is another way of increasing the productivity of agricultural products to meet the growing demand for food and to advise the government on the amount of food to be imported based on the estimated yield as outlined in the study by (Juvanna et al., 2021)

MATERIALS AND METHODS

The crop yield prediction framework integrates climatic, agronomic, and other variables (Figure 1). Key variables include temperature, precipitation, seed variety, soil water content, and soil fertility rate.

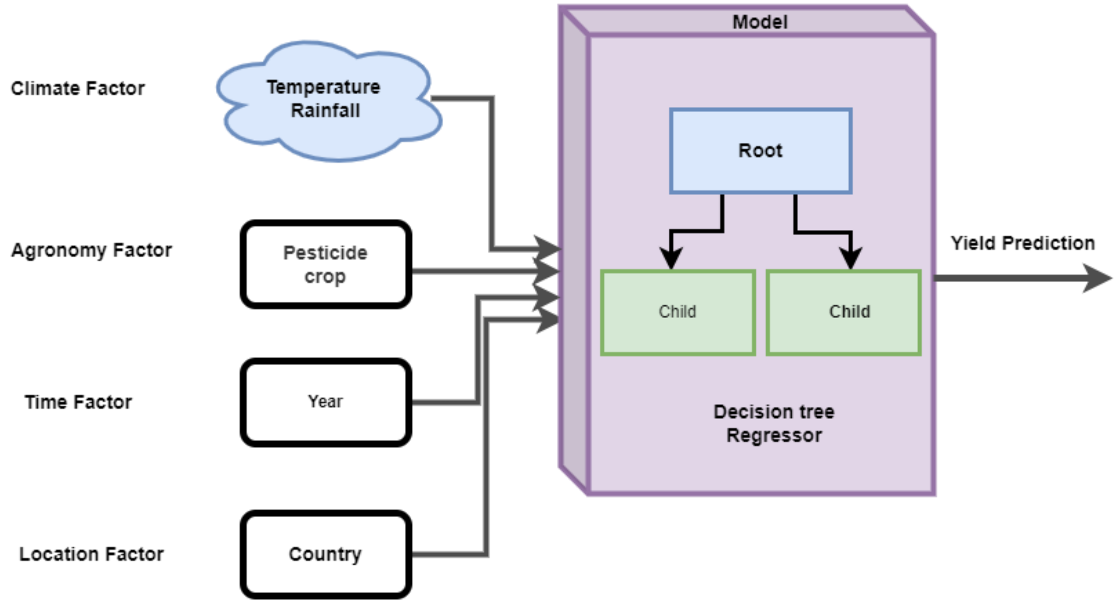

Figure 1. Crop Yield Prediction Framework.

Climatic Factors

Temperature and precipitation are pivotal to crop yield. Studies reveal that extreme climatic conditions, such as high temperatures and limited precipitation, adversely affect agricultural productivity (Beillouin et al., 2020; Guo et al., 2021).

Agronomic Factors

Pesticides play a critical role in mitigating crop losses. Research by Tudi et al. (2021) highlights that pesticide use prevents substantial losses in global fruit, vegetable, and grain production. Thus, pesticides significantly contribute to higher agricultural yields by minimizing the impact of pests and diseases.

Other Variables

Additional factors include the type of crop, the country of cultivation, and historical yield data. These variables are critical inputs for the proposed framework.

Decision Tree-Based Crop Yield Model Development

The proposed system architecture (Figure 2) outlines the stages from data collection to model training. Data preprocessing includes normalization using Min-Max scaling to ensure uniformity. The normalized dataset is split into training and testing subsets, with the decision tree (DT) regressor trained on the former. The model predicts crop yield using the trained DT algorithm.

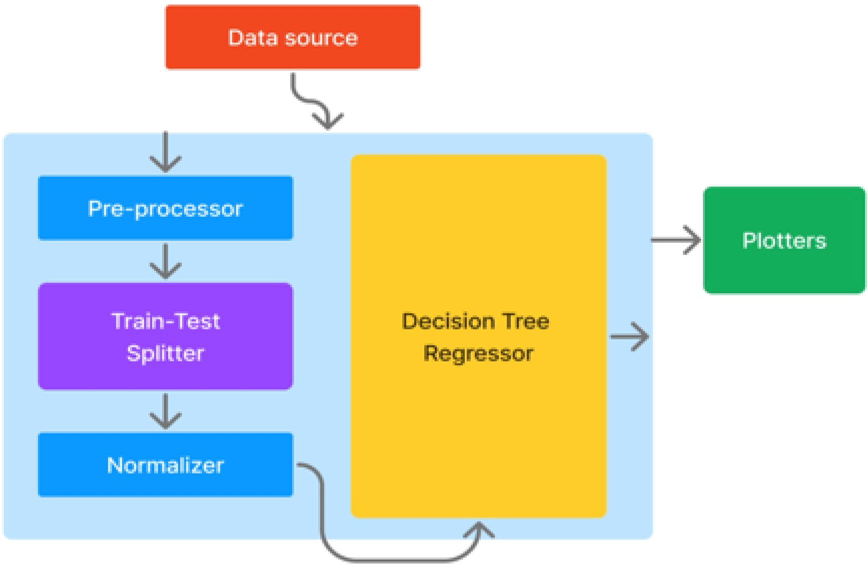

Figure 2. System Architecture for Decision Tree based Crop Prediction.

RESULTS AND DISCUSSION

The model's performance was evaluated using training and validation scores, which demonstrated its accuracy and generalization capability (Figure 3). Variable importance was assessed using the regressor feature_importances_ function, revealing that crop type is the most significant predictor of yield. Temperature and pesticide usage follow closely, with rainfall contributing less than expected (Figure 4).

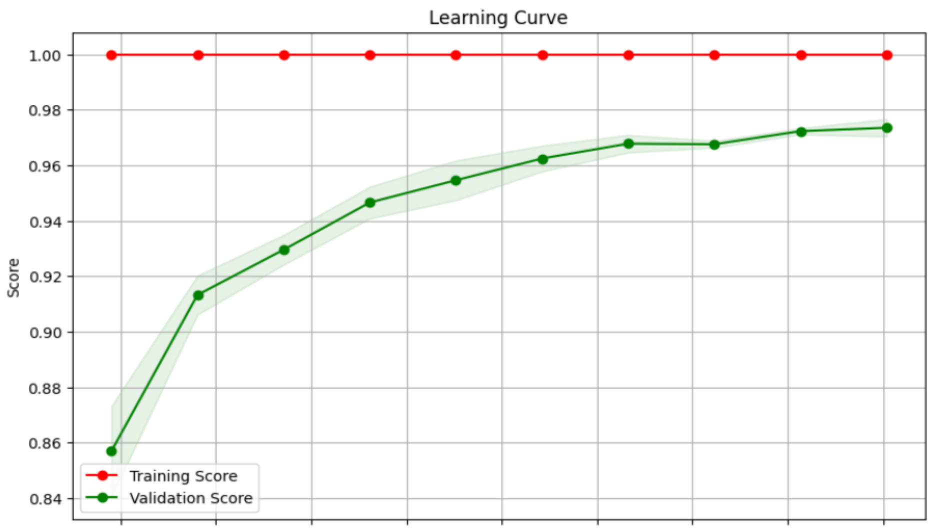

Figure 3. The Validation and the Learning Curve

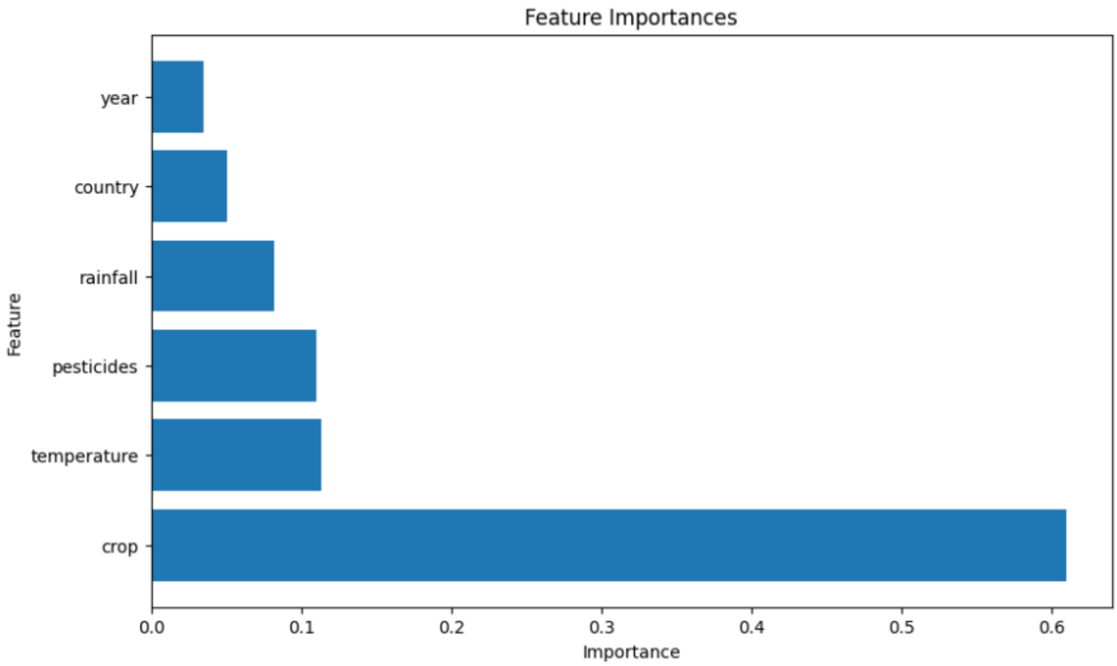

Figure 4. Bar Chart of order of Importance of the Features.

The results underscore the importance of integrating agronomic and climatic factors in predictive models. The high accuracy (97.8%) indicates the model's potential as a decision-support tool for policymakers and farmers to enhance agricultural productivity.

CONCLUSION

This study developed a decision tree-based crop yield prediction framework incorporating Min-Max scaling. The findings highlight the dominant role of crop variety, temperature, and pesticide usage in determining yield. The framework provides a reliable tool for stakeholders to optimize agricultural output and address food security challenges. Future research could explore integrating additional variables, such as soil texture and socio-economic factors, to further refine predictive accuracy.

REFERENCES

Beillouin, D.; Schauberger, B.; Bastos, A.; Ciais, P.; and Makowski, D. (2020). Impact of extreme weather conditions on European crop production in 2018: Random forest - Yield anomalies. Philosophical Transactions of the Royal Society B: Biological Sciences, 375(1810). https://doi.org/10.1098/rstb.2019.0510

Cunha, R. L. D. F.; and Silva, B. (2020). Estimating Crop Yields with Remote Sensing and Deep Learning. 2020 IEEE Latin American GRSS and ISPRS Remote Sensing Conference, LAGIRS 2020 - Proceedings, 273–278. https://doi.org/10.1109/LAGIRS48042.2020.9165608

Guo, Y. ; Xiang, H. ; Li, Z. ; Ma, F. ; and Du, C. (2021). Prediction of rice yield in east China based on climate and agronomic traits data using artificial neural networks and partial least squares regression. Agronomy, 11(2). https://doi.org/10.3390/agronomy11020282

Juvanna, I.; Yuvesh, Balaji V. G.; Sri Raam, M. A.; and Karthikeyan, T. (2021). A Novel Framework Crop Yield Prediction Using Data Mining. Turkish Journal of Computer and Mathematics Education, 12(10), 350–356.

Kavitha, M.; Anusha, K.; Santhi, M. V. B. T.; Rajesh, N.; and Choragudi, F. (2020). Predicting Crop Yield using Neural Network with Optimal. 32(2), 3028–3036.

Khaki, S.; Wang, L.; and Archontoulis, S. V. (2020). A CNN-RNN Framework for Crop Yield Prediction. Frontiers in Plant Science, 10(January), 1–14. https://doi.org/10.3389/fpls.2019.01750

Martinez-Ferrer, L.; Piles, M.; and Camps-Valls, G. (2020). Crop Yield Estimation and Interpretability With Gaussian Processes. IEEE Geoscience and Remote Sensing Letters, XX(1), 1–5. https://doi.org/10.1109/lgrs.2020.3016140

Mittal, D.; Kaur, G.; Singh, P.; Yadav, K.; and Ali, S. A. (2020). Nanoparticle-Based Sustainable Agriculture and Food Science: Recent Advances and Future Outlook. Frontiers in Nanotechnology, 2(December). https://doi.org/10.3389/fnano.2020.579954

Nishant P. S.; Venkat P.S,; Avinash B. L.; and Jabber B., ”Crop Yield Prediction based on Indian Agriculture using Machine Learning,” 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-4, DOI: 10.1109/INCET49848.2020.9154036

Nosratabadi, S.; Ardabili, S.; Lakner, Z.; Mako, C.; and Mosavi, A. (2021). Prediction of food production using machine learning algorithms of multilayer perceptron and anfis. *Agriculture* (Switzerland), 11(5) pp. 1-13. https://doi.org/10.3390/agriculture11050408

Tudi, M.; Ruan, H. D.; Wang, L.; Lyu, J.; Sadler, R.; Connell, D.; and Chu, C. (2021). Agriculture Development, Pesticide Application and Its Impact on the Environment. PP. 1–23.