Prediction of Tomato Yield in Benin Using Machine Learning Models.

S. C. Ariane Houetohossou[1], **V. Ratheil Houndji** [2,1], **K. Valère Salako**[1], **Romain Glèlè Kakaï**[1]

[1] Laboratoire de Biomathématiques et d'Estimations Forestières, Faculty of Agronomic Sciences, University of Abomey-Calavi, 04 BP 1525 Cotonou, Republic of Benin
[2] Institut de Formation et de Recherche en Informatique, University of Abomey-Calavi, 01 BP 526 Cotonou, Republic of Benin

Abstract

Tomatoes is one of the most important vegetables worldwide. Growth and yield data from greenhouse experiments were used for yield prediction with ML models. XGBoost model performs better (R^2=0.9066, MSE=0.0050, MAE=0.0420, and RMSE=0.0710).

Introduction and Objectives

- Agriculture faces many challenges in Africa: Climate variability, unpredictability, Poor access to technological innovations, low yield etc. Classical statistical methods limit: inability to capture complex, non-linear interactions between variables. **ML** offers opportunities to model complex relationships in agriculture; **crop yield prediction**.
- The study aimed to predict tomato yield, compare the models and identify the key variables of the chosen model.

Methodology

Table1: Experimental Factors

Factors	Levels	Values
Climates	Climate 1:	Tmin:]24.424 ; 26.213[Tmax:]29.214 ; 32.282[RR:]1.331 ; 3.250[
	Climate 2:	Tmin:]24.424 ; 26.213[ET: $\leq$ 1.971 Umax:]90.036 ; 93.962[
	Climate 3:	Tmin:]24.424 ; 26.213[Umax:]90.036 ; 93.962[Sun:]4.664 ; 7.528[
Fertilizers	Treatment 1	25% (Organic), 75% (Mineral)
	Treatment 2	50% (Organic), 50% (Mineral)
	Treatment 3	75% (Organic), 25% (Mineral)
	Treatment 4	100% (Organic)
	Treatment 5	100% (Mineral)

Table2: Data Collected

Type	Parameters
Growth parameters	Plant height, stem diameter, number of leaves, branches, and flowers
Yield parameters	Number of fruits, fruit weight, fruit width
Model output	Total fruit weight (g)

Table3: Data Analysis

Data Preprocessing	Data Processing
Missing data: removed	**Classical ML models**: SVM, KNN, LR, LASSO, DT
Outliers: removed (IQR)	**Ensemble models**:AdaBoost, XGBoost, GBR, LightGBM, ERT, RF
Correlation: $\geq$ 0.8	
Trasformation Label encoder,	**Hyperparameter tuning**: Grid Search
Normalisation:Min-Max Scaler	**Evaluation metrics**: R^2, MAE, RMSE
Spliting: 80/20	**Variable importance**: SHAP

Results and discussion

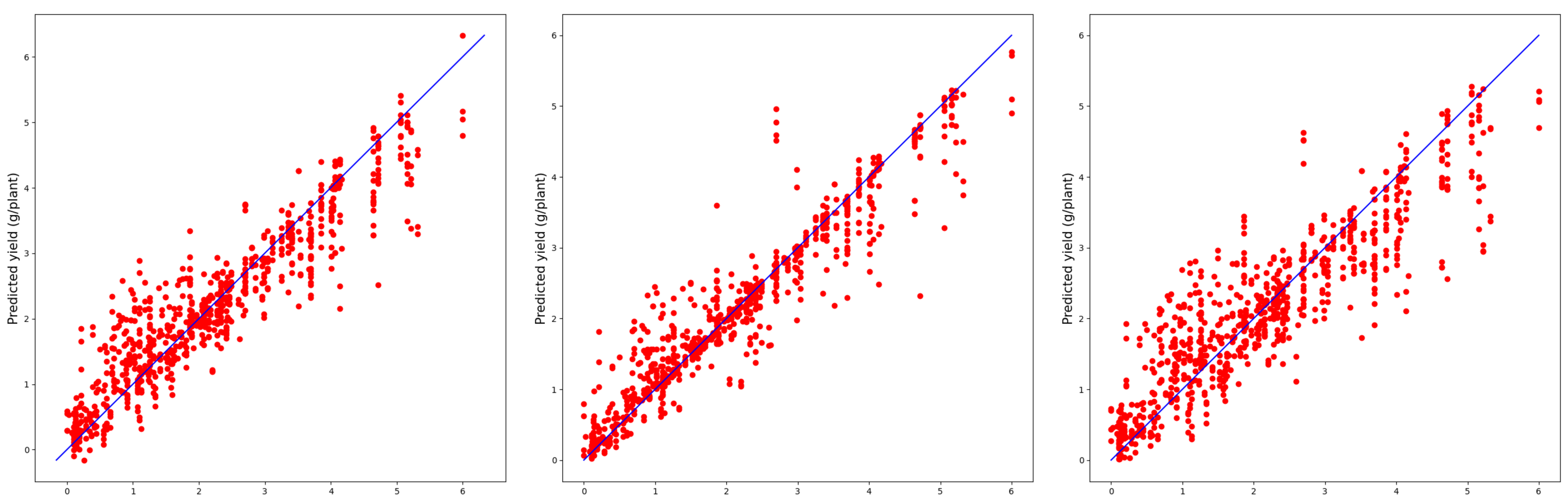
Models performance in tomato yield prediction

Table 1: Performances of the top 3 Models

Models	MSE	R^2	MAE	RMSE
GBR	0.0078	85.59	0.0636	0.0882
XGBoost	**0.0050**	**90.66**	**0.0420**	**0.0710**
Light GBM	0.0109	79.74	0.0766	0.1046

Ensemble models (XGBoost in particular) outperformed the other models. XGBoost combines boosting, regularization. It captures non-linear relationships and complex interactions between environmental and plant variables.

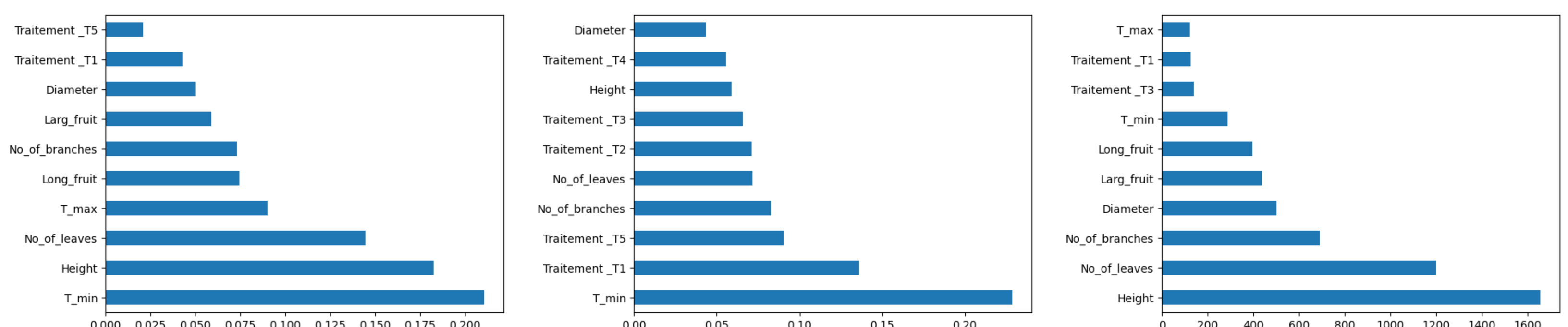
Model Predictions vs Actual Values



(a) GBoost (b) XGBoost (c) LightGBM

Figure 1. Comparison of predicted vs actual yield values

The model's performance aligns with findings from previous research: **Mariadass *et al.* (2022)** reported an R^2 of 0.98 for annual crop yield prediction in Malaysia using XGBoost, while **Li *et al.* (2023)** obtained an R^2 of 0.82 for soybean yield prediction in the USA with the same algorithm.



(a) GBoost (b) XGBoost (c) LightGBM

Figure 2. Important features.

Minimum temperature: key predictor across all models plays a crucial role in plant physiology by influencing photosynthesis, respiration, transpiration, and fruit development and ripening. **Treatment 1 (T1)**:a combination of 25% organic and 75% mineral fertilizer was highlighted by XGBoost as a significant factor. This balance optimizes nutrient availability and uptake.

Conclusions

XGBoost proved highly effective in predicting tomato yields in Benin. Most important variables: Minimum temperature and plant characteristics Future Applications encouraged to other crops.

References

[1] D. A.-L. Mariadass, E. G. Moung, M. M. Sufian, and A. Farzamnia. *Extreme gradient boosting (XGBoost) regressor and Shapley additive explanation for crop yield prediction in agriculture*. In *2022 12th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 219–224, 2022. doi:10.1109/ICCKE57176.2022.9960069.