JOSTIR Publication 2025 | ISSN: 1118-3896

Special Issue December 2025 | 306-317

Journal of Science, Technology and Innovation Research
Volume 1 Special Issue | December 2025

Development of A Weather-Based Predictive System for Optimizing Cassava and Cocoa Crops Yields for Farmers in Ondo State

Kinga M. T., Allo, O., Boyinbode O. K., and Fatoke, T. J.

Department of Information Technology, The Federal University of Technology, Akure, Ondo State, Nigeria

Correspondence: mtkinga@futa.edu.ng

ABSTRACT

Agriculture is a very crucial sector in the Nigerian economy, it provides food and income to millions of residents. However, the increasing unpredictability of climate regimes which is a direct effect of anthropogenic climate change has a highly negative impact on the total agricultural output in terms of crop yield. This paper aims at creating a predictive model based on weather, which maximizes the productivity of cassava and cocoa in Ondo State, Nigeria. Kaggle repository data on temporal crop yield were collected from 1991 to 2020 using meteorological measures provided by NASA POWER, this weather data was used to gain a calibration of forecasting models. XGBoost (Gradient Boosting) machine learning model was used to develop salient climatic factors like precipitation, temperature, and soil moisture. The findings show that the cassava model had moderate predictability ($R^2 = 0.61$) and the cocoa model had less predictability ($R^2 = 0.55$). The system seeks to provide the agricultural sector stakeholders with actionable knowledge, which will empower the stakeholders to make evidence-based decisions on how to advance agricultural production and strengthen agric-food systems during climatic disturbances. It follows that the developed predictive system have a significant contribution to sustainable agricultural development and food security enhancement across the region.

Keywords: agriculture, crop production, predictive model, XGBoost, Gradient Boosting.

Introduction

The agricultural sector forms a crucial part in world food security and supports the livelihood of millions of people, especially in the developing nations with a lower standard of living (Mukhlis *et al.,* 2022). Although it is a very important activity, farmers are often faced with challenges which hinder the achievement of maximum crop yields. The main issues are erratic weather changes, the limited availability of essential agricultural products and poor adoption of modern agricultural tools. These compounding elements contribute to low agricultural yields hence worsening poverty and food insecurity in rural areas (Gollin, 2021).

Of these obstacles, weather variability comes out as a key factor in the performance of crops as climate changes in terms of temperatures, precipitations, humidity and other peripheral climatic conditions do have extensive impacts on plant growth and their development (Huang *et al.,* 2023). The fact that it is not possible to predict weather patterns tends to increase the likelihood of lost yields which makes it difficult to institute effective agronomic practices by farmers. But then, improvement in weather prediction can radically transform the way

doi.org/10.51459/jostir.2025.1.Special-Issue.0228

people practice agriculture. Hence, this can provide farmers with a guide on when to plant cassava and cocoa, and what other factors should be considered when planting (Selz and Craig, 2023). Through the combination of deep weather records and state-of-the-art predictive systems, it becomes possible to create trusted forecasting systems that can condense the diverse meteorological data into a summary and provide brief actionable advice on how to manage cassava and cocoa crops.

Machine learning (ML) is a subtle branch of artificial intelligence that offers a powerful framework of deriving predictive information using large quantities of data (Ben-Bouallegue *et al.,* 2024). This research was conducted in Akure, a place with a tropical climate consisting of wet season running from April through October (peak in the months of June to September) and dry season, November to March, the yearly rainfall being in the range of 1,500-2,500mm. During the wet months the temperature of the air is usually about 27 °C on average, with the humidity occasionally rising to 85%. The motivation behind these warm and humid conditions is to design and deploy a weather-based predictive system that is driven by weather and that utilizes machine learning to streamline the yields of crops to farmers in the area, with cocoa as the key cash crop in the state.

Related Works

This study explores various studies that have employed machine learning techniques and climate data to enhance agricultural productivity, focusing on the methodologies, findings, and implications used for developing the weather-based predictive system that is aimed at optimizing crop yields for farmers in Ondo state. Ibrahim *et al.* (2024) predicted potato diseases in smallholder agricultural areas of Nigeria using machine learning and remote sensing-based climate data. It utilizes a Random Forest (RF) classifier combined with Multi-Criteria Classification (MCC) to predict potato disease vulnerability in Nigeria. The study highlighted that the absence of field-level management strategies, such as fungicide applications, could lead to significant errors in disease prediction. Similarly, Kuradusenge *et al.* (2023) focused on predicting maize and Irish potato yields in Rwanda using Random Forest models and weather data. It was stated that incorporating additional climatic variables like air humidity and solar radiation could enhance prediction accuracy, emphasizing the necessity of comprehensive data inclusion. Lontsi *et al.* (2022) developed various machine learning models, including Crop Multivariate Logistic Regression, Crop k-Nearest Neighbor, and Decision Tree, to predict crop yields using climatic and agricultural data. The work emphasized the effectiveness of these models, evidenced by metrics such as R^2 and Mean Absolute Error (MAE).

Further studies by Ahmed *et al.* (2023) built on this method and came up with a recommendation system to predict the crop yield in Southern Nigeria. These researches demonstrated the versatility of machine learning models in dealing with different agricultural problems in different geographical settings. An understanding of how socio-economic and climatic factors interact is a subtle concept requiring proper prediction of yields. This socio-economic and climatic factors interactions in Nigeria were studied by Tamuno-Opubo and Godwin (2023) with the help of ML algorithms, including k-Nearest neighbor (KNN), decision trees, and random forests, where additional validation of these algorithms in different climatic zones were performed to strengthen it predictive power and applicability. Despite all of these there is still complexity of the yield prediction problems in which the socio-economic and climatic factors play a central role. In Olisah *et al.,* (2024), a deep neural network was utilized to estimate the yield of corn in Enugu and Plateau states, with the weather variables and soil variables being addressed. The model was effective to identify important yield-implicating factors but found out that it is not very adaptive to unpredictable environmental changes, which is a significant gap in present ML models indicating a strong necessity to have more flexible and resilient predictive systems that can adapt to dynamic environmental conditions.

Hence, the existing gaps create possibilities to work on creating more comprehensive, adaptable, and practical predictive models that can positively influence the farmers to optimize crop yields. The literature reviewed underscores how machine-learning methods and climate statistics can be used to improve crop productivity. Nevertheless, it also points out the necessity of profound data integration, model flexibility and convenient interfaces to unite the predictive technologies and realistic farm applications. By addressing these gaps, this work creates a strong weather-driven predictive system that utilizes an efficient machine-learning methods to deliver precise and practical information to farmers on how best to cultivate cassava and cocoa, thereby enhancing agricultural productivity and resilience.

Methodology

The system architecture shown in Figure 1 shows all the processes involved in the development and training of the model.

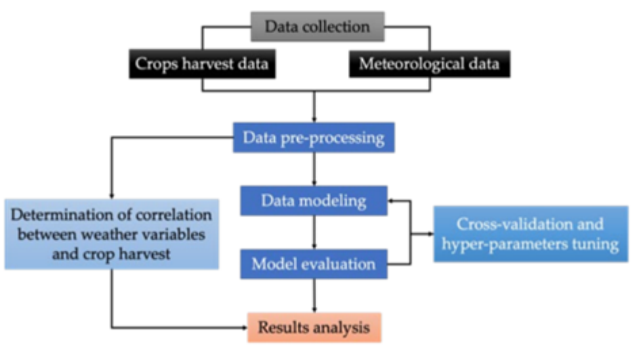

Figure 1. System Architecture

The system was developed in the form of a web-based application to serve as an online solution that is easily accessible to the farmers without having to install specific software. This architecture comprises of front-end interface and backend server.

Front-end Interface: An easy-to-use web interface, created with HTML, CSS (tailwind CSS), and JavaScript. It receives user's inputs including weather parameters and then shows the forecasted crop yields.

Back-end Server: This is a Flask-based server that processes data and makes predictions by communicating data between the front-end and the machine-learning models.

Dataset

The data used in this study has the necessary characteristics has gotten from Kaggle (http://www.kaggle.com/datasets/adekunlejoseph/ondo-state-nigeria-cocoa-cassava-and-weather-data?resource=dwonload&select=Rainfall.txt). The dataset reflects the complex interplay of climatic and crop health, supported with agronomic principles to emphasize the critical roles to the crop development and yield performance. Table 1 and Table 2 give the data of combined yield dataset and weather dataset of cassava and cocoa, respectively where results of the year, production, area, yield and weather parameters are displayed.

Data Splitting

The data were divided into training (80%) and testing (20%) sets. This is to ensure that the model is more

Table 1: Cassava Dataset

[71]:

	Year	PRODUCTION	AREA	YIELD	YEAR	CLOUD_AMT	GWETPROF	GWETROOT	GWETTOP	PRECTOTCORR	PRECTOTCORR_SUM	PS	QV2M	RH2M	T2M	T2M
0	1991	1106.2	81.34	13.599	1991	67.45	0.79	0.69	0.70	5.27	1524.02	98.76	16.97	86.44	24.72	
1	1992	1207.6	82.26	14.680	1992	65.26	0.82	0.71	0.73	5.27	1608.40	98.78	16.24	84.88	24.26	
2	1993	1310.2	87.37	14.995	1993	67.10	0.82	0.73	0.74	5.27	1835.16	98.76	16.72	86.12	24.50	
3	1994	1304.7	83.49	15.550	1994	66.79	0.87	0.77	0.78	5.27	1914.26	98.76	16.85	87.12	24.37	
4	1995	1354.3	81.54	16.609	1995	67.01	0.88	0.81	0.80	5.27	2199.02	98.73	17.09	87.31	24.62	

Table 1: Cassava Dataset

[73]: cocoa_weather

[73]:

	YEAR	PRODUCTION	AREA	YIELD	CLOUD_AMT	GWETPROF	GWETROOT	GWETTOP	PRECTOTCORR	PRECTOTCORR_SUM	PS	QV2M	RH2M	T2M	T2M_M/
0	1991	66.42	169.33	0.3920	67.45	0.79	0.69	0.70	5.27	1524.02	98.76	16.97	86.44	24.72	33.
1	1992	65.38	172.32	0.3790	65.26	0.82	0.71	0.73	5.27	1608.40	98.78	16.24	84.88	24.26	34.
2	1993	63.42	190.32	0.3330	67.10	0.82	0.73	0.74	5.27	1835.16	98.76	16.72	86.12	24.50	32.
3	1994	71.54	183.24	0.3900	66.79	0.87	0.77	0.78	5.27	1914.26	98.76	16.85	87.12	24.37	32.
4	1995	75.59	185.32	0.4080	67.01	0.88	0.81	0.80	5.27	2199.02	98.73	17.09	87.31	24.62	32.
5	1996	80.32	183.27	0.4380	67.52	0.88	0.78	0.78	5.27	1824.61	98.71	17.15	87.44	24.67	32.
6	1997	73.49	190.42	0.3860	66.74	0.87	0.77	0.77	5.27	1771.88	98.78	16.97	86.44	24.62	33.

efficient to predict accurate cassava and cocoa yields based on a particular weather condition. Although this study used an 80/20 split, other ratios can be used like 70/30 or 60/40 based on the project needs or availability of data.

Model Development

The model was trained over a training dataset which has been pre-prepared and the following hyper-parameters were optimised by grid search:

- Learning rate: cassava = 0.01, cocoa = 0.05
- Maximum depth: cassava = 5, cocoa = 2
- Number of estimators: 100

Five-fold cross-validation was used to test the model and as well avoid over-fitting ensuring a reliable and generalizable model of prediction.

The use of XGBoost as opposed to other machine-learning algorithms, such as linear regression, random forest, and neural networks, was due to the presence of better handling on structured data as well as its ability to handle missing values. The gradient-boosting framework is competent at capturing interactions, which are non-linear, which are vital in generating relationships between meteorological variables and crop performance.

Model Evaluation Metrics

Performance metrics used were Root-Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). According to Equation 1, MAE is defined as the average difference between the predicted value and observed value that is absolute.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}| \quad (1)$$

where y_i is the actual value, $\hat{y}$ *is the predicted value, and N is the number of observations*

The square root of the average squared differences between predicted and actual values is shown in

Equation 2 to generate the RMSE.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \quad (2)$$

R-squared (Coefficient of determination) as in Equation 3 is used to establish the extent to which the values predicted are very close to the data that was observed with the R^2 having a value of 0 to 1 indicating a better fit.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \quad (3)$$

These findings were contrasted with process-based crop models in order to determine the predictive potential and feasibility of the model developed.

Machine-Learning Models: Trained XGBoost regression models predict the cassava and cocoa yields, which are stored and accessed through the form of serialized files. The prediction system of crop yield pipeline is illustrated in Figure 2 as it is made up of four processes including the extraction-transformation-loading (ETL) process, the feature engineering process, the model training process, the evaluation process, and the model deployment process.To ensure a systematic way of development, the implementation process is broken into a number of systematic approach steps. This includes:

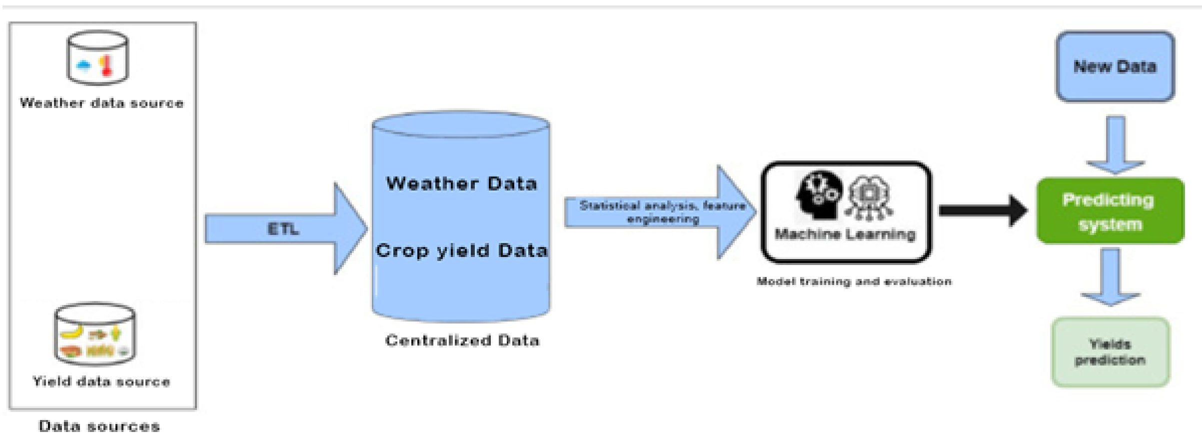

Figure 2. Pipeline overview of crops yield prediction system

Data Preprocessing and Data Integration: Figure 3 demonstrates the loading of agricultural data using Pandas.**Data Splitting**: The 80/20 split is used to create training and test sets. The training and testing

```
[8]: casava_data = pd.read_csv(r"C:\Users\OlorunfemiAllo\Downloads\Dataset for project\CASSAVA.csv")
     cocoa_data = pd.read_csv(r"C:\Users\OlorunfemiAllo\Downloads\Dataset for project\COCOA DATA.csv")
```

Figure 3. Loading the agricultural data with pandas

Data Transformation: Pivoting and merging weather information with crop yield information are presented in Figure 4.

Feature Engineering: The interaction terms and the poly features were developed to capture the non-linear relationships.

```
cassava_weather.head()
```

	Year	PRODUCTION	AREA	YIELD	YEAR	CLOUD_AMT	GWETPROF	GWETROOT	GWETTOP	PRECTOTCORR	PRECTOTCORR_SUM	PS	QV2M	RH2M	T2M	T2M
0	1991	1106.2	81.34	13.599	1991	67.45	0.79	0.69	0.70	5.27	1524.02	98.76	16.97	86.44	24.72	
1	1992	1207.6	82.26	14.680	1992	65.26	0.82	0.71	0.73	5.27	1608.40	98.78	16.24	84.88	24.26	
2	1993	1310.2	87.37	14.995	1993	67.10	0.82	0.73	0.74	5.27	1835.16	98.76	16.72	86.12	24.50	
3	1994	1304.7	83.49	15.550	1994	66.79	0.87	0.77	0.78	5.27	1914.26	98.76	16.85	87.12	24.37	
4	1995	1354.3	81.54	16.609	1995	67.01	0.88	0.81	0.80	5.27	2199.02	98.73	17.09	87.31	24.62	

```
cocoa_weather.head()
```

	YEAR	PRODUCTION	AREA	YIELD	CLOUD_AMT	GWETPROF	GWETROOT	GWETTOP	PRECTOTCORR	PRECTOTCORR_SUM	PS	QV2M	RH2M	T2M	T2M_MAX
0	1991	66.42	169.33	0.392	67.45	0.79	0.69	0.70	5.27	1524.02	98.76	16.97	86.44	24.72	33.87
1	1992	65.38	172.32	0.379	65.26	0.82	0.71	0.73	5.27	1608.40	98.78	16.24	84.88	24.26	34.68
2	1993	63.42	190.32	0.333	67.10	0.82	0.73	0.74	5.27	1835.16	98.76	16.72	86.12	24.50	32.92
3	1994	71.54	183.24	0.390	66.79	0.87	0.77	0.78	5.27	1914.26	98.76	16.85	87.12	24.37	32.65
4	1995	75.59	185.32	0.408	67.01	0.88	0.81	0.80	5.27	2199.02	98.73	17.09	87.31	24.62	32.33

Figure 4. Transformed dataset of both cocoa and cassava

of the model performance is illustrated by Figure 5.

Model Training and Evaluation

The XGBoost regression models used in the training of the models were optimally trained on cassava and cocoa with optimized hyper-parameters. A five-fold cross-validation was used in cross-validation to ensure that model performance is validated and overfitting is avoided. Hence, to determine the accuracy and reliability of performance measures, RMSE, MAE, and R- squared were used to measure the model performance.

For cassava yield prediction, the Mean Squared Error (MSE) of 5.8834 shows the average squared error of prediction, the R-squared of 0.609 shows that 60.9% of the yield variability is attributed or explained by the model and the MAE of 1.603 shows the average deviation of the prediction. By contrast, the cocoa model attained an MSE of 0.000951, indicating significantly smaller prediction errors, R-squared of 0.345, which explains a very large 34.5 percent of the variability of the yields and an MAE of 0.022. The above statistics implies that there is a high accuracy of prediction of the cocoa model. Together, the measures show that the cassava model has greater explanatory power, whereas the cocoa model offers more accuracy in the predictions, as shown in Figure 6.

Model Serialization

The trained models and the feature transformers in the form of polynomials are stored in persistent form using Joblib to be used later to make inferences in the form of Figure 7.

A Flask server was created that will receive the incoming requests and process the input data, make predictions, and provide the results. The identity of the API endpoints has been determined to send the home page request and prediction request as shown in Figure 8.

An HTML form was created, which would take weather parameters of users. The implementation of asynchronous data fetching was done through JavaScript functions, which read weather data through NASA POWER API (https://power.larc.

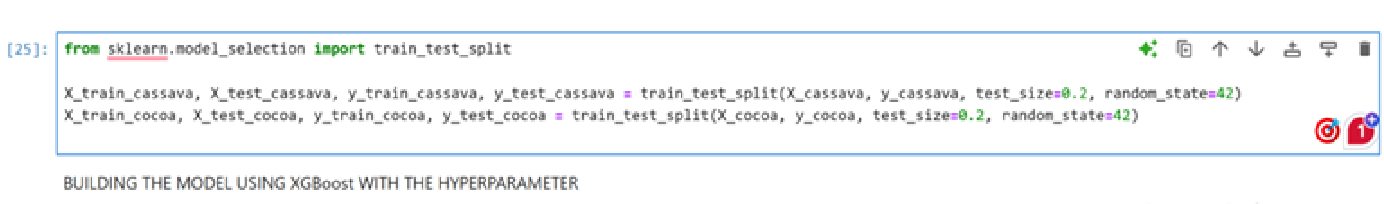

Figure 5. Splitting of dataset into training and testing sets

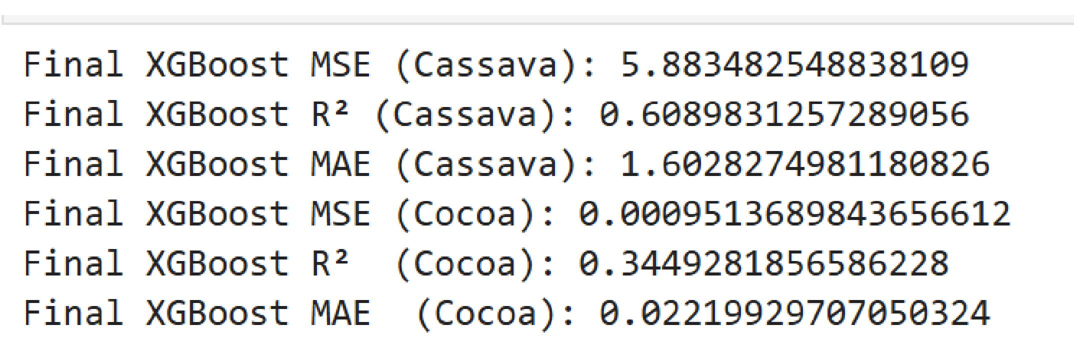

Figure 6. Performance metrics result of each model

```
# Save the trained model
joblib.dump(best_xgb_model_cassava, 'xgb_model_cassava.pkl')
joblib.dump(best_xgb_model_cocoa, 'xgb_model_cocoa.pkl')
```

Figure 7. Saving of the trained models

```
from flask import Flask, render_template, request
import joblib
import pandas as pd

app = Flask(__name__)

# Load the polynomial feature transformer and feature names
poly = joblib.load('poly_transformer.pkl')
feature_names = joblib.load('feature_names.pkl')

# Load the trained models
xgb_model_cassava = joblib.load('xgb_model_cassava.pkl')
xgb_model_cocoa = joblib.load('xgb_model_cocoa.pkl')

@app.route('/')
def index():
    return render_template('index.html')

@app.route('/predict', methods=['POST'])
def predict():
```

Figure 8. Flask setup

nasa.gov/data-access-viewer/) and dynamically change the form fields. Both frontend and backend then collaborate with each other to facilitate zero disruption between the user interface and the server to achieve accurate predictions. Functional and user experience testing has been performed to ensure that the systems are validated.

Results and Discussion

The first step is taken by the farmers is to choose the geographical location where they require yield forecasts. It is done by using input fields in which the user can key in the latitude and longitudes of the farming location. When the details about the location are typed in, the system triggers a search of the pertinent weather information. The app then uses NASA Prediction of Worldwide Energy Resources (POWER) API to retrieve detailed weather variables depending on the given location coordinates. The recovered weather information is systematically incorporated in the application form filling all the required fields of input. This automation helps lower the amount of time required by the user to source and input large volumes of weather parameters, thus making it easier to enter the data. All input fields of the form are framed by an information icon, which on clicking or passing over them will give a tooltip relating to the meaning of the particular weather variable, and its effect on crop yield.

This would be used to inform the user, especially those who are not conversant with the meteorological terms, and would thus make the application more accessible and easy to use. Once the user is satisfied with the pre-considered weather variables, the user can force the prediction process by clicking on the Predict button. The app uses the trained XGBoost regression models to run the input data and give yield forecasts on the crop of interest be it cassava or cocoa. The user is then taken to a special prediction result page where the expected yield figures are shown in a clear and understandable format. In order to gain more information on the influence of each factor of weather on crop yield, scatter plots were formed for each of the weather factors with each of the crop yields.

In the case of cloud cover, the yield of both the cassava and the cocoa is higher when the cloud cover is between 66% and 72%. Cloud cover of 66 to 68 percent is under conditions favorable to stability of yields, but still yields are not more than 0.325 tons per hectare, means that cloud cover is not a definite driver of yields.

Figure 9 and Figure 10 show cloud cover of cassava and cocoa respectively.

Cassava as well as cocoa are more likely to yield well when the soil moisture (GWETPROF) is within the range of 0.80-0.90. Cassava has certain variability in yield concerning soil moisture which indicates that

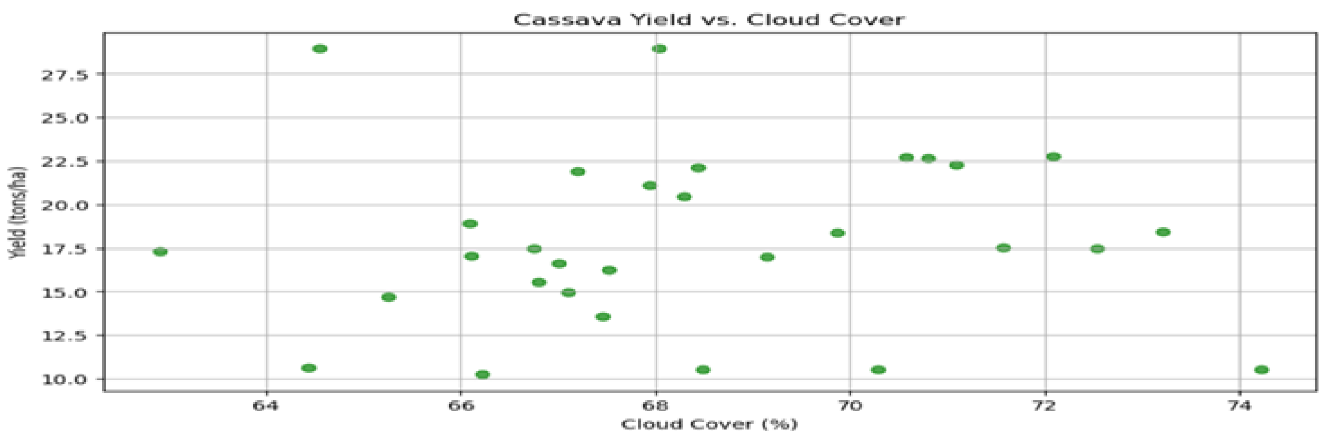

Figure 9. Scatter chart for cassava yield and cloud cover

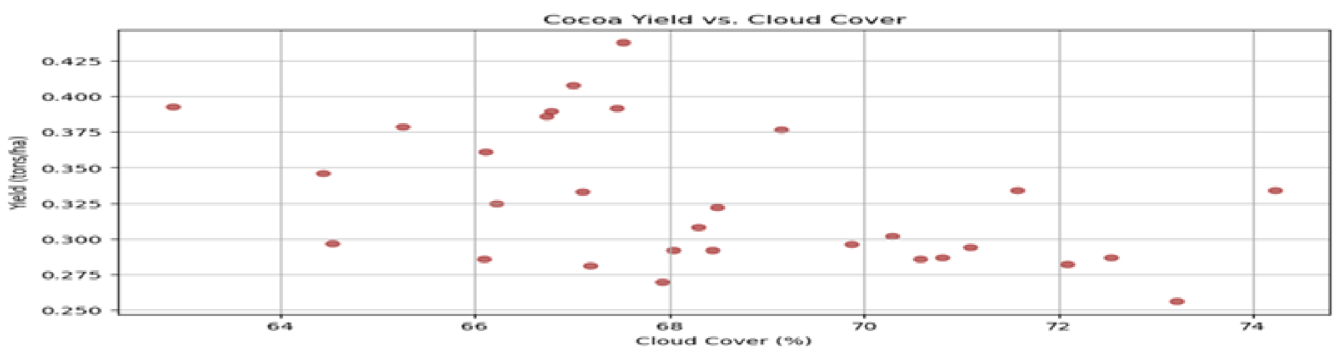

Figure 10. Scatter chart for cocoa yield and cloud cover

stable yet adequately maintained moisture levels would be favorable. The yield of cocoa is observed to be highly stable at the humidity level above 0.825 which means that stable and adequate soil moisture is the key to the optimal cocoa production. The readings are given in Figures 11 and 12 respectively.

The data density of total precipitation is greater in the range of 1400 -2000 mm/year, but yields are less than 22.5 tons/ha. This implies that the moderate level of

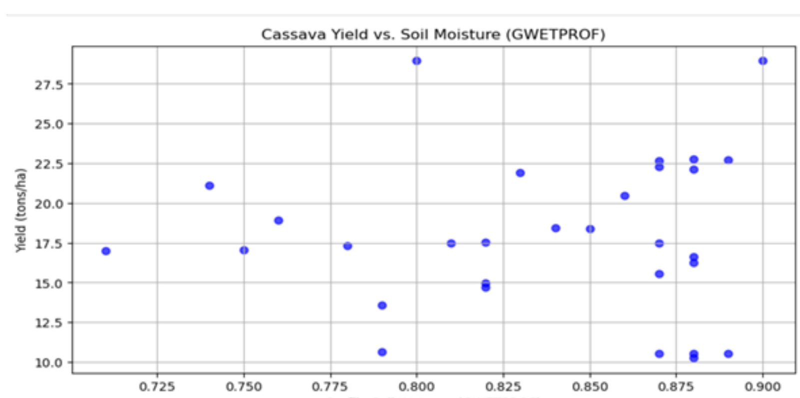

Figure 11. Scatter chart for cassava yield and profile soil wetness.

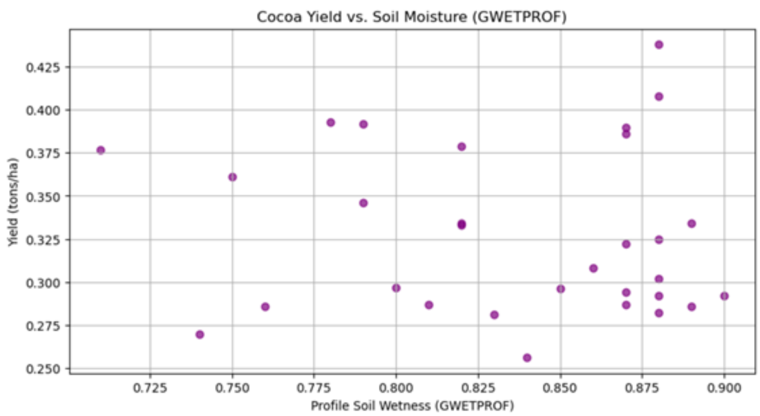

Figure 12. Scatter chart for cocoa yield and profile soil wetness.

precipitation favors stable growth, but other reasons can restrict the cassava yield at high levels. The range of precipitation suggests higher yields, which means that cocoa has a greater advantage at the range of yearly stable and adequate precipitation.

Figure 13 shows the cassava and cocoa output in comparison with the cumulative precipitation.

The data points are also more central to 24.6°C and 25.2°C, but the yields do not go beyond 22.5 tons/ha. This means that a temperature range in this range is usually good to grow but may not necessarily be the sole determinant of the maximum yield. In the case of cocoa, there are more data points between 24.6°C and 24.8°C, although the yields do not exceed 0.350 tons per hectare indicating that the temperature stability is significant but not only enough to attempt higher yields.

Figure 14 displays both cassava and cocoa yields in terms of the average temperature.

Other attributes were designed according to the domain knowledge to increase predictive capability. The cumulative precipitation (PRECTOTCORR_SUM)

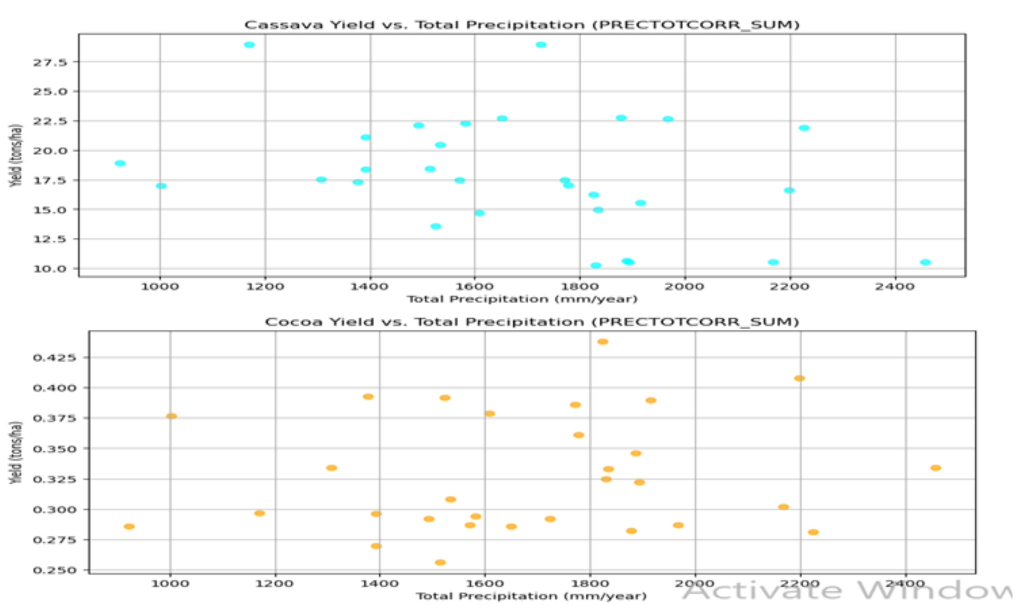

Figure 13. Cassava yield and Cocoa Yield vs Total Precipitation

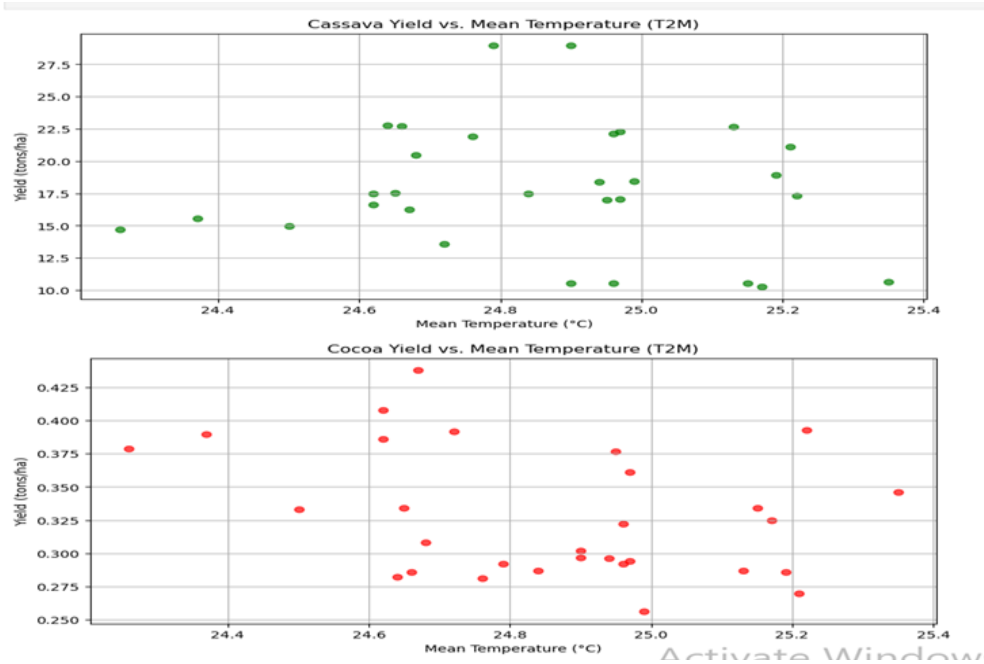

Figure 15. Cassava yield and Cocoa Yield vs Mean Temperature

which aims to capture long-term water availability to crops, and temperature range (T2M_RANGE) which is used to explain the diurnal change in temperature which can affect plant growth and yield.

Conclusion

The design of a weather-based predictive system that would help farmers maximize the size of cassava and cocoa crops is an example of the transformational power that can be unlocked by combining both sophisticated data analytics and the practical use of agriculture. This system helps farmers to be resilient economically, respond to weather-related problems, and achieve sustainable agricultural practices and food security in cassava and cocoa cultivations. It is one of the major steps in using data-driven technologies to aid agricultural practices. The current project was able to incorporate the machine learning models in a user-focused application and give farmers the actionable insights that would improve the productivity of cassava and cocoa crops. Feature works can be carried out on others crops.

References

Ahmed, A., Adewumi, S. E., and Yemi-Peters, V. (2023). Crop yield prediction in Nigeria using machine learning techniques: A case study of southern part Nigeria. *Journal of Agriculture and Agricultural Practices*, 30(8), 40-53. https://doi.org/10.56919/usci.2324.004

Ben-Bouallegue, Z., Weyn, J. A., Clare, M. C. A., Dramsch, J., Dueben, P., and Chantry, M. (2024). Improving medium-range ensemble weather forecasts with hierarchical ensemble transformers. *Artificial Intelligence for the Earth Systems*. Advance online publication. https://doi.org/10.1175/AIES-D-23-0027.1

Edwards. (2018). Seasonal climate forecasts among farmers in Nigeria: Assessment of the usability. In W. Leal Filho (Ed.), *Handbook of climate change resilience,* pp250. Springer International Publishing. https://doi.org/10.1007/978-3-319-71025-9_133-1

Gollin, D. (2021). *Agricultural productivity and structural transformation: Evidence and questions for African development* (STEG Pathfinding Paper). STEG CEPR. https://steg.cepr.org

Huang, H., Huang, J., Wu, Y., Zhuo, W., Song, J., and Li,

X. (2023). The improved winter wheat yield estimation by assimilating GLASS LAI into a crop growth model with the proposed Bayesian posterior-based ensemble Kalman filter. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. *61(1)*, pp1–18. https://doi.org/10.1109/TGRS.2023.3259742

Ibrahim, E. S., Nendel, C., Kamali, B., Gajere, E. N., and Hostert, P. (2024). Predicting potato diseases in smallholder agricultural areas of Nigeria using machine learning and remote sensing-based climate data. *PhytoFrontiers™*. American Phytopathological Society. https://doi.org/10. 1055/2023FA105663

Jiya, E. A., Illiyasu, U., and Akinyemi, M. (n.d.). Rice yield forecasting: A comparative analysis of multiple machine learning algorithms, *5*(2), 30-42. https://doi.org/10.51519/journalisi.v5i2.506

Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., Uwitonze, C., Ngabonziza, J., and Uwamahoro, A. (2023). Crop yield prediction using machine learning models: Case of Irish potato and maize. *Agriculture*, *13*(1), 225. https://doi.org/10.3390/agriculture13010225

Lontsi, S. C., Adoni, W. Y. H., Aworka, R., Zoueu, J. T., Mutombo, F. K., Krichen, M., and Kimpolo, C. L. (2022). Crops yield prediction based on machine learning models: Case of West African countries. *Applied Thermal Engineering*, *192*, 100049. https://doi.org/10.1016/j.atech.2022.100049

Mukhlis, I., Rizaludin, M. S., and Hidayah, I. (2022). Understanding socio-economic and environmental impacts of agroforestry on rural communities. *Forests*, *13*(4), 556. **https://doi.org/10.3390/f13040556**

Olisah, C., Smith, L., Smith, M., Morolake, L., and Ojukwu, O. (2024). Corn yield prediction model with deep neural networks for smallholder farmer decision support system. https://doi.org/10.48550/arXiv.2401.03768

Selz, T., and Craig, G. C. (2023). Can artificial intelligence-based weather prediction models simulate the butterfly effect? *Geophysical Research Letters*, *50(1)*, e105747. https://doi.org/10.1029/2023GL105747

Tamuno-Opubo, D., and Godwin. (2023, May 13). *How machine learning can evaluate the influence of socio-economic and climatic factors in agricultural yield: A case of Nigeria* (Master's thesis). http://repository.aust.edu.ng/xmlui/handle/123456789/5118

Tarini Prasad Das, S. and Sahoo G. H. (2024). Agronomic Principles: Fundamentals of Crop Production. (AGRO Pathfinding Paper). AGRO CEPR. https://agro.cepr.org

Appendix

Kaggle (http://www.kaggle.com/datasets/adekunlejoseph/ondo-state-nigeria-cocoa-cassava-and-weather-data?resource=dwonload&select=Rainfall.txt).

NASA POWER API (https://power.larc.nasa.gov/data-access-viewer/)

Cocoa Data

PRODUCTION	AREA	YIELD
66.42	169.33	0.392
65.38	172.32	0.379
63.42	190.32	0.333
71.54	183.24	0.390
75.59	185.32	0.408
80.32	183.27	0.438
73.49	190.42	0.386
82.58	210.36	0.393
80.47	213.57	0.377
79.24	219.23	0.361
84.65	253.57	0.334
91.57	319.25	0.287
85.74	335.42	0.256
93.27	315.14	0.296
90.59	316.24	0.286
88.27	327.48	0.2695
91.54	325.32	0.281
98.76	320.42	0.308
94.35	321.14	0.294
91.99	320.19	0.287
92.22	321.97	0.286
90.46	320.51	0.282
93.77	321.47	0.292
98.31	336.37	0.292
101.57	342.41	0.2966
119.62	345.45	0.346
110.53	340.37	0.325
105.27	348.44	0.302
125.41	375.52	0.334
123.29	382.81	0.322