# Enhancing Food Security through Seasonal Crop Yield in Nigeria Using Machine Learning Techniques

**Ahmed Abdulbasit[a], Mohammed Habib Itopa[b], Abubakar Sidikat[c], Yusuf Hassan Muhammad[d], Abubakar Barkindo[e], Abadunmi Taiwo[f], Abiodun Taiwo[g], Kunya Musa Muhammad[h], Usman Kabir[i], Usman Shamsudeen[j], Abubakar Bello Mohammed[k], Ahmed Shuaibu[l], Ibrahim Salim Suleiman[m]**

[a,b,d,e]Department of Computer Science, Ahmadu Bello University, Zaria
[c]Department of General Studies, Federal College of Animal Health and Production Technology, Moor Plantation, Ibadan
[f&m]Department of Agricultural Engineering, Ahmadu Bello University, Zaria
[g&h]Department of Computer Science, Federal University, Lokoja
[i]Department of Agricultural Extension and Management, Ahmadu Bello University, Zaria
[j]Department of Computer Science, Nuhu Bamalli Polytechnic, Zaria
[k]Department Information and Communication Technology, Federal Polytechnic Bali, Taraba State
[l]Department of Public Administration, Federal Polytechnic Ayede, Ogbomosho, Oyo State

*Corresponding email: abdulbasitahmedcs@gmail.com

**PAPER KEYWORDS**

Climate Change, Food Security, Internet of Things, Machine Learning, Smart Irrigation.

**ABSTRACT**

*Formerly, farmers were faced with the challenge of knowing the yield from their farm. Prediction is very important in agriculture. In the past, yield prediction was performed by considering farmers' experience with crops. The number of yields cannot be obtained accurately by taking into consideration yields from previous farming seasons. The traditional methods used by the farmers are very slow and unreliable, and large amounts of crops are damaged due to bacterial attacks, erosion and other natural factors. Adoption of big data and machine learning is a key tool to digitalise the agriculture sector and other sectors. Though there is a long debate on its applicability to agriculture, this study addresses how machine learning contributes to digital agriculture in terms of the prediction of crop yields. Three (3) different techniques have been adopted using support vector machine, random forest and decision tree classifiers for predicting crop yields during rainy and dry seasons after the datasets have been subjected to a series of cleansing to get the best in terms of yields and efficiency. The methodology adopted was Cross-Industry Standard Processing for Data Mining (CRISP-DM) with a series of stages before the final deployment. Southeast farming during the rainy season gives the best in terms of yields, accuracy and efficiency using a decision tree classifier with 98.61% accuracy.*

DOI: https://doi.org/10.5281/zenodo.19853536

## 1.1. Introduction

On a daily basis, the quest for agricultural produce increases as the economy develops. The Agriculture sector not only needs the use of old methods but also needs to introduce and promote the continuous improvement of agricultural output with a view to improving the quality and output of agricultural products. Cloud computing has provided a means for analysis, utilisation, and storage. As a result of this, big data technology was born. Big data is widely used in metallurgy, aerospace, agriculture, transportation, mining, medical and so on and has a positive impact on human life. Leveraging big data analytics and machine learning techniques to predict crop yield can be achieved by monitoring crop yield in a timely manner and improving the output of products from the farm. Big data is also considered a large dataset with velocity, veracity, volume, value and variety that make processing difficult while using old techniques. It can come in the form of semi-structured data, big data, unstructured data, structured data, qualitative data, and quantitative data. An advanced method can be used for interpreting, analysing, and modelling big data to reveal unknown patterns.

Timely advice for predicting crop productivity and conducting analysis should be provided to help farmers maximise crop production. Prediction yield is a problem in agriculture. In the past, farmers predicted yields from the previous year's yields. Thus, there are various algorithms or techniques, and with the help of these techniques, crop yield can be predicted. Effective use of big data analytics in agriculture is expanding in scope and importance. By analysing issues and problems such as weather, temperature and several other factors, there are no techniques and solutions to overcome the situation faced (Mayank et al., 2020).

For good crop yield, farmers require timely guidance to help predict crop yield, and analysis is to be done to help farmers utilise the full capacity in crop production. Yield prediction is a great problem. A farmer’s previous experience in a particular crop can be used to make predictions for other crops. Data Mining are methods of predicting, loading and extracting meaningful information from large data to extract some patterns and also change it into a meaningful format for further use. Applying a data mining approach to crop data and historical climate, numerous predictions can be modelled on the basis of data (Kusum & Bhushan, 2019).

## Literature Review

Putri et al. (2019) presented the applications of precision farming by understanding the Nitrogen, Phosphorous and Potassium (NPK) and leaf Chlorophyll status of soil in each of the agricultural fields as a means of improving crop yield and to equally gain high unit production costs. A link between a dependent variable, Y, yield and three predictor variables, X were used using a multiple linear regression (MLR) model (soil nutrient and crop healthiness). It was created a model to using crop healthiness and soil nutrient variability, as well as all interaction variables, to predict rice yield with an overall R2 of 0.6403.

Mayank et al. (2020) focused on yield prediction by means of a random forest based only on the existing data. Real data were adopted for testing and building the models. It made use of a test version of an interactive prediction system. It was put into practice to implement such a system with a web-based interface and user-friendly. It employs the random forest algorithm. There are no suitable solutions or technologies to deal with the scenario we are in right now, despite the analysis of all these concerns and problems, including moisture, humidity, temperature, weather, and rainfall.

Wu et al. (2018) found a new established field of study known as embedded intelligence (EI). By analysing the digital footprints people leave behind when engaging with the Internet of Things, this field seeks to shed light on individual social patterns, spatial settings, behaviours and urban dynamics (smart cards, smart cars and cameras). Knowledge extraction was included from big data and mining of already existing technologies to provide in-depth ontologies or metadata for reasoning and proactive decision-making using open data information. It also analyses the general architecture, research history, significant applications, characteristics and research challenges of EI.

Saeed and Lizhi (2019) validated the root-mean-square-error (RMSE) of 12% of the average yield and 50% of the standard deviation for the validation dataset, utilising anticipated meteorological data. The generated model was determined to have high prediction accuracy. According to computational results, the model performed better than other approaches, including Lasso, regression tree (RT) and shallow neural networks (SNN). The results also demonstrated that environmental factors had a greater impact on crop productivity than genetics.

Manoj et al. (2020) implemented the use of machine learning techniques, and a statistical model was built to provide precise and accurate decision which help farmers select the most appropriate crops to be grown depending on area and season, with minimal chances of losses.

Priya et al. (2018) studied agricultural yield based on historical data such as weather, soil conditions and previous crop yield. His work centred on using a random forest to forecast the crop yield based on the available data. The models were constructed using actual Tami lnadu data and were tested using multiple samples before planting seeds in an agricultural land.

Thomas et al. (2021) investigated and concluded that development in machine learning and sensing technologies will result to cost effective way out for the agricultural sector problem. It also conducted a systematic literature review to identify and compile the characteristics that have been employed in studies to predict crop production. We obtained 567 pertinent papers from six internet resources using our search parameters. Approximately 50 papers were chosen for additional analysis using inclusion and exclusion criteria. We thoroughly examined these chosen papers, evaluated the techniques, features applied and offered recommendations for additional study. Soil type, rainfall and temperature were the most used features, while artificial neural networks were the most widely used technique in these models. After making this remark based on an analysis of 50 publications using machine learning, we searched other electronic resources for studies using deep learning. We found 30 such papers and retrieved data from them.

Kusum and Bhushan (2019) provide a user-friendly interface for farmers that analyses the scenario of crop production forecast based on easily available datasets by applying vector support system, data mining and Bayesian network. Numerous researchers are engaged in this field and have introduced the procedures of data mining and its uses in allied sectors and agriculture. It further explained the methods for analysing historical crop output and climatic data on the basis of the data acquired, a range of forecasts was produced, and in turn helped to increase crop productivity. It further suggested a decision support system (DSS) to reduce the burden of making decisions about the soil and crop to be grown.

Bhanu et al. (2020) used a simulation and regression model to discuss and analyse crop yields supported by collected data and presented different regression models to discuss crop yield prediction using K-nearest regression, support vector regression, linear regression and decision tree regression.

Shikha et al. (2020) used regression analysis to test the effective predictions and accuracy of yield using an agricultural dataset. Linear regression was also used to study the relationship between variables like rainfall, yield and temperature. It was able to show the accuracy of crop yield while predicted values were compared with the actual quantity produced. A relationship was established between variables like rainfall, crop yield and temperature.

Guoyong and Jim (2020) simulated maize using traditional models, process-based models and machine learning algorithms to identify the strengths and weaknesses of each method.

Abhinav et al. (2021) presented reviews of machine learning applications for smart farming, and they provide insight to the research community about the deployment and adoption of digital practices in agriculture. It also did a systematic review of applications of computer vision to examine the classification of different sets of crops to monitor crop yield evaluation and quality. By integrating information gathered by collar sensors to predict and diagnose eating disorders, fertility patterns and cow behaviour, the approach was also incorporated for livestock production.

Mamunur et al. (2021) presented the use of machine learning to predict crop yield with emphasis on palm oil, using a critical evaluation of state-of-the-art machine learning prediction algorithm application of machine learning in comparative analysis and in the palm oil industry and also exploring remote sensing application, disease recognition and plant growth.

Vaishali et al. (2020) constructed a model employing five methods to forecast the yield of mustard crops. Five factors, including precision, recall, accuracy, specificity and f-score, were analysed to gauge the effectiveness of the technique. In order to determine the best reliable method for predicting mustard crop output, various experiments were conducted. KNN and ANN were shown to be the most reliable strategies for predicting mustard crop yield.

Lavanya and Parameswari (2020) developed a model using real data with soil parameters collected from the laboratory, and its used 16 parameters of soil, which comprises micro and macro nutrients. Multiple linear regression with Adam Optimisation in neural network (MLRAONN) was developed with the use of the Keras software used in deep learning. Comparison was made with the existing algorithms, and it was found that MLRAONN gave minimal error as output learning and was evaluated using RMSE, MAE and MSE.

## 1.2. Research Methods

The selected methodology enhances positive changes. The numerous problems associated with other methodologies are not suitable for the project. A lot of farmers had difficulties in carrying out their farming practices as expected. The methodology used was agile and cross-industry standard processing for data mining (CRISP-DM), as they are most widely used in industry and research institutes. The process of submitting, extracting and storing was repeated as many times as possible. The methodology adopted for the modelling was Cross Industry Standard Processing for Data Mining (CRISP-DM), and it is the most widely used in industry and research institutes. CRISP-DM has six (6) stages:

i. Business Understanding: The first stage was to understand what to design and how to gather data. Data were collected from the colleges listed above.

ii. Data Understanding: This is the stage where we identify and understand the datasets gotten from various colleges of agriculture, the relationships between each element of the dataset and the characteristics of the attributes of the dataset. This stage allows us to determine the quality of the dataset

iii. Data Preparation: We analysed and prepared our dataset obtained for modelling. Data cleaning was performed to identify errors in the dataset. The dataset was split into testing and training

iv. Modelling: After splitting our dataset into training and testing. Some portion was for predicting the modelling after evaluating their accuracy.

v. Evaluation: After a successful training model, we then evaluated the performance of the model. This is the stage where the performance evaluation of the model was done by evaluating the result to know how accurate and precise the model is. At this stage, we used the root mean square error to evaluate the accuracy of the model.

vi. Deployment: This is the stage where we will deploy the prediction model into the computer by entering values for prediction. After it was tested, evaluations were made to ensure its workability based on what it was built for before the final deployment.

## 1.3. Results

The model deployed for predicting crop yield was evaluated using Root Mean Squared Error to determine how well it performs towards achieving the goal of predicting crop yield. Below are the performance evaluations for the three algorithms used. Support Vector Machine, Random Forest and Decision Tree Classifier. Comparison between rainy and dry seasons shows that prediction in North Central using a decision tree classifier has the best in terms of accuracy and output for dry season farming. Northwest prediction using a decision tree classifier during the dry season has the best in terms of accuracy and output. North East prediction using random forest for rainy season farming has the best in terms of accuracy and output. South South prediction using a decision tree classifier for rainy season farming has the best in terms of accuracy and output. South East prediction using a decision tree classifier for the rainy season has the best in terms of accuracy and output. Finally, the southwest prediction using a decision tree classifier for rainy season farming has the best in terms of accuracy and output. South East farming during the rainy season gives the best in terms of yields, accuracy and efficiency using a decision tree classifier, and it is fast in processing and modelling with 98.61% accuracy.

## 1.4. Conclusion

The experience acquired from this paper cannot be overemphasised. It has really bridged the gap between traditional ways of predicting and automated ways using software development. Furthermore, this dissertation is still open for further studies and modifications to other machine learning techniques. Despite all the difficulties encountered, it met its aim. A prediction model was designed and tested. The new system will go a long way in proffering solutions to problems that are encountered in the traditional method of predicting seasonal crop yield.

Having presented all that is necessary for a successful implementation. After appraising the effectiveness, efficiency, and productivity of this dissertation. It is recommended for all local farmers. Regular maintenance of both the hardware and software is needed to enable the model to work properly. Finally, training of farmers in the use of a computer is highly recommended for fast and easy retrieval of records.

## 1.5. References

Abhinav, S., Arpit, J., Prateek, G., & Vinay, C. (2021). Machine Learning Applications for Precision. *International for Electrical and Electronics Engineer*, 4845-4847.

Bhanu, K. B., Priyanka, J., Sri, P. K., & Afridi, K. (2020). Crop Yield Prediction using Regression. *International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056*, 3896.

Guoyong, L., & Jim, H. W. (2020). Predicting Spatial and Temporal Variability in CROP yields: An Inter-comparison of Machine Learning, Regression and Process-Based Models. *Environmental Research* (p. 2). China: Institute of Geographic Sciences and Natural Resources Research.

Kusum, L., & Bhushan, C. (2020). Crop yield prediction using data mining techniques and machine learning models for a decision suppport system. *Journal of Emerging Technologies and Innovative Research (JETIR) Volume 6, Issue 4*, 391, 393-395.Manoj, G. S., Prajwal, G. S., Ashoka, U. R., Prashant, K., & Anitha, P. (2020). Prediction and Analysis of Crop Yield using Machine Learning Techniques. *International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181*, 19.

Lavanya, M., & Parameswari, R. (2020). A Multiple Linear Regressions Model for Crop Prediction with Adam Optimizer and Neural Network (Mlraonn). *(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 4*, 253.

Manunur, R., Bifta, S. B., Yusri, B. Y., Mohamad, A. K., & Nuzhat, K. (2021). A Comprehensive Review of Crop Yield Prediction using Machine Learning

Approaches with Special Emphasis on Palm Oil Yield Prediction. *Institute of Electrical and Electronics Engineers*, 1.

Mayank, C., Darpan, C., Chaitanya, C., & Mansing, R. (2020). Crop Yield Prediction Using Machine Learning. *International Journal of Science and Research (IJSR)*, 645.

Putri, R. E., Yahya, A., Adam, N. M., & Abd Aziz, S. (2019). *Rice Yield Prediction Model with respect to Crop Healthiness and soil fertility.* Indonesia: Food Research.

Saeed, K., & Lizhi, W. (2019). *Crop Yield Predictions using Deep Neural Network.* United States of America: Frontiers in Plant Science.

Shikha, U., Pratima, G., & Veenadhari, S. (2020). Crop Yield Prediction using Regression Model. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 269–272.

Vaishali, P., Haneet, K., Surjeet, S., Jatinder, M., & Vinod, S. (2020). Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis. *Journal of Scientific Research*, 394.

Wu, Y., Li, S., Li, L., Li, M., Li, M., Arvanitis, K. G., . . . Sigrimis, N. (2018). Smart Sensors from Ground to Cloud and Web Intelligence. *International Federation of Automatic Control Conference Paper* (p. 32). China: Elsevier.