SEASONAL CROP YIELD PREDICTION USING MACHINE LEARNING TECHNIQUES (A CASE STUDY OF NORTHERN NIGERIA)

Supported by tetfund

Abdulbasit Ahmed[1], Sunday Eric Adewumi[2], Victoria Yemi-Peters[2]
[1]Department of Computer Science, Division of Agricultural Colleges, Ahmadu Bello University, Zaria, Nigeria.
[2]Department of Computer Science, Federal University Lokoja, Kogi State, Nigeria
Corresponding author's e-mail address: abdulbasitahmedcs@gmail.com

Received: May 18, 2023 Accepted: July 10, 2023

Abstract:

Significant proportion of crops are damaged in farms by bacterial attack, erosion and lack of knowledge to follow right agricultural practices. This has negative effects on yields to be gotten from farms. Farmers prefer using previous farming experience to estimate numbers of yields, which is not a good practice. Future solutions to feeding massively expanding population involve digitalizing and automating agriculture. The use of big data and machine learning is a crucial instrument for the digitalization of the agriculture industry and other sectors. This study shows how machine learning contribute to digital agriculture in terms of crop yield prediction. Dataset were gotten from three (3) geopolitical zones of the country, modelled for the prediction using three (3) machine learning techniques using decision tree classifier, random forest and support vector so as to choose the best in terms accuracy after evaluating all the models using Root Mean Square Error as evaluation metric. Cross Industry Standard Processing for Data Mining (CRISP-DM) and Agile method were used and result shows North West yield better and accurate during dry season with accuracy of 98.257%

Keywords: Agile, Cross Industry Standard Processing for Data Mining, Decision Tree Classifier, Machine Learning, Random Forest, Root Mean Square Error and Support Vector Machine

Introduction

Deepak et al (2020) numerous factors have an impact on crop yield. These include climatic changes, farming practices and the area that will be used for farming. Like any other profession, farming has evolved past the point when a farmer can simply put his or her hands together and forecast agricultural produce. Agriculture is a culture that develops in response to environmental changes and cultivation practices. Farmers are given crop recommendations based on data analysis from crop production and farming based on previous data. This lead to loss time and resources

Aakunuri and Narsimha (2016) prediction is crucial for making decisions about different policies but it is expensive, arbitrary, time-consuming and they don't take into account variations caused by seasonal influences and cycles

Kalaiarasi and Anbarasi (2021) the seasonal cycles for crop production and agricultural phenology which depend the climate, soil conditions and weather are among the issues that farming has to face.

Mayank et al (2020) Nigeria as a nation is blessed with fertile land. To assist farmers in maximizing crop production, timely analysis and recommendations on crop productivity prediction are needed. In agriculture, yield prediction is a challenge. Farmers have historically predicted yields based on yields from the preceding year. As a result, there are numerous algorithms or methodologies, agricultural production can be predicted using these techniques. The use of algorithms effectively, the expansion of the applications and the function of big data analytics approaches in agriculture.

Kusum and Bhushan (2019) farmers need early advice to assist forecast crop productivity and analysis is needed to ensure that they are utilizing the full potential of crop production. Crop projections might be based on a farmer's historical success with that crop. Data mining techniques are ways to load, forecast and extract relevant information from enormous data sets in order to identify patterns and transform the information into a format that may be used later. Numerous predictions can be modelled using data by using data mining techniques on historical climate data and agricultural data.

Shreya (2016) the forecasting of crops is a significant issue. The amount of produce that a farmer could expect to receive was something that he paid close attention to. Traditionally, farmers make this choice based on their long-term expertise with a certain yield, plants and weather conditions. The protagonist is more concerned with produce forecast than crop prediction. Yield will be better if the correct crop is anticipated. Berry producers can earn more money by solving the crop and yield forecast problem using the modified K-means clustering method. During this process, items are grouped together so that they have a high degree of agreement with one another but are tremendously distinct from other objects in other clusters.

Maduri (2019) data that are gathered and can be divided into human, process and machine-generated categories. Displayed and described farming practices supported by technologies like thermostats, soil sensors, rain gauges and sensors which are mostly used with the help of the Internet of things (IoT). It enables the gadget to interface with each other and automate farming procedures. The addition of sensors and intelligent equipment increases the amount of data in the farm.

Raghu (2016) used historical data, Hadoop and hive, a system for providing crop pest remedies was constructed. Information acquired for the model from websites pertaining to laboratories and agriculture. It was suggested to get rid of unnecessary noise. It also described how automation is creating a new world of agriculture. Modern methods that are capable of processing complicated data are required due to the rate at which data is being generated. The impact of computers is spreading faster than ever and farming is now a part of the big data world.

Athmaja (2016) built a methodology using big data analytics, machine learning and mobile cloud computing to address the issues farmers confront. The suggested methodology helps farmers make timely decisions that are productive, hence reducing product loss.

Lucky and James (2011) analyzed projects carried out by the agri-food and agri-culture sectors to create a model that forecasts wheat production and its relationship to climate variables. It also discussed the dangers of ignoring spatial dependence and how mathematical studies of agricultural productivity could incorporate it. Particularly, methods that ignore this reliance struggle to comprehend (and foresee) the underlying occurrences. A Bayesian model for agricultural yield has been devised, which not only offers greater design flexibility but also more accurate prediction than traditional least-squares methods. Additionally, it offers precise forecasts that lessen the effects of noisy data. In order to take into consideration, the geographical non-stationarity brought on by the significant regional differences in agricultural policy and practice prior distributions are established. Numerous dimension-reduction strategies could be used to achieve this

Devdatta (2019) employed naive bayes and decision trees to examine soil supplementation using nitrogen, potassium, magnesium and iron. Three processes make up a machine learning system: the first is sampling (multiple soils with the same amount of attributes but different settings), the second is back propagation and the third is weight update. On the basis of accuracy and execution time, performance algorithms were compared.

Vikranth, (2021) guides farmers in how much fertilizer to use, water requirement, pesticide and insecticide information. Wireless Sensor Network (WSN) is a state-of-the-art technology comprised of sensor comprehension, mechanization manage, digital network transmission, information storage and information dispensation to provide effective solutions for a target range of farm applications. The efficiency, productivity and profitability of many agricultural production systems by using the Wireless Sensor Networks (WSNs) would increase precision agriculture. With the aid of the wireless network, IoT helps farmers with live updates by informing people about the crops. We can connect different sensors such as sensors for temperature, sensors for moisture, sensors for water level, and sensors for flood gauge. With the aid of these different sensors, this can provide the farmer with conscious notifications about what is continuing in the field. The development of wireless communication technology leads to improvements in relatively low cost, less power consumption and small-scale multifunctional sensors that allow short-range communication. IoT's architecture consisted of 3 strata. The front end layer gathers information about the environment generated by various agricultural activities, the middle layer acts as a layer of the gateway that connects the front end layer to the Internet and the back end layer stores all the front end data collected.

Manunur (2021) distinguishes current efforts by assembling a wide range of algorithms alongside other experimental articles and putting the focus on predicting palm oil using machine learning approaches. It was developed by gathering pertinent data from recent works and developing suggestions for subsequent development. Additionally, it covered the perspective of machine learning with regard to the prediction of palm oil, including remote sensing, disease detection, tree counting, and plant growth. To address the current crop yield issues, potential solutions were suggested

Sivanandhini and Prakash (2010) based on the proper crop parameters such as wind speed, temperature, humidity and pressure. Neural network model was utilized to estimate yield. Predictions were made using feed-forward and recurrent neural network models and root mean square error (RMSE) was utilized to assess model performance.

Sivanandhini (2020) used comparative examination of predictions made using machine learning techniques was conducted. The effectiveness of a number of techniques, including decision trees, k-nearest neighbors and support vector regression were assessed. The results show that support vector regression performs better than decision trees and k-nearest neighbors.

While attempting to automate farming and other agricultural processes, machine learning can be used on three (3) different techniques.

Material and Method

Data collection

The dataset for the prediction were gotten from three (3) different colleges of agriculture in the northern part of the country as shown in the table below:

Table 1 Dataset collected

S/no	Name of College (Farm)	Zone
1	Kabba College of Agriculture, Kabba, Kogi State	North Central
2	Samaru College of Agriculture, Zaria, Kaduna State	North West
3	College of Agriculture, Jalingo, Taraba State	North East

Description of the Proposed Model

By examining the prior data, we created an efficient recommendation system for the system. In order to forecast and analyze the answers, we processed vast amounts of data using machine learning methods. Support vector machines, random forests and decision tree models were employed in the implementation to strengthen the classifier. The input dataset was gathered from numerous agricultural colleges nationwide. Machine learning platforms received the input required for data processing from the historical dataset and it was evaluated using root mean square error. So the best was chosen for prediction in and this will offer farmers useful advice for identifying the best crop that can be grown for optimal yields.

Results and Discussion

Data Visualization

Values for rainy and dry season farming showing the efficiency and accuracy for each zone as shown in table below:

Table 2 evaluated values from the data set gotten

S/no	Zone/Season	Support Vector Machine	Random Forest	Decision Tree Classifier
1	North Central/Rainy	1412.6	432.8	332.2
2	North Central/Dry	739.6	318.8	259.1
3	North East/Rainy	825.5	260.2	637.7
4	North East/Dry	675.6	322.1	659.9
5	North West/Rainy	739.6	241.0	247.8
6	North West/Dry	699.3	182.1	175.3

Using Root means square error shows decision tree classifier perform better during dry season with accuracy of 259.1 for North Central as shown in Figure 1.

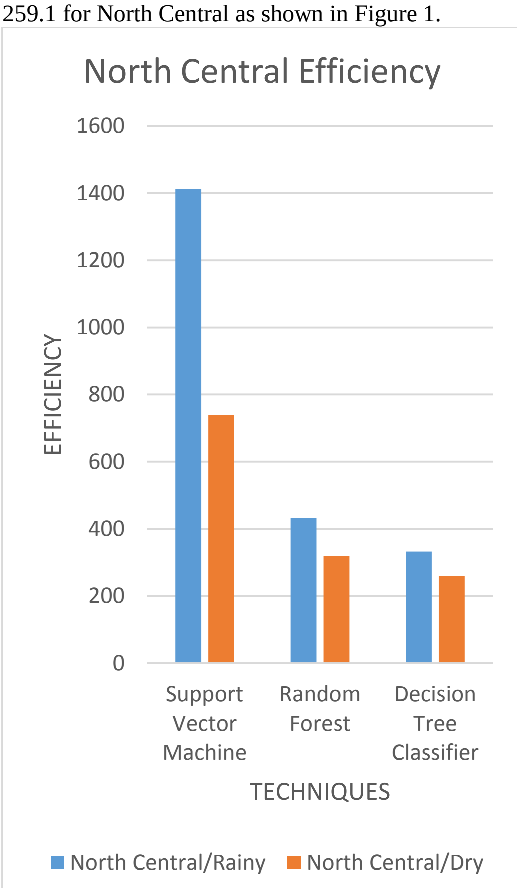

Fig. 1: Showing accuracy for North Central

Using Root means square error shows decision tree classifier perform better during dry season with accuracy of 175.3 for North West as shown in Figure 2.

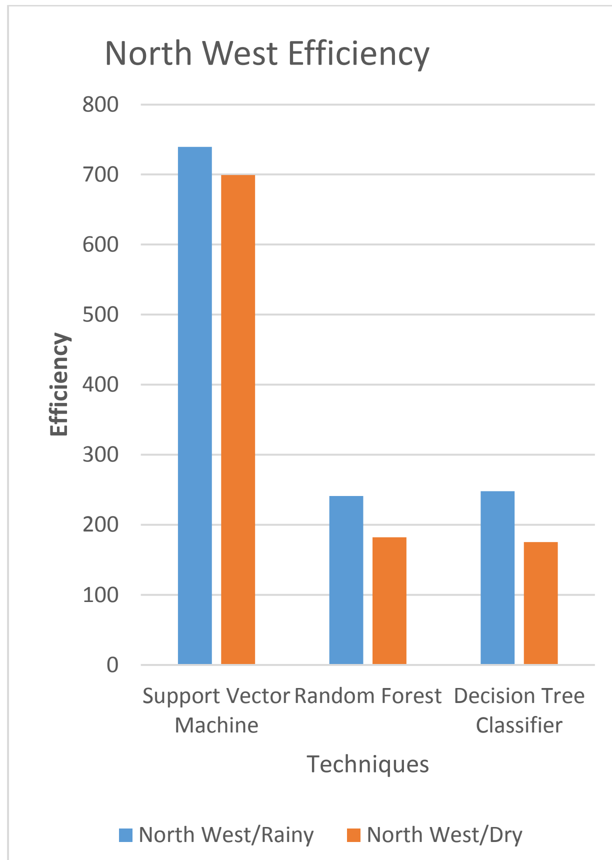

Fig. 2: Showing accuracy for North West

Using Root means square error shows random forest perform better during rainy season with accuracy of 260.2 for North East as shown in Figure 3.

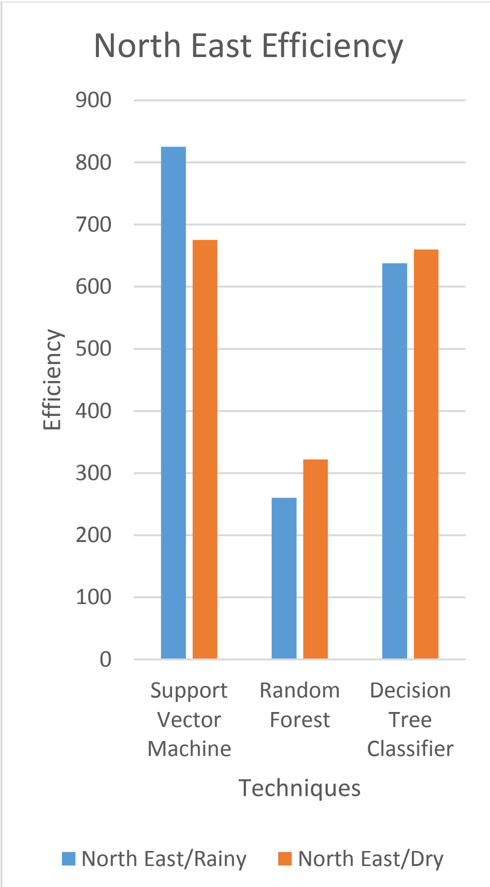

Fig. 3: Showing accuracy for North East

Using Root means square error shows decision tree classifier perform better during dry season with accuracy of 175.3 as shown in Figure 4

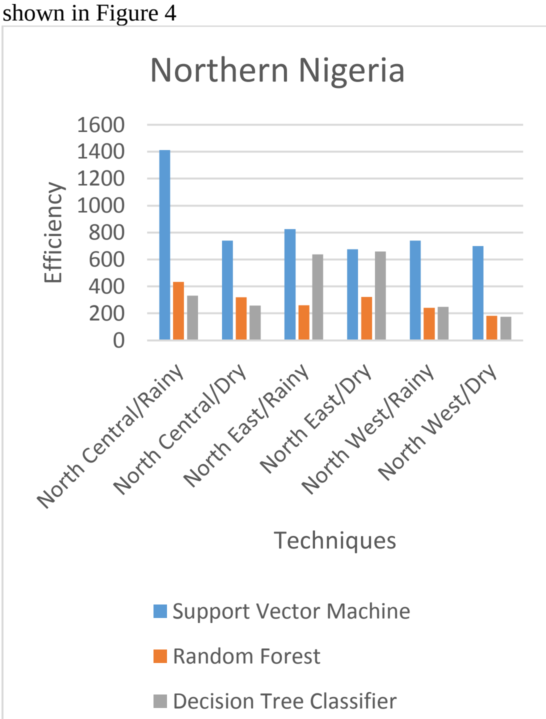

Fig. 4: Showing accuracy for whole northern region

The best model which is random forest give the best prediction while support vector machine give the worst prediction as shown in figure 5.

```
In [278]: #predicting the crop yield for plant seeds of 30kg, 2.0hectares of land and 20bags
          the_best_model_prediction= model11_rfe.predict([[30, 2.0, 20,1,1,1]])
          print(the_best_model_prediction)

          the_worst_model_prediction= model5_svr.predict([[30, 2.0, 20,1,1,1]])
          print(the_worst_model_prediction)

          [38740.]
          [27503.34567107]
```

Fig. 5: Showing the best and prediction

It allows farmers to interact with the software through a web browser that allow farmer to input the number of seeds, size of the land, number of bags, soil type, type of seeds and color of seeds as shown in figure 6.

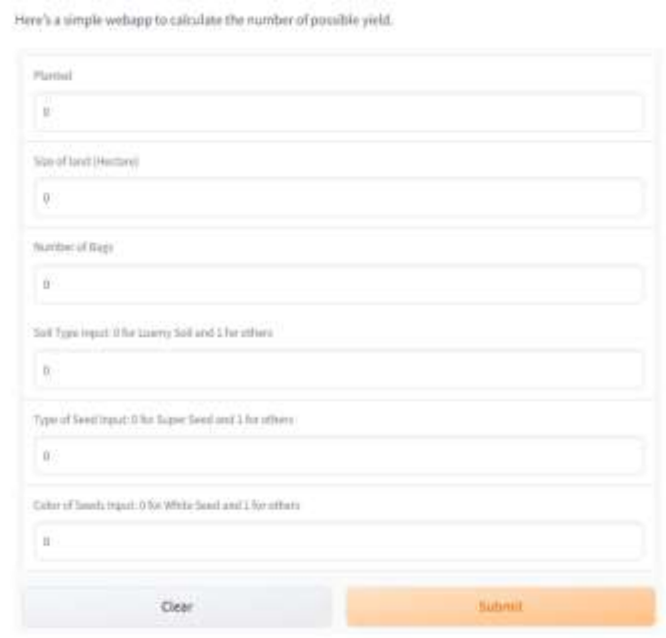

Fig. 6: Web interface of the prediction model

Conclusion

It is impossible to overstate the value of the experience gained by working on this research. It has effectively closed the gap between manual methods of prediction and automated methods utilizing software development. Additionally, is still open to new research and adjustments to other machine learning techniques. A prediction model was created and evaluated. The new technology will go a long way towards providing answers to issues related to predicting seasonal crop yield, result shows North West yield better during dry season because irrigation is mostly adopted which reduces the quantity of eroded products from the farm.

Conflicts of interest

The authors declare that there is no conflict

References

Aakunuri, M., & Narsimha, G. 2016. Crop Yield Prediction with Aid of Optimal Neural Network in Spatial Data Mining: New Approaches. *International Journal of Information & Computation Technology*, 28.

Athmaja, S., & Hanumanthappa, M. 2016. Applications of Mobile Cloud Computing and Big Data Analytics in Agriculture Sector- A Survey. *International Journal of Advanced Research in Computer and Communication Engineering*, 33.

Deepak, G., Deepika, J., Dharshini, M., & Vanathi, B. 2020. Crop Yield Prediction Based on Ensemble Model Using Historical Data. *International Journal of Advance Research and Innovative Ideas in Education (IJARIIE)*, 670.

Devdatta, A. B., & Santosh, M. 2019. Prediction of Crop Yield and Fertilizer Recommendation using Machine Learning Algorithms. *International Journal of Engineering Applied Sciences and Technology*, 371 372.

Kalaiarasi, E., & Anbarasi, A. 2021. Crop yield prediction using multi-parametric deep neural networks. *Indian Journal of Science and Technology*, 131.

Luke, B., & James, Z. V. 2011. *Efficient Stabilization of crop yield prediction in the Canadian Prairies.* Canada: Elsevier.

Madhuri, J., & Indiramma , M. 2019. Role of Big Data in Agriculture. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 3811-3812.

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

Kusum, L., & Bhushan, C. 2020. Crop yield prediction using data mining techniques and machine learningmodels for decision suppport system. *Journal of Emerging Technologies and Innovative Research (JETIR) Volume 6, Issue 4*, 391, 393-395.

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.

Sivanandhini, P., & Prakash, J. 2020. Comparative Analysis of Machine Learning Techniques for Crop Yield Prediction. *International Journal of Advanced Research in Computer and Communication Engineering*, 289.

Sivanandhini, P., & Prakash, J. 2020. Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network. *International Journal of Innovative Science and Research Technology*, 1092.

Shreya, S. B., Kalyani, A. B., Aarti, G. D., & Bhagyashree, R. G. 2016. Crop and Yield Prediction Model. *International Journal of Advance Scientific Research and Engineering Trends*, 23.

Raghu, G., & Himanshu, A. 2016. Big Data Analytics Recommendation Solutions for Crop Disease using Hive and Hadoop Platform. *Indian Journal of Science and Technology*, 1.

Vikranth, K., & Krishna, P. K. 2021. An Implementation of IoT and Data Analytics in Smart Agricultural System. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, 61.