Original Research

Volume 3, Issue 2, pp 19-27, December 2021

ACHIEVERS JOURNAL OF SCIENTIFIC RESEARCH

Opecn Access Publications of Achievers University, Owo

Available Online at *www.achieversjournalofscience.org*

A Predictive Linear Model for Outbreak of Cocoa Blackpod Disease in South West Nigeria

S. S. Olofintuyi

[1]Department of Mathematical Sciences, Achievers University, Owo, Ondo State, Nigeria

*E-mail: Olofintuyi.sundaysamuel@gmail.com

Submitted: October 20, 2021 Revised: December 1, 2021 Accepted: December 4, 2021 Published: December 13, 2021

ABSTRACT

The aim of this study is to develop a predictive linear model that notifies farmers about the outbreak of cocoa blackpod disease in South-West Nigeria. Relevant dataset of precipitation and temperature which covers five cocoa producing states (Ondo, Ekiti, Osun, Ibadan and Ogun) was collected from Nigeria Meteorological Agency (NIMET) and it spans from 1988-2018 (30 years' dataset). The predictive model was formulated with SARIMA, which was used in training the dataset and making predictions. The model was serialized with a mobile application developed using Application Programme Interface (API) so that farmers can receive monthly notification of blackpod disease. The proposed model was simulated using the python programming language. The predictive linear model was evaluated using the performance parameters: accuracy, precision, recall and Mean Square Error. Prototype implementation of the model was done with python programming language. The following evaluation results were derived from SARIMA: *Accuracy: 0.8333; Precision: 0.6316; Recall: 0.6316;* and *Mean Square Error: 0.4082*. The key contribution of this research work is the provision of a model that provides early and reliable information on the outbreak of blackpod disease. Another upgrade that was done in this research work included development of a mobile application for automatic farmer notification of possible blackpod infection through the mobile or Internet network.

KEYWORDS: Machine learning, SARIMA, Predictive linear model, Blackpod disease

1. Introduction

In the agricultural sector today, an unprecedented challenge stirring farmers across the globe in the face is the infestation of pest and crop diseases, and how to protect crops, because of lack of specialized computerized tools capable of predicting when the conditions for the outbreak of the diseases is near (Amin *et al.*, 2020). In Nigeria, cocoa, which was once a major economic crop and served as a major source of agricultural income for the country over decades, has now drastically reduced because of low production that is caused by the incidence of diseases on the crop, and farmers are highly discouraged producing cocoa for the fear of losing their crop to unforeseen outbreak of black pod disease. Nigeria used to be the leading cocoa-producing country in West Africa in the early eighties until the incidence of pest and diseases, among other factors, which are caused by changes in climate change and its untimely notification of outbreak, caused the nation to lose its position. This therefore uncovers that there is a correlation between cocoa diseases and weather condition. A succinct reflection on the impact that agriculture had on Nigeria economy is worthy relaying here. Research has it that in the 1960s out of many countries, Nigeria stood as one of the world's largest exporter of palm fruit, rubber, cotton and cocoa (Sekunmade, 2009). Even within the country, agriculture produced 40% of the Gross Domestic Product (GDP) and constituted about 70% of the working population (CIA, 2013). However even though agriculture stills stand to be the largest economic activity in the rural area where about 50% of the population lives, attention has been unfortunately shifted away from agriculture to oil industry, which had caused a drop down in Nigeria's GDP. Even where farmers are doubling their efforts to yield maximum output, it must be acknowledged that there has been continuous shortage of food supply in the country. One conspicuous reason for this are the diseases affecting crops, stunting their growth and culminating in poor production. However, for the country to take a quantum leap, she must first address the current challenge of low productivity/output and encourage competition within the agricultural sector (Ayodele *et al.,* 2013). One measure for addressing this challenge is to interrogate into the crop diseases that have made output so unpredictable. Most farmers are at the mercy of climate condition or the sensitivity of the weather which determines the output of their farming efforts. However, prediction of plant diseases is feasible by analyzing weather data using appropriate machine learning technique. A related study conducted by Chaurasia and Pal (2017) even stresses the various machine techniques for the prediction of plant diseases. Once prediction is achieved, prevention of crop diseases then becomes realistic. Having established that climate change is a cause of crop diseases, it must be added that various factors bring about changes in climate in Nigeria. Each of these factors are considered as a feature in developing a model which is handy in designing and implementing the model that farmers can utilized in forestalling diseases. Machine learning is a branch of artificial intelligence which can be used for predicting future (Olofintuyi *et al.* 2019; Olofintuyi 2021; Olofintuyi and Omotehinwa, 2021). In the meantime, various machine learning techniques have been used to improve the productivity of farmers. Alexandre *et al.,* (2020) use Convolutional Neural Network (CNN) to predict yield response based on environmental variables. Milos *et al.*, (2018) use climatic data for the prediction of cherry fruit diseases by using discriminant classification analysis and compact classification tree. Angel *et al.*, (2018) propose a Decision Support System (DSS) to enhanced food production in West Africa. Sannaki *et al.*, (2013) use k-Nearest Neighbor approach and Feed Forward Neural Network for making prediction for outbreak of disease in grape. The aim of this study is to develop a predictive linear model that notifies farmers about the outbreak of cocoa blackpod disease in South-West Nigeria. The remaining section of this work covers literature review, the methodology used, results and conclusion. Petre (2019) initiated a weather prediction with decision tree. Precisely, he applied CART algorithm with decision tree as a technique. He considered the following attributes: pressure, cloud quantity precipitation and temperature. The data were collected over the time frame of 4 years, and 48 instances were considered. After the whole experiment, the accuracy of the model was 83%. It was then concluded that the predicted model gave a good prediction accuracy. Milos *et al*., (2018) presented a data mining technique for predicting cherry fruit pathogen but the authors did not create a notification system that will notify farmers and other agricultural extension officers. Vipul *et al.,* (2017) proposed how time series forecasting could be done through clustering. The research made use of the subset of dataset, and the system model was built with this subset. The system model was built by compression of information with clustering in order to discover an inherent pattern for the data. The pattern was discovered to have the curves any time series from the dataset can follow. Linear regression was used to match the closest pattern to the time series that needed prediction. Sannaki *et al.,* (2013) used metrological parameters, such as temperature and humidity which are important in agricultural systems. The researchers proposed a model to predict weather using a modified k-Nearest Neighbor approach and Feed Forward Neural Network, and then utilised parameters such as humidity and temperature to predict the disease outbreaks in grapes. Divya *et al.,* (2014) developed a prediction model with which to understand the effect of groundnut thrips under dynamics of crop-weather-pest relations using data mining techniques. The authors worked with micro-level weather data which were obtained through wireless mote-based agrisens. Distributed sensing system and surveillance data were used to understand and quantify hidden correlation between crop pest and disease weather parameters. The authors indicated that statistical approach with use of regression mining-based correlations assisted in coming up with multivariate regression model that was used to develop an empirical prediction model to issue the forecast for population buildup, initiation and severity of thrips which will help farmers for crop productivity. Sandika *et al.,* (2014) proposed a system for severity identification of potato late blight disease from crop images captured under uncontrolled environment. The key contribution of the study was the use of an algorithm to determine the severity of potato late blight disease using image processing techniques and neural network. Vidita *et al.,* (2013) in their study utilized a fuzzy logic approach for plant disease forecasting. The authors developed a weather-based plant disease forecasting model using fuzzy logic. Hiroyuki *et al.,* (2007) focused on a prediction model of disease infection for foliar parasite on Welsh onions. The model utilized temperature and wetness duration to predict the infection of Welsh onions by rust fungus. The authors stated that rust fungus disease is the most typical disease on Welsh onions. They further mentioned that Weibull probability density function was appropriate for approximating the infection rate of the disease. Zahoor *et al.,* (2014) presented KNN data mining technique which was used for prediction of inter-annual climate. The climatic condition of a specific region in advance was forecasted with the aid of a developed system that uses historical numeric dataset and a data mining technique K-NN. The main factor considered was the temperature of a surface sea. The prediction of climate was done with the use of regression tree technique. The distance between the samples was derived using Euclidean distance metric. The dataset used for the research was collected from the Pakisthan Meteorological Department, and Nation Climate Data Capital (NCDC). SARIMA which is a statistical model, is one of the most widely used model for predicting the linearity in a given time series dataset (Box and Jenkins, 1976). SARIMA is a flexible model that is applicable on various time series dataset, and it has a high level of accuracy in forecasting with ease of implementation. The major setback for SARIMA is that it cannot predict the non-linearity in a given dataset. The major assumption of SARIMA is that there is linear relationship between the past, current and the future values.

2. Methodology

2.1. Dataset

Climatic dataset was used for simulating the proposed model. The dataset was retrieved from NIMET. The dataset spanned 30 years, from 1988-2018. It also covered five cocoa-producing states in Southwestern Nigeria. The five states include: Ondo, Ogun, Oyo, Osun and Ekiti. Two major features (Temperature and Rainfall) were considered from the dataset. After data collection, the dataset was viewed with Principal Component Analysis (PCA) in order to view the whole dataset at once in the subspace. 2 dimensional PCA was used to view the whole dataset. Figure 1 depicts the 2 dimensional PCA for the five cocoa-producing states.

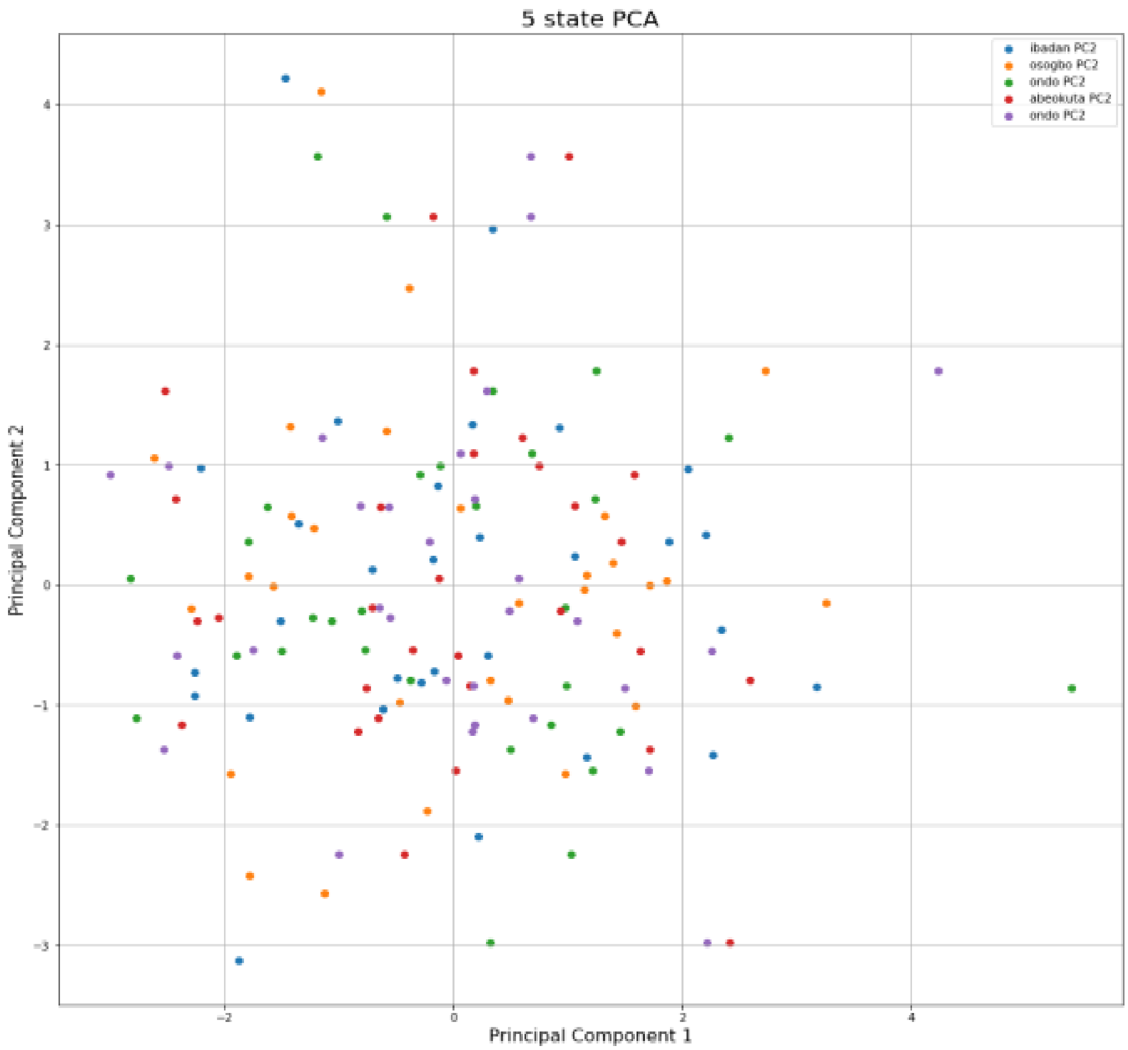

Figure 1: PCA for five states

2.2 Proposed Model

The proposed SARIMA model was trained and validated using python programming language, the proposed model represents a linear model for prediction. A 10-fold cross validation technique was applied to the model during performance testing. SARIMA popularity is due to its flexibility, and this is the reason for its adoption on varieties of time series data. The basic assumption made to use this model is that it considers only the linear component of a time series data. SARIMA can only make prediction for the linear form associated with time series data (Zhang, 2003). The following performance metrics were used: MSE, recall, precision and accuracy.

2.2.1 Mathematical Representation of SARIMA

SARIMA (a, b, c, A, B, C)$_s$ has two parts: Non-seasonal part (a, b, c) and seasonal parts (A, B, C)$_s$

Where:

1. a: order of non-seasonal AR terms
2. b: order of non-seasonal differencing
3. c: order of non-seasonal MA terms
4. A: order of seasonal AR
5. B: order of seasonal differencing
6. C: order of seasonal MA terms

SARIMA Process

SARIMA (a, b, c, A, B, C)$_s$ has the form

$$(\lambda^s)\phi_a(\lambda)(1-\lambda^s)^B(1-\lambda)^c X_t = \theta_C(\lambda^s)\theta_c(\lambda)Z_t \qquad \textit{Equation } 2.1$$

Below is the polynomial form of Equation 2.1

$$\theta_c(\lambda) = 1 + \theta_1\lambda + \cdots + \theta_c\lambda^c$$

$$\theta_c(\lambda^s) = 1 + \theta_1\lambda^s + \theta_2\lambda^{2s} + \cdots + \theta_c\lambda^{Cs}$$

$$\phi_a(\lambda) = 1 - \phi_1\lambda - \phi_2\lambda^2 - \phi_a\lambda$$

$$\phi_a(\lambda^s) = 1 - \phi_1\lambda^s - \phi_2\lambda^{2s} - \cdots \phi_a\lambda^{as}$$

Seasonal differencing

If D = 1

Then $\forall_s X_t = (1 - \lambda^s)X_t$

$$X_t - X_{t-s}$$

If D = 2

$$\forall_s^2 X_t = (1 - \lambda^s)^2 X_t = (1 - 2\lambda^s + \lambda^{2s})X_t$$

$$= X_t - 2X_{t-s} + X_{t-2s}$$

3.3 Application Programme Interface (API)

After the proposed SARIMA model has been trained and validated, the result of the disease index is then saved in the cloud. Figure 2 depicts the proposed model. A mobile application is developed which is to be install on the farmers and agricultural extension officer's mobile phone for monthly disease notification. Two API's were used which connects the model, cloud and the mobile application together. API 2 is design along with the mobile phone to collect agricultural extension officers and farmer's information after registration. Figure 3 depicts the registration interface for the mobile application developed. After information has been collected, API 2 parsed the information to API 1 which is then saved in the cloud. Once API 1 gets the date and state, it loads the data from the cloud and fits it into the *model.pkl*, which makes prediction and sends it to the farmer's and agricultural extension officer's application through API 1 for display. Figure 4 depicts the welcome interface on the mobile application after successful registration

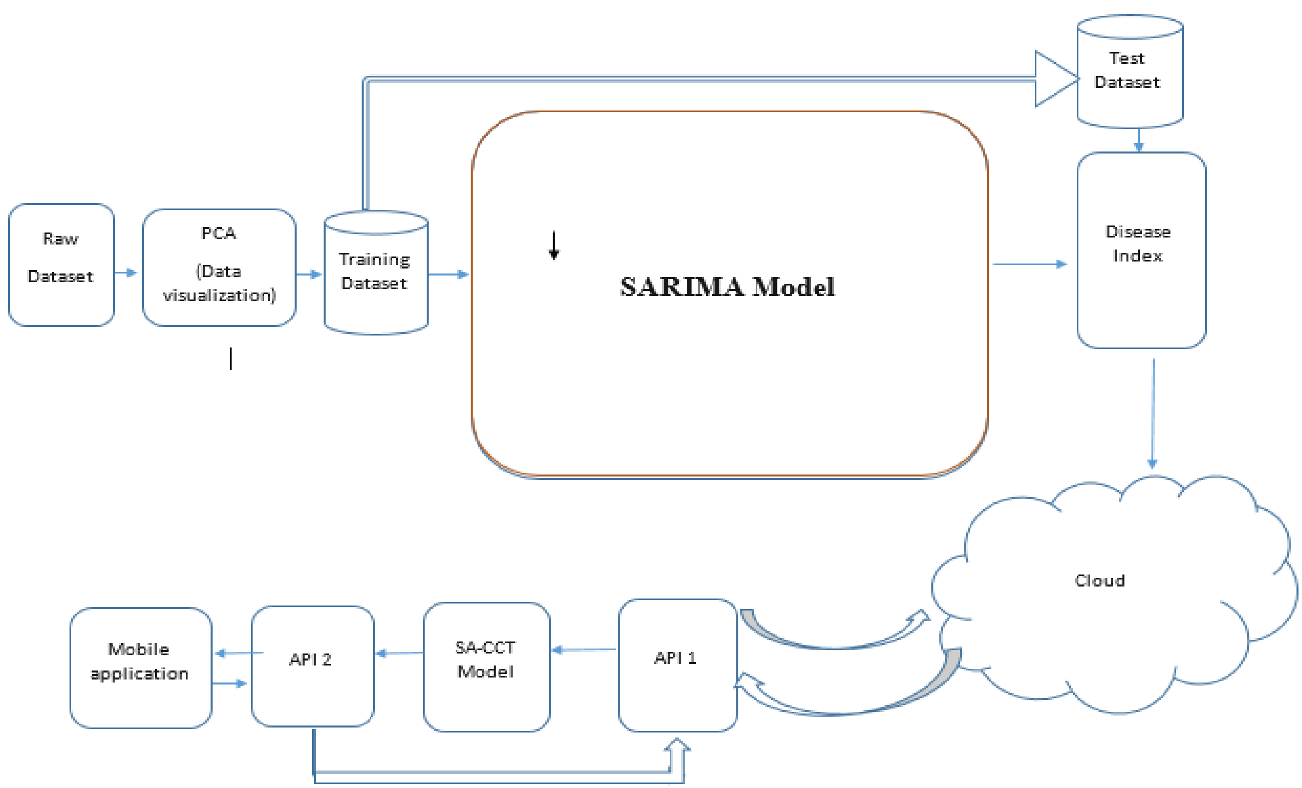

Figure 2: Proposed Predictive Linear Model

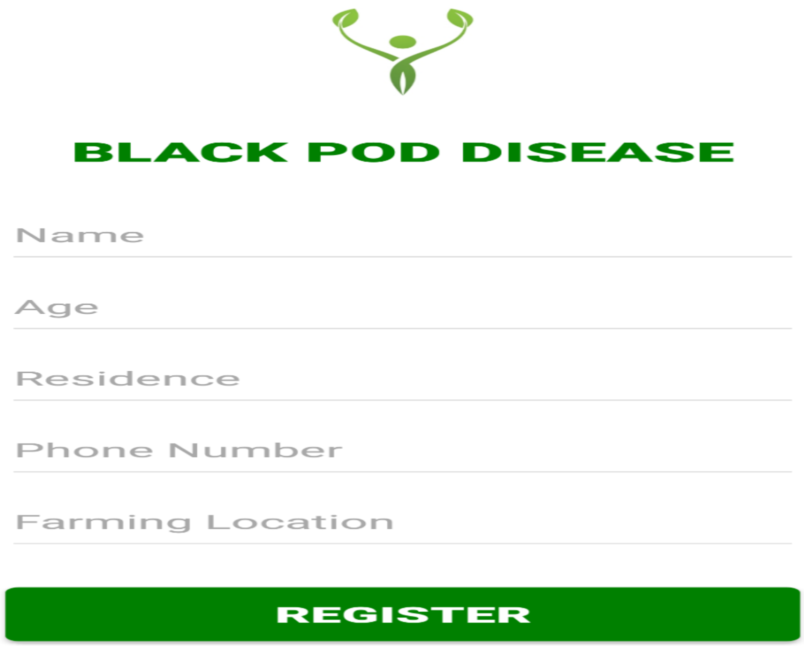

Figure 3: Registration Interface for Farmers to Register on the SA-CCT Mobile Application

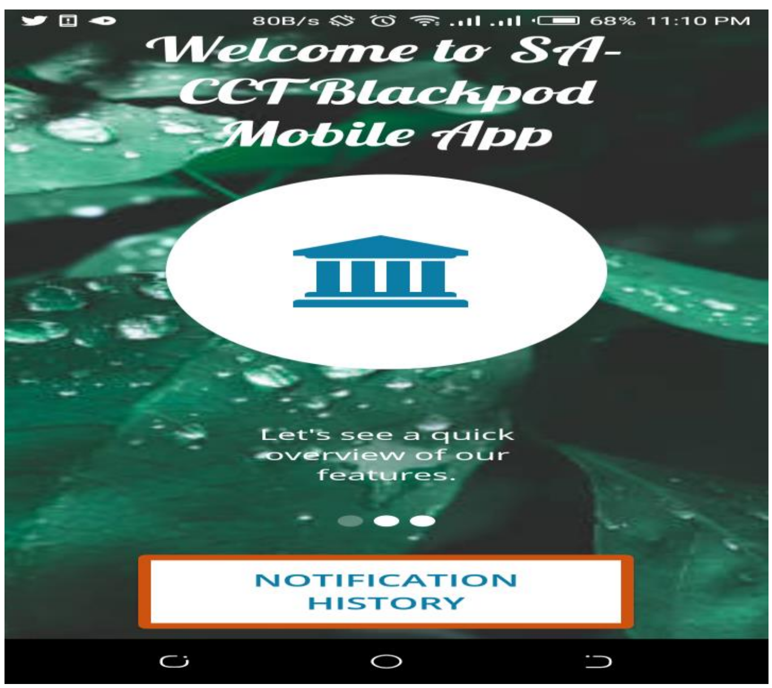

Figure 4: Welcome Interface on the SA-CCT Mobile Application after Successful Registration

4. Evaluation Result of SARIMA Model (Linear Model) Using a 10-Fold cross validation

The experiment performed with SARIMA model considered two features of the climatic dataset (Temperature and Rainfall) where A 10-fold cross validation was used and the dataset was partitioned into training and testing dataset, using a ratio 70% to 30% in python programming environment. SARIMA model gave the following results after evaluation: precision 0.6316; recall 0.6316; Mean Square Error (MSE) 0.4082 and; accuracy 0.8333. Table 1 depicts the evaluation results from SARIMA model while Figure 4.0 depicts the bar chart for the results obtained after evaluation.

5. Conclusion

The key contribution of this research work is the provision of a model that provides early and reliable information on the outbreak of blackpod disease. Another upgrade that was done in this research work included development of a mobile application for automatic farmer notification of possible blackpod infection through the mobile or Internet network. The product of this research is to be used by extension agents and farmers for the notification of blackpod disease as it bridges the gap between research and development.

Further research work can look into creating ensemble system by using more than one algorithms to make predictions. It is likely that using more than one algorithms for prediction of diseases will improve accuracy of prediction. The model can still be modified and be useful in other sectors that their mode of operation depends on climatic parameters. The developed system is meant to equip farmers and agricultural extension officers with notification and preventive measures that need to be taken against blackpod disease.

Table 1: Evaluation results from SARIMA model

Metrics	SARIMA
PRECISION	0.6316
RECALL	0.6316
ACCURACY	0.8333
MEAN SQUARE ERROR	0.4082

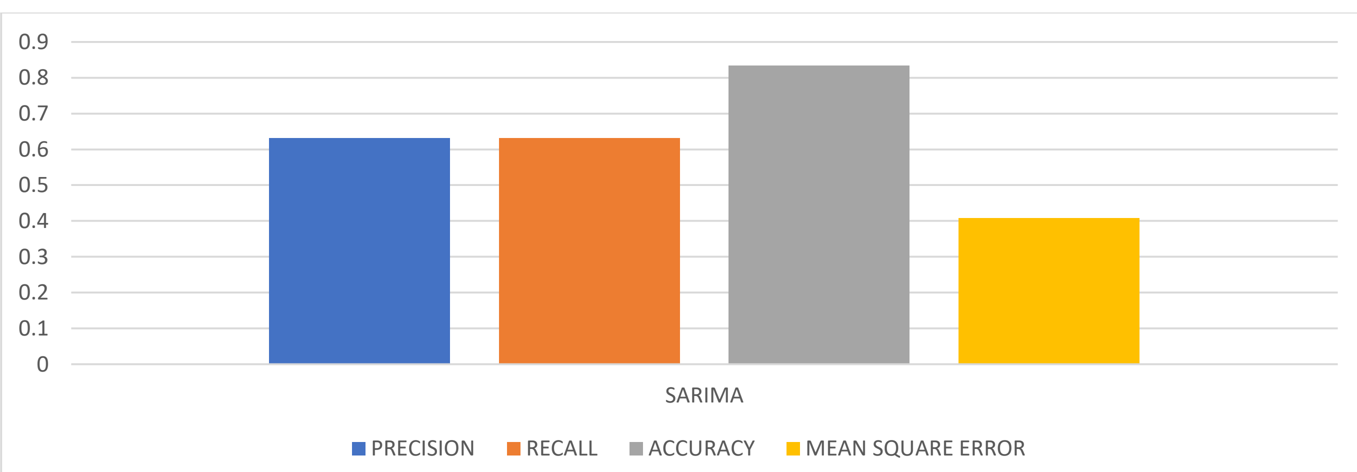

Figure 5: Evaluation results of SARIMA Model

References

Alexandre, B., Rodrigo T., Naira, H., and Nicolas, M. (2020). Modeling yield response to crop management using convolutional neural networks. Computers and Electronics in Agriculture. 170, 105-197.

Angel, U., Marco, P., Céline D., Cesar C., Abdou, A., Luigi C., and Javier C. (2018). A decision support tool to enhance agricultural growth in the Mékrou river basin (West Africa). Computers and Electronics in Agriculture. 154, 467-481.

Ayodele, O., Obafemi, F., and Ebony, F. (2013). Challenges facing the achievement of the Nigeria vision. Global Advanced Research Journal of Social Sciences, 27(2), 143-157.

Box and Jenkins. (1976). Time series analysis, forecasting and control. Holden-Day, San Francisco: CA.

CIA. (2013). World facts book. Retrieved from CIA:https://www.cia.gov/library/publications/theworld-factbook/geos/ni.html

Chaurasia, V, Pal, S. (2013). Early prediction of heart diseases using data mining techniques. Caribbean J. Sci. Technol. 1, 208–217 ISSN: 0799-3757.

Christain, H, Roberto G, Claudio, R, Tristan P, Thor I, Mogens B and Santos N (2007). Nonlinear Contauner Ship Model for Study of Parametric Roll Resonance. Modelling, identification and control, 28(4), pp. 87-103.

Divya, M., Manjunath, N. T., and Ravindra, H. S. (2014). A study on developing analytical model for groundnut pest management using data mining techniques. 2014 Sixth International Conference on Computational Intelligence and Communication Networks (pp. 691-696). IEEE

Foster W.R., Collopy F and Ungar L.H. (1992). Neural network forecasting of short, noisy time series; Computers and Chemical Engineering 16(4), 293– 297.

Hiroyuki, T, Hiromitsu, F and Seiji, C. (2007). Prediction of disease infection of welsh onions by rust fungus based on temperature and wetness duration. 2007 IEEE International Conference on control Application, 325-330

Milos, I., Sinisa, I., Srdjan J. and Stefan, P. (2018). Early cherry fruit pathogen disease detection based on data mining prediction. Computers and Electronics in Agriculture, 150, 418-425

Olofintuyi, S.S., Omotehinwa, T.O., Odukoya, O.H., and Olajubu, E.A (2019). Performance comparison of threat classification models for cyber situation awareness, Proceedings of the OAU Faculty of Technology Conference, 2019.

Olofintuyi, S.S. (2021). Cyber Situation Awareness Perception Model for Computer Network International Journal of Advanced Computer Science and Applications, 12(1), 392-397.

Olofintuyi, S.S. and Omotehinwa, T.O. (2021). Performance Evaluation of Supervised Ensemble Cyber Situation Perception Models for Computer Network. Computing, Information Systems, Development Informatics & Allied Research Journal. 11(2), Pp 1-.14. DOI Affix - https://doi.org/10.22624/AIMS/CISDI/2021/V12N1P1

Petre, I. (2019). Computational modeling technique. Retrieved April 29, 2019, from http: www.users.abo.fi/petre/compmod

Panda C and Narasimhan V. (2007), Forecasting exchange rate better with artificial neural network. Journal of Policy Modeling 29; 227–236.

Sannakki, S., Rajpurohit, V., Sumira, F. and Venkatesh, H. (2013). A neural network approach for disease forecasting in grapes using weather parameters. 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013, pp. 1-5. 10.1109/ICCCNT.2013.6726613

Sandika, B., Jagyasi, B., Singh, B., and Lal, M. (2014). Severity identification of potato late blight disease from crop images captured under uncontrolled environment. Pp. 1-5. 10.1109/IHTC. 2014.7147519.

Sekunmade, A. (2009). The effects of petroleum dependency on agricultural trade in Nigeria: An error correlation modeling. Scientific Research and Essay, 4(11), 1385-1391

Vidita, T, Patel, J and Bhatt C. (2013). Weather based plant diseases forecasting using fuzzy logic, 2013 Nirma University International Conference on Engineering (NUiCONE), Ahmedabad, 2013, pp.1-5. 10.1109/NUiCONE.2013.6780173

Vipul, K., Vamsidhar, T., and Kamalakar, K. (2017). Time series forecasting through clustering.

Zahoor, J., Abrar, M., Shariq, B., and Anwar, M. (2014). Seasonal to inter-annual climate prediction using data mining knn technique. Berlin Heidelberg: Springer -Verlag.

Zhang G., Patuwo B.E., and Hu M.Y. (1998). Forecasting with artificial neural networks: the state of the art; International Journal of Forecasting 14; 35– 62.