International Multilingual Journal of Science and Technology (IMJST)
ISSN: 2528-9810
Vol. 9 Issue 12, December - 2024

Comparative Analysis Of Machine Learning Models For The Determination Of The Optimal Yield Of A Palm Kernel Oil Extractor Machine

Emmanuel Udama Odeh[1]

Department of Mechanical and Aerospace Engineering
University of Uyo, Akwa Ibom State, Nigeria
emmanuelodeh@uniuyo.edu.ng

OLALEYE, O. Olukayode[2]

Marine Engineering Department,
Maritime Academy, Oron, Akwa Ibom State, Nigeria
kayola_nan@yahoo.com

Emem Sunday Ezekiel[3]

Department of Mechanical and Aerospace Engineering
University of Uyo, Akwa Ibom State, Nigeria
ememezekiel@gmail.com

Abstract **— In this study, comparative analysis of machine learning models for the determination of the optimal yield of a palm kernel oil (PKO) extraction machine is presented. The machine learning models used are Random Forest (RF), XGBoost (XGB), and Support Vector Regression (SVR) as well as ensemble model developed from the three models. Comparison of the Mean Absolute Error (MAE), the Mean Square Error (RMSE), and the R-Squared (R^2) for the four models show that the RF model has the best results with the lowest MAE of 1.07 E-15, the lowest MSE of 2.14 E-30 and the best R^2 of 1. The Support Vector Regression model has the worst results with the highest MAE of 0.084175, the highest MSE of 0.007832 and the least R^2 of 0.992293. The XGBoost model is in the second position while the ensemble model is in the third position based on their prediction performance metrics values. Furthermore, the RF, XGBoost and Ensemble model all predicted the same optimal confirmation values of 1.5 mm for the cone gap, 8 % for the palm kernel moisture content, 18 rpm for the main shaft speed and 43.4 % for the optimal oil yield. However, the Support Vector Regression (SVR) model differed in the predations for optimal oil yield with a value of 42.7 % and shaft speed of 20 rpm. In all, the RF model is the best and hence the preferred model among the four models given its outstanding prediction performance metrics values.**

Keywords — Machine Learning Models, Regression Problem, Optimal Yield, Palm Kernel Oil Extraction Machine, Ensemble Model

1. Introduction

Increasingly, machine learning models are being applied in the manufacturing sector [1,2,3]. Machine learning models are used in materials design and selection [4,5], manufacturing process optimization [6,7] and also in inventory management [8,9]. Apart from these three listed areas, there are endless applications of machine learning models in the industrial sectors. Automation of process plants and overall management of industrial outfits are relying heavily on machine learning based decision making processes [10,11,12].

In this study, the focus is on the application of machine learning models for optimal configuration of a palm kernel oil (PKO) extracting plant [13,14]. The essence of the study is to determine the specific input parameter settings of the machine that will guarantee optimal PKO yield at all times. Several machine learning models can be applied for such study. However, the essence of this work is to determine from a selected number of machine learning models which particular model is most suitable for the case study 10-ton PKO extractor plant located in Akwa Ibom State in Nigeria. The study unstilted selected input parameter dataset obtained from the plant to train and evaluate a number of machine learning models after which performance analysis of those models is conducted to select the best model for the case study PKO extractor plant. The outcome of the study is relevant for the researchers and operators of such plant as it provides requisite insights that will guide in input parameter selection and configuration to ensure optimal yield. The manufacturers of the PKO extractor plant will also benefit from the study as it provides key machine parameters that influence its performance of the machine.

2. Methodology

2.1 Development of the Ensemble Model

The research focus is on the application of machine learning models to optimize the palm kernel oil (PKO) yield of 10-ton PKO extraction machine located at Ikpe in Akwa Ibom State, Nigeria. The detailed procedure adopted in the research is captured in the system model of Figure 1 with six key sub-modules while the seventh step is the output of the system from the different machine learning models considered in the research. The target of the research work is to develop machine learning solution that will provide the operators of the case study extractor machine with the appropriate parameters setting mechanism such that the machine will give optimal PKO yield. Essentially, when the palm kernel moisture content is known, the machine learning model-based parameter setting mechanism will enable the operator to select the energy efficient main shaft crushing speed along with the cone gap value that will enable the PKO extractor to give the maximum yield. In this case the dependent variable or the model output is the optimal PKO yield while the independent variables are the main shaft crushing speed, cone gap and the palm kernel moisture content.

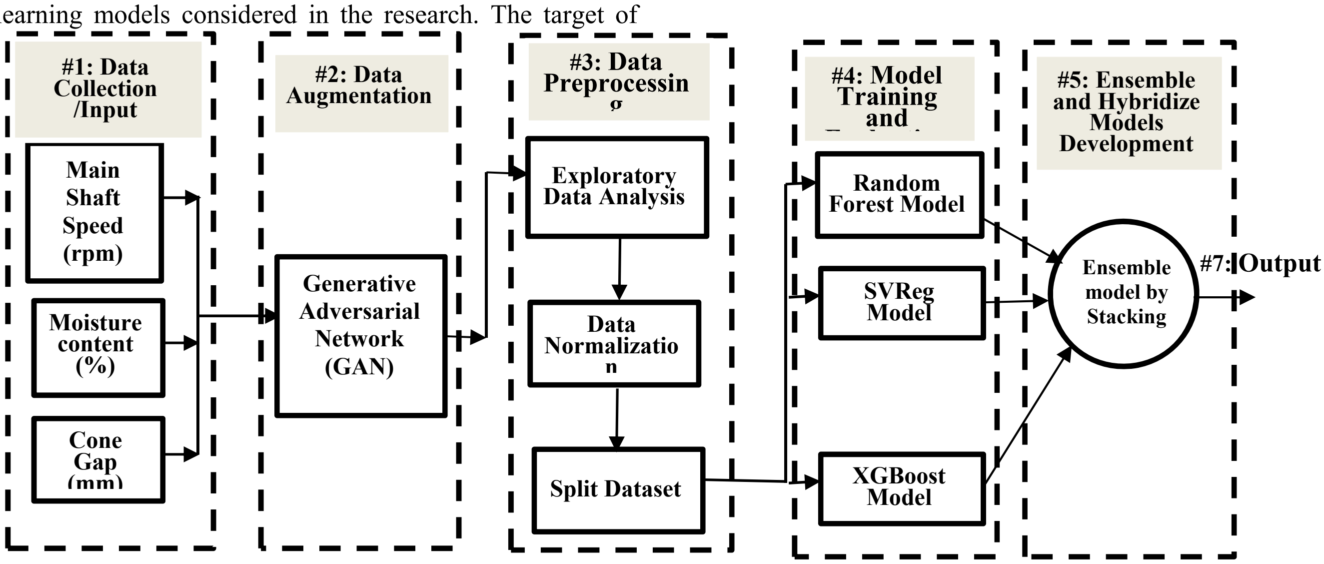

Figure 1 The system model with the detailed procedure adopted in the research

The key steps taken in carrying out the research based on the research procedure in Figure 3.1 are as follows;

i. Data collection from the case study 10 ton palm kernel oil extraction machine located at Ikpe in Akwa Ibom State, Nigeria
ii. Data argumentation
iii. Data preprocessing
iv. Machine learning model development and evaluation
v. The ensemble model development
vi. The comparative evaluation of the models

The dataset used was originally made up of 125 data records but after the Generative Adversarial Network (GAN) approach was used to augment the data thereby generating additional data records that provided a working dataset with about 5000 records. The data was further normalized and split 80% by 20 % training and validation sets. Each of the listed component machine learning models was trained and validated after which the ensemble model was developed from the individual models, namely, Random Forest (RF), XGBoost (XGB), and Support Vector Regression (SVR). The study here focused on the details of the ensemble model development.

2.2 Stacking the Ensemble Model

The ensemble learning method combines predictions from multiple base models—Random Forest (RF), XGBoost (XGB), and Support Vector Regression (SVR)—to improve accuracy and robustness. The stacking method is implemented with a meta-learner, while the surrogate function helps refine the final predictions.

In stacking, multiple base models (level-0 learners) generate predictions, which are then fed into a meta-learner (level-1 model) that optimally combines these predictions. Let X be the input feature matrix:

$$X = [shaft\ speed, cone\ gap, moisture\ content]$$

y is the actual output (oil yield), f_{RF}, f_{XGB}, and f_{SVR} are the prediction functions of Random Forest, XGBoost, and SVR, respectively. Each base model predicts oil yield:

$$\hat{y}_{RF} = f_{RF}(X) \quad (1)$$

$$\hat{y}_{XGB} = f_{XGB}(X) \quad (2)$$

$$\hat{y}_{SVR} = f_{SVR}(X) \quad (3)$$

These prediction forms a new dataset;

$$X_{meta} = [\hat{y}_{RF}, \hat{y}_{XGB}, \hat{y}_{SVR}] \quad (4)$$

A meta-learner gX_{meta} is trained on this dataset to produce the final ensemble prediction:

$$\hat{y}_{ensemble} = g(\hat{y}_{RF}, \hat{y}_{XGB}, \hat{y}_{SVR}) \quad (5)$$

Where, g is a linear regression model that learns the optimal combination of the base model outputs.

2.3 The Surrogate Function for the Model Improvement

To refine the ensemble model, surrogate function was introduced so that it minimizes prediction errors by learning a transformation of the base models' predictions. The residual error for each base model is defined as:

$$e_{RF} = y - \hat{y}_{RF} \quad (6)$$

$$e_{XGB} = y - \hat{y}_{XGB} \quad (7)$$

$$e_{SVR} = y - \hat{y}_{SVR} \quad (8)$$

The surrogate function $S(X_{meta})$ is modelled as:

$$S(X_{meta}) = w_1\hat{y}_{RF} + w_2\hat{y}_{XGB} + w_3\hat{y}_{SVR} + b \quad (9)$$

Where, w_1, w_2, and w_3 are optimized weight coefficients, while b is the bias term. The objective is to minimize the total squared error:

$$\min_{w_1,w_2,w_3,b} \sum\left(y - S(X_{meta})\right)^2 \quad (10)$$

The model in Equation 10 ensures that the ensemble model adapts dynamically to different parameter settings (shaft speed, cone gap, moisture content), hence, reduces overfitting and improving accuracy. The final predicted oil yield is given by:

$$\hat{y}_{final} = S(\hat{y}_{RF}, \hat{y}_{XGB}, \hat{y}_{SVR}) \quad (11)$$

The hyperparameters used in each individual model are maintained in the ensemble model.

2.4 Model Performance Evaluation

The performance of the machine learning models were evaluated using the following metrics:

i. Mean Square Error (RMSE)
ii. Mean Absolute Error (MAE)
iii. R-Squared (R^2)

The RMSE is expressed as:

$$\text{MSE}= \frac{\sum_{i=1}^{i=n}\left(x_{Act(i)}-x_{Pred(i)}\right)^2}{n} \quad (12)$$

Where n denotes the number of data items in the dataset, $x_{Act(i)}$ denotes the ith value of the actual data and $x_{Pred(i)}$ denotes the ith value of the model predicted data.

The R-Squared (R^2) of the model prediction is expressed as:

$$R^2 = 1- \frac{\sum_{i=1}^{i=n}\left(x_{Act(i)}-x_{Pred(i)}\right)^2}{\sum_{i=1}^{i=n}\left(x_{Act(i)}-x_{MeanAct}\right)^2} \quad (13)$$

Where,

$$x_{MeanAct} = \frac{\sum_{i=1}^{i=n}\left(x_{Act(i)}\right)}{n} \quad (14)$$

3. Results and discussion

3.1 The Prediction Performance Results for the Models

The prediction performance of the ensemble model in terms of MAE, MSE and R^2 versus epochfor are shown in Table 1 and Figure 2. The results show that the MAE is 0.016519 and MSE is 0.000443 while the coefficient of correlation is 0.999543. In addition, the line chart of the actual versus predicted oil yields for the Ensemble model is shown in Figure 3 which shows a highly correlated actual and predicted values.

Table 1: The Results of the Error Metrics over Epochs for the Ensemble model

Epoch	MAE	MSE	R^2
0	0.016519	0.000443	0.999543
20	0.016519	0.000443	0.999543
40	0.016519	0.000443	0.999543
60	0.016519	0.000443	0.999543
80	0.016519	0.000443	0.999543
100	0.016519	0.000443	0.999543

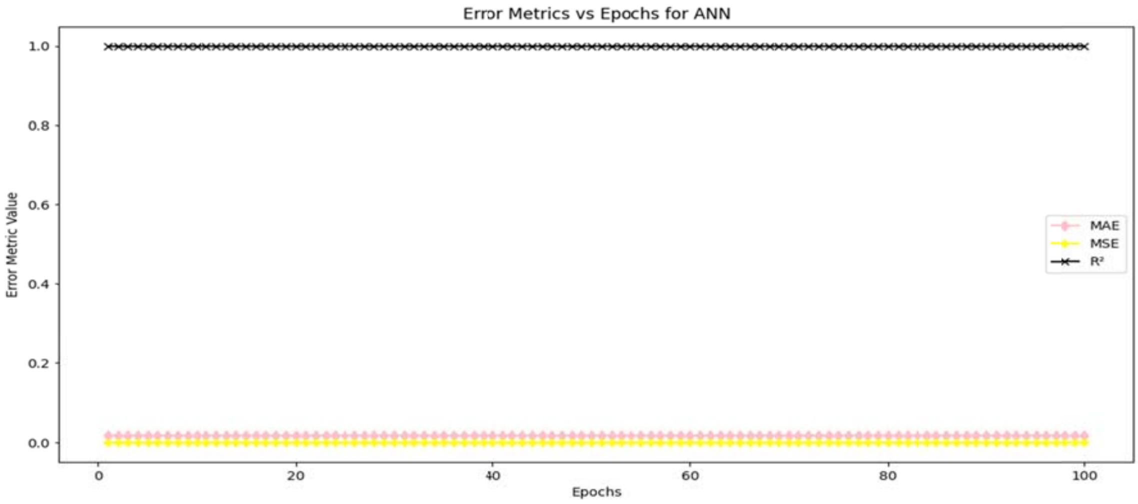

Figure 2 The Plot of the Error Metrics Over Epochs for the Ensemble model

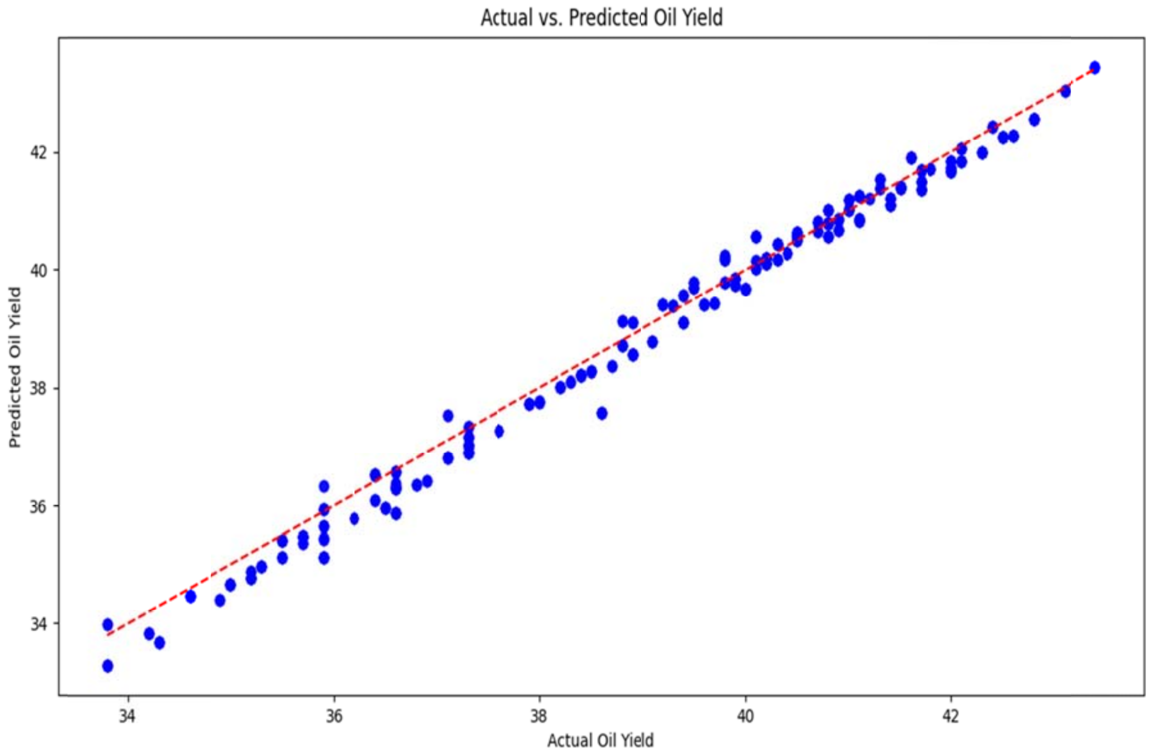

Figure 3 : The Line Chart of the Actual Versus Predicted Oil Yields for the Ensemble Model

Comparison of the MAE, the MSE and the R^2 for the four models are presented in Figure 6 to Figure 6. In all, the Random forest model has the best results with the lowest MAE of 1.07E-15, the lowest MSE of 2.14E-30 and the best R^2 of 1. The Support Vector Regression model has the worst results with the highest MAE of 0.084175, the highest MSE of 0.007832 and the least R^2 of 0.992293. The XGBoost model is in the second position while the ensemble model is in the third position based on their prediction performance metrics values.

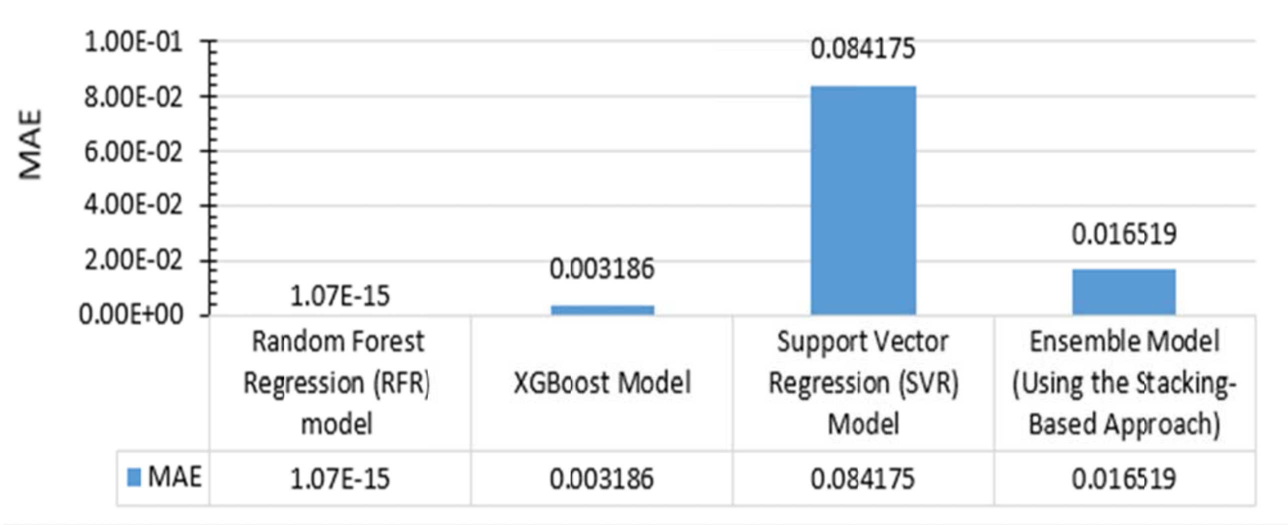

Figure 4 : Comparison of the MAE for the four models

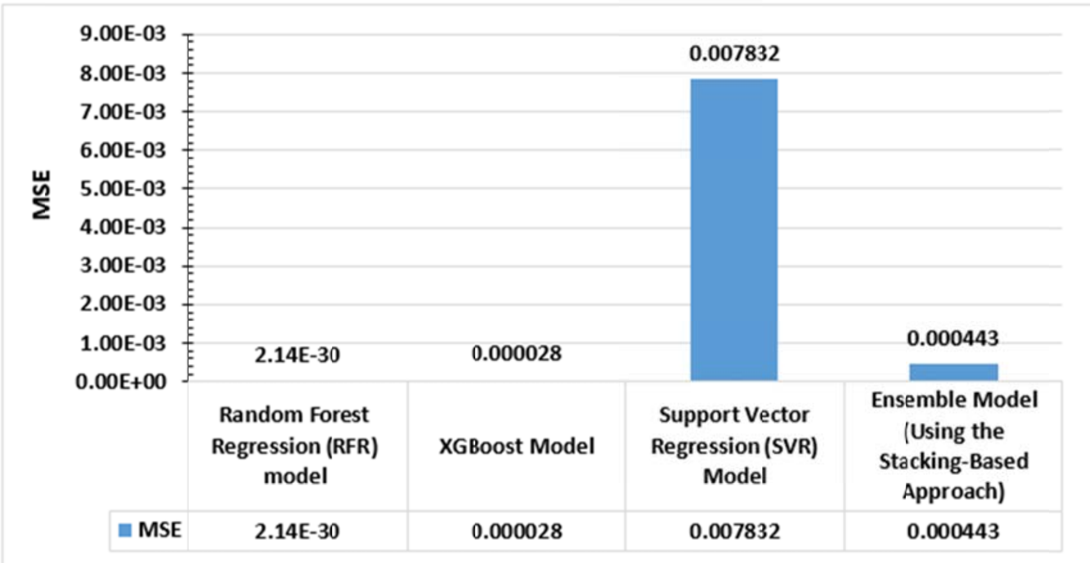

Figure 5 : Comparison of the MSE for the four models

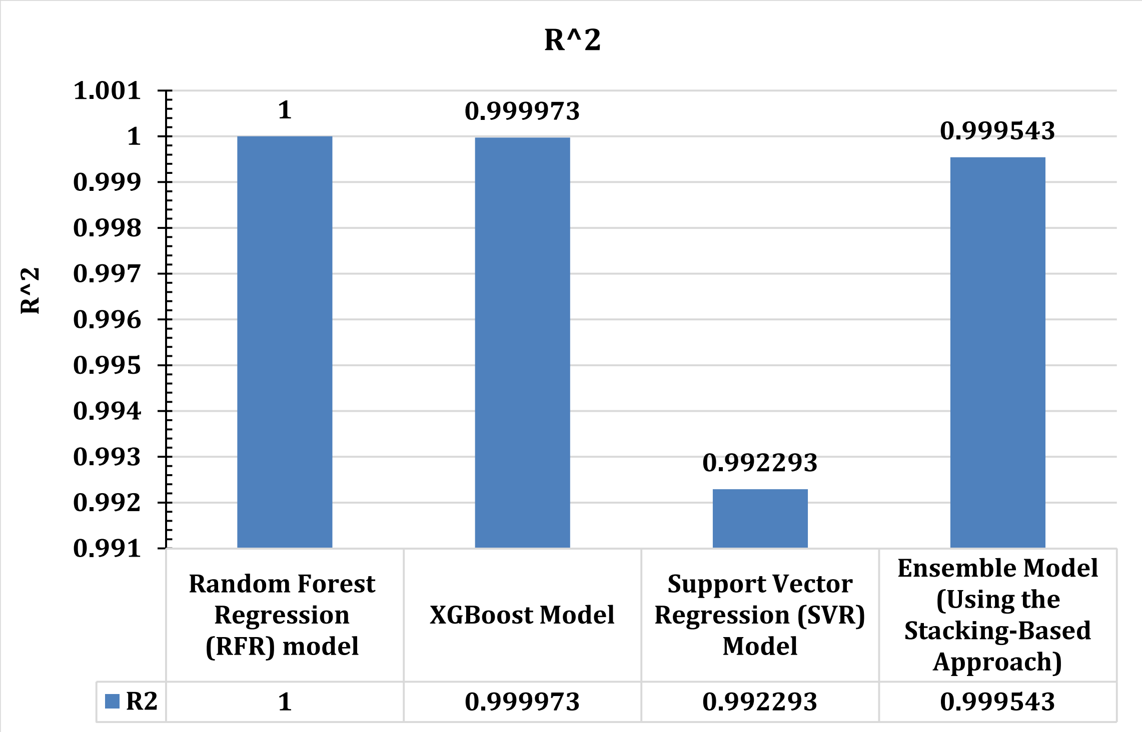

Figure 6 : Comparison of the R^2 for the four models

4.6.2 The Results of the Oil Yield for Various Input Variables Configurations for the ENSEMBLE Model

3.2 The Results for the Optimal Oil Yield Prediction

The results of the oil yield for various input variables configurations for the four models are presented in Table 2. Also, comparison of the optimal oil yield predictions and the shaft speed predictions for the four models are presented in Figure 7 and Figure 8 respectively. The results showed that the Random Forest Regression (RFR), XGBoost model and Ensemble model all predicted the same optimal confirmation values 1.5 mm for the cone gap, 8 % for the palm kernel moisture content, 18 rpm for the main shaft speed and 43.4 % optimal oil yield. However, the Support Vector Regression (SVR) model differed in the predations for two parameters. The SVR predicted optimal oil yield of 42.7 % (Figure 7) and shaft speed on 20 rpm (Figure 8). In all, apart from the SVR model, the Random Forest Regression (RFR), XGBoost model and Ensemble model can be used for the prediction of the optimal oil yield. However, the RFR model is the best and hence the preferred model among the four models given its outstanding prediction performance metrics values.

Table 2: The results of the oil yield for various input variables configurations for the four models

	Random Forest Regression (RFR)	XGBoost model	Support Vector Regression (SVR) model	Ensemble model
Cone gap (mm)	1.5	1.5	1.5	1.5
Moisture content (%)	8	8	8	8
Shaft speed (rpm)	18	18	20	18
Optimal oil yield (%)	43.4	43.4	42.7	43.4

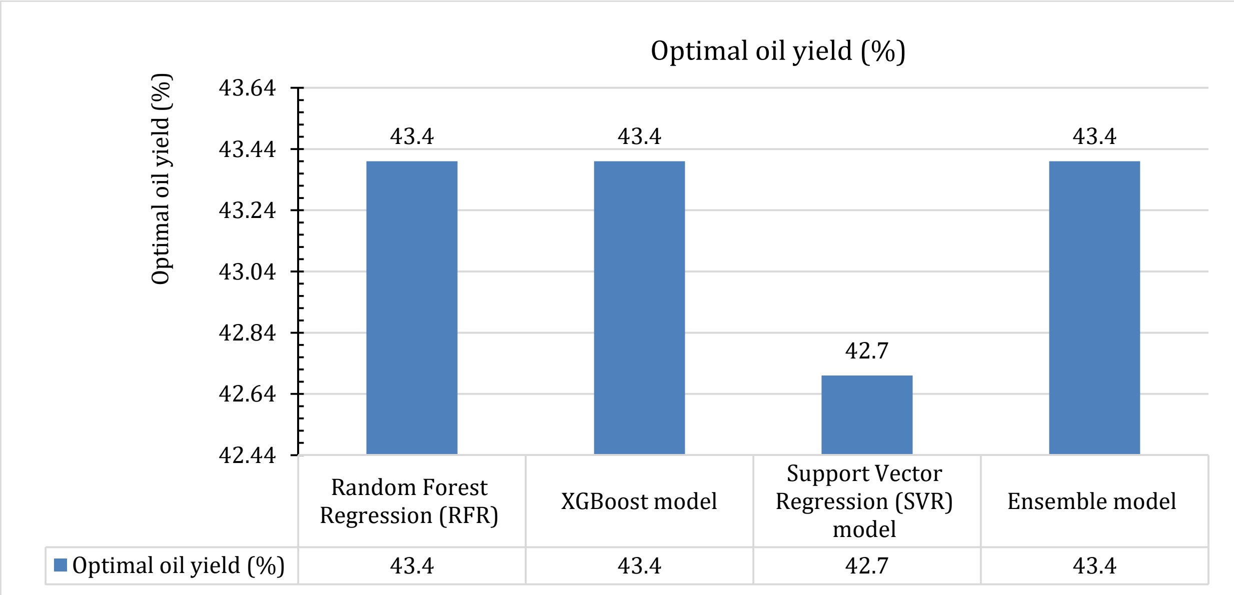

Figure 7 : Comparison of the Optimal Oil Yield Predictions for the four models

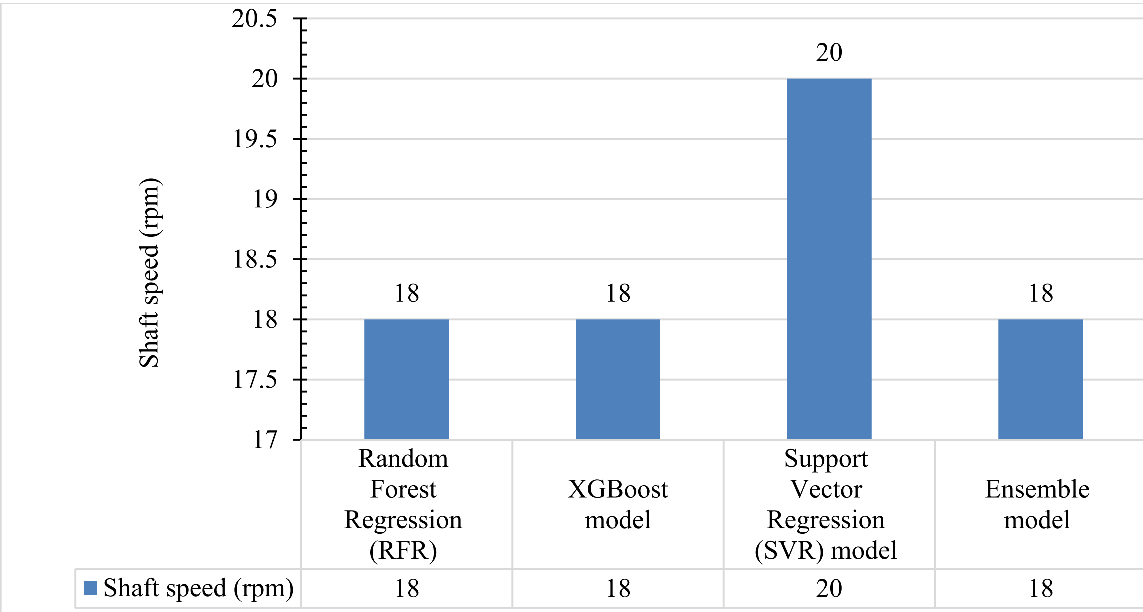

Figure 8 : Comparison of the Shaft Speed Predictions for the four models

4. Conclusion

Determination of the input parameter configuration for optimal yield of a palm kernel oil (PKO) extracting machine in Akwa Ibom State Nigeria is presented. The solution is carried out using three machine learning models which are then ensemble to make up the fourth model. The three models are Random Forest (RF), XGBoost (XGB), and Support Vector Regression (SVR). Although the Random Forest (RF), XGBoost (XGB) and the ensemble model gave the correct optimal oil yield prediction, the results showed that Random Forest (RF) is the best model with the lowest prediction errors. On the other hand, the Support Vector Regression (SVR) did not accurately predict the optimal oil yield.

References

1. Rai, R., Tiwari, M. K., Ivanov, D., & Dolgui, A. (2021). Machine learning in manufacturing and industry 4.0 applications. *International Journal of Production Research*, *59*(16), 4773-4778.
2. Chopra, V., & Priyadarshi, D. (2019). Role of machine learning in manufacturing sector. *Int. J. Recent Technol. Eng*, *8*(4), 2320-2328.
3. Solke, N. S., Shah, P., Sekhar, R., & Singh, T. P. (2022). Machine learning-based predictive modeling and control of lean manufacturing in automotive parts manufacturing industry. *Global Journal of Flexible Systems Management*, *23*(1), 89-112.
4. Kopper, A., Karkare, R., Paffenroth, R. C., & Apelian, D. (2020). Model selection and evaluation for machine learning: deep learning in

materials processing. *Integrating Materials and Manufacturing Innovation*, *9*(3), 287-300.

5. Vasudevan, R., Pilania, G., & Balachandran, P. V. (2021). Machine learning for materials design and discovery. *Journal of Applied Physics*, *129*(7).
6. Weichert, D., Link, P., Stoll, A., Rüping, S., Ihlenfeldt, S., & Wrobel, S. (2019). A review of machine learning for the optimization of production processes. *The International Journal of Advanced Manufacturing Technology*, *104*(5), 1889-1902.
7. Suzuki, Y., Iwashita, S., Sato, T., Yonemichi, H., Moki, H., & Moriya, T. (2018, December). Machine learning approaches for process optimization. In *2018 international symposium on semiconductor manufacturing (ISSM)* (pp. 1-4). IEEE.
8. Praveen, K. B., Kumar, P., Prateek, J., Pragathi, G., & Madhuri, J. (2020). Inventory management using machine learning. *International Journal of Engineering Research & Technology (IJERT)*, *9*(06), 866-869.
9. De Santis, R. B., De Aguiar, E. P., & Goliatt, L. (2017, November). Predicting material backorders in inventory management using machine learning. In *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)* (pp. 1-6). IEEE.
10. Adekunle, B. I., Chukwuma-Eke, E. C., Balogun, E. D., & Ogunsola, K. O. (2021). Machine learning for automation: Developing data-driven solutions for process optimization and accuracy improvement. *Machine Learning*, *2*(1).
11. González Rodríguez, G., Gonzalez-Cava, J. M., & Méndez Pérez, J. A. (2020). An intelligent decision support system for production planning based on machine learning. *Journal of Intelligent Manufacturing*, *31*(5), 1257-1273.
12. Al Shahrani, A. M., Alomar, M. A., Alqahtani, K. N., Basingab, M. S., Sharma, B., & Rizwan, A. (2022). Machine learning-enabled smart industrial automation systems using internet of things. *Sensors*, *23*(1), 324.
13. Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. *IEEE access*, *9*, 63406-63439.
14. Aqbar, K., & Supomo, R. A. (2023). Performance analysis of lstm and xgboost models optimization in forecasting crude palm oil (cpo) production at palm oil mill (pom). *International Journal of Computer Applications*, *185*(17), 37-44.