PREDICTION OF OPTIMAL YIELD FOR A 10-TON PALM KERNEL OIL EXTRACTION PLANT USING XGBOOST MODEL

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_man@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 work, prediction of the yield for a 10-ton palm kernel oil (PKO) extraction plant using XGBOOST model is presented. The study was meant to determine the optimal PKO yield of the case study machine. The study was carried out using 125-records original dataset obtained from a case study 10-ton PKO extractor machine located in Akwa Ibom State Nigeria. The dataset has shaft speed value range of 14 to 18 rpm, cone gap setting range of 0.5 to 2.5 mm and palm kernel moisture content range of 6 % to 14 %. The PKO yield of the extractor machine for different combinations of the three input parameters were captured and used in the model training. Data augmentation was conducted using Generative Adversarial Network (GAN) model to enhance the data records number and thereby enhance the model prediction performance. The XGBoost model was trained and validated using 80 % /20 % training to validation data split. The prediction performance results show that the XGBoost model has Mean Absolute Error (MAE) of 0.007832 and Mean Square Error (MSE) of 0.000028 while the coefficient of correlation (R2) between the actual and the predicted results was 0.999973. The predicted optimal PKO was 43.4 % and it occurred at shaft speed of 18 rpm, cone gap of 1.5 mm and moisture content of 8 %. .**

Keywords— Palm Kernel Oil Extraction, Optimal Oil Yield, Palm Kernel Oil Extraction, Moisture Content, Feature Importance

1. Introduction

In recent years, owners of industrial machines are increasing turning to artificial intelligent (AI)-based data driven modelling to optimize the operations of their machines [1,2,3]. This due to the superior performance and cost effective nature of many AI models when compared with the traditional approaches to machine operations enhancements [4,5,6]. The AI models can also be tuned to suit the specific machine operational dataset [7,8].

In addition, the era of explainable AI has also enabled the users of AI models to understand the logics behind the decisions that are made by the AI models in arriving at the optimal solution for the machine under study [9,10,11]. In addition, the explainable AI models use the feature importance to identify key features that contribute significantly to the decision making process [12,13]. In this way, the user can focus on collecting the data on the few identified important features which will cut down on the cost and labour committed to data collection and processing. Moreover, the AI models will also utilize the fewer number of features to realize high prediction accuracy thereby minimizing the model execution time while maintaining high accuracy due to the use of fewer number of features identified through feature importance analysis.

In view of these salient reasons, this work is focused on applying XGBoost model in the determination of the machine input parameters configurations that will give the optimal output from the machine. Specifically, a palm kernel oil (PKO) extractor machine is studied and the XGBoost model is trained and evaluated using the empirically collected dataset from the case study machine [14,15]. At the end, the XGBoost model is used to determine the combination of the input parameter settings that will give the optimal yield of the PKO from the machine. The outcome of the study will minimize losses and enhance productivity and profitability of the machine.

2. Methodology

2.1 Development of the XGBoost Model

The XGBoost (Extreme Gradient Boosting) is an advanced machine learning algorithm which is based on gradient boosting, optimized for speed and accuracy. In this work, the XGBoost model (with the architecture as captured in Figure 1) is used to model the relationship between shaft speed, cone gap, and moisture content to predict the oil yield for a palm kernel oil (PKO) extracting machine.

Generally, the XGBoost builds an ensemble of weak learners (decision trees) in a stage-wise manner, optimizing the loss function using gradient descent approach. The XGboost model predicts the oil yield $(\hat{y})$ as the sum of multiple decision trees as follows;

$$\hat{y}_i = \sum_{t=1}^{T} f_t(x_i), \ f_t \in F \quad (1)$$

Where, $\hat{y}_i$ is the predicted oil yield, $f_t(x_i)$ is the decision tree at iteration$t$, F is the space of all regression trees, and T is the total number of trees. The XGBoost optimal solution can be expressed as:

$$L = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{t=1}^{T} \omega(f_t) \quad (2)$$

Where, $l(y_i, \hat{y}_i)$ is the loss function measuring error, $\omega(f_t)$ is the regularization term to control complexity.

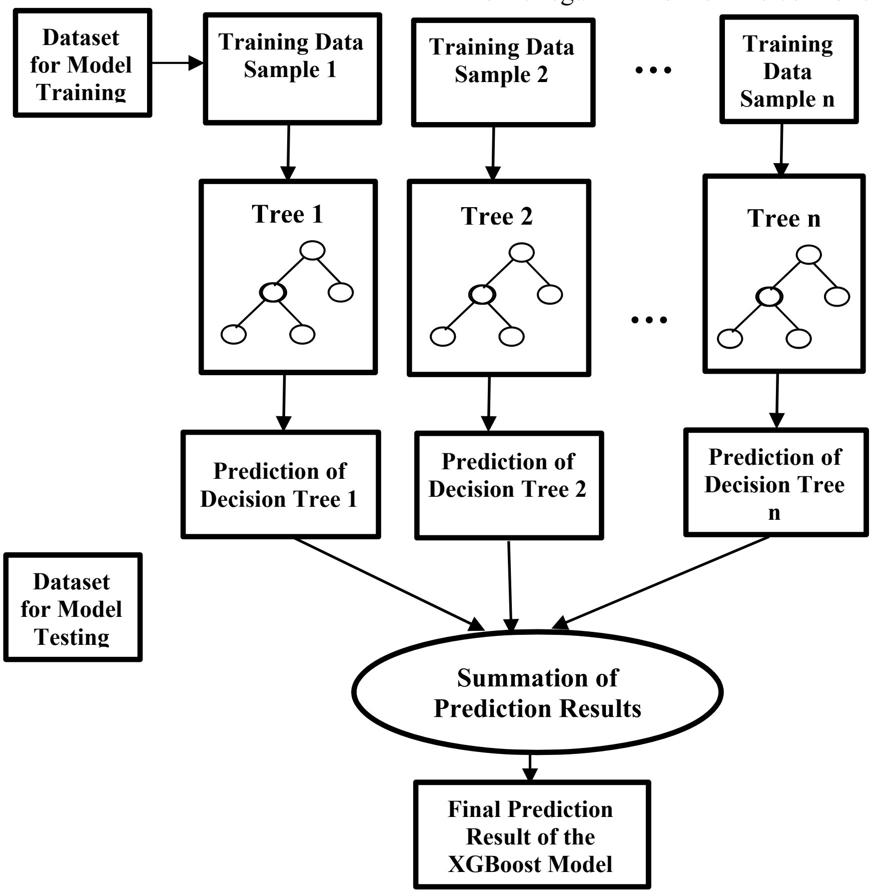

Figure 1 The XGBoost (Extreme Gradient Boosting) Model Architecture

For regression problems like oil yield prediction, the XGBoost typically uses Mean Squared Error (MSE) as the loss function, where:

$$l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \quad (3)$$

The regularization term is computed as:

$$\omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{L} w_j^2 \quad (4)$$

Where, T denotes number of leaf nodes in the tree, w_j is the leaf weight, γ is the complexity penalty for each leaf, and $\lambda = L2$ regularization coefficient (prevents overfitting). This balances prediction accuracy and model complexity. The gradient boosting update rule utilized in the study is expressed with respect to the learning rate as η, as follows:

$$\hat{y}_i^{(t+1)} = \hat{y}_i^{(t)} + \eta f_t(x_i) \quad (5)$$

Where, $f_t(x_i)$ is used to indicate the new tree trained on residuals. Each new tree corrects the residuals from previous trees.

Also, the study also considered the feature importance of the input parameters for the PKO yield prediction. Particularly, the XGBoost assigns importance scores to the features (shaft speed, cone gap, and moisture content) by measuring:

i. Gain (how much a feature improves the model)
ii. Coverage (how many times a feature is used in splits)
iii. Weight (how frequently a feature appears in trees)

These scores help identify the most influential parameters in oil yield prediction. Given the input parameters: x_1 = Shaft Speed, x_2 = Cone Gap and x_3 = Moisture Content, then the XGBoost constructs a function as follows:

$$\hat{y} = f(x_1, x_2, x_3) \quad (6)$$

which is learned through gradient boosting, iteratively refining predictions. The XGBoost uses a second-order approximation (Taylor Expansion) to optimize the loss as follows:

$$L^{(t)} \approx \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \omega(f_t) \quad (7)$$

Where, $g_i = \frac{\partial l(y_i,\hat{y}_i)}{\partial \hat{y}_i}$ (first derivative), and $h_i = \frac{\partial^2 l(y_i,\hat{y}_i)}{\partial \hat{y}_i^2}$ (second derivative). This makes training more efficient. The summary of the XGBoost model's hyperparameters and their values are presented in Table 1.

Table 1 The Hyperparameter settings for the XGBoost Model

Hyperparameter	Value
N_estimators	100
Maximum depth	6
Learning rate	0.05
Random state	42

The study was carried out using 125 original data records (Table 2) obtained from a case study 10-ton PKO extractor machine located in Akwa Ibom State Nigeria. The data has shaft speed range of 14 to 18 rpm, cone gap setting rage of 0.5 to 2.5 mm and palm kernel moisture content range of 6 to 14 %. The PKO yield of the extractor machine for different combinations of the three input parameters were captured and used in the study. Data augmentation was conducted to enhance the data records number and thereby enhance the model prediction performance.

Table 2 The 125 Original Data Records Obtained from the Case Study 10-Ton PKO Extractor Machine

S/No,	Main Shaft Speed (RPM)	Cone Gap(mm)	Moisture Content (%)	Oil Yield (%)	S/No,	Main Shaft Speed (RPM)	Cone Gap(mm)	Moisture Content (%)	Oil Yield (%)
1	14	0.5	6	35.9	63	18	1.5	10	43.1
2	14	0.5	8	37.3	64	18	1.5	12	42.8
3	14	0.5	10	36.6	65	18	1.5	14	42.5
4	14	0.5	12	35.9	66	18	2	6	40.4
5	14	0.5	14	33.8	67	18	2	8	41.7
6	14	1	6	36.6	68	18	2	10	41.4
7	14	1	8	38	69	18	2	12	41.1
8	14	1	10	37.3	70	18	2	14	40.8
9	14	1	12	36.6	71	18	2.5	6	38.9
10	14	1	14	35.9	72	18	2.5	8	40
11	14	1.5	6	37.3	73	18	2.5	10	39.7
12	14	1.5	8	38.7	74	18	2.5	12	39.4
13	14	1.5	10	38	75	18	2.5	14	39.1
14	14	1.5	12	37.3	76	20	0.5	6	39.8
15	14	1.5	14	36.6	77	20	0.5	8	41.3
16	14	2	6	36.6	78	20	0.5	10	41
17	14	2	8	37.3	79	20	0.5	12	40.7
18	14	2	10	36.6	80	20	0.5	14	40.4
19	14	2	12	35.9	81	20	1	6	40.9
20	14	2	14	35.2	82	20	1	8	41.6
21	14	2.5	6	35.2	83	20	1	10	41.3
22	14	2.5	8	35.9	84	20	1	12	41
23	14	2.5	10	35.2	85	20	1	14	40.7
24	14	2.5	12	34.2	86	20	1.5	6	41.4
25	14	2.5	14	33.8	87	20	1.5	8	42.4
26	16	0.5	6	35.9	88	20	1.5	10	42.1
27	16	0.5	8	38.6	89	20	1.5	12	41.8
28	16	0.5	10	36.4	90	20	1.5	14	41.5
29	16	0.5	12	35.5	91	20	2	6	40.8
30	16	0.5	14	34.6	92	20	2	8	42.6
31	16	1	6	36.8	93	20	2	10	42.3
32	16	1	8	37.1	94	20	2	12	42
33	16	1	10	36.4	95	20	2	14	41.7
34	16	1	12	35.7	96	20	2.5	6	40.2
35	16	1	14	35	97	20	2.5	8	42

36	16	1.5	6	37.3	98	20	2.5	10	41.7
37	16	1.5	8	38.3	99	20	2.5	12	41.4
38	16	1.5	10	37.6	100	20	2.5	14	41.1
39	16	1.5	12	36.9	101	22	0.5	6	39.5
40	16	1.5	14	36.2	102	22	0.5	8	39.8
41	16	2	6	36.5	103	22	0.5	10	39.5
42	16	2	8	37.1	104	22	0.5	12	39.2
43	16	2	10	36.4	105	22	0.5	14	38.9
44	16	2	12	35.7	106	22	1	6	40.3
45	16	2	14	35	107	22	1	8	40.8
46	16	2.5	6	35.3	108	22	1	10	40.5
47	16	2.5	8	35.9	109	22	1	12	40.2
48	16	2.5	10	35.5	110	22	1	14	39.9
49	16	2.5	12	34.9	111	22	1.5	6	41.1
50	16	2.5	14	34.3	112	22	1.5	8	42
51	18	0.5	6	38.8	113	22	1.5	10	41.5
52	18	0.5	8	40.1	114	22	1.5	12	41
53	18	0.5	10	40.3	115	22	1.5	14	40.5
54	18	0.5	12	39.8	116	22	2	6	40.1
55	18	0.5	14	39.3	117	22	2	8	40.5
56	18	1	6	39.4	118	22	2	10	40.2
57	18	1	8	41.2	119	22	2	12	39.9
58	18	1	10	40.9	120	22	2	14	39.6
59	18	1	12	40.5	121	22	2.5	6	38.4
60	18	1	14	40.1	122	22	2.5	8	38.8
61	18	1.5	6	42.1	123	22	2.5	10	38.5
62	18	1.5	8	43.4	124	22	2.5	12	38.2
63	18	1.5	10	43.1	125	22	2.5	14	37.9

3. Results and discussion

3.1 The Results of the impact of the inputs and the predicted oil yields for the XGBoost

The augmented dataset from the PKO extractor machined was normalized and then used to train and validate the XGBoost model using 80/20 % training to validation data split. The results of the average impact of the inputs on the XGBoost model output are shown in Figure 3. It shows that moisture content has the highest impact on the XGBOOST model output. Also, the results of the error metrics over epochs for the XGBOOST model are shown in Table 3 and Figure 3. The line chart of the actual versus predicted oil yields for the XGBOOST model is shown in Figure 4. The results show that the Mean Absolute Error (MAE) is 0.007832 and Mean Square Error (MSE) is 0.000028 while the coefficient of correlation (R2) between the actual and the predicted results is 0.999973.

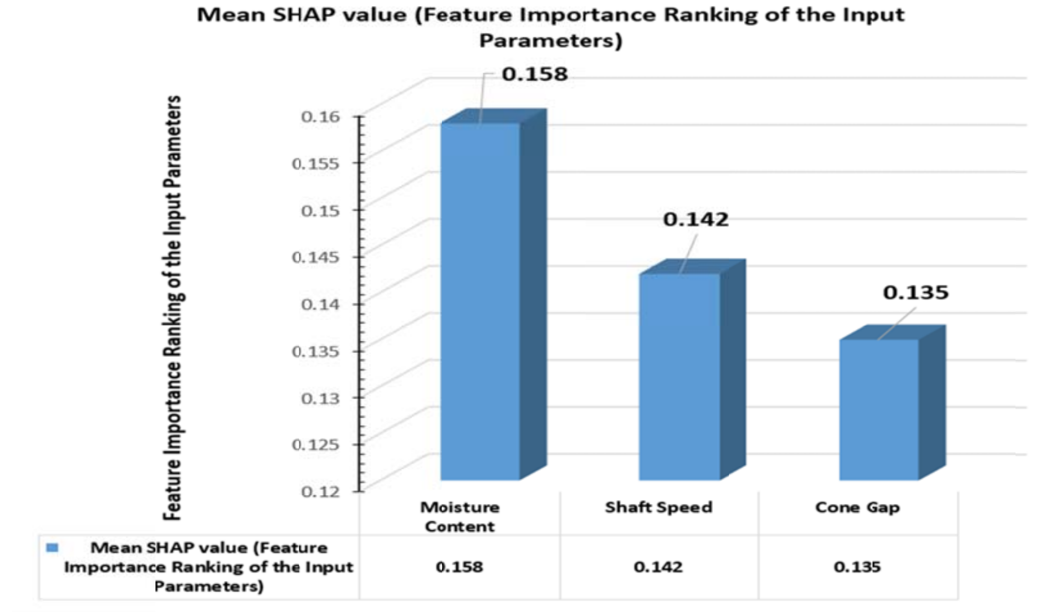

Figure 2: Average impact of the inputs on the XGBoost model output

Table 3 The Results of the Error Metrics over Epochs for the XGBoot Model

Epoch	MAE	MSE	R2
0	0.003186	0.000028	
20	0.003186	0.000028	0.999973
40	0.003186	0.000028	0.999973
60	0.003186	0.000028	0.999973
80	0.003186	0.000028	0.999973
100	0.003186	0.000028	0.999973

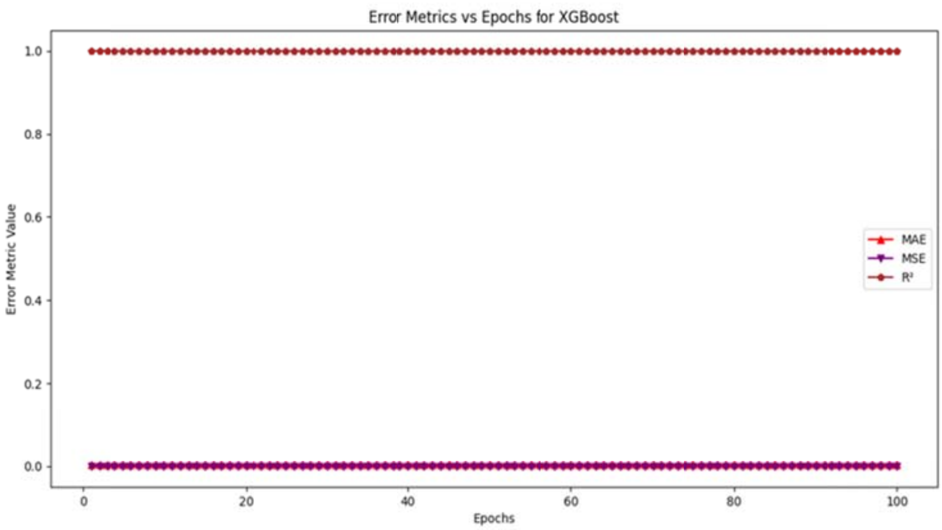

Figure 3 The Plot of the Error Metrics Over Epochs for the XGBoost model

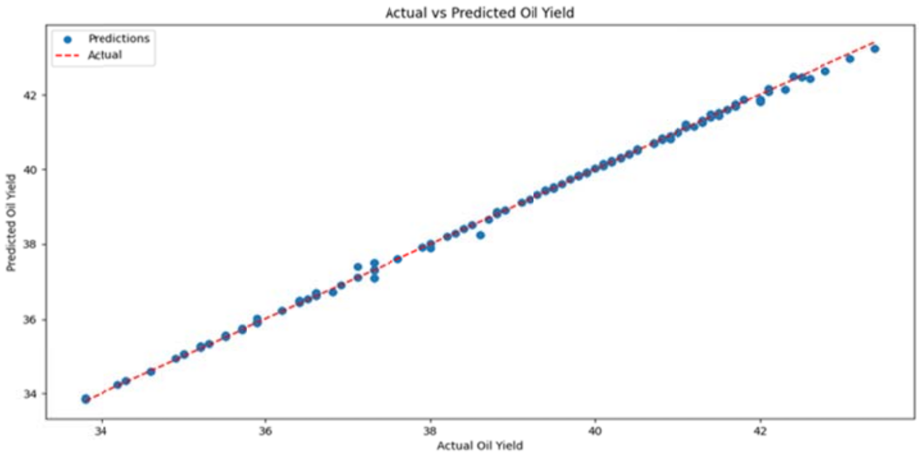

Figure 4: The Line Chart of the Actual Versus Predicted Oil Yields for the XGBOOST Model

3.2 The Results of the Oil Yield for Various Input Variables Configurations for the XGBoost Model

The results obtained from the XGBOOST model predictions as presented in Figure 5 to Figure 9 show that the optimal PKO is 43.4 % and it occurred in Figure 7 having the input parameter setting with shaft speed of 18 rpm, cone gap of 1.5 mm and moisture content of 8 %.

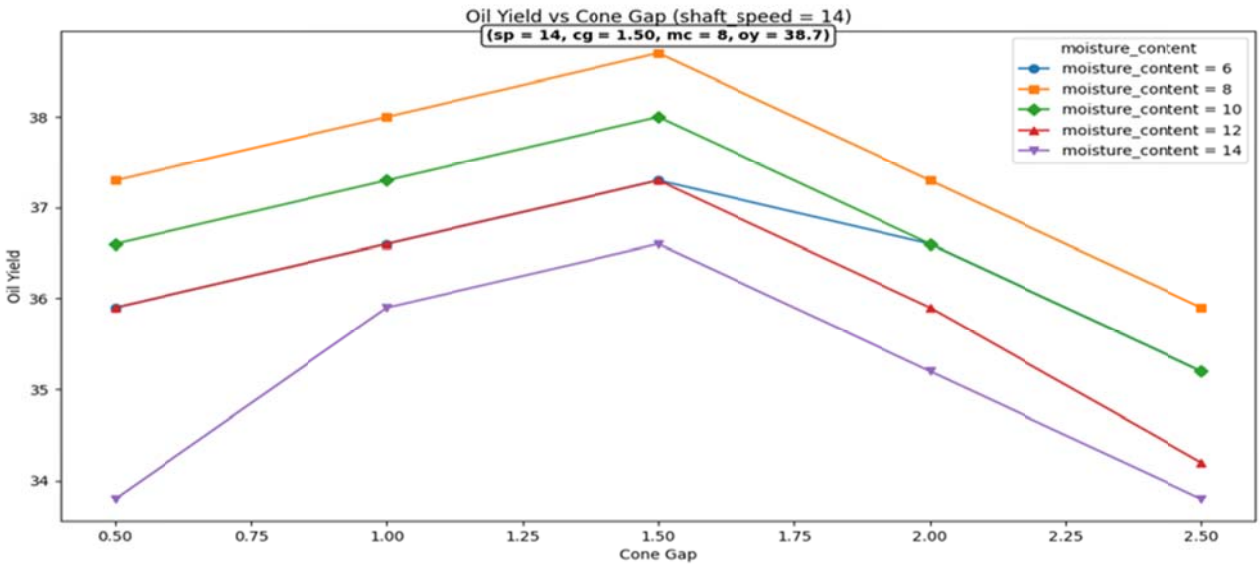

Figure 5: Oil yield versus cone gap at shaft speed = 14rpm and varying moisture content

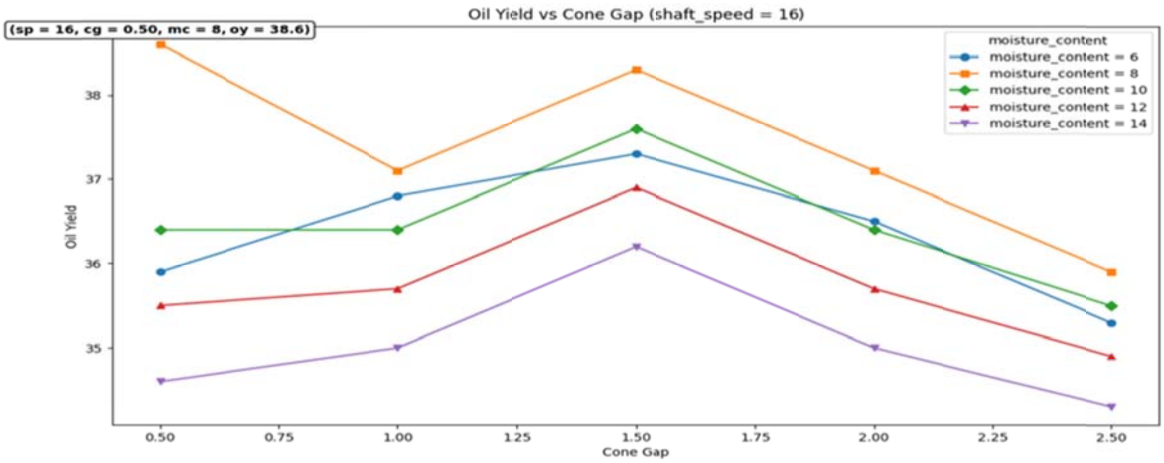

Figure 6: Oil yield versus cone gap at shaft speed = 16rpm and varying moisture content

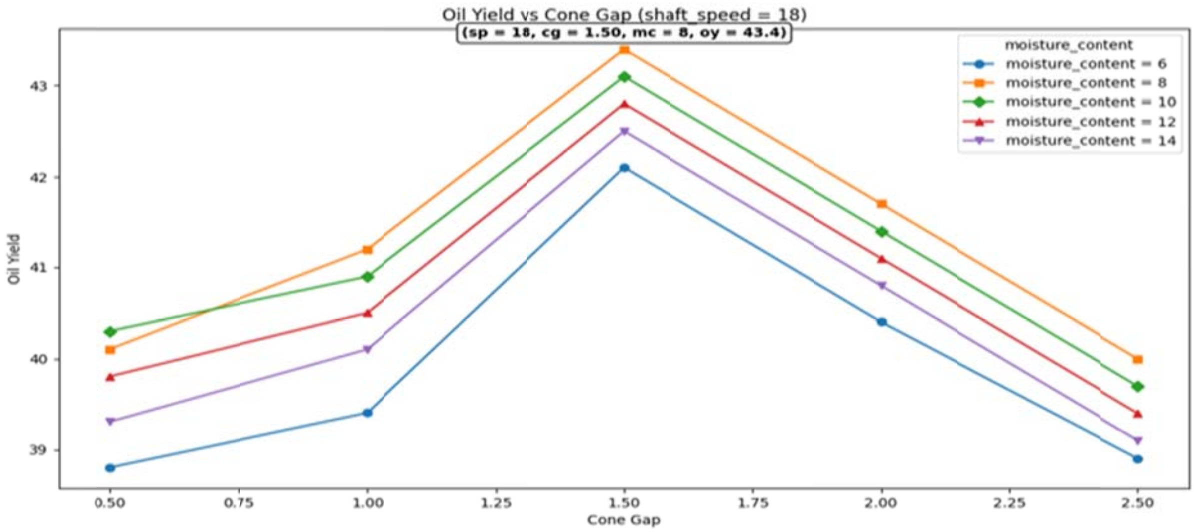

Figure 7: Oil yield versus cone gap at shaft speed = 18rpm and varying moisture content

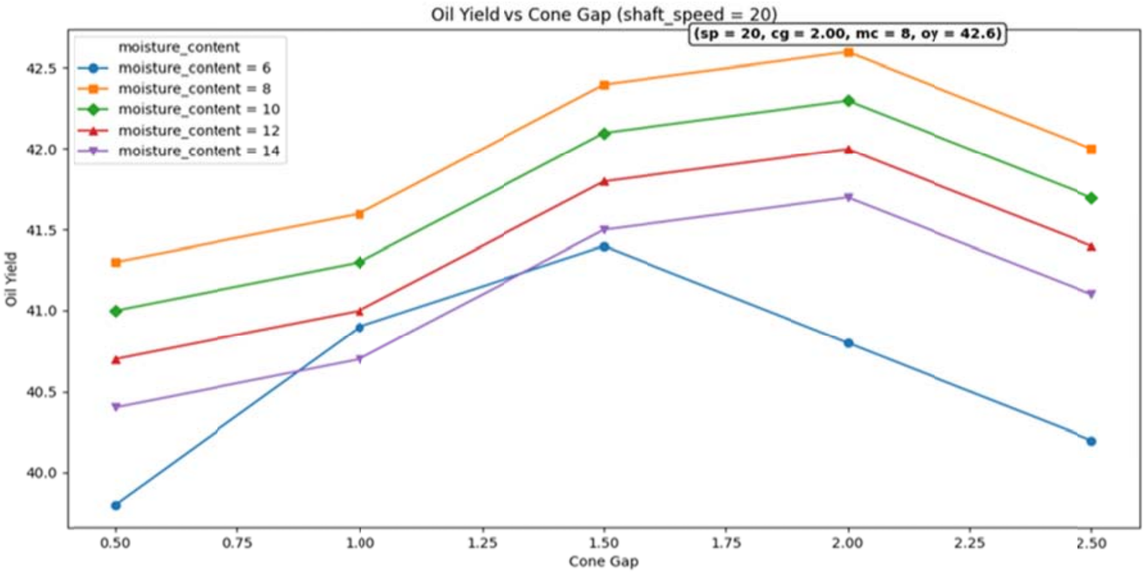

Figure 8: Oil yield versus cone gap at shaft speed = 20rpm and varying moisture content

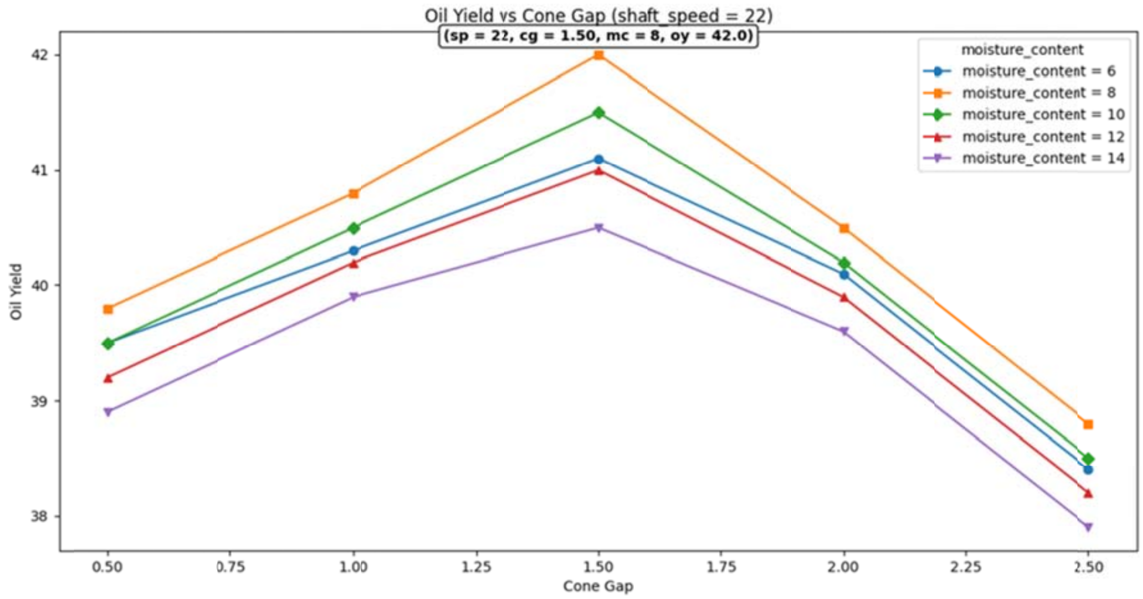

Figure 9: Oil yield versus cone gap at shaft speed = 22rpm and varying moisture content

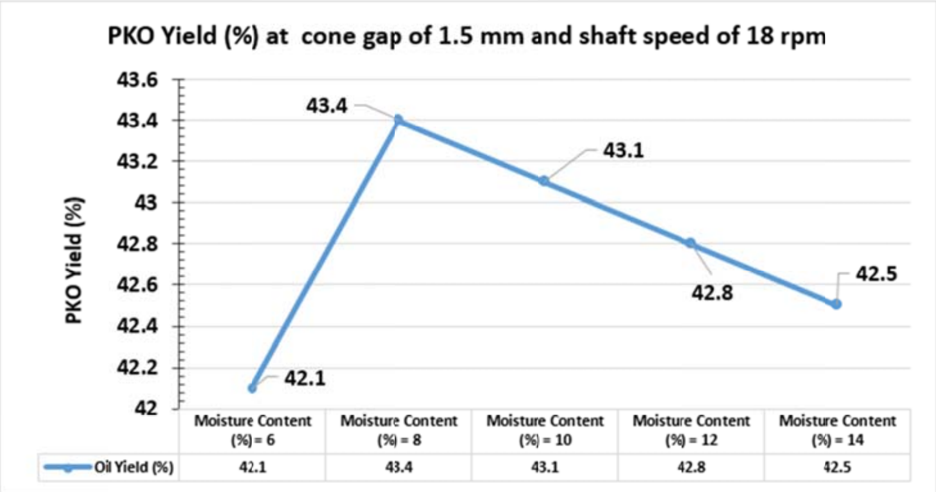

Figure 10 PKO yield versus palm kernel moisture content at cone gap of 1.5 mm and shaft speed of 18 rpm

4. Conclusion

The XGBoost model is presented for prediction of the palm kernel oil (PKO) yield from a machine design for PKO extraction. The case study machine has capacity of 10-ton palm kernel oil extraction with selected study parameters as shaft speed, cone gap and moisture content. The XGBoost model was trained with the augmented data originally obtained empirically from the 10-ton machine. The results showed the optimal PKO yield along with the input parameters setting that gave the optimal point. Also, the feature importance analysis indicated the moisture content as the most significant input parameter in the determination of the optimal oil yield.

References

1. Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2022). Artificial intelligence applications for industry 4.0: A literature-based study. *Journal of Industrial Integration and Management*, *7*(01), 83-111.
2. Yang, T., Yi, X., Lu, S., Johansson, K. H., & Chai, T. (2021). Intelligent manufacturing for the process industry driven by industrial artificial intelligence. *Engineering*, *7*(9), 1224-1230.
3. Lazaroiu, G., Androniceanu, A., Grecu, I., Grecu, G., & Neguriță, O. (2022). Artificial intelligence-based decision-making algorithms, Internet of Things sensing networks, and sustainable cyber-physical management systems in big data-driven cognitive manufacturing. *Oeconomia Copernicana*, *13*(4), 1047-1080.
4. Mishra, S. (2022). Reducing Points of Failure-a Hybrid and Multi-Cloud Deployment Strategy With Snowflake. *International Journal of AI, BigData, Computational and Management Studies*, *3*(1), 66-78.
5. Nti, I. K., Adekoya, A. F., Weyori, B. A., & Nyarko-Boateng, O. (2022). Applications of artificial intelligence in engineering and manufacturing: a systematic review. *Journal of Intelligent Manufacturing*, *33*(6), 1581-1601.
6. Li, B. H., Hou, B. C., Yu, W. T., Lu, X. B., & Yang, C. W. (2017). Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers of Information Technology & Electronic Engineering*, *18*(1), 86-96.
7. Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research*, *6*(4), 379-391.
8. Zhao, Z., Wu, J., Li, T., Sun, C., Yan, R., & Chen, X. (2021). Challenges and opportunities of AI-enabled monitoring, diagnosis & prognosis: A review. *Chinese Journal of Mechanical Engineering*, *34*(1), 56.
9. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, *6*, 52138-52160.
10. Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, *55*(5), 3503-3568.
11. Jagatheesaperumal, S. K., Pham, Q. V., Ruby, R., Yang, Z., Xu, C., & Zhang, Z. (2022). Explainable AI over the Internet of Things (IoT): Overview, state-of-the-art and future directions. *IEEE Open Journal of the Communications Society*, *3*, 2106-2136.
12. Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, *55*(5), 3503-3568.
13. Zacharias, J., von Zahn, M., Chen, J., & Hinz, O. (2022). Designing a feature selection method based on explainable artificial intelligence. *Electronic Markets*, *32*(4), 2159-2184.
14. Ezeoha, S. L. (2020). Effects of some kernel factors on palm kernel oil extraction using a screw press. *Agricultural Engineering International: CIGR Journal*, *22*(1), 156-161.
15. Ezeoha, S. L., & Akubuo, C. O. (2021). Influence of palm kernel variables on the yield and quality of oil expressed using an expeller. *Research in Agricultural Engineering*, *67*(2).