RESEARCH ARTICLE OPEN ACCESS

# Oil Formation Volume Factor Prediction Using Advanced Machine Learning Method

Mbachu Ijeoma Irene

(Petroleum and Gas Engineering Department, University of Port Harcourt, Rivers, Nigeria)

----------------------------------------************************----------------------------------

## Abstract:

Oil Formation Volume Factor(OFVF) is a very important fluid property in reservoir engineering computations. It is the volume of oil (and dissolved gas) at reservoir pressure and temperature required to produce one stock tank barrel of oil at the surface. It can be obtained either by conducting laboratory study on reservoir fluid samples or estimated, using empirically derived PVT correlations. Although laboratory results give better prediction where controlled conditions are imposed but in situation where the experimental data are not available, artificial intelligence and published empirical correlations are used. Unfortunately, the development of published empirical correlations have many drawbacks and limitations as they were originally developed for certain ranges of reservoir fluid characteristics. This research work aimed at using Extreme Gradient Boost (XG Boost) to predict oil formation volume factor as to address the limitations of empirical correlations and some of the existing artificial intelligent tools. 1402 data set which was obtained from PVT report from Niger-Delta was used for the study, 70% were used to train the model, 20% for testing and 10% for validation. Statistical analysis was carried out as to compare the predictions of newly trained artificial intelligent tool with the predictions from neural networks, support vector machine and some published empirical correlation techniques. The result revealed that the XG Boost performed better than the popularly used Feed Forward Back Propagation ANN, support vector machine and the empirical correlations in terms of quantitative and qualitative analysis employed. The model gave better prediction with highest correlation coefficient of 0.9849, Mean Absolute Error (MAE) of 1.2003, the best rank of 1.3149 with a superior performance plot.

***Keywords*** —Artificial Neural Network,Extreme Gradient Boost, Oil formation volume factor, statistical analysis, Support Vector Machine, Niger- Delta

----------------------------------------************************----------------------------------

## I. INTRODUCTION

The oil formation volume factor ($B_o$) plays a vital role in oil and gas engineering and is also one of the very important PVT properties. PVT properties are very crucial in estimation of reserves, determination of oil reservoir performance, recovery efficiency, production optimization and design of production systems at different conditions of pressure and temperature. Oil formation volume indicates the change in the volume of produced oil from the reservoir to surface conditions. In fact, the volume of oil that enters the stock tank under surface conditions is less than the volume of oil produced in reservoir conditions that enter the production well. The oil volume change (from reservoir to surface conditions) is most affected by the significant pressure reduction below the bubble point and the resultant release of dissolved gases in oil, especially in large amounts of solution gases. Therefore, the oil formation volume factor defined as below is always equal to or greater than 1 ([1], [2]). Oil formation volume factor is predicted using conventional methods such experimental tests, correlations, and Equations of State models but lately artificial intelligences are adopted because of their higher prediction accuracy. As a substitute to conventional black oil methods, the compositional oil method has been recently used for accurately predicting the oil formation volume factor. Although oil composition is essential for estimating this parameter, it is time-consuming and cost-intensive to obtain through laboratory analysis. Therefore, the inputparameter of dissolved gas in oil has been used as a representative of the number of light componentsin oil, which is an effective factor in determining oil volume changes, along with other parameters, including pressure, API gravity, and reservoir temperature. Several correlations have been proposed for determining crude oil formation volume factor such as [3], [4], [5], [6], [7] and [8].

To find the relationship between the input and output data driven from experiment, a powerful method than traditional modelling is necessary, hence, computational intelligence techniques, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Fuzzy logic, deep learning, super-leaners, Neural-Fuzzy and Extreme gradient boosting among others, have been applied recently.

Investigators has recognized that the neural network can serve the petroleum industry to create a more accurate PVT model ([9], [10], [11], [12], [13], [14]). [10] published neural network models for estimating bubble point pressure and oil formation volume factor for Middle East crude oils. They used two hidden layers neural networks to model each property separately. The bubble point pressure model had eight neurons in the first layer and four neurons in the second. The formation volume factor model had six neurons in both layers. Both models were trained using 498 data sets collected from the literature and unpublished sources. The models were tested by other 22 data points from the Middle East. The results showed improvement over the conventional correlation methods with reduction in the average error for the bubble point pressure oil formation volume factor.[11] presented an Artificial Neural Network (ANN) for estimation of PVT properties of compounds. The data set was collected from Perry's Chemical Engineers' Handbook. Different training schemes for the back propagation learning algorithm; Scaled Conjugate Gradient (SCG), Levenberg-Marquardt (LM) and Resilient back Propagation (RP) methods were used. The accuracy and trend stability of the trained networks were tested. The LM algorithm with sixty neurons in the hidden layer proved to be the best suitable algorithm with the minimum Mean Square Error (MSE) of 0.000606. ANN is one of the best estimating methods with high performance used in forecasting the PVT properties.

[13] developed a new artificial neural network model to predict oil formation volume factor using 802 data sets collected from the Niger Delta Region of Nigeria. One-half of the data was used to train the ANN models, one quarter to cross-validate the relationships established during the training process and the remaining one quarter to test the models to evaluate their accuracy and trend stability during the training process. Both quantitative and qualitative assessments were employed to evaluate the accuracy of the new model to the empirical correlation. The authors reported that the new ANN outperformed the best existing correlation by the statistical parameters used with a rank of 0.85 and better performance plot. Their trained model was also used in testing the accuracy of the new XG Boost developed in this study. [14] researched on building an artificial neural network (ANN) model to predict oil formation volume factor for the different API gravity ranges. The new models were developed using combination of 448 published data from the Middle East, Malaysia, Africa, North Sea, Mediterranean basin, Gulf of Persian fields and 1389 data set collected from the Niger Delta Region of Nigeria. The data was divided into the following four different API gravity classes: heavy oils for API $\leq$ 21, medium oils for 21>API$\leq$26, blend oils for 26>API$\leq$35 and light oils for API> 35. The data set was randomly divided into three parts of which, 60% was used for training, 20% for validation, and 20% for testing for each API grade. The ANN models outperformed the existing empirical correlations by the statistical parameters used with the best rank and better performance plots.

Through the literature, it can be found that Artificial Neural Network (ANN) model has gained ground but recently, some authors have started investigating other artificial intelligent as to address the shortcomings of ANN tool in predicting PVT properties. Lately, some researchers have investigated new machine learning algorithm such as XG Boosts, Neuro-Fuzzy Inference System (ANFIS), random Forest, Supper learn and Lightgbm etc rather than Artificial Neural Network model.[15] researched on the use of Support Vector Machine (SVM) to predict oil formation volume factor as to address the limitations of empirical correlations and Neural Network models. The authors used 1402 data set from PVT report from Niger-Delta. They used 70% to train the model, 20% for testing and 10% for validation. Samuel and Mbachu compared their work with some of the existing correlations and ANN model. Their result revealed that the support Vector Machine Model performed better than the popularly used Feed Forward Back Propagation ANN and the empirical correlations in terms of quantitative and qualitative analysis used. The new trained SVM performed best for the Niger-Delta Crude with highest correlation coefficient of 0.9812, Mean Absolute Error (MAE) of 1.2895, the best rank of 0.606267 with a better performance plot.

[16] did a study on oil formation volume factor using 1241 PVT data from Iran's oil reservoirs. This study created machine learning models utilizing Gradient Boosting Decision Tree (GBDT) techniques, which also incorporated Extreme Gradient Boosting (XGBoost), GradientBoosting, and CatBoost. The authors compare the results with recent correlations and machine learning methods adopting a compositional approach by implementing tree-based bagging methods: Extra Trees (ETs), Random Forest (RF), and Decision Trees (DTs), is then performed. From their statistical and graphical analysis, they reported that the XGBoost model outperforms the other models in estimating the oil formation volume factor across the reservoir pressure region.[17] used multi-layer neural network to prediction oil formation volume factor via the anaconda programming environment 2994 dataset from the Niger Delta region and open literature was used. The dataset was randomly divided into three parts of which 60% was used for training, 20% for validation, and 20% for testing. Itoro reported that the new model performed better than the existing correlations by the statistical parameters used for the same set of field data with mean average error of 0.05. It can be found from the literature that petroleum and gas engineers are still exploring the best way to predict PVT properties. This study aims at using an advanced machine learning method to predict oil formation volume factor using Niger Delta formation data.

## 2. EXTREME GRADIENT BOOSTING (XG Boost)

Extreme Gradient Boosting also known as XG Boost is the machine learning algorithm adopted in this study. Machine learning is major aspect of Artificial Intelligence (AI) that gives it capability to learn a pattern. Extreme Gradient Boost is scalable, distributed gradient-boosted decision tree (GBDT) machine learning library that helps to understand data and make better decisions [18]. The learning process is achieved by using algorithms to discover patterns and generate insights from the original or measured data they are exposed to (Fig. 1). It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking problems. The XG Boost algorithms builds on supervised machine learning, decision trees, ensemble learning, and gradient boosting.

Supervised machine learning uses algorithms to train a model to find patterns in a dataset with labels and features and then uses the trained model to predict the labels on a new dataset's features Fig. 1.

Decision trees create a model that predicts the label by evaluating a tree of if-then-else true/false feature questions and estimating the minimum number of questions needed to assess the probability of making a correct decision. Decision trees can be used for classification to predict a category, or regression to predict a continuous numeric value.

A Gradient Boosting Decision Trees (GBDT) is a decision tree ensemble learning algorithm similar to random forest, for classification and regression. Ensemble learning algorithms combine multiple machine learning algorithms to obtain a better model. Both random forest and GBDT build a model consisting of multiple decision trees. The difference is in how the trees are built and combined. The term "gradient boosting" comes from the idea of "boosting" or improving a single weak model by combining it with several other weak models to generate a collectively strong model. Gradient boosting is an extension of boosting where the process of additively generating weak models is formalized as a gradient descent algorithm over an objective function. Gradient boosting sets targeted outcomes for the next model to minimize errors. Targeted outcomes for each case are based on the gradient of the error (hence the name gradient boosting) with respect to the prediction. GBDTs iteratively train an ensemble of shallow decision trees, with each iteration using the error residuals of the previous model to fit the next model. The final prediction is a weighted sum of all the tree predictions. Random forest "bagging" minimizes the variance and overfitting, while GBDT "boosting" minimizes the bias and underfitting.

XG Boost is a scalable and highly accurate implementation of gradient boosting that pushes the limits of computing power for boosted tree algorithms, being built largely for energizing machine learning model performance and computational speed. With XG Boost, trees are built in parallel, instead of sequentially like GBDT. It follows a level-wise strategy, scanning across gradient values and using these partial sums to evaluate the quality of splits at every possible split in the training set.

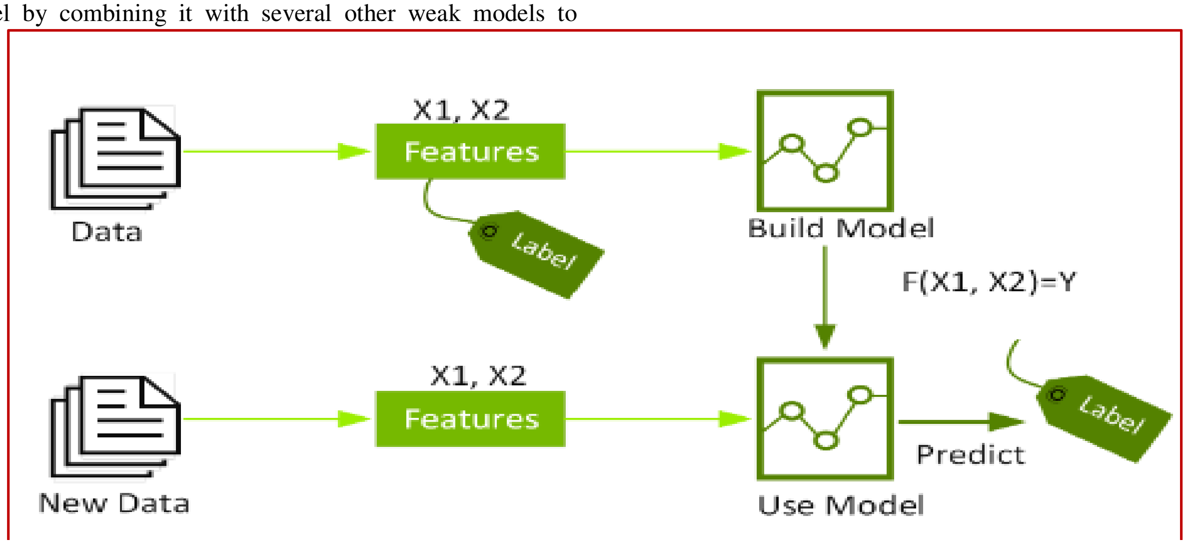


*Fig. 1.Machine learning Process*

## 3. Methodology

### 3.1 Data Description

The data used was obtained from conventional PVT reports that derive the various fluid properties through liberation process from the Niger-Delta Region of Nigeria. The data parameters include Oil density ($\gamma_o$), Reservoir temperature ($^oF$), Solution Gas Oil Ratio ($R_s$-scf/stb), Gas density ($\gamma_g$) and Oil Formation Volume Factor (OFVF). The maximum, minimum and mean values of data used for training, test and validation are shown in Tables 1, 2 and 3.

**TABLE 1**. Summary of maximum, minimum and mean values of training data for oil formation volume factor

| PVT Properties | Maximum Values | Minimum Values | Mean Values |
|---|---|---|---|
| $\gamma_g$ | 1.3710 | 0.531 | 0.68952 |
| $\gamma_o$ | 0.9667 | 0.7553 | 0.867082 |
| T(°F) | 264.00 | 104.40 | 174.2746 |
| Rs (scf/stb) | 3299.0 | 7.0000 | 369.557 |
| Ofvf | 3.6708 | 1.0130 | 1.335471 |

**TABLE 2**. Summary of maximum, minimum and mean values of validation data for oil formation volume factor

| PVT Properties | Maximum Values | Minimum Values | Mean Values |
|---|---|---|---|
| $\gamma_g$ | 1.294 | 0.5310 | 0.613763 |
| $\gamma_o$ | 0.9570 | 0.7949 | 0.894722 |
| T(°F) | 218.0 | 122.3000 | 151.1921 |
| Rs (scf/stb) | 389.0 | 14.000 | 162.6112 |
| Ofvf | 1.1500 | 1.0810 | 1.113211 |

**TABLE 3.** Summary of maximum and minimum values of test data for oil formation volume factor

| PVT Properties | Maximum Values | Minimum Values | Mean Values |
|---|---|---|---|
| $\gamma_g$ | 1.3710 | 0.5630 | 0.6688 |
| $\gamma_o$ | 0.9667 | 0.7553 | 0.8671 |
| T(°F) | 264.00 | 104.40 | 166.60 |
| Rs (scf/stb) | 2948.8 | 7.000 | 246.722 |
| Ofvf | 2.8905 | 1.0165 | 1.2566 |

### 3.2 Data Validation

Before any experimental PVT data are used for design or study purposes, it is necessary to ensure that there are no error or major inconsistencies that would render any subsequent work useless. Two such means of data validation are the Campbell diagram (Buckley plot) and the Mass Balance Diagram which are otherwise known as cross plot. These techniques were used to validate the data set used in this work.

### 3.3 Modelling Technique

Support vector machine regression was used to build the Oil Formation Volume Factor model using cost functions which incorporate first-order Taylor information procedure using MATLAB (2021) version. The importation and selection of data are as follows.

Import the data; the input data was imported to the MATLAB environment using the import command. The following variables: Solution GOR (scf/stb), Reservoir temperature ($^o$F), Gas relative density, Oil relative density, were imported into the MATLAB environment.

Select the variables; this is to arrange a set “P input” vector and “T output” vectors as columns into first and second matrix in the MATLAB workspace as follows.

(P) Input data = [GOR;T; $\gamma_g$ ; $\gamma_o$ ] (1)

(T) Target data = [OFVF] (2)

The Extreme Gradient Boosting (XGBoost) algorithm, which was developed in this studywas introduced by[19] belongs to modern Machine Learning techniques based on Gradient Boosting Decision Trees. The algorithm focuses to minimize errors and maximize adaptability by creating many trees as to approximate the estimated value as closely as possible. By combining weak learners, the algorithm builds a strong learner. However, in this algorithm, weak learners are constructed through residual fitting [19, [20]). The XGBoost model extends the cost function by incorporating first-order Taylor information and presenting second-order derivative information. This enhancement enables faster convergence during the learning process. In addition,the XGBoost algorithm includes a regularization component in the cost function, which helps control complexity and reduces the risk of overfitting. For a more detailed understanding of the general process of the XGBoost algorithm [19].

### 3.3 Correlation Comparison

To compare the performance and accuracy of the new model to other empirical correlations, two forms of analyses were performed which are quantitative and qualitative screening. For quantitative screening method, statistical error analysis was used. The statistical parameters used for the assessment were percent mean relative error (MRE), percent mean absolute error (MAE), percent standard deviation relative

(SDR), percent standard deviation absolute (SDA) and correlation coefficient (R).

For correlation comparison, a new approach of combining all the statistical parameters mentioned above (MRE, MAE, SDR, SDA and Rank) into a single comparable parameter called Rank was used [7]. The use of multiple combinations of statistical parameters in selecting the best correlation can be modelled as a constraint optimization problem with the function formulated as;

$$Min \quad Rank \quad = \sum_{j=1}^{m} S_{i,j} q_{1,j} \tag{3}$$

Subject to $$\sum_{i=1}^{n} S_{i,j} \tag{4}$$

With $$0 \le S_{ij} \le 1 \tag{5}$$

Where $S_{i,j}$ is the strength of the statistical parameter j of correlation i and $q_{ij}$, the statistical parameter j corresponding to correlation i. j = MRE, MAE, …. $R^1$, where $R^1 = (1\text{-}R)$ and $Z_i$ is the rank, (or weight) of the desired correlation. The optimization model outlined in equations 3 to 5 was adopted in a sensitivity analysis to obtain acceptable parameter strengths. The final acceptable parameter strengths so obtained for the quantitative screening are 0.4 for MAE, 0.2 for R, 0.15 for SDA, 0.15 for SDR, and 0.1 for MRE. Finally, Equation 5 was used for the ranking. The correlation with the lowest rank was selected as the best correlation for that fluid property. It is necessary to mention that minimum values were expected to be best for all other statistical parameters adopted in this study except R, where a maximum value of 1 was expected. Since the optimization model (Equations 3 to 5) is of the minimizing sense a minimum value corresponding to R must be used. This minimum value was obtained in the form (1-R). This means the correlation that has the highest correlation coefficient (R) would have the minimum value in the form (1-R). In this form the parameter strength was also implemented to 1-R as a multiplier. Ranking of correlations was therefore made after the correlations had been evaluated against the available database.

For qualitative screening, performance plots were used. The performance plot is a graph of the predicted versus measured properties with a $45^o$ reference line to readily ascertain the correlation's fitness and accuracy. A perfect correlation would plot as a straight line with a slope of $45^o$.

## 4. RESULTS AND DISCUSSION

After training XG boost model with training data, its performance was tested with 20% of test data (280 data) points that were not previously used during training and validation. These data were randomly selected by the XG boost tool to ascertain the accuracy and stability of the model. The performance of the XG boost model was compared with predictions from Artificial Neural Network model trained byAzubuike and Ikiensikimama[12], Support Vector Machine trained by Samuel and Mbachu [15],and some selected empirical correlations such as [5], [6], [7] and [3].These predictive correlations were carefully selected, having been developed specifically for the prediction of oil FVF and [8] and [5] were developed specifically for Niger Delta.

### 4.1 Quantitative Results

The results of the statistical assessment as presented in Table 4 and Fig. 2 gives the statistical accuracies for all the oil formation volume factor correlations and ANN model examined. The results show that the XG Boost algorithm has both reliable and efficient performance as to compare to other existing correlations and ANN model and support vector machine. Table 4 shows the numerical values of the models accessed with XG Boost having the best rank of 1.3149 with Mean absolute Error (MAE) of 1.2002 and correlation coefficient (R) of 0.9849 followed by [15] which has the Rank value of 2.5036 and correlation coefficient of 0.9603. Among the three machine learning algorithms studiedANN [12] ranked the least with the numerical values of 4.0695. All the artificial intelligent evaluated performed better than all the empirical correlations studied. Among all the empirical correlations studied [7] which is Niger Delta oil formation volume factor model performed better than others with a Rank value of 8.4840 and correlation coefficient of 0.959.[5]is also oil formation volume factor correlation developed using Niger Delta data. Its performance was not impressive, which might be from the data set they employed. This study recommends[7]as a good predictive model for oil formation volume factor for Niger Delta region in absence of the machine learning algorithm developed in this study.

[6] and [3]are foreign correlations accessed. Between these two foreign correlations,[6]performed better than the other evaluated equations with a rank of 9.2742 and[3] has theRank of 10.5689.[6]correlation can be used to forecast oil formation volume factor for Niger Delta region in absence of this newly developed intelligent model and [7] correlation. This study showed again the supremacy of machining learning in predicting reservoir PVT properties particularly in applying XG Boosts algorithm.

  

**TABLE 4.** Statistical Accuracy of Oil Formation Volume Factor Using Niger-Delta Data

| AUTHORS | %MRE | %MAE | %SDR | %SDA | R | Rank |
|---|---|---|---|---|---|---|
| XG Boost | 0.43140 | 1.2002 | 2.1772 | 1.6062 | 0.9849 | 1.3149 |
| SVM - Samuel and Mbachu (2021) | 3.38610 | 3.4865 | 2.2412 | 2.8310 | 0.9603 | 2.5036 |
| ANN - Azubuike and Ikensikimama (2012) | 6.14050 | 6.1405 | 3.2311 | 3.2311 | 0.9700 | 4.0695 |
| Ikiensikimama (2009) | 11.2109 | 11.2109 | 9.4601 | 9.4601 | 0.9594 | 8.4840 |
| Omar and Todds (1993) | 12.0765 | 12.0765 | 10.643 | 10.643 | 0.9570 | 9.2742 |
| Standing (1947) | 14.8781 | 14.8781 | 10.3531 | 10.3531 | 0.9760 | 10.5689 |
| Obomanu and Okpobiri (1987) | 25.5150 | 25.515 | 20.515 | 20.515 | 0.6400 | 19.272 |

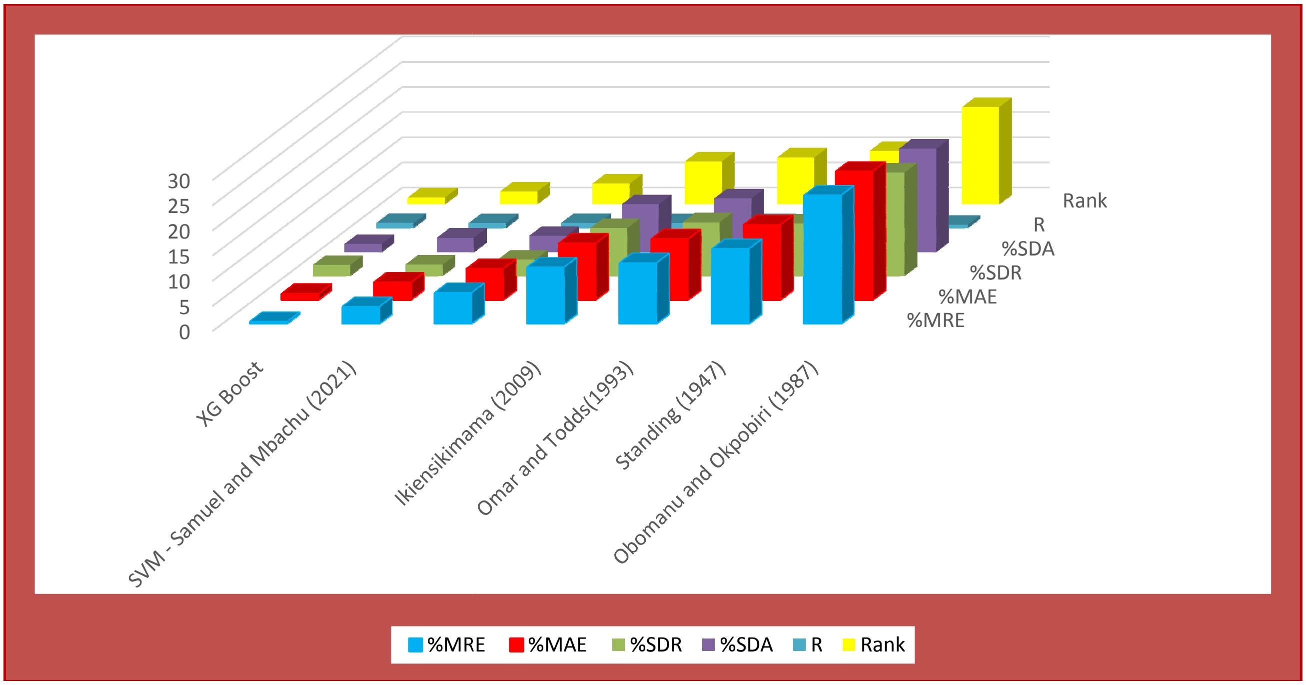


Fig. 2. Comparison of the Statistical Accuracy for Different correlations using Niger -Delta Data

## 4.3 Performance Plot Results

Figs. 3 to 7 illustrate cross plots of the predicted versus experimental oil formation volume factor (OFVF) values. A cross plot is graph of predicted versus measured properties with a 45° reference line to readily ascertain the correlation's fitness and accuracy.

Compared to other cross plots, Fig.7which is Extreme gradient boost model shows the tightest cloud of points around the 45° line with very good clusters at low band, indicating the excellent agreement between the experimental and the calculated data values when compared to Figs. 3 to 6 which are predictions from other models. In addition, this indicates the superior performance of XG Boost to Support vector Machine, artificial neural network model over empirical correlations evaluated. The accuracy of the model indicates that the Extreme gradient boost model does not over fit the data, which implies that it was successfully trained.

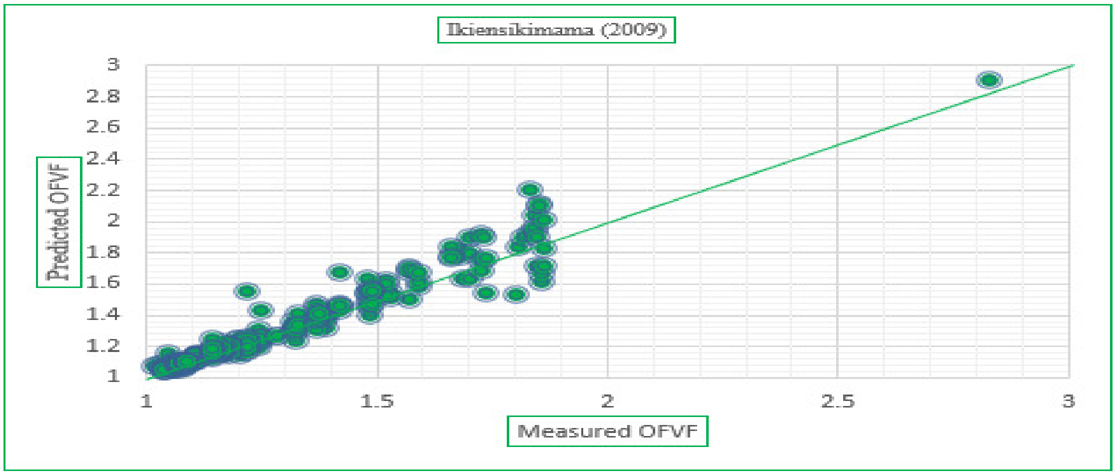


**Fig. 3.** Cross plot of Ikiensikimama (2009) Correlation using Test Data

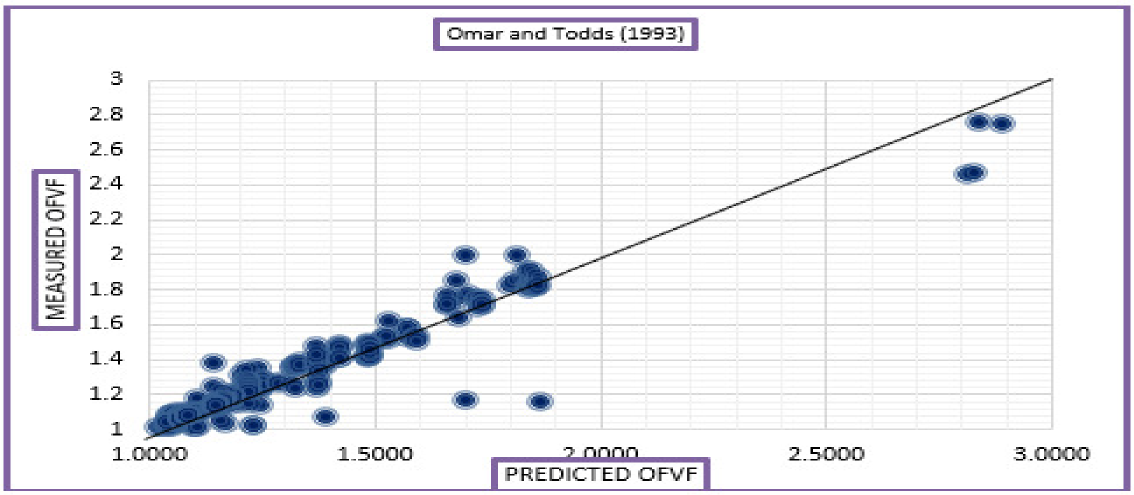


**Fig. 4.** Cross plot of Omar and Todd (2013) correlation using Test Data

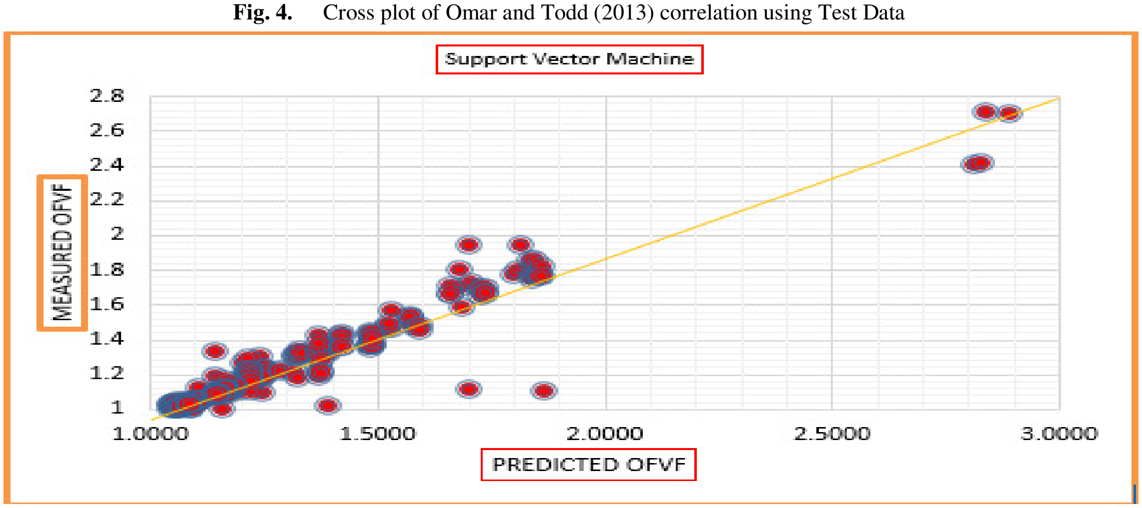


**Fig. 5**. Cross plot of Support Vector Machine- Samuel and Mbachu(2021) using Test Data

  

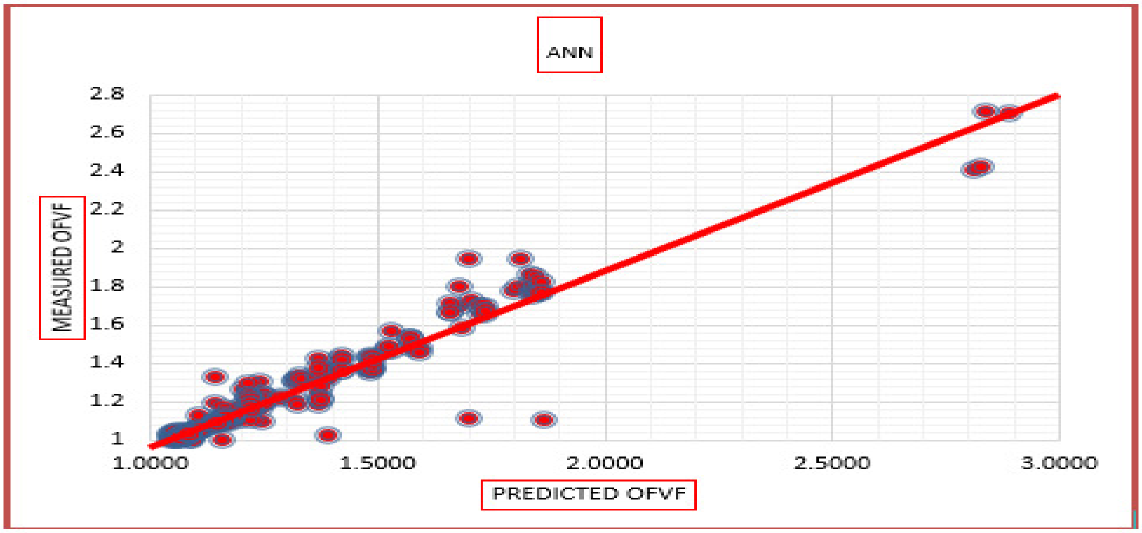


Fig. 6. Cross plot of ANN- Azubuike and Mbachu (2012) using Test Data

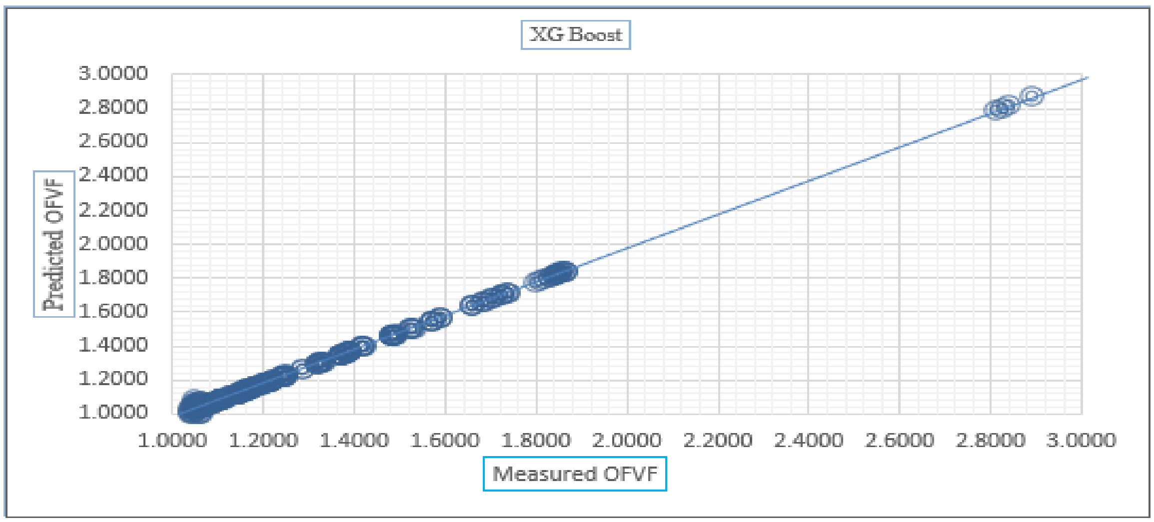


**Fig. 7.** Cross plot of XG Boost (This Study) using Test Data

## 5. CONCLUSION

The newly proposed Extreme Boost model for predicting crude oil formation volume factor for Niger-Delta region was developed in this study using MATLAB 2021 Version. The cost function imbedded in the Extreme gradient boost was used to estimate the model parameters. The new tool outperformed the existing correlations as well as the other Artificial intelligent model studied by the statistical parameters used. It shows a best rank with a numerical value of 1.3149and better performance plot as compared to the existing empirical correlations for those regions where the data was used. This leads to a bright light of support vector machines modeling and will assist petroleum exploration engineers to estimate various reservoir properties with better accuracy, leading to reduced exploration time and increased productions.

.

## Reference

[1]. McCain, W. D. (1993). *Properties of Petroleum Fluids*.

[2]. Dake, L. P. (1983). *Fundamentals of Reservoir Engineering*, Elsevier

[3]. Obomanu,D. and OkpobiriG. A, (1987). Correlation the PVT Properties of the Nigeria Crudes, *Trans. ASME Vol. 109*, 14-24.

[4]. Al-Marhoun, M. A. (1988). "PVT Correlations for Middle East Crude Oils," *J. Pet. Tech,*, pp. 650-666.

[5]. Obomanu, D.A. and Okpobori, G.A. (1987). Correlating the PVT Properties of NigerianCrudes. Trans. ASME 109, 214–224.

[6]. Omar,M. I. and Todd, A. C., (1993). Development of Modified Black oil Correlation for Malaysian Crudes, *Paper SPE 25338, presented at the SPE Asia Pacific Oil and Gas Conference, Singapore*.

[7]. Ikiensikimama, S. S (2009), *Reservoir Fluid Property Correlations*, Chi Ikoku Petroleum Engineering Series, IPS Publications, Port Harcourt..

[8]. Al-Marhoun, M. A. and E. A. Osman, Using Artificial Neural Networks to Develop New PVT Correlations for Saudi Crude Oils, *Paper SPE 78592, presented at the 10th Abu Dhabi International Petroleum Exhibition and Conference (ADIPEC), Abu Dhabi, UAE*, 2002.

[9]. Gharbi,R. B and E. A El-sebakhy, (1997). "Neural-Network Model for Estimating the PVT Properties of Middle East Crude Oils," *SPE 37695 presented at the SPE Middle East Oil Show and Conference, Bahrain*.

  

[10]. Mohammad, R. and Mohammad, A new heuristic model for estimating the oil formation volume factor. Faculty of Petroleum Engineering, Amirkabir University of Technology (Tehran Polytechnic), Hafez Avenue, Tehran, Iran.

[11]. Omole, O. A. Falode, and A. D. Deng,(2009). Prediction of Nigerian Crude Oil Viscosity Using Artificial Neural Network, *Petroleum and Coal ISSN 151(3),* 181-188.

[12]. Azubuike, I. I. and S. S. Ikiensikimama, (2012). Modelling Approach for Niger Delta Oil Formation Volume Factor Prediction using Artificial Neural Network" *Refereed Proceedings, Paper SPE 162987. 36th Annual International Conference of the SPE Nigerian Council,, Expo Centre, Eko Hotel and Suites, Victoria Island, Lagos, Nigeria.* 2012.

[13]. Azubuike, I. I. and S. S. Ikiensikimama, Forcasting Oil Formation Volume Factor for API Gravity Ranges Using Artificial Neural Network, *Advances in Petroleum Exploration and Development Journal Vol. 5, No. 1,* 2013, pp 1-9.

[14]. Anifowose, F. (2011). "Prediction of Oil and Gas properties using Support Vector Machine" IPTC 14514.

[15]. Shaibu, S. and Mbachu I. I. (2021). "Modeling Approach for Niger Delta Oil Formation Volume Factor Prediction using Support Vector Machining Learning". American Journal of Engineering Research (AJER), vol.10, Is. 12, p. 80 -88.

[16]. Parsa, K. E., Kiana, P. A. L and Mohammad, R. K. M. (2023). Enhanced machine learning—ensemble method for estimation of oil formation volume factor at reservoir conditions, Scientific Reports, https://doi.org/10.1038/s41598-023-42469-4

[17], Itoro, K. (2021). A Deep Learning Approach for the Prediction of Oil Formation Volume Factor, SPE-208627-STU,DOI: 10.2118/208627-STU

[18]. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232.

[19]. Chen, T. Yang,H. Li, Q. and Yu, Y. (2013)."General functionalmatrix factorization using gradient boosting. In Proceedingof 30th International Conference on Machine Learning". (ICML'13), volume 1, pp. 436{444.

[20]. Xiao, Z. and Luo, A. L. (2019). XGBoost based stellar spectral classification and quantized feature. *Spectrosc. Spectral Anal.* **39**, 3292–3296.

  