Original Research Article

PREDICTING BUBBLE POINT PRESSURE AND GAS-OIL RATIO PVT PROPERTIES USING BAGGING ENSEMBLE TECHNIQUES

Abstract

An enhanced accurate predictive model has been developed for the estimation of reservoir oil Pressure-Volume-Temperature (PVT) properties of Bubble Point Pressure (BPP) and Gas-Oil Ratio (GOR) using Bagging Ensemble machine learning. To develop the Bagging ensemble (BE) model, three different estimators, Decision Tree Regressor, Random Forest Regressor, and Extra Trees Regressor, were used as the base estimators. An averaging method to finally predict the model performance was done using a voting regressor to fit the base estimators. Hyper-parameter tunings for optimization were determined using cross-validation grid search and the implementation of Bagging ensemble described. The ensemble methods were compared with those developed using Artificial Neural Network (ANN) and some selected empirical correlations. Their performances were evaluated using Average Absolute Percentage Relative Error (AAPRE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and graphical cross-plot error analysis Correlation Coefficient (R^2). Result showed improved performances, notably the R^2 for BPP, the BE is 96.6%, ANN is 89.04%, and the best empirical model is 88.1%. For the GOR, the R^2 for BE and ANN are 94.1% and 89.0%, respectively while best empirical model is 88.1%.

Keywords: Machine Learning, Ensemble, PVT, Bagging

1.0 Introduction

Many researchers have in recent times applied Artificial Intelligence (AI)/Machine Learning (ML) algorithms in oil and gas industry, such as in areas of drilling and completion optimization, reservoir characterization and simulation, reservoir management, facility maintenance, production forecasting and optimization. Ensemble machine learning technique combines the predictions from multiple learners or machine learning models that are weak to produce a better prediction on a new sample either by voting or averaging. Ensemble learning when compared to other machine learning models, give an improve reliable and stable prediction performance, Louppe and Geurts [1]. The types of Ensemble learning are Bagging, Boosting, and Stacking. Bagging works by taking random subsets in an original dataset, with replacement so that the individual data points can be chosen more than once to fit either a classifier or regressor, to each subset dataset, applicable for classification and regression purpose. Boosting Ensemble deploys an iterative process where the training set is reweighted at each interaction based on errors a weak learner made in the first model. Stacking ensemble method creates a linear combinations of different predictors to improve the accuracy of a prediction.

Pressure-Volume-Temperature (PVT)

PVT analysis is the process of determining and predicting fluid behaviors and properties of oil and gas samples of an existing well and it is an integral part in understanding flow of hydrocarbon fluids from the well Standing [2]. The properties can be determined either experimentally in the laboratory or by empirical PVT correlations. The laboratory measurements are often not available or too expensive to obtain, while the empirical correlations differ from one geographical location to another. Tables 1. and 2. shows some selected regional and global empirical Pressure-Volume-Temperature (PVT) empirical correlations in the literature for fluid properties to determine reservoir performance, estimate reserves, make real-time decision. Some published work using machine learning to predict PVT properties include the works of Oladipo and Johnathan [3] where a novelty approach design was implemented to predict crude-oil PVT properties with a more cost effective and accurate result obtained. Ehsan *et. al.* [4] used advanced computational frameworks and ANN to model the viscosity of light and dead oil, and compare them with existing empirical models. Mohammed *et. al.* [5] used 369 data points to combine neural network and Neuro-fuzzy logic to model to predict coefficient of isothermal oil compressibility. Kingsley and Akinsete [6] developed a predictive model to improved PVT properties of Niger Delta crude oil using KNN, Random Forest, and Linear regression. Isemin and Akinsete [7] presented the use of Support Vector Machine to predict gas-saturated and undersaturated oil viscosities using data collected from the Niger Delta field. Yingxian *et. al*. [8] leverage machine learning algorithms to predict fluid properties based on Big Data using data from 100 wells on Bohai Oilfield.

This work focuses on the application of Bagging Ensemble models to predict the bubble point pressure and gas-oil-ratio fluid properties of a Niger Delta crude oil. A standalone ANN algorithm is also used to develop a predictive model for predictions, the results compared with some selected empirical correlations.

Bagging Ensemble Machine Learning

Bagging ensemble method also known as bootstrap aggregating, is an ensemble method that is designed by Leo Breiman [9] to improve the stability, accuracy, prediction, and performance of machine learning algorithms. The bootstrap method uses either the Out-of-Bag technique or the Leave-One-Out Bootstrap technique. Typically, a Bagging Ensemble machine learning process include cleaning and preparing data, choosing base models, splitting the data into training and test sets, instantiating and training models, and finally evaluating the model. The Bagging algorithm by Breiman is as follows;

Step 1: construct a bootstrap sample $(X_1^*, Y_1^*), \dots., (X_n^*, Y_n^*)$ by randomly drawing n times with replacement from the data $(X_1^*, Y_1^*), \dots., (X_n^*, Y_n^*)$

Step 2: compute the bootstrapped estimator $\hat{y}^*(\cdot)$ by the plug-in principle: $\hat{y}^*(\cdot) = h_n((X_1^*, Y_1^*), \dots., (X_n^*, Y_n^*))(\cdot)$.

Step 3: repeat steps 1 and 2 M times, where M is often chosen as 50 or 100, yielding $\hat{y}^{*k}(\cdot)$ $(k = 1, \dots, M)$. The bagged estimator is $\tilde{\hat{y}}\, Bag(\cdot) = M^{-1} \sum_{k=1}^{M} \hat{y}^{*k}(\cdot)$.

Artificial Neural Network (ANN)

ANN in machine learning are computational algorithm that are biologically inspired, non-digital systems designed to simulate in same manner how the human brain actually process and analyzes information. Artificial neural network computing tools Is made up of neurons consisting of simple interconnected elements. The neurons are a simple adaptive processing unit and consists of an information-processing unit which is fundamental to the operation of a neural network with nonlinearities in computations, interconnected neurons that forms a large network.

Mathematically, for an ANN model, the relationship between an input and output is given as

$$y_k = f_o\left[\sum_j W_{kj}\, f_h\left(\sum_i W_{ji} X_i + b_j\right) + b_k\right]$$

Where,

X = input vector
W_{ji} = connection layer in the hidden layer
b_j = threshold value or bias of the jth hidden neuron
W_{kj}= connection weight from the jth neuron in the hidden layer to the kth neuron in the output layer
b_k = bias of the kth output neuron
$f_o,\ f_o$ = activation functions applied to the weighted sum of the inputs

Table 1. – Existing Empirical Correlations for Bubble Point Pressure

Author	**Empirical Correlation**	**Origin of Samples**	**Oil Mixture Ranges**
Standing (1947)	$P_b = \phi\left(\left(\frac{GOR}{\gamma_g}\right)^{0.83} \cdot \frac{10^{0.00091\,T}}{10^{0.0125\,API}}\right)$	California oil fields	16 – 64^{O}API
Vazquez & Beggs (1980) [10]	$P_b = \left[A\left(\frac{R_s}{\gamma_{gc}}\right)10^{a}\right]^{c}$	Worldwide	$\gamma \leq 30^{o} API\ and\ \gamma_o \leq 30^{o} API$
Al-Marhoun (1988) [11]	$p_b = 5.38088 \times 10^{-3} R_S^{0.715082} \times \gamma_g^{-1.877840} \times \gamma_o^{3.143700} \times T^{1.326570}$	Middle East	14 – 45^{o} API
Petrosky & Farshad (1988) [12]	$p_b = 112.727\left(\frac{R_S^{0.5774}}{\gamma_g^{0.8439}} \times 10^{x} - 12.34\right)$	Gulf of Mexico	16 – 45^{o} API
De Ghetto (1994) [13]	$P_b = 31.7648\left[\left(\frac{R_s}{\gamma_g}\right)^{0.7857} \cdot \frac{10^{0.0009 \cdot T}}{10^{0.0148 \cdot API}}\right]$	Mediterranean Basin, Africa & Persian Gulf	10 < ^{O}API < 22.3
Dindoruk and Christian (2004) [14]	$P_{bp} = a_8\left(\frac{R_s^{a9}}{\gamma_g^{a10}}10^{A} + A_{11}\right)$	Middle East	
Ikiensikimama & Ogboja (2009) [15]	$\rho_b = \frac{(\rho_b^{*})\,(T + X10)}{\gamma_g}$	Niger Delta	

Table 2. – Existing Empirical Correlations for GOR

Author	Empirical Correlation	Sample Origin	Oil Mixture Ranges
Standing (1947)	$R_S = \left(\left(\frac{p}{18.2}+1.4\right).\frac{10^{0.0125}\gamma API}{10^{0.0091}T}\right)^{\frac{1}{G}}.G$	California oil fields	16 – 64^{O}API
Vazquez & Beggs (1980)	$R_s = C_1\gamma_g p^{c_2} exp\left(C_3\left(\frac{\gamma_{API}}{T+459.7}\right)\right)$	Worldwide	$\gamma \leq 30^o API\ and\ \gamma_o \leq 30^o API$
Al-Marhoun (1988)	$B_o = 0.497069 + 0.862963 \times 10^{-3}T + 0.812594 \times 10^{-2}F + 0.318099 \times 10^{-5}F^2$	Gulf of Mexico	14 – 45^{o} API
Petrosky and Farshad (1988)	$R_s = \left(\left(\frac{p}{112.727}+12.34\right)\times\gamma_g^{1.541}\times 10^x\right)^{1.73184}$	Gulf of Mexico	16 – 45^{o} API
De Ghetto (1994) [16]	$R_s = 0.101347 \cdot (\gamma_{gcorr})^{0.3873}\cdot P_b^{1.1715} \cdot 10^{(12.753\cdot API/(T+460))}$	Mediterranean Basin, Africa & Persian Gulf	10 < ^{O}API < 22.3
Dindoruk and Christian (2004)	$A = \frac{a_1T^{a2}+a_3API^{a4}}{\left(a_3+\frac{2R_s^{a6}}{\gamma_g^{a7}}\right)^2}$	Gulf of Mexico	
Obomanu and Okpobiri (1987) [17]	$R_s = \frac{0.03008P^{0.927}\gamma_g^{2.15}\left(\frac{141.5}{\gamma_o}-131.5\right)^{1.27}}{1O^{0.811}(1.8T-459.67)^{0.497}}$	Niger Delta	$0.811 \leq \gamma_o \leq 0.966$

2.0 Methodology

2.1 Bagging Ensemble Technique

Leo Breiman method is used in this study to predict reservoir fluid properties of bubble point pressure and gas oil ratio (GOR). A bagging algorithm with the base models is first created by developing a predictive model on the dataset, then hyper-parameters tunings are then defined, and is implementation done. The estimators that are used to fit base models on training data are: Decision Tree Regressor (DTR), Random Forest Regressor (RFR), and Extra Tree Regressor (ETR). A total of 3424 data points were used with 80% for training and 20% for testing.

Using the Python tool Scikit-learn [18], a bagging regressor is implemented so as to fit base model on a random subsets of the original data. Each of the subset sample from the training dataset was selected with replacement and then fit a regressor to each subset. The DTR creates a predictive model that will be able to predict (i) target variable value "bubble point pressure" from data features "solution GOR, specific gravity, temperature, and API" and (ii) target variable value GOR from the data features "bubble point pressure, gas specific gravity, AP gravity, and temperature". The training set size was set at 80% for training and 20% for testing. To instantiate and train models, the created base regressors were initialized to fit the model into the training data using the bootstrap aggregating method. The bagging meta-estimator aggregates individual performance to form a final prediction (fig. 1.).

The Bagging Algorithm Implementation

1. From the dataset, create multiple samples of n numbers of subset with replacement
2. Create a weak learner (or base model) m, with m1, m2, and m3 representing each of the DTR, RFR, and ETR models.
3. Fit and build a model for each of the subsets in parallel, $m_{i1}(\cdot), m_2(\cdot), and\ m_3(\cdot),$
4. Run all models in parallel
5. aggregate model predictions into a single prediction by voting regressor
6. predict bagging ensemble, $\hat{p}_{bagging} = p_1(X) + \hat{p}_2(X) + \hat{p}_2(X)$. where subscript 1,2, and 3 rep individual bagged DTR, RFR, and ETR model predictions respectively.

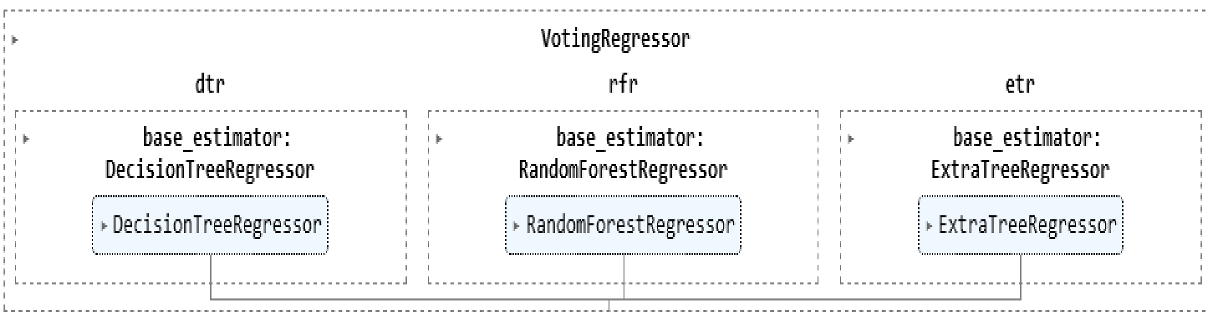

Figure 1. – Bagging technique: A voting regressor to fit base estimators

The Decision Tree Regressor creates a model, observes the features, then train the model in the form of a tree to enable prediction. The tuning parameters used are max_depth which represent the maximum tree depth; min_samples which represent the minimum number of samples required to split an internal node; splitter to choose each node split; and the random state to control the randomness of the estimator use. The attributes are n_features to account for the number of features seen during fit; n_output for the number of outputs when fit is performed i.e. to fit model and return the prediction or transformed data.

The Random Forest Regressor fits regression tree regressors on the dataset sub-samples whilst using an averaging method to improve predictions and controls overfitting. Hyperparameter tuning used are maximum depth of tree and random state.

Extra Tree Regressor, the fit method builds a number of randomized decision trees from the training set and it is dependent on parameters such as the sample weight and the input variables.

2,2 Artificial Neural Network (ANN)

For the ANN standalone modeling, the process involved data preparation, splitting data randomly, generating random architecture for ANN, compiling the model, evaluating and fine tuning, trained the ANN using the weight and bias, model testing, model evaluation, and finally prediction.

Dataset used in this study are representative of the Niger Delta wells collected from nine different reservoirs with 3424 data points. Data preprocessing was performed to help transform the raw data to ensure the quality of the data in terms of accuracy and completeness.

The statistical error analysis used to evaluate metrics are coefficient of correlation (R^2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Average Absolute Percentage Error (AAPE). More emphasis on the predictive models' performance and error minimization measured by R^2 and AAPRE. The lower the MAE and RMSE prediction values, the better the accuracy of the prediction.

Correlation of Coefficient (R^2) explicitly describe how much the model performance has improved and measures the goodness of fit of our model. We will classify the R^2 model predictions in the following ranges: Weak: 0-3; moderate 0.3 to 0.5; strong: 0.5 to 0.7; and very strong 0.7 to 1.0.

Mathematically expressed as:

$$R^2 = \sqrt{1 - \frac{\sum_{i=1}^{n}[(y_{est} - y_{exp})]^2}{\sum_{i=1}^{n}[(y_{exp} - y_i)]^2}}$$

Where,

n = total no. of data points

y_{est} = estimated values

y_{exp} = expected values

i = variable i

Mean absolute error (MAE) measures the relative error, describes the fitness of the model, and accounts for the difference between the predicted values from the machine learning model and the actual target value while dividing it with the number of testing sample values`

Expressed as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{y}_i$$

Where,

n = total no. of data points

y_i = prediction

$\hat{y}_i$ = observation

i = variable i

Root Mean Squared Error (RMSE)

The RMSE is a performance metrics used in this study to measure the model efficiency. It takes the error of the predictive modeling and measure the average difference between actual and predicted values. The lower the RMSE prediction value, the better the accuracy of the prediction.

$$RMSE = \sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right]}$$

Average Absolute Percentage Error

The AAPE measures the relative error by taking the average of the absolute percentage differences between actual and predicted values of the machine learning model. It measures the prediction accuracy of a model.

$$AAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left(y_{est} - y_{exp}\right)}{y_{exp}}$$

Tools

The machine learning models are built using Python Anaconda version 3.11.5, an interpreted, object-oriented, high level programming language with dynamic semantics, and it is used for machine learning via the machine learning libraries and its framework such as the Scikit-learn, Panda, and NumPy.

3.0 Result and Discussion

For Bubble Point Pressure

Table 3. shows the train and test result of the predictive modeling of each regressors used. Results shows a higher R^2 of 0.9310 and 0.9292 for the bagged extra tree (ETR) and the bagged decision tree (DTR) as compared to the DTR with 0.8744. However, the test score shows a higher R^2 for RFR at 86.72% compared to the DTR of 85.58% and ETR of 82.748%. For the MAE, the DTR and ETR produced a lower train values of 165.8964 and 170.3479 compared to the RFR with a higher value of 246.6362. Again, during the testing the RFR returns with a better lower value at 413.8698 compared to the 417.6217 for DTR and 430.1111 for the EFR. The MAE and root mean squared error result shows a somewhat high values for all base models and this is largely due to the independent variables values in the dataset. The RMSE train score for the ETR is 269.4790 better compared to that of DTR at 273.0141 and RFR at 363.5071 while during the testing predicted 614.2398, 629.2637, and 634.3476 in that order. For the AAPE, bagged ETR showed reduced percentage error at 9.94% during the training better compared to that of the bagged DTR at 10.40% and bagged RFR at 16.30%. upon testing the individual bagged models, the relative percentage errors were 15.02%, 15.38%, and 15.68% for the RFR, DTR, and ETR respectively

Aggregating all base predictions using a voting regressor predicted a higher R^2 for bagging ensemble (BE) with R^2 of 96.65% while *MAE, RMSE,* and *AAPRE* are 309.2188, 285.1020, and 14.98%, respectively. The standalone machine learning model ANN predicted results are R^2 of

98.04%, *MAE* as 861.335, *RMSE* of 673.710, and *AAPRE* as 0.219. the best performing empirical correlations is the one by Petrosky and Farshad model with R^2 *of* 88.10%, while the Standing correlation outperformed all empirical correlation in terms of *MAE*, *AAPRE and RMSE*. The final hyper-parameter tuning used are: numbers of base estimators, 10 for the model to aggregate together; random state, 22 to control random sampling. Fig. 2. shows the prediction accuracy of actual values vs predicted value while figure 3. shows the statistical analysis of Bagging, ANN, and Empirical correlations for *pb*.

Table 3. - Statistical analysis for *pb* prediction

	Data set	**R^2**	**MAE**	**RMSE**	**AAPE**
DTR	Train score	0.9292	165.8964	273.0141	0.1040
	Test score	0.8558	417.6217	629.2637	0.1538
RFR	Train score	0.8744	246.6362	363.5071	0.1630
	Test score	0.8672	413.8698	614.2398	0.1502
ETR	Train score	0.9310	170.3469	269.4790	0.0994
	Test score	0.8274	430.1111	634.3476	0.1568
BE	Train score	0.9836	154.7002	257.0547	0.1190
	Test score	0.9665	309.2188	285.1020	0.1498

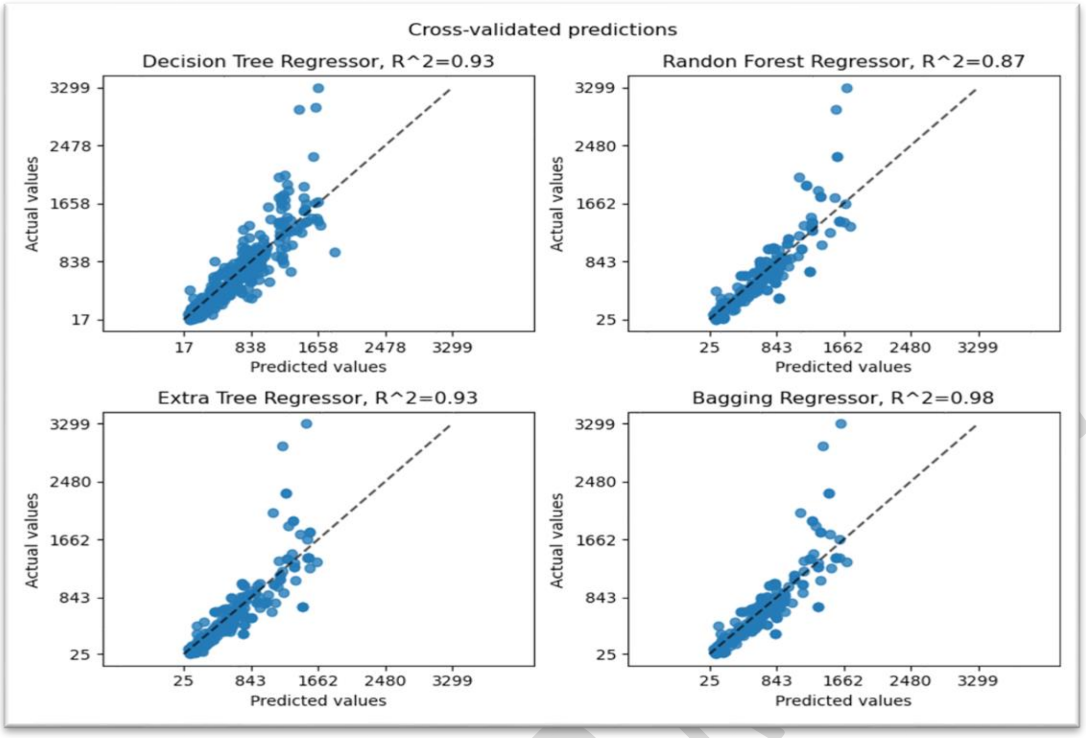

Figure 2. - Prediction accuracy showing actual values vs predicted values of bubble point pressure with Bagging Ensemble technique

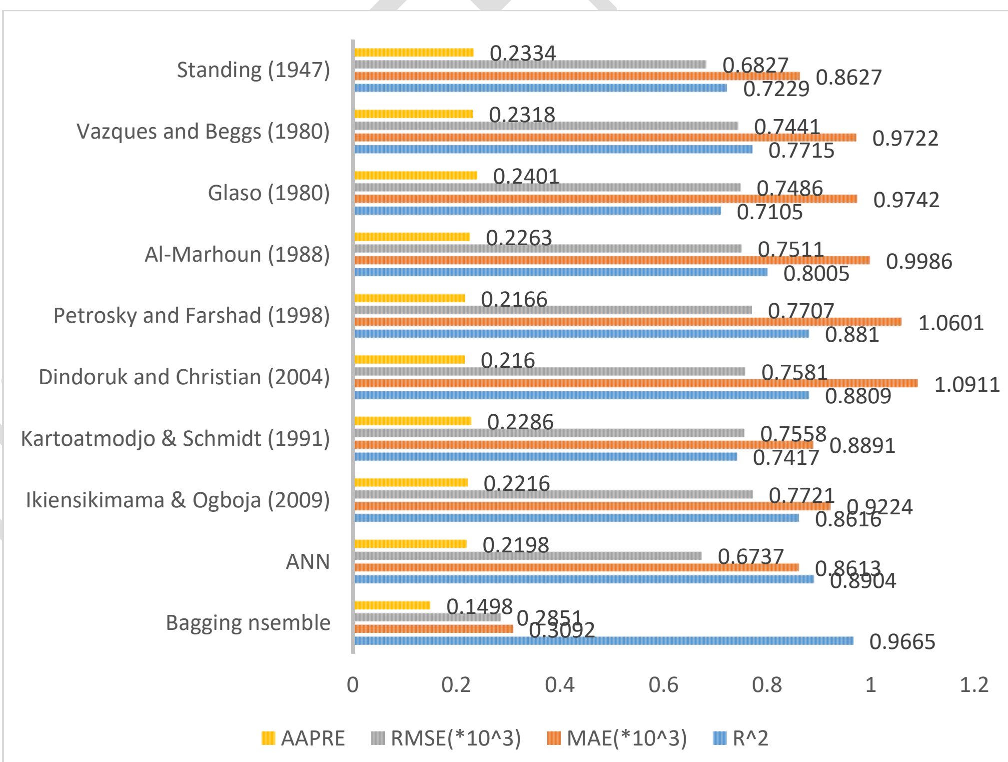

Figure 3. - Performance evaluation of Bagging, ANN, and Empirical correlations for *pb*

For Gas Oil Ratio (GOR)

Table 4.0 shows results of the individual estimators' train and test predictions alongside the statistical error analysis for the predictions of the gas oil ratio which describes the amount of gas present in stock tank oil. The individual estimators predicted a R^2 of 96.14% for the DTR during the training phase and achieving an 88.19% improvement during the testing; for the RFR the train score predicted is 93.63% with an 85.29% model accuracy during prediction; while the ETR which is a randomized decision tree predicted a train score of 96.63% with a test score of 89.97% a bit higher that the two previous individual estimators. In terms of the *MAE*, the DTR estimator train score of 111.9103 and a test score of 270.7617 was better compared to the predictions of the other estimators with a train score of 114.5613 and a test score of 299.9169 for ETR and the bagged RFR train score of 167.643475 with a test score preferably better than that of the ETR at 278.3475. Result of the model efficiency in terms of the *RMSE* shows a somewhat good prediction between the actual and predicted values on individual basis with train score of 226.0532 for the ETR compared to the DTR and RFR at a score of 246.0478 and 374.2689 respectively but a twist here is the reverse test score which seems to be less better for the ETR at 581.6251 compared to the DTR of 475.8154 and RFR of 489.5754. in predicting the relative percentage error using the AAPE evaluation metrics, the bagged ETR predicted a better train score of 15.66% compared to those of the ETR at 17.40% and RFR at 18.85% while the test score shows DTR to be more accurate and reduce model error at a rate of 16.85% better compared to the RFR at 17.03% and ETR at 18.27% respectively.

Testing results from aggregating all base predictions using a voting regressor obtained R^2 of 94.14%. In describing how much the model fit on the training data this is represented in the gas solubility predicted *MAE* train score of 279.90, while the model efficiency also in terms of *RMSE* predicted a test score of 475.51 with a AAPE of 17.82%. The standalone machine learning model ANN, predicted R^2 of 89.004%, *MAE* is 522.514, *RMSE* of 592.980, and *AAPRE* OF 0.1821. The best performing empirical correlations in terms of the are the ones developed by Obomanu and Okpobiri at 87.066%. Fig. 4 shows prediction accuracy of actual values vs predicted value and fig. 5 shows evaluation metrics of Bagging, ANN, and Empirical Correlations for GOR.

Table 4. – Statistical Analysis for Gas Oil Ratio Prediction

	Data set	**R^2**	**MAE**	**RMSE**	**AAPE**
DTR	Train score	0.9614	111.9103	246.0478	0.1740
	Test score	0.8819	270.7617	475.8154	0.1685
RFR	Train score	0.9363	167.6431	374.2689	0.1885
	Test score	0.8529	278.3475	489.5754	0.1703
ETR	Train score	0.9663	114.5613	226.0532	0.1566

	Test score	0.8997	299.9169	581.6251	0.1827
BE	Train score	0.9704	124.6622	258.7323	0.1721
	Test score	0.9414	279.9002	475.5120	0.1782

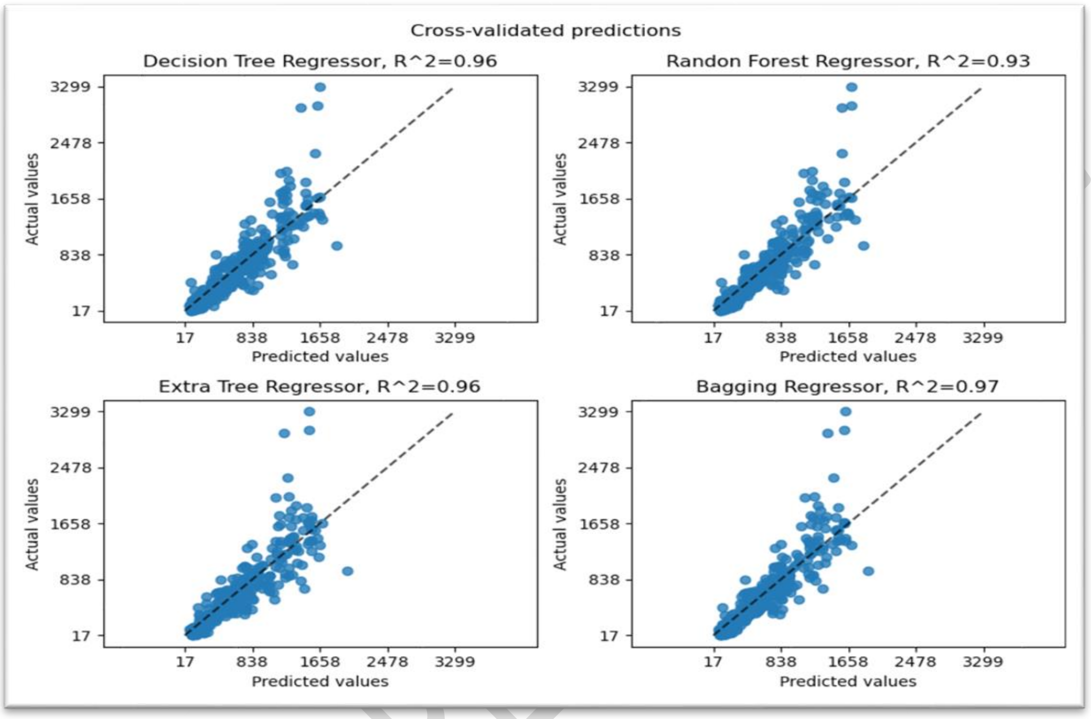

Figure 4 - Prediction accuracy showing actual values vs predicted values of GOR with Bagging Ensemble technique

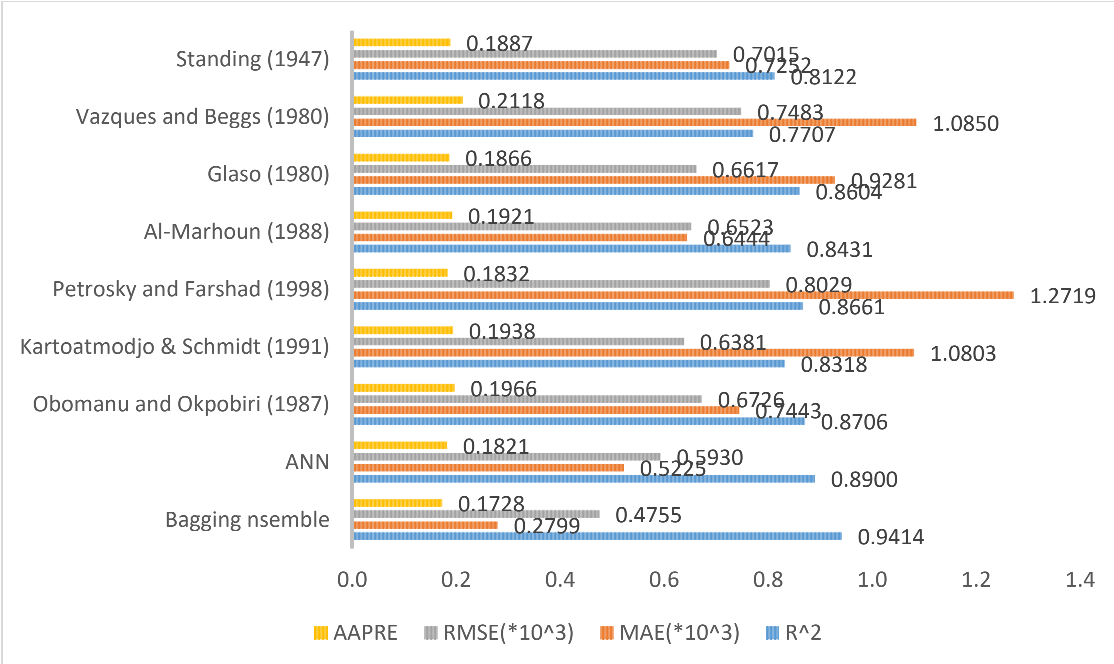

Figure 5 - Performance evaluation of Bagging, ANN, and Empirical Correlations for GOR

Conclusion

A predictive model has been developed to predict reservoir oil PVT properties using Bagging Ensemble machine learning. The dataset used was divided into subsets. Each subset is used to fit data in the final regressor using averaging method. The results were compared to the ones developed using ANN model and selected existing empirical correlations in the literature. The Bagging ensemble minimizes overfitting of data and improves the model accuracy while also predicting better results in terms of the statistical error analysis used for comparisons,

Disclaimer (Artificial intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

References

1. Louppe G.and P. Geurts, Ensembles on Random Patches. Machine Learning and Knowledge Discovery in Databases, 346-361, 2012.
2. Standing, M.B. 1947. A Pressure-Volume-Temperature Correlation for mixtures of California Oils and Gases, Drill. &Prod. Prac. API, 275 – 287.
3. Oladipo, O. and Johnathan, M. 2020. A Novel ALGORITHMIC Design and Implementation for Predicting Crude Oil PVT Properties. Paper 203716

4. Ehsan Khamehchi, Mohammad Reza Mahdiani, Mohammad Amin Amooie, and Abdolhossein Hemmati-Sarapardeh. 2020. Modeling viscosity of light and intermediate dead oil systems using advanced computational frameworks and artificial neural networks. J. Pet. Eng. Volume 193, October 2020, 107388
5. Mohammed A. A., Siti N.Z., Mysara E. E., Ku E. H., and Quosay Ahmed. 2020. Revisiting the Coefficient of Isothermal Oil Compressibility Below Bubble Point Pressure and Formulation of a New Model using Adaptive Neuro-Fuzzy Inference System Technique. Paper presented at the IPTC Dhahran, Kingdom of Saudi Arabia, January 2020. Paper IPTC-20293-
6. Kingsley Uzogor and Oluwatoyin O. Akinsete. 2020. Improved Correlations and Predictive Models for Nigerian Crude Oil PVT Properties Using Advanced Regression and Intelligent Techniques. Published 11th August for the SPE NAICE
7. Isemin, I.A. and Akinsete, O.O. (2021). Support Vector Machine Model for Predicting Gas Saturated and Undersaturated Crude Oil Viscosity of Niger Delta Oil Reservoir. Journal of Engineering Research and Reports,
8. Yingxian Liu, Cunliang Chen, Hanqing Zhao, Yu Wang, and Xiaodong Han. 2021. A RobustMethod to Predict Fluid Properties Based on Big Data and Machine Learning Algorithms. Paper presented at the International Petroleum Technology Conference, Virtual, March 2021. Paper IPTC-21356-MS
9. Breiman, Leo. Bagging predictors. Machine Learning, 24(2), 123-140,1996
10. Vazquez, M., & Beggs, H. D. 1980. Correlations for fluid physical property prediction. *JournaL of Petroleum Technology*, *32*(06), 968–970
11. Al-Marhoun, M.A. 1988. PVT Correlation for Middle East Crude Oils. Journal of Petroleum Technology (JPT), May 1988: 650–665
12. Petrosky Jr, G. E., & Farshad, F. 1998. Pressure-volume-temperature correlations for Gulf of Mexico crude oils. SPE Reservoir Evaluation & Engineering, 1(05), 416–420, paper 51395-PA.
13. De Ghetto Giambattista., Paone, F., and M. Villa. 1995. Pressure-Volume-Temperature Correlations for Heavy and Extra Heavy Oils. paper SPE 30316 presented at the SPE International Heavy Oil Symposium, Calgary, 19 -21 June
14. Birol Dindoruk and Peter G. Christman. 2004. PVT Properties and Viscosity Correlations for Gulf of Mexico Oils. SPE Reservoir Evaluation & Engineering 7(6):427-437. SPE-89030-PA.
15. Ikiensikimama S. Sunday and Olu Ogboja. 2009. New Bubble Point Pressure Empirical PVT. Correlation. SPE NAICE. SPE-128893-MS
16. De Ghetto Giambattista., F. Paone, and M. Villa. 1994. Reliability Analysis on PVT Correlations. SPE paper 28904, presented at the European Petroleum Conference of the SPE London, October 25-27
17. Obomanu D. A. and Okpobiri G. A. 1987. Correlating the PVT Properties of Nigerian Crudes. *J. Energy Resour. Technol.* Dec 1987, 109(4): 214-217
18. Data manipulation with Pandas, Retrieved