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 model, three different estimators, Decision Tree Regressor, Random Forest Regressor, and Extra Trees Regressor, were used as the base estimators, then an averaging method to finally predict the model performance was done by 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. Statistical error analysis formed the basis for which this models were evaluated. Result showed that the Bagging ensemble predicted better results in terms of the statistical error analysis used for comparisons.

Keywords: Machine Learning, Ensemble, PVT, Bagging,

1.0 Introduction

Ensemble machine learning technique combines the predictions from multiple learners or machine learning models that are weak so as 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. While Boosting Ensemble deploys iterative process where the training set is reweighted at each interaction based on errors a weak learner made in the first model, Stacking Ensemble is a method for creating a linear combinations of different predictors to improve the accuracy of a prediction.

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. Some published work using machine learning to predict PVT properties include the works of Oladipo and Johnathan [2], Ehsan *et. al.* [3], and Mohammed *et. al.* [4], Kingsley and Akinsete [5], Isemin and Akinsete [6], and Yingxian *et. al*. [7].

Bagging Ensemble Machine Learning

Bagging ensemble method also known as bootstrap aggregating, is an ensemble method that is designed by Leo Breiman [8] to improve the stability, accuracy, prediction, and performance of machine learning algorithms. Generally, 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. 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.

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

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 [9]. 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 1.0 and 2.0 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.

Table 1.0 – 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 2.0 – 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 Methods

2.1 Bagging Ensemble Technique

This method designed by Leo Breiman 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 it's 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).

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. 1.0).

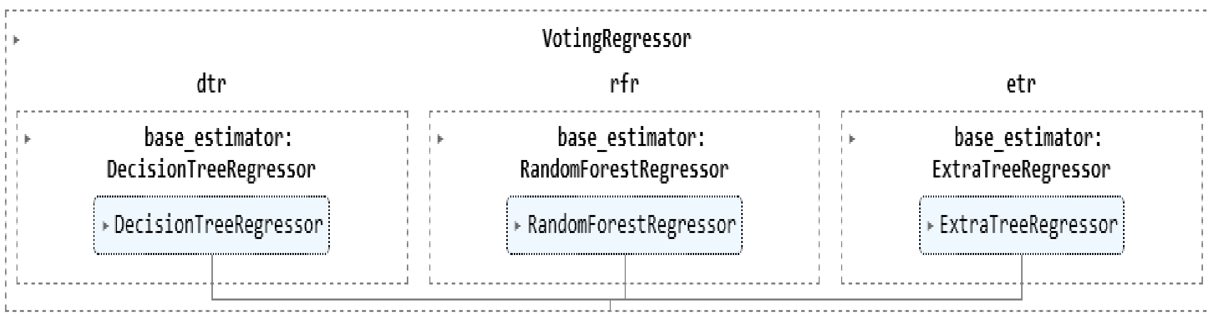

Figure 1.0 – Bagging technique: A voting regressor to fit base estimators

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 into clean data before feeding into the algorithm 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).

3.0 Result and Discussion

For Bubble Point Pressure

Table 3.0 shows the train and test result of the predictive modeling of each regressors used. Aggregating all base predictions using a voting regressor predicted a higher R^2 for bagging ensemble (BE) with test score of 96.65% while *MAE*, *RMSE*, and *AAPRE* are 309.2188, 285.1020. and 14.98% respectively. Fig. 2.0 shows the prediction accuracy of actual values vs predicted value.

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* as shown in figure 3.0. 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 and figure

Table 3.0 - 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

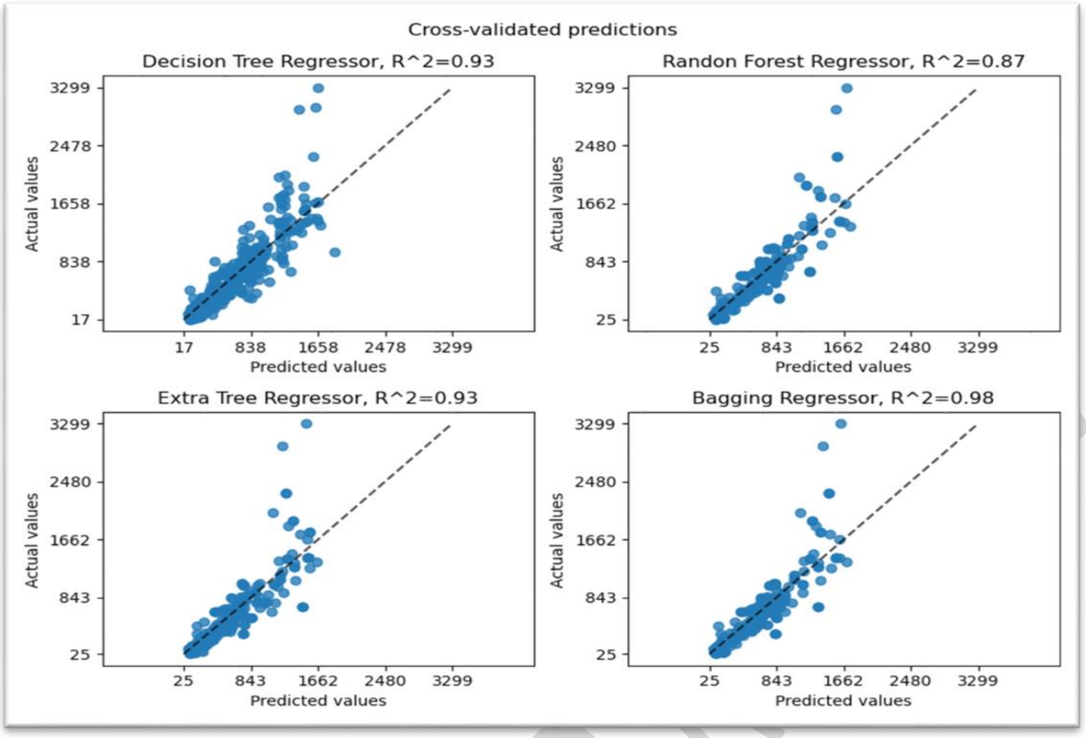

Figure 2.0 - Prediction accuracy showing actual values vs predicted values of bubble point pressure with Bagging Ensemble technique

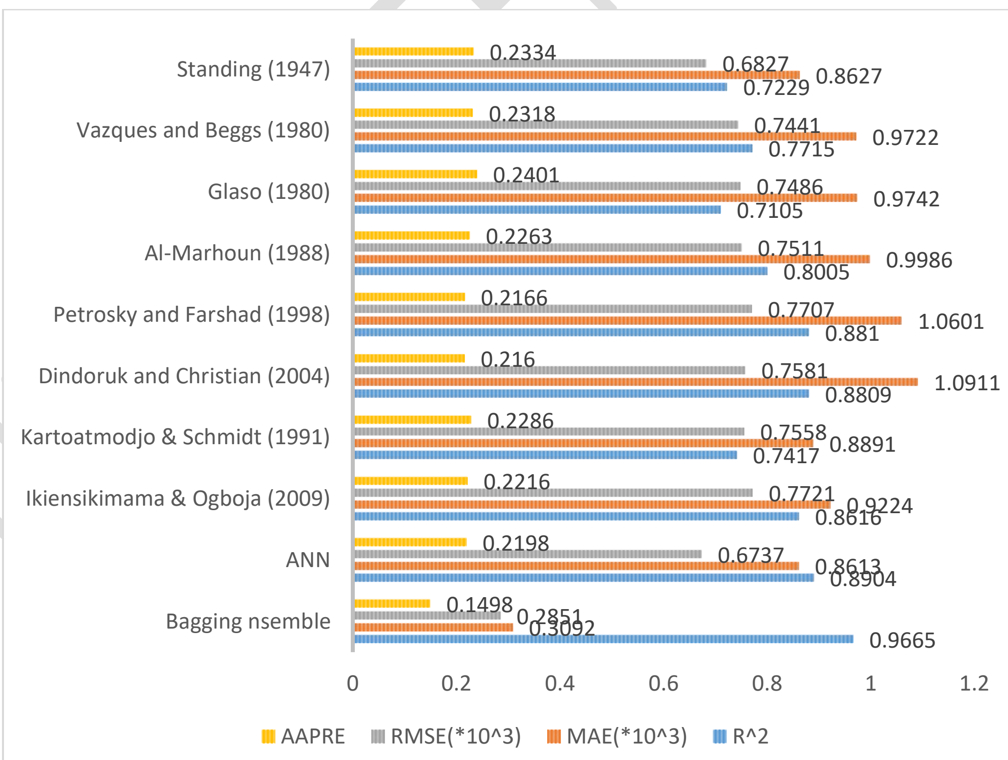

Figure 3.0 - Performance evaluation of Bagging, ANN, and Empirical correlations for *pb*

For Gas Oil Ratio (GOR)

Table 4.0 shows results of the individual estimators' train and test predictions. Upon testing, the obtained R^2 is 94.14% showing improvement performance. In describing how much the model fit on the training data this is represented in fig. 5.0 showing prediction accuracy of actual values vs predicted value. The gas solubility predicted *MAE* train score of 279.9002, while the model efficiency also in terms of *RMSE* predicted a test score of 475.5120The bagging ensemble AAPE shows a train score of 17.21% and test score of 17.82% relative error. 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. 5.0 shows performance evaluation of Bagging, ANN, and Empirical Correlations for GOR

Table 4.0 – 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

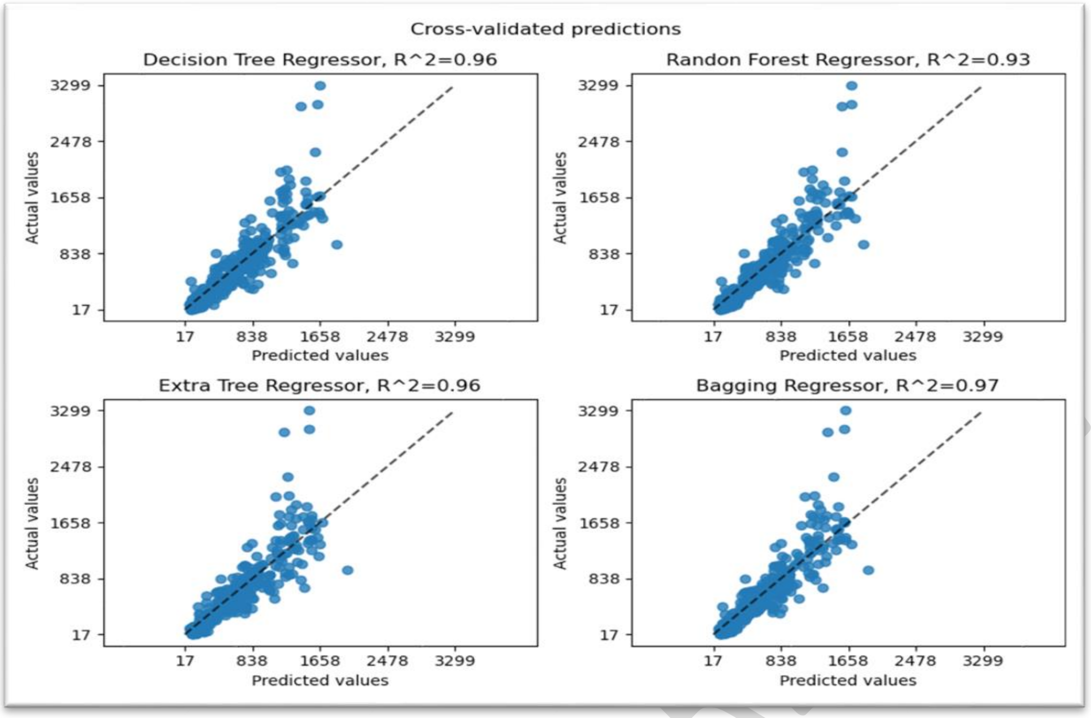

Figure 4.0 - Prediction accuracy showing actual values vs predicted values of GOR with Bagging Ensemble technique

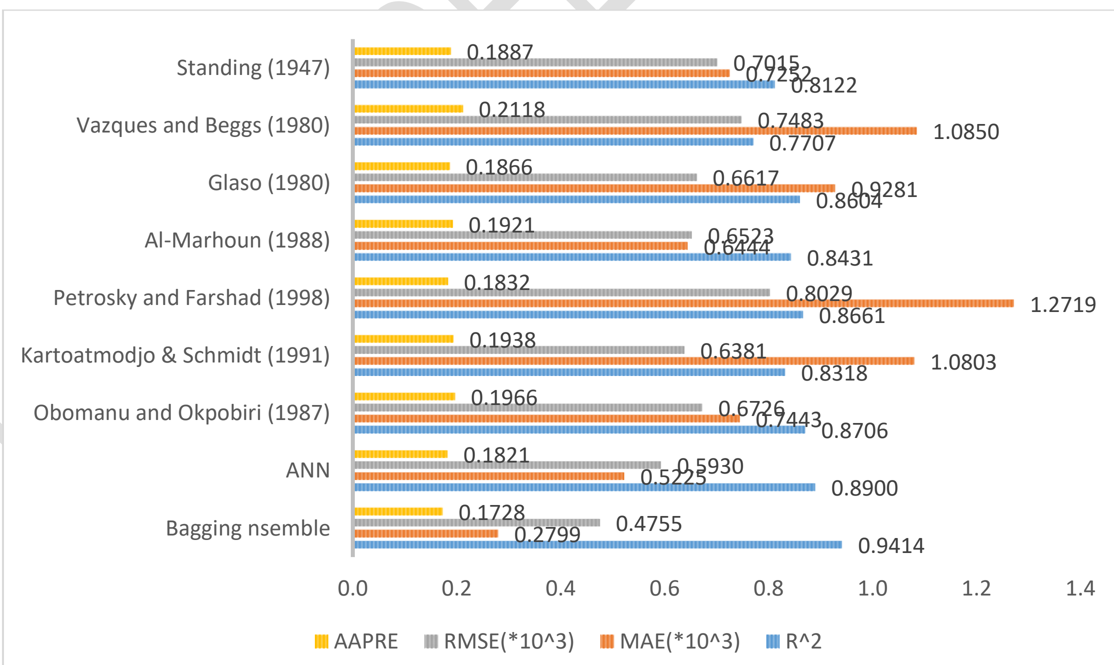

Figure 5.0 - Performance evaluation of Bagging, ANN, and Empirical Correlations for GOR

Conclusion

A predictive model has been developed to predict reservoir oil PVT properties using Bagging Ensemble machine learning. The results were compared to the ones developed using ANN model and selected existing empirical correlations in the literature. The Bagging ensemble predicted better results in terms of the statistical error analysis used for comparisons.

References

1. Louppe G.and P. Geurts, Ensembles on Random Patches. Machine Learning and Knowledge Discovery in Databases, 346-361, 2012.
2. Oladipo, O. and Johnathan, M. 2020. A Novel ALGORITHMIC Design and Implementation for Predicting Crude Oil PVT Properties. Paper 203716
3. Ehsan Khamehchi, Mohammad Reza Mahdiani, Mohammad Amin Amooie, and Abdolhossein
4. 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-
5. 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
6. 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,
7. 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
8. Breiman, Leo. Bagging predictors. Machine Learning, 24(2), 123-140,1996
9. Standing, M.B. 1947. A Pressure-Volume-Temperature Correlation for mixtures of California Oils and Gases, Drill. &Prod. Prac. API, 275 – 287.
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