International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Application of Machine Learning as an Alternative to Multiphase Flow Meters for Oil Production Rates Estimation.

Seth Uba Wadike, Isaac Eze Ihua-Maduenyi, Jovi Kelvin Oghenetega
Department of Petroleum Engineering, Rivers State University, Nkpolu - Oroworukwo, Rivers State, Nigeria

Abstract— *Critical evaluation of the rate of oil production is vital for accounting, optimum production and timely estimate of the well life. Numerical and measurement methods have been widely adopted but time-consuming and rigorous and require a fast and accurate approach for future prediction. This work utilizes machine learning technique to reduce dependence on multiphase flow meters for oil production rate prediction using Niger Delta field data. Three models, Linear regression, Multivariable linear regression and random forest model were implemented and the performance of the model in predicting oil flow rate evaluated using Coefficient of determination R2 method, Regression plots and Root mean square error method (RMSE). The random forest model had an R2 score of 0.9394 and RMSE value of 23.9 which represented the highest level of accuracy among the models being compared ,the Linear regression model had its highest R2 score and lowest RMSE values to be 0.5386 and 580.16 with the use of a single dependent variable Well head pressure , this was significantly improved with the addition of Well head temperature in the multivariable linear regression model with an R2 score of 0.8034 and RMSE value of 187. The best split between training and test data to obtain the most accurate performance for the model predictions was also established in the course of the project to be between 70% to 90% of training data, and 10% to 30% of test data size.*

Keywords— *Machine learning, Oil production, Linear regression, Multiple regression, Radom regression.*

I. Introduction

The use of machine learning techniques in the oil and gas industry is made possible today due to the availability of large data sets obtained from various measuring devices with sensors that transmit real time data of process variables. Multiphase flow meters (MPFM) are currently used in deep offshore production systems on an individual well basis to estimate oil/gas and water production flow rates. Test separators fitted with volumetric flow meters are also used in deep offshore production systems to obtain individual oil and gas production flowrates. Well's flow rates play a crucial role in oil and gas field monitoring, management, and optimization, making them indispensable in the context of operational decision-making. The flow rates act as the key indicators for confirming the anticipated performance of the reservoir and the wells, allowing for informed operational choices concerning field management. Flow rate monitoring is a crucial, vital and critical task which must be correctly and strictly done to evaluate other production parameters (Olivares, 2012). Additionally, the financial implications of flow rates significantly impact tax distribution. Regulatory bodies mandate the use of flow rates to accurately distribute production and carry out financial allocation. A numerical approach of flow rates measurement based on pressure versus time has emerged (Ferreira et al., 2012; Sundaram et al., 2012). Temperature, pressure and are the most vital parameters affecting production process from reservoir to production and separation unit and should be given priority in production rate prediction (Mollaiy,2011). Thus, well production parameters (temperature, pressure and flow rates) are major primary variables for production forecast whether numerically, machine learning or Artificial intelligence approach. Artificial intelligence based models have been considered as one of the most favorable numerical and inverse tactics which can be applied in well testing to predict the future flow rate by using other production parameters (Weldu et.al., 201; Grujic et.al.,2010). Mohammed et al. (2012) used Fuzzy logic, Artificial Neural Network(ANN), and Imperialist Competitive algorithms to build a model for the prediction of oil flow rate and it their model had a better performance. Mohammad & Zhangxing (2019) compiled several machine learning algorithms to predict porosity and permeability through the inclusion of petro-physical logs for better production performance, but their model took an excessive amount of time for accuracy and output. A deep belief network (DBN) model was developed by Wang et al. (2021) for predicting the production of unconventional wells and the model was reliable and accurate. Al yateem &Almri (2012) opined that for accuracy and faster prediction of production rate, transition from Multiphase flow meters (MPFM) and Test separators fitted with volumetric flow meters method to machine learning and artificial intelligence approach is eminent. Hence, this work will utilize machine learning technique for oil production rate prediction.

II. Materials and Methods

2.1 Materials, Data and Location of field

The materials used were Microsoft excel, Python Programming language using Anaconda work environment, Jupyter notebook and the data were Multiphase flow meter data (Oil flow rate, Oil density, gas density, water flowrate , gas flow rate, water cut) from a Production well, and Well bore data (well temperature and pressure).The field studied is a deep-water field in West Africa with a subsea wellhead at a depth of 1350m below sea-level and located at 4.13km from the riser base. The FPSO (Platform) stands in 1345m of water with the topside located 50m above sea-level. The field is situated in Oil

Seth Uba Wadike, Isaac Eze Ihua-Maduenyi, and Jovi Kelvin Oghenetega, "Application of Machine learning as alternative to multiphase flow meters for Oil Production Rates Estimation," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 1, pp. 174-181, 2025.

Mining Lease 130, 150km south of Port Harcourt. Water depth over the field is in the range 1300m – 1450m.

2.2 Model Algorithm and Selection

Linear regression, Multivariable linear regression and random forest model were selected and well bore, multiphase flow meter and production data imported using panda library to the jupyter workspace in preparation for analysis. The input variable for the model was selected to be Well head pressure and Oil flow rate, well head pressure was assigned to the X axis and the Y axis value which was predicted was assigned as Oil flow rate in the model. The selected data was split into test and training data. The value of the test and training data was varied to obtain the best sample size that gave the most accurate prediction with an initial size of 0.3 used and later varied to obtain the most suited value using sensitivity analysis. A scatter plot of well head pressure Vs oil flowrate was created to visualize the linearity of the data set selected. There was a linear relationship between the wellhead pressure and oil flow rate as a reduction in well head pressure led to increase in the oil flowrate making the data suitable for the linear regression model. The data was fitted into the imported regression model. The data reshape code was used to reshape the data and fit them into a 2-dimensional array which were then used for prediction.

For model performance and evaluation, Coefficient of determination ($\boldsymbol{R}^2$) and Root Mean Square Error (RMSE) were used.

Mathematically, RMSE is the square root of the mean of the squared differences between predicted and observed values and presented as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}}$$

Where; Pi is the predicted value for the (i)-th observation, Oi is the observed value for the (i)-th observation, and n is the number of observations. A lower RMSE value indicates a better fit of the model to the data, meaning the predictions are closer to the actual values.

III. Results

3.1 Linear Regression Model Performance evaluation

Figure 1 and 2 shows more accurate predictions at lower head pressures and higher oil flowrates, and can be attributed to the lack of sufficient data of oil flowrates at higher well head pressures. The model output shows a high level of deviation between the predicted and actual values.

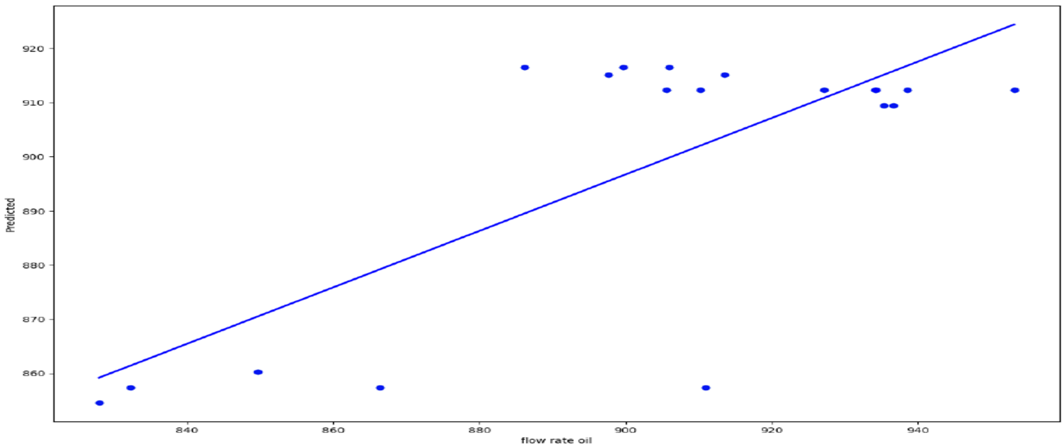

Fig. 1. Actual vs predicted oil flow rates of the linear regression model with test size 0.2

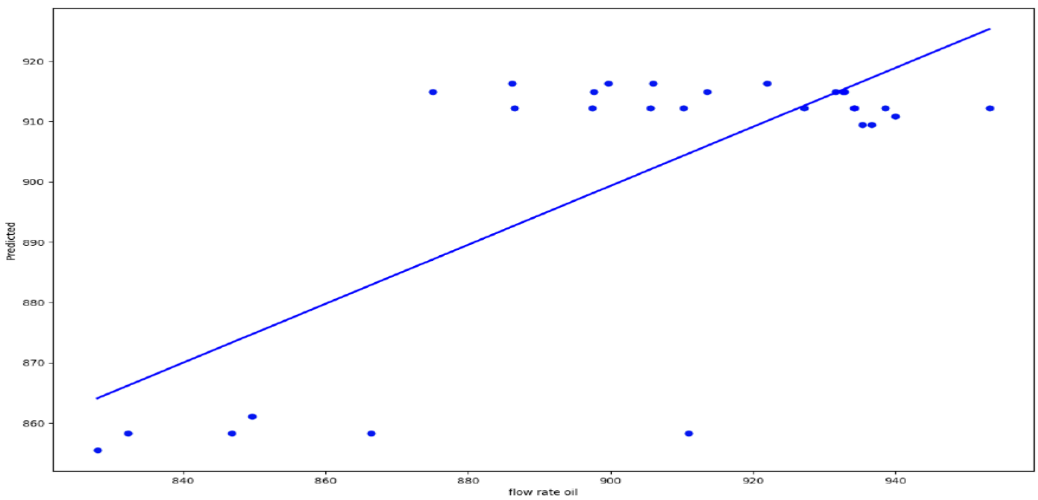

Fig.2. Actual vs predicted oil flow rates of the linear regression model with test size 0.3

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

The Coefficient of determination r2 score and Root-mean-square deviation RMSE score at varying test size are presented in figure 3 and 4. Figure 3 and 4 shows that the highest r2 score obtained was 0.5386 and the lowest RMSE obtained was 580.16, which represents the accuracy of the linear regression model in predicting oil flow rate at varying Well head pressure using a test size of 0.2 and 0.3.

3.2 Multivariable Linear Regression Model Performance evaluation

Figure 5 shows that the predicted data from the model appears closer to the actual variable and indicates the closeness of the predicted values to the actual values of oil flowrate at varying well head pressure for 0.2. Figure 6 shows that at test size 0.3, there was a higher deviation between predicted and actual oil flow rates, hence the model output of predicted oil flowrate was best with lower test size.

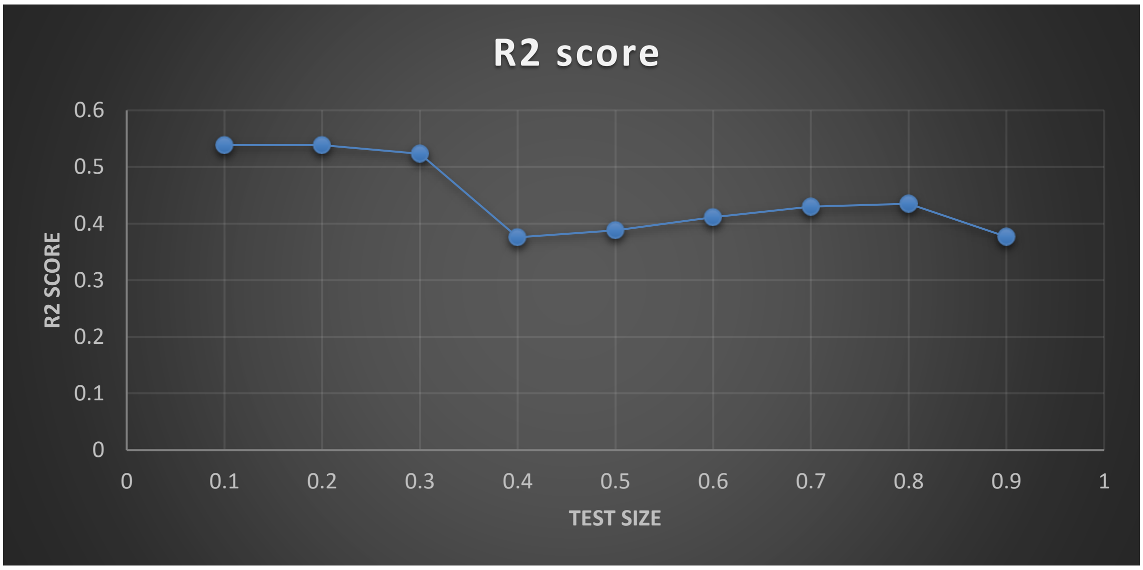

Fig.3. Effect of test size against R2 score for the linear model

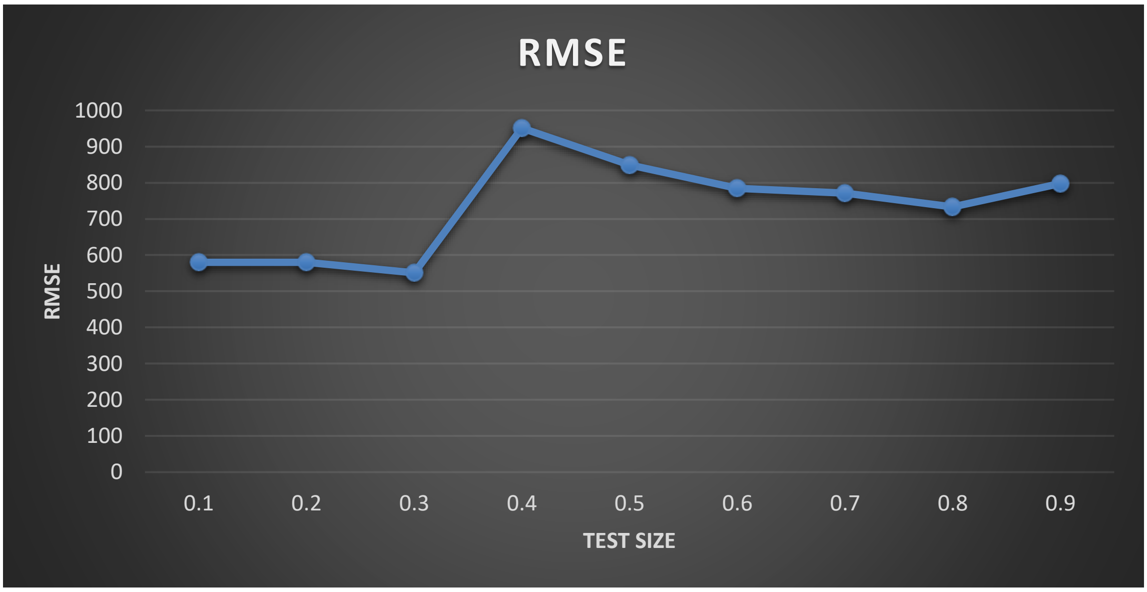

Fig.4. Effect of test size against R2 score for the linear model

Seth Uba Wadike, Isaac Eze Ihua-Maduenyi, and Jovi Kelvin Oghenetega, "Application of Machine learning as alternative to multiphase flow meters for Oil Production Rates Estimation," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 1, pp. 174-181, 2025.

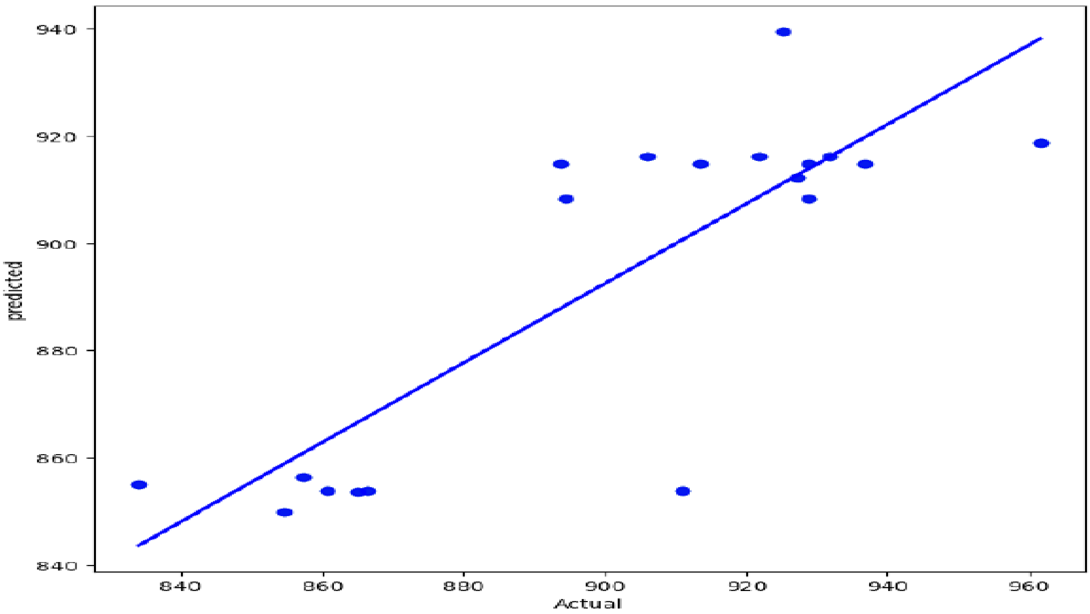

Fig.5 . Actual vs Predicted values of the multivariable linear regression model with test size 0.2

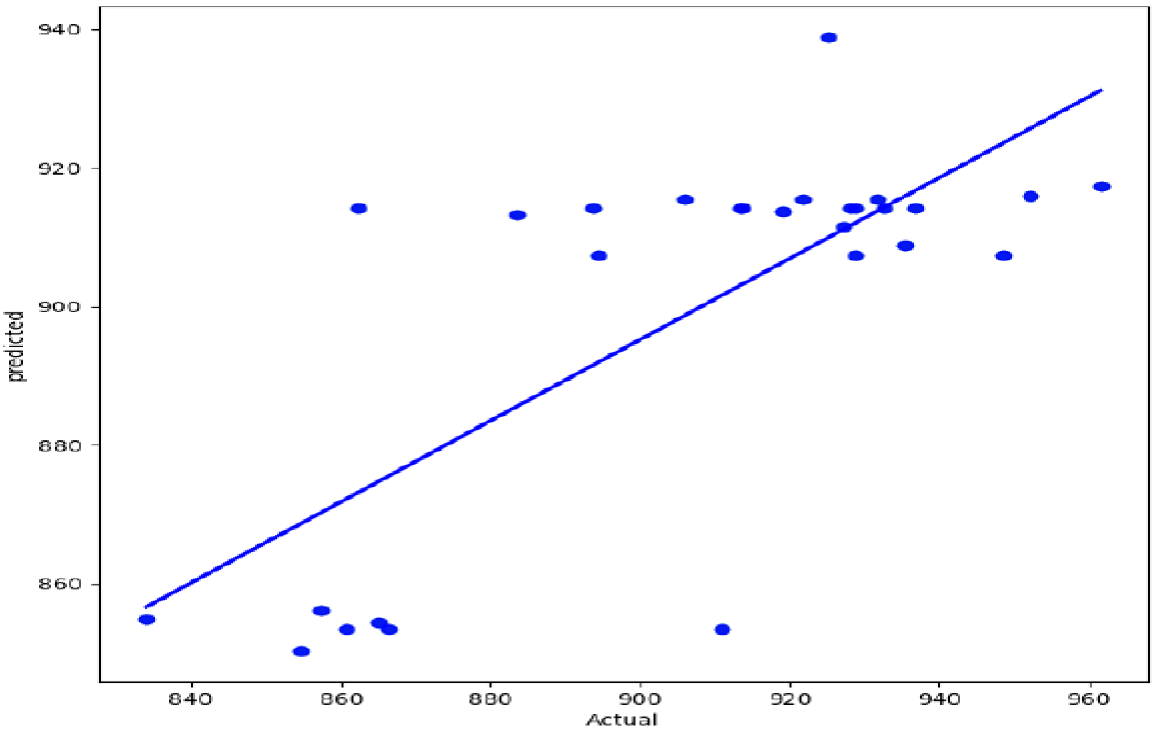

Fig.6. Actual vs Predicted values of the multivariable linear regression model with test size 0.3

The Coefficient of determination r2 score and Root-mean-square deviation RMSE score at varying test size are presented in figure 7 and figure 8. With the combination of Well head pressure and well head temperature, the highest R2 score and lowest RMSE score was obtained to be 0.8034 and 187 respectively which indicates a very high level of accuracy of the multivariable linear regression model thus validating the model. The model performance was the best at lower test sizes of 0.1 and 0.2 as there was 80-90% of the data used for training of the model, with a reduction in the data available for training of the model and a massive reduction in models' ability to predict accurate oil flowrates

3.3 Decision tree from the random forest model

Figure 9 shows the decision-making process of a single tree in the random forest model and the predicted oil flow rate was based on decisions made on the lower branches using the other parameters, density, water flow rate, gas density, well head pressure and well head temperature.

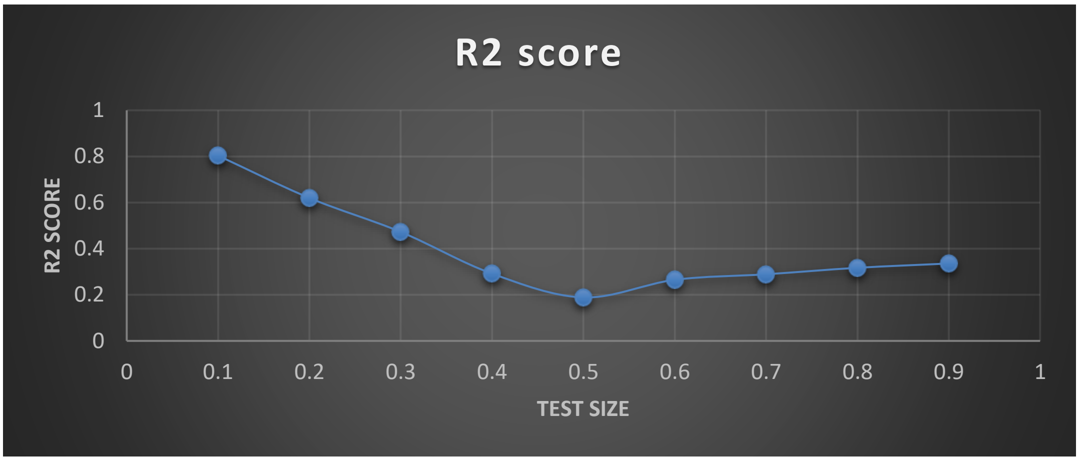

Fig. 7. Test size against R2 score of multivariable linear regression model

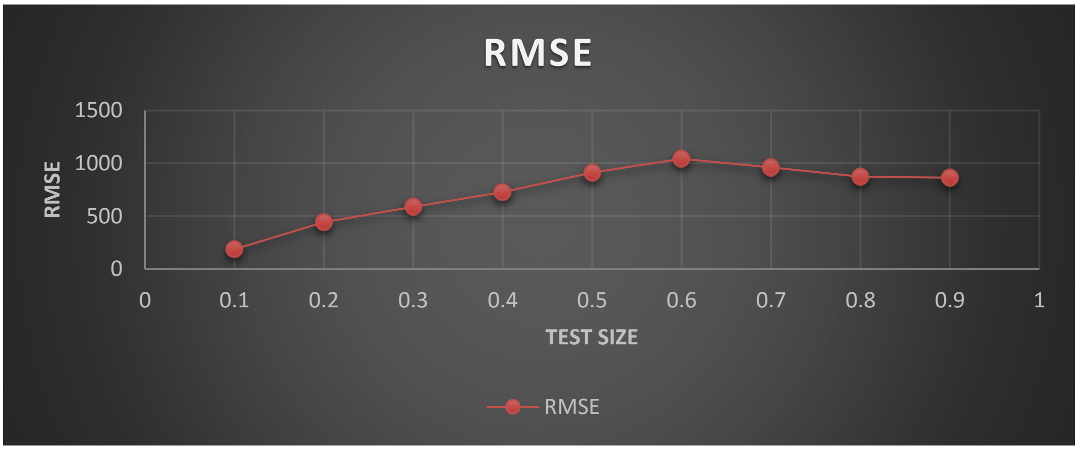

Fig.8. Test size against RMSE value of the multivariable linear regression model

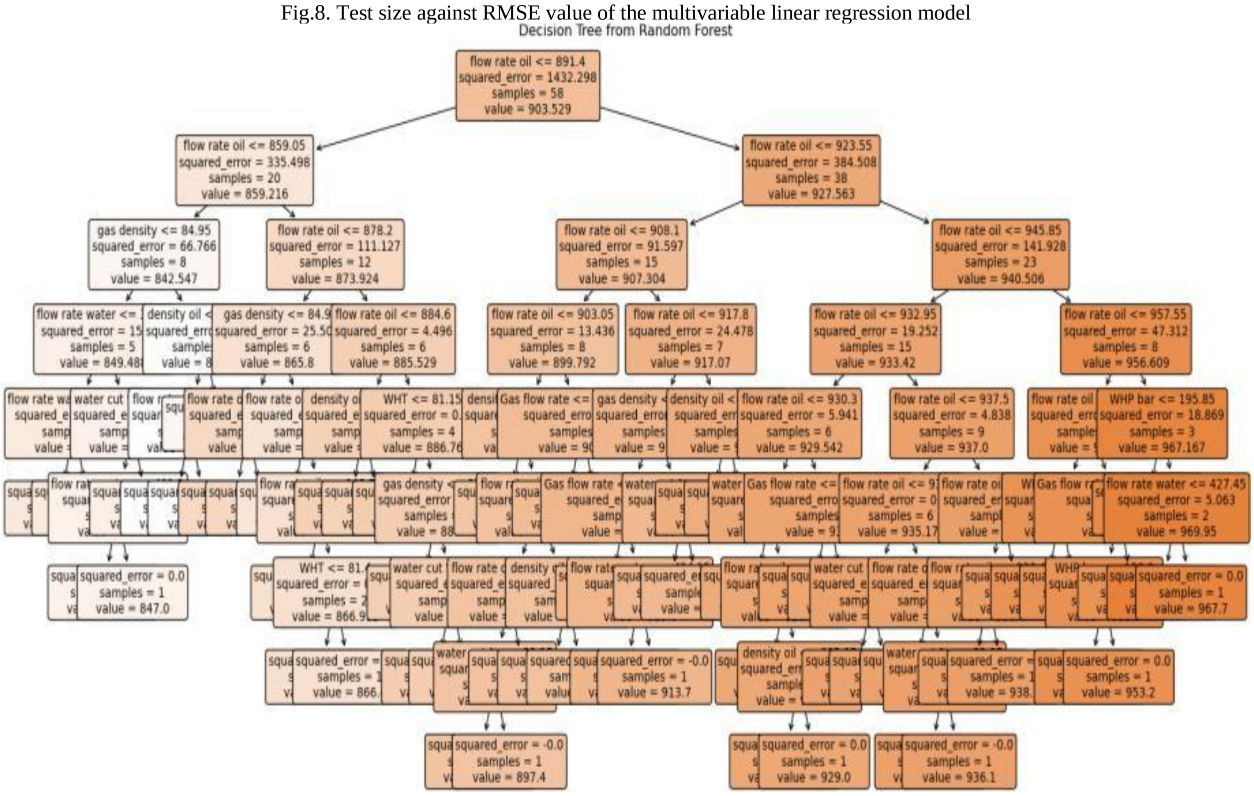

Fig.9. Decision tree from random forest

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

3.4 Random Forest Model Performance evaluation

The Actual against Predicted values of the Random Forest model with test size 0.2 and 0.3 are shown in figure 10 and 11. The results shows a very high accuracy of the random forest model, as predicted values have little deviations from actual values.

Figure 12 and 13 shows the R2 and RMSE score for the random forest model. The highest R2 score and lowest RMSE score was 0.93944 and 23.9 respectively which indicates a high level of accuracy of random forest model. Prediction accuracy of the random forest model was higher at lower test size, as more of the data was used in the decision tree process. The accuracy of predictions were as high as 93.9% using a test size of 0.1 and it implies that predicted oil flowrates were much closer to actual oil flowrates using the random forest algorithm thus validating the model.

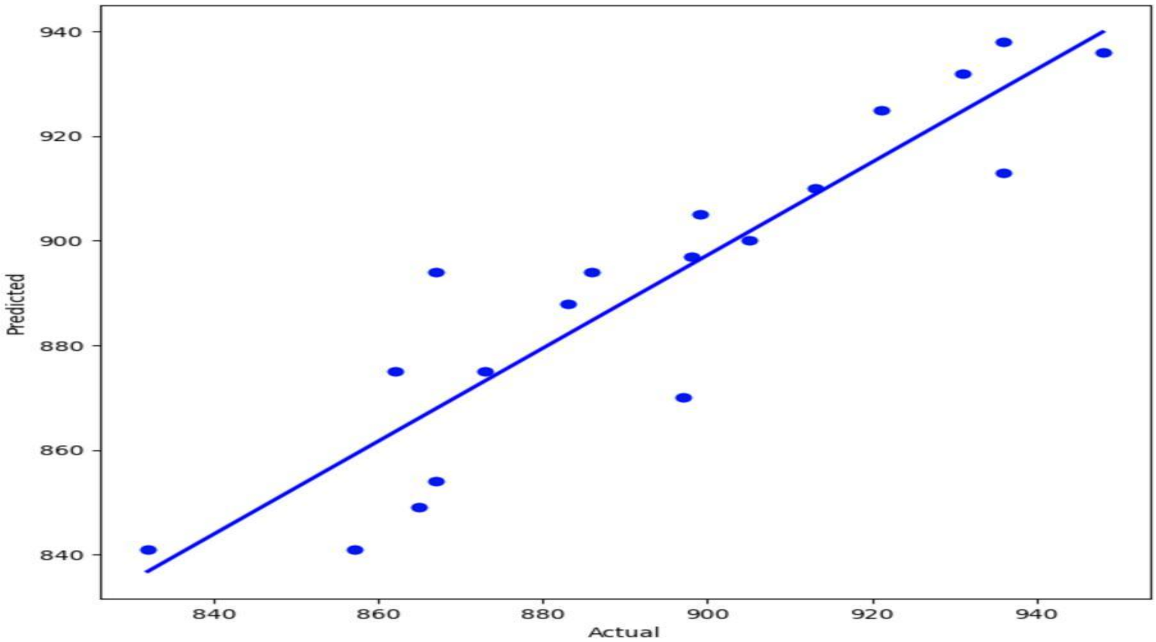

Fig.10. Actual vs Predicted values of the Random forest model with test size 0.2

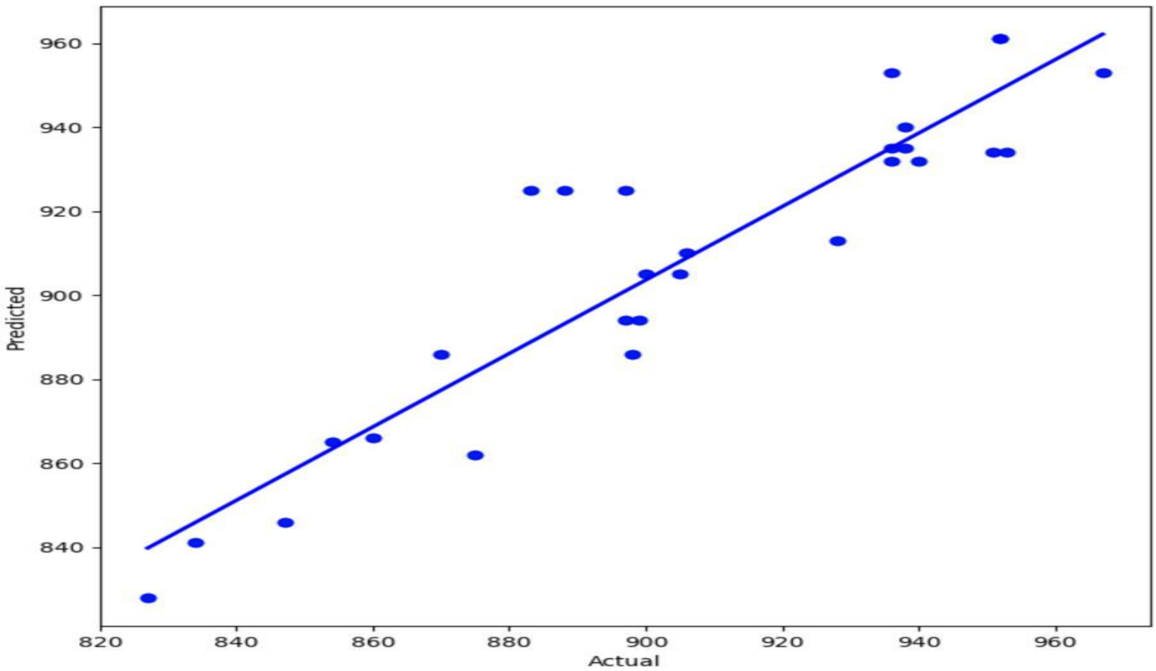

Fig.11. Actual vs Predicted values of the Random forest model with test size 0.3

Seth Uba Wadike, Isaac Eze Ihua-Maduenyi, and Jovi Kelvin Oghenetega, “Application of Machine learning as alternative to multiphase flow meters for Oil Production Rates Estimation,” *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 1, pp. 174-181, 2025.

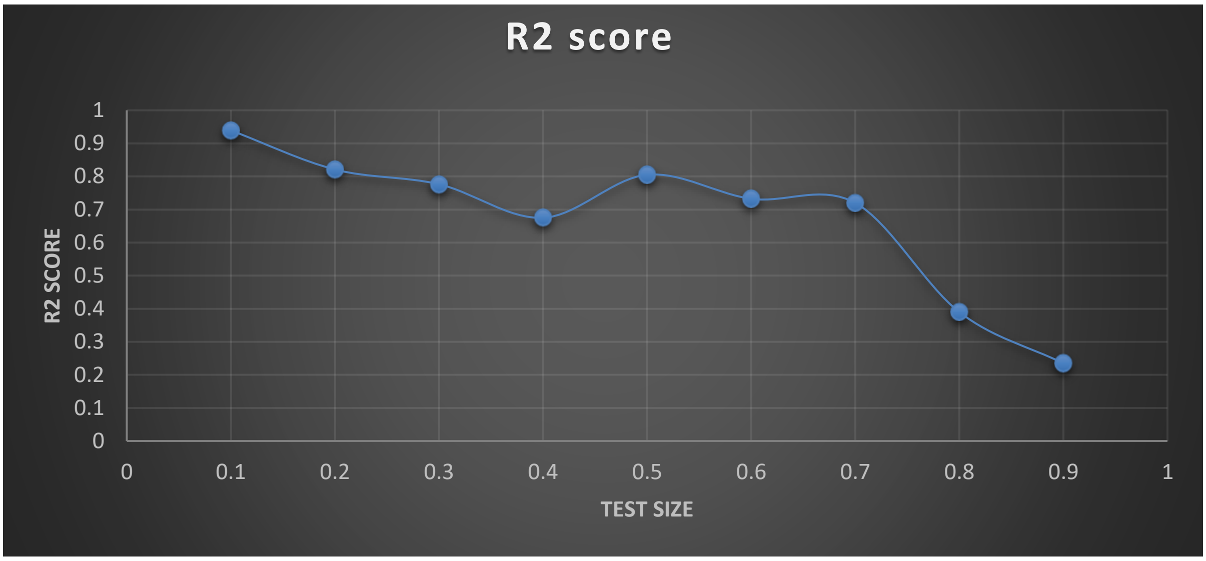

Fig. 12 . Test size against R2 score random forest model

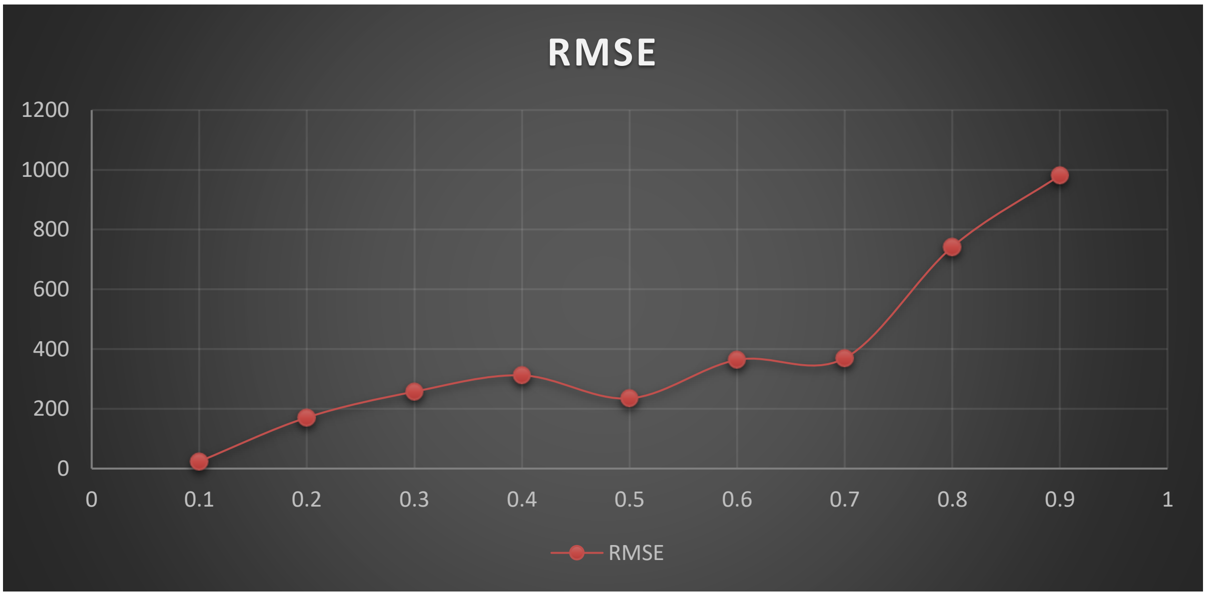

Fig.13. Test size against RMSE for the Random Forest model

IV. Conclusion

This work analyzed real world production MPFM data from a complex deep-water field in Niger delta using Machine learning techniques with objective of predicting oil production rates in the event of failure of flow meters using other well parameters such as well head pressure and temperature with their own dedicated sensors that are independent of the failed MPFM. A total of three machine learning techniques (Linear regression, Multivariable linear regression and Random Forest decision tree model) were applied to the production data and the model output analyzed by comparing actual oil flowrate values vs predicted oil flowrates using regression plots, and the errors and performance of each model evaluated using the coefficient of determination(R2), and the Root Mean Square Error method. The most accurate machine learning technique was the Random Forest model which showed the highest R2 score of 0.9394 and the lowest RMSE value of 23.9 and represents a 94% prediction accuracy of the target variable oil flowrate. The Regression plot of the random forest model output also showed the highest linearity between predicted oil flowrate and actual flowrates and high level of accuracy validates the use of the random forest for oil production rates predictions in the brown fields. The use of multiple variables including well head pressure and temperature improved the prediction accuracy of the linear regression model. The single variable linear regression model showed higher deviation between predicted oil flowrates and actual oil flowrates showing higher error margins. The best prediction results for each of the models were obtained at test size 0.1-0.3 (Training size 70 % to 90%) which indicates that as more data is available for training the model accuracy largely improves.

References

[1]. Al Yateem, S.K.& Al Amri, N., (2012). Moving toward Intelligent Field Applications: MPFM for Production Rate Testing and Beyond, SPE Intelligent Energy International, 27–29 March 2012, Utrecht, The Netherlands

[2]. Ferreira, A., Vidal, J., Abi-Ramia, R., Jouti, I., Flores, S., Ca˜nas, J., Rey, F., Landinez, G., Maizeret, P., Villanueva, G., Cabanilla, J., Alves, R.,(2012). Innovative Work Scheme: Real Time Analysis During Well Test Operations, SPE Intelligent Energy International, 27–29 March 2012, Utrecht, The Netherlands.

[3]. Grujic O., Mohaghegh, S.D., Bromhal, G., (2010). Fast Track Reservoir Modeling of Shale Formations in the Appalachian Basin. Application to Lower Huron Shale in Eastern Kentucky, SPE Eastern Regional Meeting, 12–14 October 2010, Morgantown, West Virginia, USA.

[4]. Mohammad, A. A., Mohammad, E., Amin, S. & Seyed, M. J. M.(2012). Evolving artificial neural network and imperialist competitive algorithm for prediction oil flow rate of the reservoir. *Appl. Soft Comput.* **13**(2), 1085–1098.

[5]. Mohammad, A. A. & Zhangxing, C. (2019)Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. *Petroleum.* **5**(3), 271–284.

[6]. Mollaiy,B.S.(2011) An imperialist competitive algorithm artificial neural network method to predict oil flow rate of the wells, International Journal of Computer Applications 10, 47–50.

[7]. Olivares,G., Escalona,C., Gimenez,E., (2012).Production Monitoring Using Artificial Intelligence, SPE Intelligent Energy International, 27–29 March 2012, Utrecht, The Netherlands.

[8]. Sundaram,K.M., Kumar, N., Nangia,V., Desphande,V., Nahar,S., Jackson,R.(2012), Estimating Well Deliverability in Deepwater Wells With Interval Pressure Transient Testing (IPTT): A Case Study, SPE Oil and Gas India Conference and Exhibition, March 28–30 2012, Mumbai, India, .

[9]. Wang, S., Qin, C., Feng, Q., Javadpour, F. & Rui, Z.(2021). A framework for predicting the production performance of unconventional resources using deep learning. *Appl. Energy*. **295**, 117016.

[10]. Weldu,T., Ghedan,S., Al-Farisi,O.,(2010) Hybrid Artificial Intelligence and Conventional Empirical Approach for improved Prediction of Log-Derived Permeability of Heterogeneous Carbonate Reservoir, SPE Production and Operations Conference and Exhibition, 8–10 June 2010, Tunis, Tunisia .