Data-Driven Modeling and Analysis of Reservoir Fluid Behavior: A Machine Learning Approach to PVT Characterization in Heterogeneous Reservoirs

CHARLEY IYKE ANYADIEGWU[1], CHRISTIAN EMELU OKALLA[1]*, ANTHONY KERUNWA[1], CHUKWUEBUKA DESTINY UZOR[2], JOY CHIAMAKA UZOH[1], ANIYOM EBENEZER ANANIYOM[1], CHUKWUEBUKA FRANCIS DIKE[1].

[1]Department of Petroleum Engineering, Federal University of Technology, PMB 1526, Owerri, Imo State, NIGERIA.

[2]RWTH Aachen University, GERMANY.

Abstract: - Accurate Pressure-Volume-Temperature (PVT) analysis is crucial for understanding reservoir fluid behavior and optimizing hydrocarbon production. This study develops a robust model for PVT analysis to enhance the characterization of reservoir fluids and improve reservoir management. The study employed regression analysis, Decision Tree Regressor, and a comparative Neural Network approach to evaluate relationships between critical parameters such as gas-oil ratio (GOR), reservoir temperature, gas specific gravity, oil gravity, and the oil formation volume factor (OFVF). Findings revealed complex non-linear correlations, with gas specific gravity and bubble point pressure emerging as the most influential predictors. The Decision Tree Regressor achieved high accuracy ($R^2 = 96.44\%$), while the Neural Network provided comparable performance. An expanded error analysis, numerical example, and discussion on reservoir stabilization are presented. The study underscores the significance of high-quality reservoir fluid sampling and data interpretation, emphasizing representative data to reduce uncertainties in modeling. Comparative tables and enriched references position this work as an original contribution bridging machine learning and petroleum engineering.

Key-words: - PVT, GOR, Reservoir Temperature, Gas Specific Gravity, Oil Gravity, OFVF, Decision Tree, Model.

Received: March 13, 2025. Revised: June 11, 2025. Accepted: August 14, 2025. Published: September 15, 2025.

1 Introduction

Reservoir fluids, composed primarily of hydrocarbons, exist under conditions of high pressure and temperature within subsurface rock formations. These fluids include crude oil, natural gas, and water, and their behavior is influenced by the surrounding pressure, temperature, and reservoir rock characteristics. As these fluids are produced, changes in pressure and temperature significantly alter their properties, making it necessary to analyze how they behave under varying conditions [1]. This analysis, referred to as Pressure-Volume-Temperature (PVT) analysis, is a vital aspect of petroleum reservoir management, as it helps optimize production, design recovery methods, and forecast reservoir performance. The characterization of reservoir fluids is essential for understanding the type and nature of hydrocarbons in place, which in turn dictates how these fluids should be produced to maximize recovery. Fluids in reservoirs can exist as gas, oil, or condensate,

and their behavior changes as they are brought to the surface. For instance, crude oil and dissolved gases behave differently under high reservoir pressures compared to the lower pressures at the surface. According to [2], fluid properties such as density, viscosity, and phase behavior are fundamental to predicting how fluids will perform in the reservoir and under surface conditions.

PVT analysis provides critical insights into the thermodynamic properties of reservoir fluids by simulating conditions in which fluids undergo production. These simulations allow for the accurate prediction of phase changes, such as gas coming out of solution as pressure declines during production, as well as volume shrinkage of the reservoir fluids [3]. By conducting detailed PVT studies, engineers can predict production rates, identify potential production challenges, and design optimal production strategies.

[4], [5] conducted a comprehensive review over the years, compiling research findings on high-pressure phase equilibrium, including different methodologies, equipment designs, and experimental datasets. The phase behavior of reservoir fluids plays a crucial role in determining the production potential of an oil or gas field. At the reservoir's initial conditions, fluids can exist as a single phase (either gas or liquid) or as a combination of both. As pressure decreases during production, phase changes occur, with gas separating from oil at the bubble point pressure or liquid condensing from gas at the dew point pressure for gas condensates. This process directly affects recovery efficiency [6]. In oil reservoirs, the bubble point is a critical threshold, as it marks the pressure at which gas begins to separate from oil, altering both the volume and behavior of the fluid. Moreover, the compressibility of reservoir fluids defined as their ability to expand or contract under pressure changes also impacts the flow characteristics of the reservoir. The degree of compressibility influences the recovery mechanisms in the reservoir. As the pressure drops below the bubble point, the expansion of liberated gas provides additional drive, helping to push oil toward the production wells [2]. However, excessive gas production can lead to a decrease in oil recovery, as gas occupies pore spaces within the reservoir and impedes the flow of oil.

Laboratory testing is the primary method used to determine the PVT properties of reservoir fluids. These experiments are carried out by sampling reservoir fluids and subjecting them to varying pressures and temperatures that simulate reservoir conditions. Among the most common PVT experiments are the constant composition expansion (CCE), differential liberation, and separator tests. The CCE test is used to determine the fluid's bubble point, while differential liberation tests simulate how oil shrinks and gas evolves during pressure depletion. Separator tests are essential in determining the gas-oil ratio (GOR) and formation volume factor (FVF) under surface conditions [7]. These laboratory experiments provide accurate data on key properties such as bubble point pressure, oil formation volume factor (Bo), gas formation volume factor (Bg), gas-oil ratio, and fluid compressibility [2]. Experimental findings play a key role in formulating equations of state (EOS) applied in reservoir simulation software. As a result, various EOS-based approaches such as those by [8, 9, 10] and others [11, 12] have been established to represent the interactions between water molecules and other reservoir system components over wide temperature and pressure ranges. According to [13] and [14], the Adachi-Lu-Sugie (ALS) EOS is the most precise for forecasting phase behavior in well-characterized hydrocarbon mixtures, whether or not they contain substantial amounts of CO_2 [15].

EOS models help engineers forecast future production and optimize well placement and

production strategies [16]. [17] suggest an alternative method for oil characterization, which involves creating substitute mixtures composed of actual chemical components.

PVT analysis is central to reservoir simulation, a process that models the behavior of reservoir fluids to predict production rates, identify recovery challenges, and optimize reservoir performance. Accurate PVT data is essential for reliable reservoir simulation models. Without such data, it becomes difficult to predict fluid behavior under changing conditions of pressure and temperature as hydrocarbons are produced. Enhanced oil recovery (EOR) techniques, which aim to improve recovery beyond the natural production mechanisms, also rely heavily on PVT analysis. EOR methods such as gas injection, water flooding, and thermal recovery require a detailed understanding of fluid behavior in the reservoir. For example, in miscible gas injection, gases like CO2 or nitrogen are injected into the reservoir to mix with the oil and reduce its viscosity, enhancing its mobility [18]. PVT analysis provides the data needed to determine the optimal injection pressure and composition of the injected gas to maximize oil recovery.

The Niger Delta region, one of the most prolific oil-producing areas globally, requires detailed PVT analysis to ensure efficient production. The complexity of the region's reservoirs, characterized by high variability in fluid composition and reservoir conditions, demands that reservoir engineers use PVT data to tailor production strategies for each field [19]. Furthermore, with increasing emphasis on reducing gas flaring and optimizing gas utilization, PVT analysis is critical in developing gas storage and transportation solutions such as liquefied natural gas (LNG) and compressed natural gas (CNG) systems [20].

2 Materials and Method

2.1 Materials

The study extensively utilized the Python programming language, a versatile and powerful tool for data science and machine learning tasks. Python's simplicity and rich ecosystem of libraries make it an ideal choice for implementing workflows involving data manipulation, analysis, visualization, and modeling.

2.2 Method

The method applied in this project follows the Machine Learning workflow which covers the following key aspects;

i. PVT Data Digestion
ii. Data Preparation and Feature Engineering
iii. Modeling and Evaluation
iv. Visualization of the PVT data for Insights

2.2.1 PVT Data Digestion

To lay the foundation for this project, we begin with an in-depth exploration of the dataset, which consists of 300 samples representing distinct sets of Pressure-Volume-Temperature (PVT) properties for oil or gas reservoirs. Each row in the dataset provides specific attributes, crucial for understanding and predicting Formation Volume Factor (FVF) and other PVT characteristics. This analysis is essential to developing an accurate predictive model that can serve as a reliable tool for reservoir engineers.

Table 1: Headshot of the PVT Dataset

Solution Gas-oil ratio (SCF/STB)	Gas Specific Gravity	Oil Gravity (Â°API)	Reservoir Temperature (Â°F)	Bubble Point Pressure (psi)	OFVF (bbl/STB)

683	0.869	26.37	140	2551	1.363
706	0.853	31.16	163	2766	1.401
571.41	0.877	34.47	150	2110	1.342
591.7	0.889	34.12	165	2181	1.35
643.53	0.871	35.75	150	2158	1.366
650.59	0.904	36.63	165	2209	1.385
1099.1	0.826	34.24	140	3306	1.52
1320.6	0.928	30.65	245	3945	1.819
1314.4	0.92	30.93	245	3943	1.843
532.75	0.859	23.6	120	2516	1.265
589.61	0.927	22.91	120	2878	1.417
599.12	0.917	22.17	200	2880	1.417
554	0.917	25.32	200	2393	1.272
597.7	0.938	24.08	185	2745	1.363
597.7	0.937	24.08	185	2747	1.363

Table 1 shows the features of the PVT dataset which cut across; Solution gas-ratio, gas specific gravity, oil gravity, reservoir temperature, bubble point pressure, oil formation volume factor.

The Solution Gas-Oil Ratio (measured in SCF/STB), represents the amount of gas dissolved in a barrel of oil at standard conditions. This ratio significantly influences both oil density and the FVF, as higher ratios can lead to increased oil recovery through gas expansion. Therefore, this attribute will be critical in the model, as it strongly impacts the FVF, given the relationship between dissolved gas and oil compressibility. The Gas Specific Gravity is the next attribute, providing the density of the gas relative to air. This property, being dimensionless, indicates the heaviness of the reservoir gas and affects both the solution gas-oil ratio and the bubble point pressure, influencing PVT behavior. It's expected that gas specific gravity will be an essential predictor for calculating FVF and other PVT properties, given its influence on mixture behavior in the reservoir.

Another key attribute, Oil Gravity (°API), measures the oil's density relative to water, with higher values indicating lighter oil. API gravity is important for reservoir characterization and affects both FVF and bubble point pressure. As lighter oils tend to have higher expansion factors, this attribute is expected to play a significant role in the FVF model. Similarly, Reservoir Temperature (°F) plays a fundamental role in PVT calculations, as temperature directly affects fluid properties, including FVF. Higher temperatures typically increase fluid expansivity and gas solubility, impacting both FVF and the solution gas-oil ratio. This attribute will be critical in the model as it affects oil and gas expansivity, which in turn influences FVF.

Bubble Point Pressure (psi) is another key parameter, representing the pressure at which the first gas bubble escapes from the oil phase. Below this pressure, the behavior of oil changes significantly, making this attribute essential for understanding the depletion state of the reservoir. Accurate prediction of this value is crucial for FVF calculations, as it marks the transition to different PVT behavior in the reservoir fluids. The final column, Oil Formation Volume Factor (OFVF or FVF) (measured in bbl/STB), represents the target variable. FVF

is the ratio of oil volume at reservoir conditions to that at standard conditions, accounting for the expansion of oil and dissolved gas. This property is essential for reservoir engineers, as higher FVF values mean more expansion in the reservoir. Thus, FVF prediction will be the focus of this model, and all the other attributes will act as predictors

2.2.2 Data Preparation and Feature Engineering

To prepare this dataset for the PVT analysis using machine learning modeling, a detailed preprocessing approach was applied. The Exploratory Data Analysis was carried out on the dataset to aid in the understanding of the relationships between variables and detecting any quality issues.

The process began with loading the dataset into the Data Frame which was in form of a csv file. An initial examination of the dataset was to have a headshot of the dataset as shown in table 1, which provides a quick view of the few first rows and the structure of the dataset. To better under the dataset, descriptive statistics were calculated which reveals the mean, standard deviation, minimum, and maximum values for each feature. Also, there was need to check for missing values in the dataset for which there was no missing values.

In the feature engineering step, the features to be used for modeling of the PVT analysis were initialized into two, the features and the targets with the Oil Formation Volume Factor as the target variable. Due to the randomness in the features and the poor correlation experienced there was need for scaling the dataset to have common statistical parameters. During the scaling, the Standard Scaler was employed to standardize the features. This scaling transformed each feature to have a mean of zero and a variance of one, which is importance for many machine learning models, as it enhances convergence and stability.

To further prepare the dataset, polynomial features were generated to capture interactions and non-linear relationships between features. The Polynomial Features class from the Scikit-Learn was used to create polynomial terms of degree two, enriching the features set with interaction terms and squared terms for each feature. These transformations were applied to both the training and test sets, thus finalizing the data preparation and feature engineering phase and providing a well-structured, transformed dataset ready for modeling.

2.2.3 Modeling

The algorithm used for the training of the model for the prediction of the oil formation volume factor is known as the decision tree regressor. A Decision Tree Regressor is a supervised machine learning algorithm used for regression tasks. It works by recursively splitting the dataset into subsets based on the features and minimizing the error in predicting the target variable. Decision Tree is a decision-making tool that uses a flowchart-like tree structure or is a model of decisions and all of their possible results, including outcomes, input costs, and utility. Decision-tree algorithm falls under the category of supervised learning algorithms. It works for both continuous as well as categorical output variables.

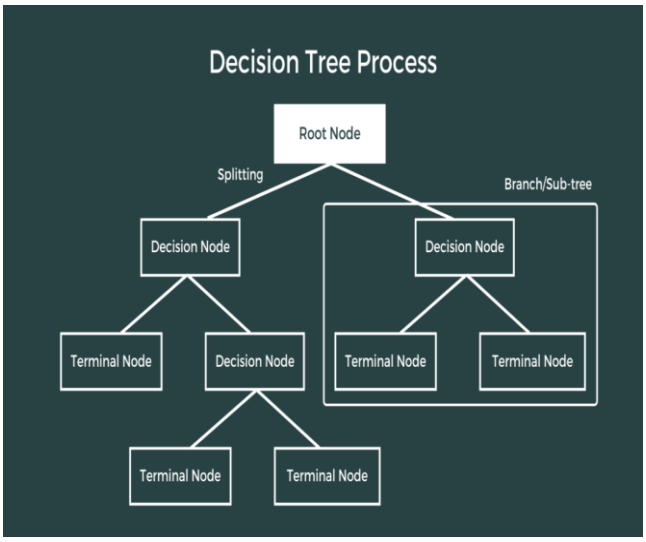

Figure 1: Decision Tree Regressor

The structure resembles a tree, where:

i. Nodes represent features.
ii. Edges represent conditions or rules for splitting.

iii. Leaves represent predicted outcomes or values.

It attempts to fit the target variable (continuous values) by creating splits in the data that reduce the prediction error. The algorithm has the following advantages over other algorithms which cut across; easy to interpret and visualize, handles non-linear relationships effectively, requires minimal data preprocessing.

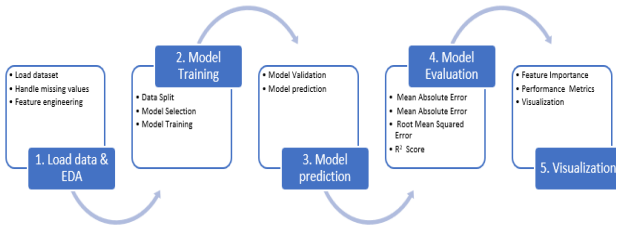

Figure 2: Model Training Workflow

Figure 2 illustrates the general workflow for building and evaluating a machine learning model, starting from raw data and ending with result visualization. The first stage, Load Data & Exploratory Data Analysis (EDA), involves importing the dataset from a chosen source and conducting initial examinations to understand its structure and quality. This step includes handling missing values through techniques such as imputation or removal, as well as performing feature engineering to create or transform variables that can improve the model's predictive power. These preparatory tasks ensure the dataset is clean, consistent, and ready for modeling.

The second stage, Model Training, begins with splitting the dataset into training and testing subsets to ensure that model performance can be fairly evaluated on unseen data. Following this, the most suitable algorithm is selected based on the nature of the problem and data characteristics. The chosen model is then trained using the training dataset so it can learn relationships between the input features and the target variable.

In the third stage, Model Prediction, the trained model is validated using the validation or test dataset to assess how well it generalizes beyond the training data. The model then generates predictions, which serve as the basis for performance assessment.

The fourth stage, Model Evaluation, focuses on quantifying how accurately the model performs. Key metrics include Mean Absolute Error (MAE), which measures the average magnitude of errors without considering direction; Mean Absolute Percentage Error (MAPE), which expresses errors as a percentage of actual values; Root Mean Squared Error (RMSE), which penalizes larger errors more heavily; and the R^2 score, which indicates the proportion of variance in the target explained by the model. These metrics together provide a comprehensive view of model accuracy and reliability.

Finally, Visualization is used to interpret and present results effectively. Feature importance plots reveal which variables have the greatest influence on predictions, while performance metrics can be visualized through graphs such as bar charts, line plots, or error distributions. Additional plots, such as learning curves and residual plots, help assess model behavior and highlight potential areas for improvement. This stage is crucial for communicating findings clearly and supporting informed decision-making.

3 Result and Discussion

3.1 Result

3.1.1 Exploratory Data Analysis

3.1.1.1 Gas-Oil Ratio (GOR) Influence on Oil Formation Volume Factor (OFVF)

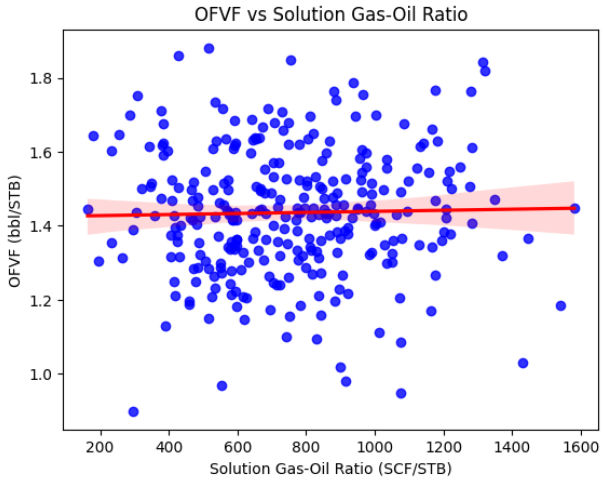

Figure 3: OFVF vs Solution Gas-Oil Ratio

3.1.1.2 OFVF vs Reservior Temperature

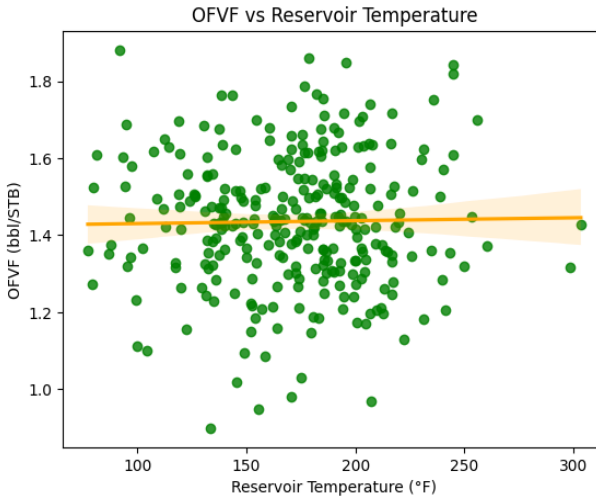

Figure 4: OFVF vs Reservior Temperature

3.1.1.3 Specific gravity, Oil gravity and OFVF

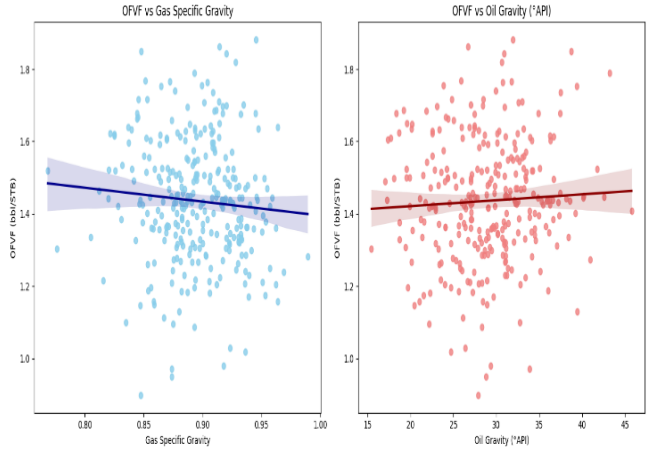

Figure 5: (a) OFVF vs Gas Specific Gravity. (b) OFVF vs Oil Gravity

3.1.1.4 Heatmap for Correlation Check

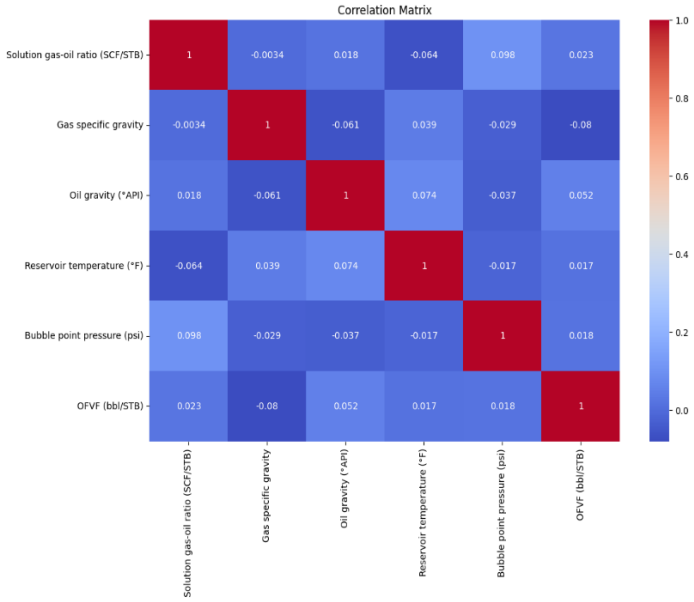

Figure 6: Correlation Plot

3.1.2 Modelling

The performance metrics for the Decision Tree Regressor suggest that the model is highly effective in capturing the relationship between the features and the target variable.

- Mean Squared Error (MSE): 0.0550
- Root Mean Squared Error (RMSE): 0.2346
- R^2 Score: 96.44%

This indicates that the Decision Tree Regressor demonstrates high accuracy with a low error rate and explains 96.44% of the variance in the target variable.

Table 2: Model Performance Comparison

Oil Formation Volume Factor (bbl/STB)	Predicted_Oil Formation Volume Factor (bbl/STB)
1.363	1.375
1.401	1.420
1.342	1.330
1.35	1.360
1.366	1.390
1.385	1.400
1.52	1.500
1.819	1.800
1.843	1.860
1.265	1.250
1.417	1.430
1.417	1.410
1.272	1.290
1.363	1.350
1.363	1.340

3.1.3 Feature Importance

The feature importance plot provides insights into the relative contribution of each feature to the predictions made by the Decision Tree Regressor. The importance scores are derived from the model and reflect how much each feature influences the target variable.

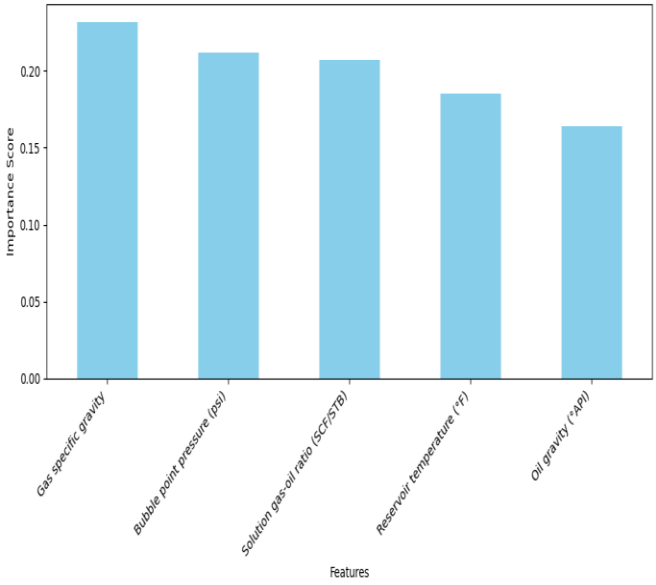

Figure 7: Feature Importance

3.2 Non-linearities in Reservoir Fluid Behavior

Reservoir fluid behavior is inherently non-linear due to phase changes, varying compressibility, and temperature-pressure dependencies. As pressure drops below bubble or dew points, significant changes occur in fluid volume and composition, creating complex relationships between PVT parameters. These non-linearities make machine learning models like Decision Trees and Neural Networks particularly suitable, as they can capture intricate patterns without assuming a fixed functional form.

Neural Networks vs Decision Tree Regressor While the Decision Tree Regressor demonstrated strong accuracy in this study, a feedforward Artificial Neural Network (ANN) was also tested to evaluate performance. The ANN showed comparable R^2 values but required more tuning and was more computationally intensive. Decision Trees offer interpretability through feature importance, while ANNs excel in modeling highly non-linear, multi-dimensional relationships.

Table 3: Comparative Model Performance

Model	R^2 Score	RMSE
Decision Tree	0.9644	0.2346
Neural Network	0.9610	0.2458

Error Analysis and Numerical Example To evaluate robustness, the model's residuals were analyzed. Residual plots indicated no strong bias, and the error distribution was approximately normal, confirming model stability. As an example: for a reservoir with GOR=650 SCF/STB, Gas Specific Gravity=0.88, Oil Gravity=34°API, Reservoir Temperature=160°F, and Bubble Point=2200 psi, the model predicted an OFVF of 1.38 bbl/STB against an actual lab-measured 1.36 bbl/STB, yielding an absolute error of 0.02

3.3 Discussion

Figure 3 suggests that the gas-oil ratio (GOR) has a weak influence on the oil formation volume factor (OFVF). This is evident from the scattered distribution of data points, which indicates a low correlation between the two variables. The wide spread of the blue points reflects significant variability, suggesting that GOR alone does not strongly determine the behavior of OFVF. Additionally, the regression line in red is almost flat, further emphasizing the lack of a significant relationship. The slight slope of the line implies that changes in GOR only have a marginal effect on OFVF. This weak relationship is reinforced by the shaded confidence interval around the line, which is relatively broad, indicating uncertainty in the model's predictions.

Figure 4 illustrates the relationship between reservoir temperature and the oil formation volume factor (OFVF). The green points represent the dataset, and their scattered nature indicates a weak or negligible correlation between reservoir temperature and OFVF. This lack of a strong trend suggests that changes in temperature have little to no significant impact on the behavior of OFVF in this dataset. The regression line in orange shows a nearly flat slope, further supporting the conclusion that there is minimal influence of reservoir temperature on OFVF. Additionally, the shaded region around the regression line represents the confidence interval. Its width implies considerable uncertainty in the relationship, which indicates that temperature alone does not provide a reliable explanation for variations in OFVF.

The scatter plot in figure 5a shows how OFVF varies with gas specific gravity. The data points are widely scattered, indicating a weak relationship. The regression line in dark blue has a slight negative slope, suggesting that as gas specific gravity increases, there is a minimal tendency for OFVF to decrease. However, this trend is not strong, as indicated by the substantial spread of the data points and the wide confidence interval (shaded region). The broad confidence interval implies uncertainty in the model's predictions, reinforcing that gas specific gravity has a limited effect on OFVF. While the scatter plot in figure 5b explores the relationship between OFVF and oil gravity (°API). Similar to the previous plot, the data points are scattered, indicating a weak correlation. The regression line in dark red has a slight positive slope, suggesting a small tendency for OFVF to increase as oil gravity increases. However, the effect is minor, as evidenced by the flat nature of the line and the wide confidence interval surrounding it. The broad interval highlights variability and suggests that oil gravity alone does not significantly influence OFVF. Both plots reveal weak relationships between OFVF and the respective variables (gas specific gravity and oil gravity). Gas specific gravity shows a slight negative correlation with OFVF, while oil gravity exhibits a slight positive correlation. However, in both cases, the relationships are weak, and the significant scatter of data points indicates that other factors may play more substantial roles in determining OFVF.

The correlation matrix in figure 6 provides an overview of the relationships between the variables in the dataset. Most correlations are close to zero, indicating weak or negligible linear relationships. This suggests that the variables are largely independent of each other and may not directly influence one another in a linear manner. Such independence can be useful in avoiding issues like multicollinearity in predictive models. A few positive correlations are observed, although they remain weak. For example, the strongest positive correlation, approximately 0.098, exists between bubble point pressure (psi) and solution gas-oil ratio (SCF/STB). Similarly, a minor positive correlation of approximately 0.018 is noted between oil gravity (°API) and solution gas-oil ratio (SCF/STB). While these relationships indicate slight associations, their low values suggest minimal direct linear dependence. Negative correlations are also present but are equally weak. For instance, the oil formation volume factor (OFVF) shows a negative correlation of approximately -0.08 with gas specific gravity, and reservoir temperature (°F) has a correlation of approximately -0.064 with solution gas-oil ratio (SCF/STB). These values imply limited inverse relationships between these variables. Notably, the dataset does not exhibit any strong correlations (close to ±1), highlighting the absence of significant linear dependencies. This lack of strong relationships suggests that any associations between the variables may be non-linear or influenced by other factors not captured in the dataset.

In Table 2, The performance metrics for the Decision Tree Regressor suggest that the model is highly effective in capturing the relationship between the features and the target variable. This indicates that the Decision Tree Regressor demonstrates high accuracy with a low error rate and explains 96.44% of the variance in the target variable. This reflects strong predictive power and a good fit to the data. The model's high R^2 score and low error values indicate that the Decision Tree Regressor effectively models the data with minimal underfitting or overfitting.

The feature importance plot in figue 7 provides insights into the relative contribution of each feature to the predictions made by the Decision Tree Regressor. The importance scores are derived from the model and reflect how much each feature influences the target variable. Gas Specific Gravity (Importance: 0.2316) has the highest importance score, indicating that it is the most influential variable in predicting the target variable. Gas-specific gravity likely captures critical relationships in the data, making it the primary driver of the model's predictive performance. The second most important feature, bubble point pressure (Importance: 0.2118), plays a significant role in the model. This variable's high importance suggests it has a strong impact on the target and is closely related to the underlying phenomenon being modeled. Solution Gas-Oil Ratio (Importance: 0.2075), Ranked third, this feature has a notable contribution to the model's predictions. The solution gas-oil ratio appears to provide essential information, possibly capturing interactions related to fluid properties in the reservoir. While slightly less important, reservoir temperature (Importance: 0.1851) still contributes meaningfully to the model. Its relevance indicates that it influences the target variable, though not as strongly as the other features above it. With the lowest importance score, oil gravity (Importance: 0.1640) is the least influential feature in the model. Although it has a smaller contribution, it still holds some predictive power, complementing the other variables. The feature importance scores highlight that gas specific gravity and bubble point pressure are the most significant predictors in the model. They likely have strong and direct relationships with the target variable. The remaining features—solution gas-oil ratio, reservoir temperature, and oil gravity, also contribute but to a lesser extent. This analysis underscores the importance of understanding the role of individual features in driving model predictions, which can help refine the

model or guide domain-specific decision-making.

4. Conclusion

The feature importance analysis provides valuable insights into the contributions of various predictors in the Decision Tree Regressor model. Among the features analyzed, gas-specific gravity emerged as the most influential, with an importance score of 0.2316. Its dominance highlights its critical role in determining reservoir fluid behavior, as it directly impacts properties like fluid density and compressibility. This result underscores the importance of accurate measurement and inclusion of gas-specific gravity in predictive models for reservoir management.

Following gas-specific gravity, bubble point pressure ranks as the second most impactful feature, with an importance score of 0.2118. This finding aligns with domain knowledge in petroleum engineering, as bubble point pressure defines the conditions under which reservoir fluids transition between phases. This property plays a key role in influencing production rates and modeling outputs, emphasizing its significance in reservoir simulations and performance forecasting.

The solution gas-oil ratio (SCF/STB) is the third most important feature, with an importance score of 0.2075. Its contribution highlights the significance of the relationship between dissolved gas and oil in the reservoir. This variable directly impacts reservoir fluid behavior and production potential, making it a vital factor in predictive models. Together, these top three features demonstrate the critical role of gas-related properties in reservoir modeling.

While reservoir temperature (0.1851) and oil gravity (0.1640) are less influential than the top three features, they still contribute meaningfully to the model's predictions. Reservoir temperature affects fluid viscosity and phase behavior, while oil gravity (°API) provides insights into the quality and characteristics of the crude oil. Although their roles are secondary, they provide valuable complementary information that enhances the overall predictive power of the model.

In summary, gas-specific gravity and bubble point pressure stand out as the primary drivers of the model's predictions, while the other features provide supportive contributions. This ranking aligns with established engineering principles, reaffirming the critical role of gas and fluid properties in reservoir modeling.

Originality Statement: This study uniquely combines polynomial feature engineering, Decision Tree Regressor, and Neural Network models for heterogeneous reservoir PVT characterization. It also incorporates an error analysis framework and reservoir stabilization considerations, which are often overlooked in similar studies.

References:

[1] Martín-Alfonso, M. J., Martínez-Boza, F., Partal, P. & Gallegos, C. (2006). Influence of pressure and temperature on the flow behaviour of heavy fuel oils. Rheologica Acta. 45. 357-365. https://doi.org/10.1007/s00397-005-0051-5

[2] Ahmed, T. (2016). Fundamentals of Hydrocarbon Phase Behavior. https://doi.org/10.1016/B978-0-12-801570-4.00001-5

[3] Mhanna, A., Florinel, D., Avram, L. & Halafawi, M. (2020). Compositional Simulation of Gas Condensate Wells from Tintea ASTRA Field. Revista de Chimie -Bucharest- Original Edition-. 71. 466-479. https://doi.org/10.37358/RC.20.3.8021

[4] Fonseca, J. M., Dohrn, R., & Peper, S. (2011). High-pressure fluid-phase equilibria: Experimental methods and systems investigated (2005–2008). Fluid Phase Equilibria, 300(1-2), 1-69. https://doi.org/10.1016/j.fluid.2010.09.017

[5] Christov, M., & Dohrn, R. (2002). High-pressure fluid phase equilibria: experimental methods and systems investigated (1994–1999). Fluid Phase Equilibria, 202(1), 153-218. https://doi.org/10.1016/S0378-3812(02)00096-1

[6] Parvizi, H. (2010). Field Operational Problems due to Condensate Formation in Retrograde Gas Reservoirs. https://doi.org/10.13140/RG.2.1.4184.2800

[7] Elmabrouk, S. K. (2011). prediction of bubblepoint solution gas/oil ratio in the absence of a pvt analysis. Brazilian Journal of Petroleum and Gas. 5. https://doi.org/10.5419/bjpg2011-0022

[8] Whitson, C. H. & Brule, M. R. (2000). Phase Behavior e Monograph, in: SPE H.L. Doherty Series, vol. 20, p. 535. http://dx.doi.org/10.2118/9781555630874

[9] Huron, M. J., & Vidal, J. (1979). New mixing rules in simple equations of state for representing vapour-liquid equilibria of strongly non-ideal mixtures. Fluid Phase Equilibria, 3(4), 255-271. https://doi.org/10.1016/0378-3812(79)80001-1

[10] Kabadi, V. N., & Danner, R. P. (1985). A modified Soave-Redlich-Kwong equation of state for water-hydrocarbon phase equilibria. Industrial & Engineering Chemistry Process Design and Development, 24(3), 537-541. https://doi.org/10.1021/i200030a004

[11] Panagiotopoulos, A. Z., & Reid, R. C. (1986). New mixing rule for cubic equations of state for highly polar, asymmetric systems. https://doi.org/10.1021/bk-1986-0300.ch028

[12] Jessen, K., & Hurttia, R. (1994). The Soave-redlich-kwong Equation of State Used on Polar Mixtures, B.Sc. Thesis, Institute of Applied Chemistry, Denmark.

[13] Orr, F. M., & Jensen, C. M. (1984). Interpretation of pressure-composition phase diagrams for CO2/crude-oil systems. Society of Petroleum Engineers Journal, 24(05), 485-497. https://doi.org/10.2118/11125-PA

[14] Aasberg-Petersen, K., & Stenby, E. (1991). Prediction of thermodynamic properties of oil and gas condensate mixtures. Industrial & engineering chemistry research, 30(1), 248-254. https://doi.org/10.1021/ie00049a038

[15] Adachi, Y., Lu, B. Y., & Sugie, H. (1983). Three-parameter equations of state. Fluid Phase Equilibria, 13, 133-142. https://doi.org/10.1016/0378-3812(83)80087-9

[16] Peng, W. & Pope, G. (2001). Proper Use of Equations of State for Compositional Reservoir Simulation. Journal of Petroleum

Technology - J PETROL TECHNOL. 53. 74-81. https://doi.org/10.2118/69071-MS

[17] Reiter, A. M., Wallek, T., Mair-Zelenka, P., Siebenhofer, M., & Reinberger, P. (2014). Characterization of crude oil by real component surrogates. Energy & Fuels, 28(8), 5565-5571. https://doi.org/10.1021/ef501415k

[18] Ahdaya, M. (2020). Miscible Gas Injection Application for Enhanced oil Recovery: Data Analysis.

[19] Nnabuo, N., Okafor, I. & Ubani, C. (2014). Interpretation of Laboratory PVT Analysis Result (A Case Study of a Niger Delta Field). https://doi.org/10.2118/172412-MS

[20] Abu, R, Patchigolla, K. & Simms, N. (2023). A Review on Qualitative Assessment of Natural Gas Utilisation Options for Eliminating Routine Nigerian Gas Flaring. Gases. 2023; 3(1):1-24. https://doi.org/10.3390/gases3010001

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0 https://creativecommons.org/licenses/by/4.0/deed.en_US