© OCT 2025 | IRE Journals | Volume 9 Issue 4 | ISSN: 2456-8880

# Predicting The Flowing Bottom-Hole Pressure of a Vertical Well Using Surface Pressure and Well Parameters

IKEH LESOR[1], OSIGWE UCHE STANLEY[2]
*University of Port Harcourt, Faculty of Engineering, Department of Petroleum and Gas Engineering, East-West Road, Choba, Port Harcourt, Nigeria, Choba.*

***Abstract- Obtaining the bottomhole flowing pressure of a producing well from readily available surface pressures had been a significant concern for operators in the petroleum industry, as accurate knowledge of this pressure is crucial for determining the most efficient recovery methods and lifting procedures. Although many existing correlations aim to achieve this, their predictive capabilities are limited due to the inability of current models to account for sand particles in the flow stream and the need to shut in the well for bottomhole pressure predictions, which seems counterproductive. This study introduces a data-driven approach to determine of the flowing bottomhole pressure of a vertical well using surface and well parameters. Existing models and correlations provide insights into the relationship between flowing bottomhole pressures and wellhead pressure, while artificial feedforward neural networks, random forest decision trees and support vector machine algorithms are employed to develop regression models based on available field data. Evaluation metrics such as mean squared error and mean absolute error are used to assess the performance of these machine learning models. The artificial neural network performed best on both training and testing data-sets, predicting the flowing bottomhole pressures with a mean squared error of 7.5% and a mean absolute error as low as 3.9% on the test set. This model offers advantages in estimating flowing bottomhole pressure from real-time surface pressures and well data compared to empirical models that rely on simplifying assumptions.***

***Keywords: Flowing Bottom-Hole Pressure, Wells, Reservoir, Hydrocarbon, Well Completions, Production***

## I. INTRODUCTION

In hydrocarbon reservoir exploitation, reservoir fluids (liquids and gases) flow simultaneously in any direction or pattern, which is generally called multiphase flow. The simultaneous flow of liquids and gases in the production system is much more difficult than single-phase flow because an interface, which may be smooth or not depending on the flow regime and pattern, can exist between the gases and liquids. Interpretation of data from well test analysis has traditionally been based on the assumption that the reservoir is a homogeneous single layer.

However, the true petroleum reservoir consists of layers with distinct interlayer characteristics. These layers are typically separated by interfaces that can be either permeable or impermeable. Pressure behavior in such a vertically heterogeneous system differs from that of a single-layered system and rarely reflects more than the average properties of the entire system. It is in this context that this study becomes necessary. Well completion in these systems would be more formative and improve reservoir and production engineering practices if detailed layer information is available. The petroleum industry, however focused on accurately calculating the pressure losses that occur during multiphase flow in tubing and pipelines. Accurate prediction of pressure losses enables proper system design. Additionally, pressure measurement in a production system is crucial the petroleum industry because it supports efficient oil and gas extraction from the reservoir. Amongst these, bottom-hole pressure is particularly vital because knowing it helps determine many parameters essential for optimal production and prevents early reservoir depletion. It can also be used to avoid formation damage caused by early sand production reservoir. Surface pressures can often be converted to bottom-hole values if sufficient information about the wellbore system is available

### 1.1 VERTICAL MULTIPHASE FLOW

Much has been written in the literature regarding the multiphase flow of fluids in pipes. This problem is much more complex than single phase-flow because it involves the simultaneous flow of both liquid (oil or water) and vapor (gas). The mechanical energy equation form the basis for methods used to estimate the pressure drop in multiphase flow. However, challenge lies in accurately determining the velocity, friction factor, and density to be used for the multiphase mixture in calculations. Additionally, the problem becomes more complicated as velocities, fluid properties, and the liquid fraction change as the fluid flows to the surface due to pressure variations. Many researchers (Adekomaya *et al.,* 2014, Guo, 2001, and Omohimoria *et al.*, 2014) have proposed methods to estimate pressure drops in multiphase flow. Each approach is based on a combination of theoretical, experimental, and field data, which has led some researchers to relate the pressure-drop calculations to flow patterns. Flow patterns or flow regimes describe to the distribution of each fluid phase inside the pipe. This means that pressure calculation depend on the predicted flow pattern. There are four main flow patterns in the simplest classification of flow regimes.

## 1.2 OVERVIEW OF BOTTOM HOLE FLOWING PRESSURE

The petroleum industry aims to in accurately calculate pressure losses in multiphase flow within tubing and pipelines. Precise predictions of pressure losses enable proper pipe design. Additionally, determining pressure in a production system is crucial in the petroleum industry because it helps optimize oil and gas extraction from reservoirs. Among all, the most critical measurement is the flowing bottom-hole pressure, which is the pressure recorded at or near the producing formation's depth. Although surface pressures can often be converted to bottom-hole pressure values if sufficient information about the wellbore system is available (Aggour *et al.,* 2015), Knowing this pressure is essential for selecting the most effective recovery and lifting methods. However, there is less information available about these pressures than about any other part of the broader issue of oil production (Adekomaya *et al.,* 2014).

As mentioned earlier, the bottom-hole pressure can be determined from surface pressures like the wellhead pressures if sufficient information about the production system is available, which can be easily obtained from well testing operations. Since the wellhead pressure and the relevant parameters are readily obtained from pressure transient analysis, and the success of this analysis depends on the accurate measurement or estimation of bottom-hole pressure, it is therefore important and necessary to determine the bottom-hole pressure from this data (Guo, 2001). This will be done to further emphasize the benefits of having adequate knowledge of a reservoir's bottom-hole pressure.

It is also well known that the knowledge of bottom-hole pressure is necessary for determining the well productivity index, which is derived from the curve of the inflow performance relationship, that is, the plot of bottom-hole pressure against flow rate. Therefore, the ability to monitor bottom-hole pressure is very important because it offers many advantages for reservoir management. Its monitoring capabilities can prevent severe damage to the well, which could lead to early breakthrough, early well intervention, or even premature abandonment of the well before its intended lifespan. (Clinton *et al.,* 2020).

It is essential to study ways to correct this problem quickly and cost-effectively. Therefore, this work aims to at determine the bottom-hole pressure of a vertical well from surface pressure and parameters by modifying of the general energy equation to include considering only the frictional pressure term. (Omohimoria *et al.*, 2014)

Since the inception of the original work on multiphase flow by Poetmann and Carpenter (1952), several authors (Ayub *et al.,* 2014; Medhat *et al.,* 2015; Guo, 2011) have developed various correlations and models. Poetmann and Carpenter (1952), Ros, N.C. (1961), and Orkiszewski, J. (1967) developed models for pressure drop or pressure gradient along the tubing, which may only provide approximate solutions. This means they might not offer accurate information about the pressure conditions at the bottom of the well caused by the fluid column containing two or more fluid phases. Their models treated the liquid and gas as a homogeneous single-phase flow without largely considering dissolved gas in oil. The developed models and correlations can be categorized into three main types: empirical models, mechanistic models and artificial neural networks. (Ayub *et al.,* 2014).

The empirical model or correlation uses measured experimental production data based on mathematical equations obtained from research facilities. While most early pressure drop calculations relied on this correlation due to its direct applicability and reasonable accuracy within the data range used, the model generation, was limited by the data range and its applicability to all for all types of fluids and conditions encountered in oil and gas fields. Beggs and Brill (1973) developed a widely used model for estimating pressure drops in horizontal, inclined, and vertical flows. The model also considered several flow regimes in multiphase flow and can be used to predict liquid holdup. The parameters used include gas flow rate, liquid flow rate, pipe diameter, inclination angle, liquid hold up, pressure gradient, and horizontal flow regime.

The mechanistic model, also known as the semi-empirical model, helps in determining and estimating of pressure drop holdup in pipes by addressing the physical phenomena of multiphase flow. This work provides a way to predict pressure drop in situations that cannot be modeled in a laboratory and where reliable and calculable empirical correlations are not availcale. Mechanistic models are generally considered more reliable and versatile because they incorporate important flowparameters. (Medhat *et al.,* 2015).

Several studies were conducted by Guo, B (2001) in various areas of oil and gas well drilling and production technology, requiring bottom-hole pressure estimations. He developed a model to simulate four-phase flow (gas, oil, water and solid particles) in underbalanced drilling practices. Later, this model was found to be useful in simulating the simultaneous flow of gas, water, and coal particles in coal-bed methane production wells, (Guo, 2011). However, the artificial neural network model became popular several years ago, as it has been applied in the industry for many purposes, such as PVT properties prediction, enhanced oil recovery, and more. It has been proven that empirical and mechanistic models do not provide convincing and reliable tools for estimating of pressure in multiphase flow wells, as high errors are usually associated with these models. The artificial neural network demonstrates a better performance compared to the conventional empirical and mechanistic models. Ayoub developed an artificial neural network model for estimating bottom-hole flowing pressure and pressure drop in vertical multiphase flow, showing the power of neural networks in solving complex engineering problems. This model could simulate the actual physical process of determining bottom-hole pressure and outperform all existing models (Ayoub *et al.,* 2015).

Flowing Bottom-Pressure Importance and the Electrical Submersible Pump (ESP) Symtem

Electric Submersible Pumping (ESP) is the second most commonly used method for well production/and fluid lifting in the oil and gas industry. It accounts for the highest volume of total fluids produced - both oil and water by any artificial lift method and is especially suitable for wells with high water cuts. Centrifugal pumps can be designed as single-stage or multi- stage units. Single-stage pumps are typically used when low to medium discharge pressure is needed, while multi-stage pumps are built to handle higher discharge pressures. This is the case with ESPs used in the petroleum industry, where fluids must be lifted from deep formations. (zhang *et al.,*2016)

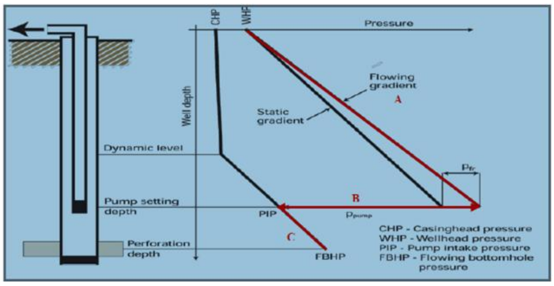


Fig. 1. ESP Well System and Typical Pressure Drop Profile Diagram (Ayub et al., 2014)

An ESP is typically installed at the end of the production tubing string, which is inserted inside a larger casing pipe. Usually, the ESP installation depth is shallower than the formation (producing zone) depth. The pressure drop schematic of a flowing oil well with an ESP is shown in Figure. 1. The pressure drop lines of interest in this study are the lines drawn in red and labeled as A, B, and C. The pressure at the top of line A is the well-head pressure, and the end of this line represents the pump discharge pressure. Line A indicates the pressure drop caused by the hydrodynamic multiphase flowing column and frictional losses in the tubing. Line B shows the difference between the discharge and intake pressures of the pump, essentially representing the total pressure developed by the ESP. Line C illustrates the pressure drop between the pump intake and the perforations at the producing formation, due to the hydraulic column and frictional losses in the casing below the pump. The top of line C is the pump intake pressure, and its lower end is the well flowing bottom-hole pressure. It is a standard practice to have online pressure measurements at the well-head, pump discharge, and pump intake. In the fields studied, these measurements are recorded every 15 minutes. In this study, re-sampled data for daily records have been used. Permanent pressure gauges installed within the ESP assembly typically measure the pressure at thye pump discharge and intake. However, the formation bottom-hole pressure (FBHP) at the perforations has no permanent measurements, and in our case, we have almost no records of the FBHP for an ESP well due to access difficulties and other restrictions. Therefore, this work is limited to estimating the pressure drops along lines A and B only. Estimating the pressure drop along line C (i.e. the pressure difference from FBHP to the pump intake pressure) is less complex compared to the drop along line A, because the flow is more homogeneous and frictional losses are negligible. Adittionally, it is very challenging to evaluate this due to nthe lack of FBHP records. The flowing bottom-hole pressure of a well is the pressure measured or calculated at or near the producing formation at the bottom of the well while the well is producing hydrocarbons. It's always higher than the surface flowing pressure, but lower than the shut- in bottom-hole pressure.

Knowing the bottom-hole pressure of an oil well can help forecast the well's potential throughout its life-cycle. In other words, well production monitoring and artificial lift optimization can be performed, which are key objectives maximizing oil production and reducing operational costs. Bottom-hole pressure data can also provide information on pore pressure, which is used for safety calculations when drilling development wells in the area. This data is especially critical for drilling operations, particularly underbalanced drilling. It also helps in selecting the accurate kill fluid weight. Additionally, this data can improve accuracy of under- or over-balance decisions before perforation.

Tubing pressures and casing pressures in flowing wells have always been key factors in well operation, and their importance increases under restricted production. Changes in these pressures, related to well age or production, provide valuable information about the well's conditions, sand presence, bore-hole conditions through the sand, and whether the equipment in the hole is functioning correctly. A broad study of bottom-hole pressures across an entire field directly applies to the operation of a specific lease or individual well. Field-wide bottom-hole pressure surveys offer data that can help make more accurate early estimates of when wells will need artificial lift and how much fluid they will. Knowing roughly when wells will require pumping is highly useful.

### 2.5 MACHINE LEARNING ALGORITHMS

This work aims to predict wellbore flowing pressure from surface pressure and well parameters using a machine learning model. The target variable is the bottom-hole pressure, while oil flow rate, gas flow rate, total gas rate, water rate, bottom hole and surface temperatures, oil gravity in API, and wellhead pressure serve as input features. This is a regression problem because the target variable is continuous. Due to the nature of these variables, a thorough study was conducted to select appropriate machine learning algorithms for this work. The considered algorithms were:

i) Artificial Neural Network
ii) Decision tree algorithm
iii) Support vector machine

#### 2.5.1 ARTIFICIAL NEURAL NETWORK (ANN)

An artificial neural network (ANN) is a linear model inspired by brain architecture, developed to transfer learning ability to a computer system (Castro *et al.,* 2017; Souza *et al.,* 2019).

Artificial neural networks can perform complex learning and adaptation tasks by mimicking the functions of biological neural systems. Unlike knowledge-based techniques, they do not require explicit knowledge for application. Their primary strength is the ability to learn complex functional relationships by generalizing from a limited amount of training data. Neural networks can thus serve as a black-box model for nonlinear systems and can be trained using input and output data observed in the system. The mathematical model simulates the functionality of biological neurons (called artificial neurons) at various levels of detail. Essentially, it is a static function with multiple inputs (representing dendrites) and one output (the axon). Each input has an associated weight factor. The weighted inputs are summed and then passed through a nonlinear activation function, which produces the neuron's output.

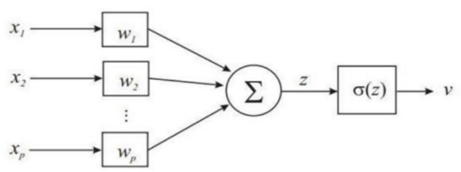


Fig. 2. Artificial Neuron (Babushka, 2010)

Networks with multiple layers are called multi-layer neural networks, compared to single-layer networks that have only one layer. They consist of one input layer, one output layer and several hidden layers in between. The layers are made up of simple nonlinear processing units called neurons. These neurons are interconnected through adjustable weights. The information relevant to the input and output mapping of the net is stored in these weights, store information related to the input and output mapping of the network. In a feedforward neutral network, information lows only in one direction, from the input layer to the output layer. Israel (2021) pointed out that each node calculates the sum of the products of the weights and inputs, and if this value exceeds threshold (typically 0), the neuron fires and outputs an activated value (usually 1); otherwise, it outputs a deactivated value (usually −1).

The number of nodes in the output layer depends on whether a regression or classification model is being built. For a regression model, the output layer has one node, as it's expects only a single output. In Contrast, for classification, the output layer of the ANN has a number of nodes equal to the number of classes being classified.

2.5.2 The Decision Tree Algorithm

The decision tree is a practical, fast, and robust method for supervised inductive learning (Maimon *et al*., 2010). It effectively aids in extracting previously unknown information from analyzing large datasets. Examples of applications using decision trees include landslides (Alkhasawneh *et al*., 2014), classification and identification of natural minerals (Akkas *et al*., 2015), and image classification (Loussaief, 2018). Essentially, a decision tree consists of a series of if-else statements organized through a nodes and leaves. When applied to database records, it classifies data and proves to be a resilient method forhandling noisy or nonstandard data (Sáez, 2013). Configurations such as maximum tree depth, number of features for the best split, maximum number of nodes, maximum number of leaves, and the functions used for division and node selection can be defined and optimized during training.For node division and selection, methods such as Gini impurity, entropy, information gain, and chi-square are available

2.5.3 SUPPORT VECTOR MACHINE

The Support Vector Machine is a supervised machine learning algorithm used mainly for classification and, to a lesser degree, regression problems (Cortes & Vapnik, 1995). It seeks to find the optimal decision boundary, also called a hyperplane that best separates the data into different classes (Bishop, 2006). The ideal hyperplane is the one that maximizes the margin, which is the distance between the hyperplane and the closest data points from each class, known as support vectors (Cortes & Vapnik, 1995). When the data is linearly separable, SVM treats a straight hyperplane that clearly divides the classes (Hastie, Tibshirani, & Friedman, 2009). Mathematically, this involves solving a convex optimization problem to minimize the norm of the weight vector, ensuring a wide margin between classes (Bishop, 2006). If the data is not linearly separable, SVM adds slack variables and a regularization parameter C to allow some misclassifications while still aiming to find the best separation (Hastie et al., 2009).

2.6 MACHINE LEARNING MODELS FOR BOTTOM HOLE PRESSURE DETERMINATION

Guo *et al*. (2016) pioneered the investigation of lithology identification using support vector machines (a traditional machine learning algorithm). In their research, they developed a model to predict bottom hole pressure based on four surface parameters. These parameters include oil rate, gas rate, water rate and wellhead pressure. They served as inputs to their model. A total of 100 data points were used to develop the model, with bottom hole pressure as the target variable. The algorithms used included the random forest decision tree, linear regression and support vector machine. Evaluation metrics such as mean squared error, mean absolute error, and coefficient of determination assessed the performance of the machine learning algorithms. Their study showed that a random forest decision tree was the best for predicting bottom hole flowing pressure, with a mean absolute error of 3%. The analysis of feature importance indicated that wellhead pressure played a crucial role predicting bottom hole flowing pressure.

Nagham and Ibrahim (2021) developed three machine learning models for predicting the multiphase flowing bottom hole pressure using three different algorithms. They employed an artificial neural network, a random forest, and a K-nearest neighbors algorithm. Results showed that the an artificial neural network model achieved an error of 2.5% in estimating the flowing bottom hole pressure, which was lower than the errors of 3.6% and 4% for the random forest and K-nearest neighbors models respectively. The machine learning model were built using surface production data, making it possible to predict the flowing bottom hole pressure with actual field data. The accuracy of the models was validated by comparing their effectiveness. Overall, the study demonstrates the potential of artificial intelligence in predicting one of the most complex parameters in multiphase petroleum production.

1. Method and Procedures

This study aims to develop a model to predict the bottom hole flowing pressure using well and surface parameters.

The different steps involved in developing this AI-based model includes: data gathering, data preprocessing, selection of algorithms and model training, and model evaluation, validation, and prediction. To develop the model for bottom-hole pressure prediction, three different machine learning algorithms were used. The models built are regression models since the target (bottom-hole pressure) is a continuous variable. The developed machine learning models can predict the flowing bottom-hole pressure from various well and surface parameters. The algorithms used include: random forest regressor, support vector machine and linear regressor. It is important to note that the steps in developing the model using any of these algorithms are essentially the same. Machine learning model building follows these generic steps as mentioned above. To simplify and reduce ambiguity, the steps for developing this AI -based proxy modelcan be grouped into five major phases. These are:

- Data Acquisition / Input Data Selection
- Data Preprocessing
- Network Development
- Network Training
- Network Validation

DATA ACQUISITION / DATA SELECTION

A total of 206 multiphase flow data points collected from Niger Delta fields for vertical wells were obtained. These wells are flowing naturally without any artificial lift process. During the measurements, the well bottom-hole flowing pressure was recorded using down-hole pressure gauges located just above the perforations. The dataset includes 9 production-related variables used to predict the bottom hole flowing pressure, FBHP (psia) as well as the flowing oil rate Qo (bbl/day), flowing gas rate, Qg (Mscf/day), flowing water rate, Qw (bbl/day), production tubing internal diameter, ID (inches), perforation depth, (ft), oil gravity, API, surface temperature, ST (ºF), bottom-hole temperature, BT(ºF), and wellhead pressure, Pwh (psia). The output is the measured flowing bottom-hole pressure FBHP, (psia).

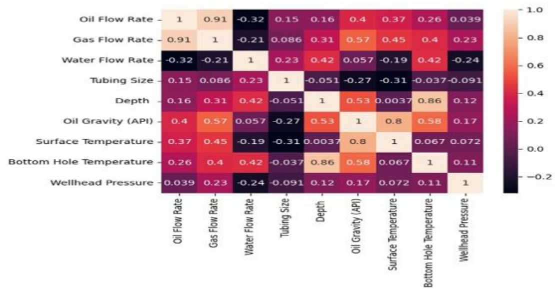

Fig. 3. The correlation plot of all attributes in the dataset.

## II. DATA PREPROCESSING AND FEATURE ENGINEERING

Data preprocessing, feature engineering, and exploratory data analysis are essential initial steps after collecting data.

Data preprocessing includes to data integration, analysis, data cleaning, transformation, and dimension reduction. It is the process of cleaning and preparing raw data to facilitate feature engineering.

Feature engineering, on the other hand, involves techniques such as adding or removing relevant features, handling missing data, encoding data, and managing categorical variables. The performance of a machine learning model is heavily depents on both the quantity and quality of the training data.Although we have little controlover the amount of data, ensuring data quality is solely the responsibility of modeldeveloper. In this work, data analysis and preprocessing are carefully performed to ensure the training date-set quality. To reduce the number of input parameters and improve the model's efficieny, fewer data points were used.

Feature selection is allowed to train machine learning models in a way that is as honest as possible based on the data. A dataset includes nine independent variables, or input attributes. It is important to select the right number of these variables because using too many during model training can cause lengthy processes and overfitting.Strong correlations between independent variables can lead to collinearity and multicollinearity issues in linear machine learning models, which reduce the models' prediction accuracy (Munqith *et al.,* 2017. Although collinearity does not affect the chosen machine learning techniques, feature selection is performed to remove redundant variables. The correlation between variables was calculated using Pearson's correlation coefficient, a popular method for measuring the linear relationship between two data variables, which can range from -1 to 1.

NETWORK DEVELOPMENT/ TRAINING

Traditional algorithms were the only ones used for the models development due to limited data. A total of 206 tests were collected from different wells and divided into two sets to train and validate the model. The training dataset account for 80% (165 data points) were used to prepare and train the models, while the remaining 20% of the data (41 data points) was reserved. The testing set was used to validate the trained model and assess the prediction capabilities of the developed models.

Both models rely on scientific libraries like pandas and numpy for data manipulation and matrix operation. The visualization library was also useful for data visualization. The data was loaded into the Python environment as a DataFrame using pandas, and visualized with matplotlib and seaborn. For the traditional algorithms, the scikit- learn library was used to import pre-developed model from their library. Support vector machines and random forest regressors were the chosen algorithms. The training process for both involved hyperparameter tuning, which helps optimize the model's parameters. Gridsearchcv was used to enhance the performance of both algorithms.

NETWORK VALIDATION

Validating a model involves comparing its results to actual data. The final model was validated by testing it against different data sets that were not used during training. Evaluation metrics such as mean squared error and mean absolute error assesses the model's performance. The mean squared error is defined as the average of the squared of the differences between actual and predicted values of the target variable. It indicates how well the model performs, with lower values being better, since high values suggest poor performance. Conversely, the mean absolute error measures the average of absolute differences between the actual and predicted target variable.

## III. RESULTS

Before data is used for model development, extensive questions are asked about the data and potential solutions are provided at stages where the investigation reveals a loophole. Exploratory data analysis is essential for understanding quantitative variables in a dataset and serve as a visual tool for high-dimensional data. The goal of this analysis is to identify patterns within the dataset.

Table 1. Statistical Description of Dataset

| | Oil rate | Gas rate | Water rate | Tubing size | Depth | Oil Gravity | ST | BT | Pwh | Pwf |
|---|---|---|---|---|---|---|---|---|---|---|
| Count | 206 | 206 | 206 | 206 | 206 | 206 | 206 | 206 | 206 | 206 |
| mean | 6321.51 | 3416.07 | 2700.00 | 3.833 | 6359.86 | 33.7723 | 117.733 | 203.640 | 321.077 | 2489.03 |
| Std | 4835.15 | 3068.43 | 2793.08 | 0.387 | 566.278 | 2.3179 | 30.7934 | 16.9572 | 153.563 | 302.165 |
| min | 280.00 | 33.600 | 0 | 1.995 | 4550.00 | 30.00 | 76.00 | 157.00 | 80.00 | 1227.00 |
| 25% | 2543.75 | 1051.6025 | 3.250 | 3.813 | 6299.75 | 32.60 | 90.00 | 208.00 | 210.00 | 2288.25 |
| 50% | 4761.50 | 2454.525 | 1834.50 | 3.958 | 6509.5 | 32.60 | 90.00 | 212.00 | 280.00 | 2500.00 |
| 75% | 9576.00 | 4918.515 | 5033.50 | 3.958 | 6712.75 | 36.50 | 155.00 | 212.00 | 390.00 | 2700.50 |
| max | 19618.00 | 13562.20 | 11000.0 | 4.00 | 7100.00 | 37.00 | 160.00 | 215.00 | 960.00 | 3217.00 |

ST: surface temperature.
BT: Bottom hole temperature.
Pwh: Wellhead pressure. , Pwf: Bottom hole pressure

The table above provides a detailed description of the data set. Descriptive data analysis was performed on the ten variables. The first nine columns represent the input features, while the last column is the target feature. Descriptive data analysis is essential during exploratory data analysis because it helps the analyst gain a thorough understanding of the data distribution. The rows of the table offer various descriptions of the data. The count indicates the total number of observations in the data-set. From Table , row one, it can be inferred that there are no missing value in the data, as all entries have the same count row (206). This means missing data points were not an issue during the modeling process.

Graphically, outliers can be tested using a box plot as shown in Figures 4 and 5. The figures (Figures 4 and 5) display few or no outliers in the oil flow rate and the gas flow rate. This was also true for all the variables used in this model development.

There were four basic questions about the data asked before starting the data analysis. The first concerns to whether the data is discrete or continuous. The second involves examining the symmetry of the distribution to identify if skewness exists in the dataset. The third question relates to the upper and lower boundaries of the data, and the final question assesses the likelihood of observing extreme values in the distribution. Most of the data used is continuous with values within a finite interval, except for the facies variable. The density plot (Figure 4) was used to display the variables with skewed data points, and normalization was performed on the data-set before using it to develop the model.

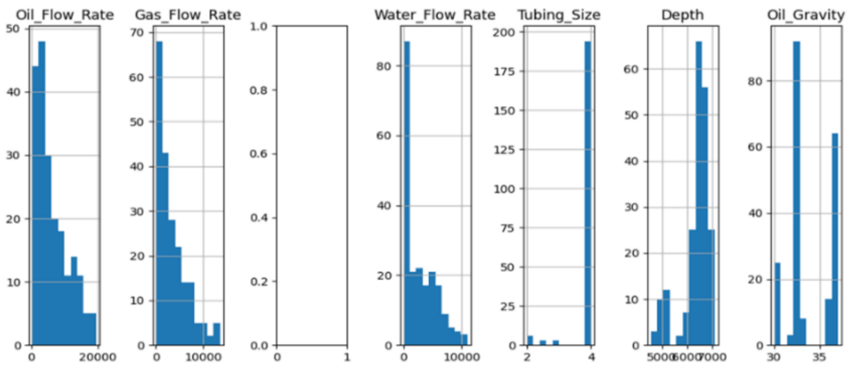

Fig. 4. Density plot showing the distribution of the input feature.

Proper attention was given to normalizing or scaling these skewed input features before developing the model.It was also necessary to understand the pairwise relationships between variables. A bivariate distribution in the form of a pair plot was created. This plot is shown in Figure 5.The pair plot shows illustrates the different variables relate to each other, providing more insight into the data. From the pair plot (Figure 5), it can be observed that gas flow rate, surface temperature, and bottom -hole temperature exhibit a very strong positive relationship with other input variables. This relationship among input features is called multicollinearity and could affect the performance of the models, especially those built with traditional machine learning algorithms.

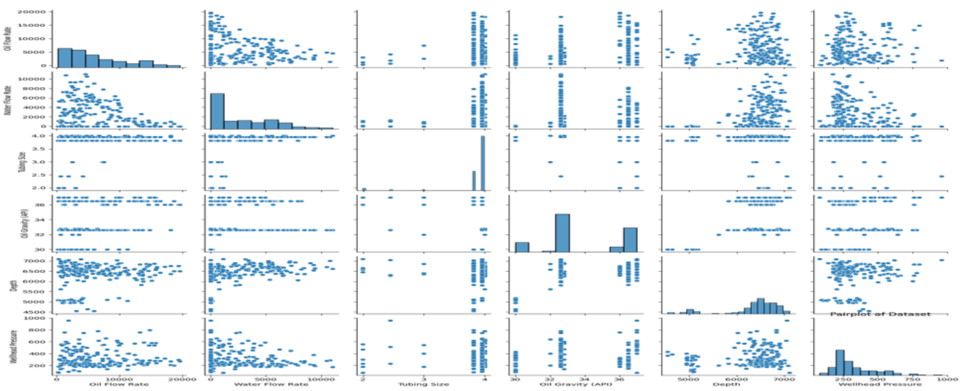
Fig. 5. Pair plot showing bivariate relationship between variables in the dataset

The correlation plot (Figure 6) below is a good way of showing these positive correlations that exist among input features. Figure 6 shows the correlation plot of the variable and the magnitude of their correlation coefficient.

Due to the issues of multicollinearity and its purported impact on model performance, those input features with strong correlation were dropped and the model was developed with only features which correlates only with the target feature (bottom hole flowing pressure) .

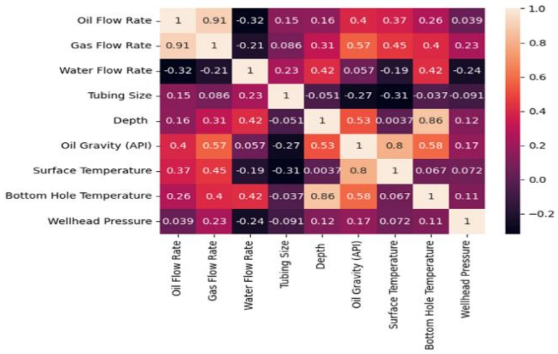


Fig. 6. Correlation Plot with Multicollinearity

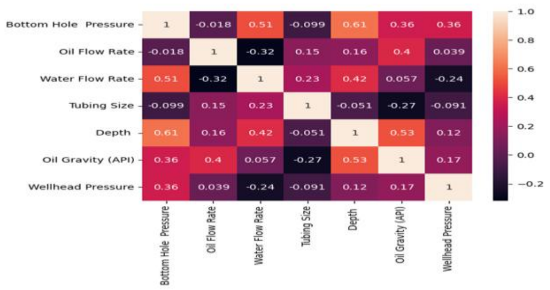


Fig. 7. Correlation Plot After Handling Multicollinearity

The distribution of the target data set was also examined before developing the model, as shown in Figure 8. This was done to check for skewness in the target variable. The histogram below (Figure 8) was used for this purpose and it shows a normal distribution for the model development.

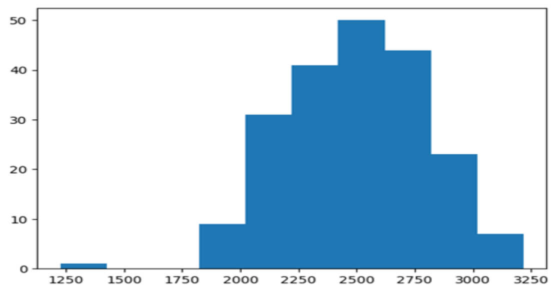

Fig. 8. Distribution of the target data

Again, four different algorithms were used to develop a model for predicting the bottom- hole pressure. The model consisted of one deep learning algorithm and two traditional algorithms. The algorithms included an artificial neural network, a decision tree algorithm, and a support vector machine. After removing the blind data set for model validation, about ten percent of the remaining data was used to test the model's performance. Additionally, the relationship between the target variable (bottom-hole pressure) and other input features examined and represented graphically before model development. Although not all input features were plotted, the depth and oil flow rate showed a direct relationship, as shown in Figure 9 and Figure 10.

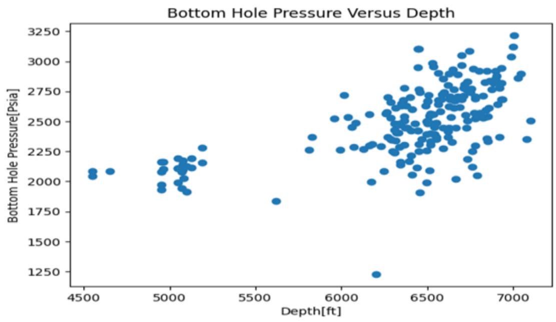


Fig. 9. Scatter Plot of Bottom Hole Pressure against Depth

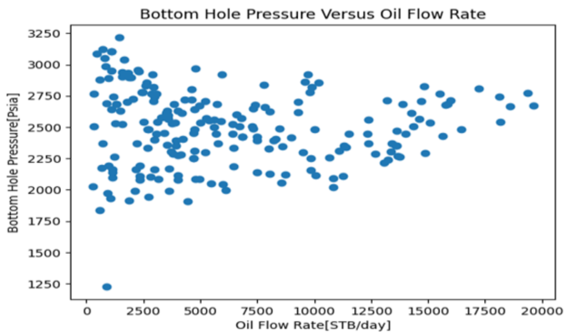


Fig. 10. Scatter Plot of Bottom Hole Pressure against Oil flow rate.

After the exploratory data analysis, the data-set was normalized and then split into two parts for model development. The first part of clean data (80% of the total data) was used to train the algorithms, while the second part (20%) was used to tune and validate the performance of the trained model.

## IV. DISCUSSION

This work aims to predict the bottom hole flowing pressure of a vertical well using a machine-learning model based on surface parameters. To maximize the performance of various models, the hyperparameters within the algorithms were tuned to their optimal values. For the artificial neural network, the keras Tuner from Tensor Flow was used to find the best hyperparameters. The GridsearchCV (a package in sklearn) was employed to identify the optimal hyperparameters for the traditional algorithms (support vector machine and decision tree). The table below shows the hyperparameters used for model development.

Table 2. Hyper parameter for various Algorithms.

| S/N | Hyper parameter | ANN | Support vector | Decision tree |
|---|---|---|---|---|
| 1 | Hidden Layer | ✓ | — | — |
| 2 | Batch Normalization | ✓ | — | — |
| 3 | Optimizer | ✓ | — | ✓ |
| 4 | Learning rate | ✓ | — | — |
| 5 | Max-depth | — | — | ✓ |
| 6 | Min-samples-split | — | — | ✓ |
| 7 | Min-sample-leaf | — | — | ✓ |
| 8 | Regularization parameter (C) | — | ✓ | — |
| 9 | Gamma | — | ✓ | — |
| 10 | Kernel | — | ✓ | — |

The three algorithms were trained, tuned, and validated independently using the necessary data-set. The ANN model was optimized with three hidden layers, while the support vector machine and decision tree repressor were optimized by adjusting their hyperparameters as shown in Table 2 above.

STATISTICAL ANALYSIS

After developing the model, it was used to predict the bottom -hole pressure based on test input data. This step is common in every machine learning workflow to assess how well the model can predict data it hasn't seen to during training. This is known as the concept of generalization. To evaluate the model's predictions and compare them with actual values, statistical analysis was performed on the forecasted data set for the developed models. Table 3 and Table 4 shows a brief summary of these metrics after the models were used to predict the training and test data set

Table 3. Performance of Models on Training Data

| Model | MSE | MAE |
|---|---|---|
| ANN | 5.2 | 3 |
| Decision Tree | 7.4 | 3.7 |
| Support Vector Machine | 8.3 | 4.2 |

Table 4. Performance of Models on Testing Data

| Model | MSE | MAE |
|---|---|---|
| ANN | 7.5 | 3.9 |
| Decision Tree | 9.4 | 5.2 |
| Support Vector Machine | 10.7 | 5.6 |

Model Comparison

The developed models were evaluated using their respective mean squared error and mean absolute error when predicting the test and training data sets. The results from the evaluation are shown uin the bar charts in Figure 11. From Figure 11, it can be deduced that the ANN model outperformed the support vector machine and decision tree regressor. This could be due to the inherent ability of deep learning algorithms. In general, ANN, which are invariants of deep learning algorithms, have the ability to detect patterns more easily compared to traditional algorithms like support vector machine and decision tree algorithms.

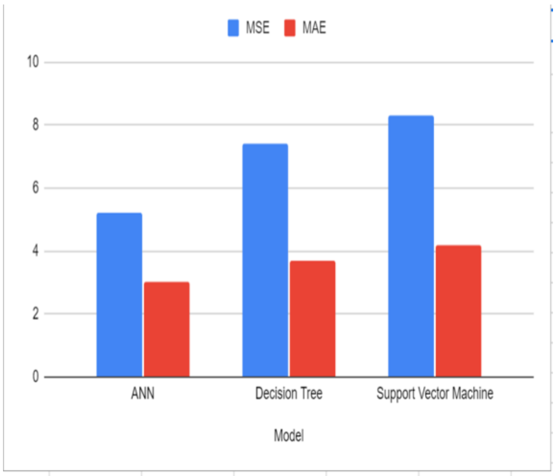


Fig. 11. Model Performance Comparison on Train Data Set

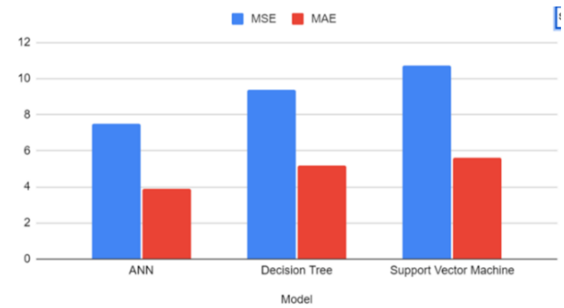


Fig. 12. Model Performance Comparison on Test Data Set

## V. CONCLUSION

The determination of the flowing bottom -hole pressure of a vertical well from surface pressure and well parameters has been carried out using a machine learning approach. From the model predictions and analysis, the following can be deducted:

The accuracy of the suggested correlations is reduced by modeling and calculating the flowing bottom-hole pressure in multiphase oil well flow, which involves to numerous assumptions. Bottom-hole pressure is the most important factor for reservoir and production engineering.

The FBHP is determined by the oil industry through empirical and/or mathematical relationships. One branch of artificial intelligence that demonstrate promising results is machine learning (ML). This advanced field of study supports complex problem-solving for humans. From this study, bottom -hole pressure during multiphase well production can be accurately predicted using machine learning methods. The most ideal model among the three machine-learning models that were built was ANN. These findings demonstrate artificial intelligence's capacity to forecast the most intricate aspects of the oil and gas sector. This effort highlights the importance the significance of data-driven computational models for production planning in the petroleum industry.

## VI. ACKNOWLEDGMENTS

The authors would like to thank the Department of Petroleum and Gas Engineering at the University of Port Harcourt for providing research facilities for this work.

Conflicts of Interest

The authors declare that there are no conflicts of interest

## REFERENCES

[1] Guo, B. (2001): An Analytical Model for Gas-Water Coal Particle Flow in Coal Bed-Methane Production Wells. Paper SPE 72369, presented at the SPE Eastern Regional Meeting held 17-19th October 2001 in Canton, Ohio, 2001.

[2] Babushka, L (1987): An introduction to computing with neural nets" vol 4, 1987

[3] Ardil, T. Manjunath, H. Mehta, and A. Mehta, (2007): A Multi-Layer Artificial Neural Network Architecture Design for Load Forecasting in Power Systems" International Journal of Applied Mathematics and Computer Sciences, 4(4), 227-228, 2007.

[4] Omohimoria C.U. and C.O. Ayodele, (2013): A Simple Analytical Model for Estimating Bottom-hole Pressure in Gas Condensate Wells. International Journal of Advancements

in Research & Technology, 2013, 2(12), 148-155.

[5] Millikan, C.V. and Sidwell, C.V. (1930): Bottom-hole Pressures in Oil wells. Society of Petroleum Engineers, 1930, 92(01), 194-205.

[6] Beggs D. and J.P. Brill, (1973): A study of two-phase flow in inclined pipes" Journal of. Petroleum Technology, 25(5), 607-617, 1973.

[7] Beggs D.H. and Brill. (1973): A Study of Two-Phase Flow in Inclined Pipes. J. Pet. Technol., 1973,25, 607-617.

[8] Osman, M.A. Ayoub E.A. and Aggour, M.A. (2005): Artificial Neural Network Model for Predicting Bottom-hole Flowing Pressure in Vertical Multiphase Flow. SPE 93632 presented at the 14$^{th}$ SPE Middle East Oil & Gas Show and Conference held in Bahrain International Exhibition Centre, Bahrain, 12–15 March 2005

[9] Poettmann F.H. and Carpenter, P.G. (1952): Multiphase Flow of Gas, Oil and Water through Vertical Strings with Application to the Design of Gas Lift Installation. Drilling Production Practice. 1952, 257.

[10] Orikiszewski, J.(1967): Predicting Two-Phase Pressure Drop in Vertical Pipe. Journal of Petroleum Technology,1967, 19, 829-838

[11] Souza J.M and Blanco A. (1995): An empirical procedure to obtain fuzzy rules using neural networks" July 1995

[12] Vogel, J.V. (1968): Inflow Performance Relationships for Solution-Gas Drive Wells. JPTQw, 83- 87; Trans., AIME, 1968, 243.

[13] Aziz, G. W. Govier, K. and Fogarasi, M. (1972): Pressure Drop in Wells Producing Oil and Gas" Society of Petroleum Engineers, 11(3), 1972.

[14] Gomez, L.E. Shoham, O. Schmidt, R.N. Chokshi Z. and Northug, T. (2000): Unified Mechanistic Model for Steady State Two-Phase Flow: Horizontal to Vertical Upward Flow. SPE J., 2000, 5, 339-350.

[15] Ayoub, M.A. (2004): Development and Testing of an Artificial Neural Network Model for Predicting Bottom-hole Pressure in Vertical Multiphase Flow. M.Sc. Thesis, King Fahd University of Petroleum and Minerals, Dahran, Saudi Arabia, 2004.

[16] Musaab M.A. and Ayoub, M.A. (1979): A Comprehensive Study on the Current Pressure Drop Calculation in Multiphase Vertical Wells; Current Trends and Future Prospective. Journal of Applied Sciences, 2014, 14(23), 3162-3171.

[17] Economides, M.J. (1979): Shut-in and Flowing Bottom-hole Pressure Calculation for Geothermal Steam Wells, Proc. Fifth Workshop Geoth. Res. Eng., 1979, 139-152.

[18] Maimon k. and Ranka S.(1998): A decision tree classifier for large datasets, conference on knowledge discovery and data mining" August 1998

[19] Medhat Awadalla, Hassan Yousef, Ahmed Al-Hinai1, and Ali Al-Shidani (2016): Prediction of Oil Well Flowing Bottom-hole Pressure in Petroleum Fields", International Conference on Industrial Engineering and Operations Management, Kuala Lumpur, Malaysia, March 8-10, 2016

[20] Ros, N.C.J. (1961). Simultaneous Flow of Gas and Liquid as Encountered in Well Tubing. Transaction AIME, 1961, 222, 1037.

[21] Adekomaya, A.S. Fadairo, O. and Falode, O. (2008): Predictive Tool for Bottom-Hole Pressure in Multiphase Flowing Wells" Petroleum & Coal, 50 (3), pp. 67-73,

[22] Adekomaya, F.A.S. Adesina O. and Olugbenga, F. (2008): Predictive Tool for Bottom-Hole Pressure in Multiphase Flowing Wells. Petroleum & Coal, 2008, 50(3), 67-73.

[23] Adewale, O. (2010): Optimization of natural gas field development using artificial neural networks" Msc Thesis, The Pennsylvania State University, USA, 2010.

[24] Omotara O. Oluwagbenga and Abdulwahab GIWA, (2020): Determination of Flowing Bottom-Hole Pressure of a Well Using Modified Guo's Model. Journal of Environmental Science Computer Science and Engineering & Technology · April 2020 DOI: 10.24214/jecet.C.9.2.27086

[25] Qiang-han Feng, Shi-chen Gao , Jia-chao Zhang1 and Xiao-chuang Ye (2016): Well bottom-hole flowing pressure evaluation method in 48 Block of S Gas field" International Conference on Energy and Environmental Protection (ICEEP 2016)

[26] Bikbulatov, M. Khasanov, A. Zagurenko, S. (2005): Flowing Bottom-hole Pressure calculation for a pumped well under Multiphase Flow. (Unpublished), 2005.

[27] Zhang Xuanqi, Chen Wei, Li Jianpeng, (2014): density of iteration method to calculate bottom-

hole flowing pressure. (Journal of Xian University (Natural science Edition), 2014

[28] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297

[29] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.