SPE-198762-MS

Bottom-Hole Pressure Estimation from Wellhead Data Using Artificial Neural Network

Oluwatoyin Akinsete and Blessing Adetoye Adesiji, University of Ibadan

This paper was prepared for presentation at the Nigeria Annual International Conference and Exhibition held in Lagos, Nigeria, 5–7 August 2019.

Abstract

Accurate pressure losses prediction for flow in tubing installations is of great importance in the petroleum industry. Historically, the Bottom-Hole Pressure (BHP) determination was obtained using down-hole pressure gauges, because of the economic disadvantage and redundancy, this procedure seems to be less effective, this led to the adoption of the BHP prediction process in estimation. The wide acceptance of data-driven analytics makes the estimation procedure a valid approach in the industry today. Recently, the Artificial Neural Network (ANN), a technique which has been widely accepted because the model proved to predict better than the conventional correlations.

This present work aims to develop a prediction model for BHP based on input and output data obtained from the Volve production field in Norway. Machine learning algorithm based on ANN was used to predict and further improve the accuracy of the prediction while considering a large production dataset from different wells of the field. In developing the model, the initial dataset was processed to about 2,500 data points; the model was trained, tested and cross-validated based on the parameters from the data. Results affirmed that ANN has the ability to handle large dataset, results also revealed that ANN outperformed other models, with a Coefficient of Determination of 0.99997, Root Mean Squared Error of 0.07405 and Mean Absolute Error of 0.02657, which shows high predictability of the model. These results indicated that the ANN model gives a better prediction of BHP when compared to other mechanistic models.

Finally, this work supports the claim, that Production engineers can accurately predict the pressure at the sand-face of a producing well without the use of expensive BHP gauge.

Introduction

Pressure is the continuous physical force exerted in a perpendicular direction to the surface of an object per unit area. There are various forms of pressure in solid, gas and liquid, fluid pressure was considered in this present work. The pressure exerted by a static fluid depends upon the depth, density, and the acceleration of gravity, this also applies to the pressure exerted by flowing fluid down-hole. The latter is what this present work will be focused on, determining flowing bottom-hole pressure.

The petroleum reservoir is made up of layers with special interlayer properties which contain any of the three fluid phases – water, oil, and gas. The flow of these fluids in the phases they may occur is governed by the individual layers which are separated by an interface that could either be permeable or impermeable. The concurrent flow of water, oil, and gas in the vertical pipe setting is of great importance in the Petroleum Industry. The multiphase flow in Petroleum installations have been studied for various years and a large number of research materials have been published. The Industry is interested in determining accurate pressure losses that occur for multiphase flow in tubing installations. For production optimization, it is important that the bottom-hole pressure (BHP) is accurately determined. Historically, the BHP data have been obtained by the use of down-hole gauges for direct measurements. This has resulted in a lot of redundancy because the down-hole equipment is exposed to mechanical failure.

However, it is not economical and practical to deploy down-hole equipment (pressure gauge) to obtain BHP data directly. Engineers have developed ingenious models to determine BHP directly from surface readings using multiphase correlations or mechanistic methods. Technological advancements have contributed to the accurate predictions of BHP data because machine learning algorithms have been developed to better determine BHP data from surface measurements. Traditionally, the flow pressure gradient for a single-phase flow is easily obtainable using Colebrook equations (Colebrook, 1939) and the Moody chart (Moody, 1944). The flow prediction complexity increases when there is more than one fluid flow in the medium resulting in the multiphase flow. In pipe configuration, there can be various types of flow phases which are known as flow regime or flow pattern, this depends on the physical forces and interaction acting on different phases (Li et al., 2014).

Multi-phase correlations analysis is the key to determining bottom-hole pressure from wellhead data because the pressure gradient of the multiphase flow pattern is obtained for a particular length of tubing. However, the prediction is not a single estimation of pressure gradient; the flow pattern has to be actively considered. There are several multiphase correlations, mechanistic models and machine learning algorithms that have been used to predict bottom-hole pressure in multiphase flow. The most common models are the method of Poettman and Carpenter (1952), Baxendell (1961), Fancher and Brown (1963), Hagedorn and Brown (1965), Duns and Ros (1963), Orkiszewski (1967), Beggs and Brill (1973).

Technological advancements have contributed to the complexity of multiphase correlations and there is a need for improved prediction accuracy of a pressure gradient model to determine bottom-hole pressure. With advances in Drilling and Completions operations, complex completion design and various wellbore trajectories which result in different pipe configuration and changing inclination, the multiphase correlation is affected. A number of machine learning algorithms are available to determine the BHP and is mostly Artificial Neural Networks algorithms (ANN). Previous ANN models were done by; Ternyik et al (1995), Osman et al. (2005) and Al-Shammari (2011). These models were arrived at by feeding the algorithm with several input variables of pressure losses and the pressure gradient is directly set as the target variable of the model which is combined with the desired length to obtain the BHP from the wellhead. Also, the above ANN models outperform the multiphase correlations and mechanistic modeling, thus generating high prediction accuracy.

The focal point of this study is to determine bottom-hole pressure using Artificial Neural Network while considering data set from Volve field, on the Norwegian continental shelf (NCS) by Equinor(formerly StatOil), Norway.

Recent advancements in pressure prediction methods revealed that most of the mechanistic models failed to produce the accuracy desired, and major adjustments are still needed. Machine learning models have been of great help in this area of research as recent developers can boast of the desired accuracy in their work. The most common prediction model used to solve difficult engineering problems is Artificial Neural Networks as they have been widely accepted in the Petroleum Industry.

The success of intelligent models in the industry can be traced back to twenty years ago, and numerous research work has been published for the improvement of past pressure predictions. When compared to the correlations and mechanistic models, intelligent models have shown a quality application in multiphase flow area, researchers have applied these techniques to solve problems arising from flow pattern predictions, liquid hold-up correlation, and gas & liquid superficial velocities determination.

Artificial Neural Network Techniques

The roots of all research on artificial neural networks are in neurobiological examinations that go back to about a hundred years ago. Until the mid-twentieth century, application of neural networks especially in areas like; Behavior of nerves in electrical stimulation, communication of nerve cells in impulse stimulation etc., seems ambiguous. System-builders were worried about inquiries regarding whether a neuron model is adequately broad to empower adapting a wide range of capacities, while being anything but difficult to execute, without requiring inordinate calculation inside every neuron. Most neural system learning standards have their foundations in measurable connection examination and in gradient descent methods. The way to make use of an artificial neural network technique is through a training procedure.

In this present study, the input layer of the neural network is: wellhead pressure, pipe length, chokes size, flow rates, Annulus pressure etc. There are corresponding weights that multiply each neuron and pass the result through an activation function, unto the next layer. The estimation of bottom-hole pressure is obtained from the output layer which can be compared to the actual bottom-hole pressure, using statistical metrics. The advantage is that the weights can be adjusted until the estimated value is very close or equal to the actual value. Then the network can now learn from this process for further predictions. (Mehrotra et al, 1997)

Application in the Petroleum Industry

Large datasets are acquired in each area of Petroleum Industry which makes the applicability of Artificial Neural Networks techniques worthwhile. In multiphase flow prediction, researchers have successfully used the technique in prediction as explained below.

Osman et al (2005) presented an Artificial Neural Network (ANN) for predicting flowing bottom-hole pressure and consequently the pressure drop in multiphase flow. The model considered 206 field data sets after data pre-processing which reduced 386 data sets from Middle East fields to what was used in the model development. The datasets were divided into training, validation and testing sets. To develop the model and tune the network weights, the training sets were used; the validation set is used to monitor the generalization of the developed network, and to examine the final performance of the developed model. Osman et al's model achieved a correlation coefficient of 0.9735, the maximum absolute relative error of 7.1401% and average absolute percent error of 2.1654%. The research demonstrated the use of an Artificial Neural Network model in providing solutions to multiphase flow problems. (Osman et al, 2005) Jahanandish and Salimifard (2011) also presented an Artificial Neural Network model for predicting bottom-hole flowing pressure and as a result the pressure drop in vertical multiphase flowing wells. After considering a wide range of data sets, 413 field data sets collected from the research work, the model outperformed the conventional models with correlation coefficient 0.9222 and 3.5% absolute average percent error. The above statistical results were derived from implementing a feed-forward neural networks hidden layer in the predictive model. (Jahanandish and Salimifard, 2011)

Li et al (2014) presented a combined approach involving a calculation procedure using multiphase correlation and Artificial Neural Network models. Back Propagation neural network models were incorporated into a piece-wise calculation approach of determining pressure gradient; this resulted in higher prediction accuracy and broadens the prediction range (Li et al., 2014). The model gave the lowest average absolute percent error to be 3.1%. Comparing to the multiphase correlations, the final combined approach has an average absolute percent error of 23.0%. The model was however published with a user interface bottom-hole pressure calculator deployed.

Model Development

ANN Theoretical Framework

First, it should be understood that this algorithm works in a way, similar to the human brain. The human brain takes input, which goes through some transient states in the neurons between the sensory organs and the brain, to generate output. The ANN algorithm works in a similar manner, but sophisticated than the input-output flow of a machine. The architecture of the algorithm stemmed from the neural network residing in the human brains. It operates on the transient state called "Hidden states/layers", which are similar to neurons. With a probabilistic behavior, each grid of such state acts as a bridge between the input and the output. Finally, the intermediate latent states allow the algorithm to learn from every prediction.

Data Acquisition and Analysis

The data used in this research was obtained from the disclosed subsurface production data from a field on the Norwegian continental shelf (NCS) by Equinor, Norway. The operator released the datasets of the field known as "Volve field", a decommissioned field in September 2016 after 8.5 years in operation, it is found 200 kilometers west of Stavanger at the Southern end of the Norwegian division. Initially, the original datasets from Equinor contained 15,635 data points from 7 Wells. Furthermore, the datasets were processed to have 2,447 data points. They were divided into two parts, one part was used for training the networks with 1713 data points, and the other 734 data points were used for testing the trained network.

Feature Selection and Engineering

Using the parameters obtained from the datasets, the features were engineered and selected based on past research work. Li et al (2014) advised that to reduce errors, less external features should be used in bottom-hole pressure prediction. This can produce some errors compared to the actual field properties. In data science, this is known as domain expertise selection. The features used in the algorithm training are: Well Depth, Average Downhole Pressure, Average Downhole Temperature, Average Tubing pressure differential, Average Annulus Pressure, Average Choke Size, Average Well-Head Pressure, Average Well-Head Temperature, Pressure Differential in Chokes, and Bore Oil Volume. The features were used as obtained from the Norwegian field; the pressure readings are measured in Bars.

```
Data columns (total 10 columns):
Well Depth                   2447
AVG_DOWNHOLE_PRESSURE        2447
AVG_DOWNHOLE_TEMPERATURE     2447
AVG_DP_TUBING                2447
AVG_ANNULUS_PRESS            2447
AVG_CHOKE_SIZE_P             2447
AVG_WHP_P                    2447
AVG_WHT_P                    2447
DP_CHOKE_SIZE                2447
BORE_OIL_VOL                 2447
```

Figure 1—List of selected features and data-points

Data Description

From the datasets, a descriptive statistical summary was generated, which include the central tendency, dispersion, percentiles, and standard deviation, as shown in Table 1.

Table 1—Statistical Summary of the Training and Testing Dataset

	count	mean	std	min	25%	50%	75%	max
Well Depth	2217.0	3649.655841	1278.574498	347.00000	3750.00000	3750.00000	4685.000	4685.000
AVG_DOWNHOLE_PRESSURE	2217.0	239.735789	26.373755	193.18700	206.92300	249.97200	263.124	285.248
AVG_DOWNHOLE_TEMPERATURE	2217.0	102.699840	3.139834	94.19200	99.66500	100.82600	106.379	107.044
AVG_DP_TUBING	2217.0	207.334180	29.012916	153.68300	181.55200	218.24300	231.277	243.705
AVG_ANNULUS_PRESS	2217.0	18.493564	3.851948	7.08400	14.67400	20.04800	21.879	27.034
AVG_CHOKE_SIZE_P	2217.0	65.408385	40.463483	0.75599	12.78188	97.91667	100.000	100.000
AVG_WHP_P	2217.0	32.401619	5.549192	15.70200	29.52400	31.47200	33.460	48.553
AVG_WHT_P	2217.0	69.453203	24.652247	17.26500	42.07600	35.48500	87.948	89.578
DP_CHOKE_SIZE	2217.0	5.049078	4.323476	0.58400	2.19500	3.09900	5.691	18.108
BORE_OIL_VOL	2217.0	361.508345	254.943427	4.00000	178.00000	249.00000	500.000	999.000

When dealing with the data, it is imperative that we understand how the variables are distributed. As shown in the figures below, the univariate distribution of each feature is plotted as a histogram and fitted with a kernel density estimate (KDE). The KDE is a non-parametric way to estimate the probability density function of the variables, most of the distributions are not normal, as some are bimodal and others skewed to the left. Hence, there is a need for the data to be normalized using data transformation, to enable the model to perform better. To understand the relationship between the variables, with a great observation on the target variable, it is important that the correlation heatmap be visualized. In this research, it helped to understand the relationship of all features with Average down-hole pressure, as there are negative and positive correlations with the target variable.

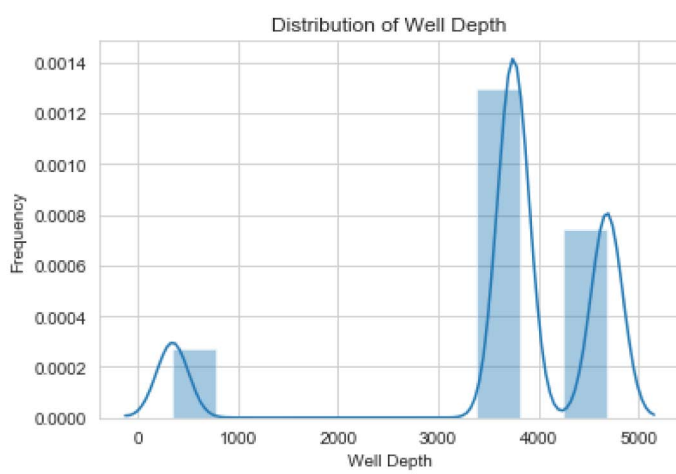

Figure 2a—Distribution Plot of Well Depth

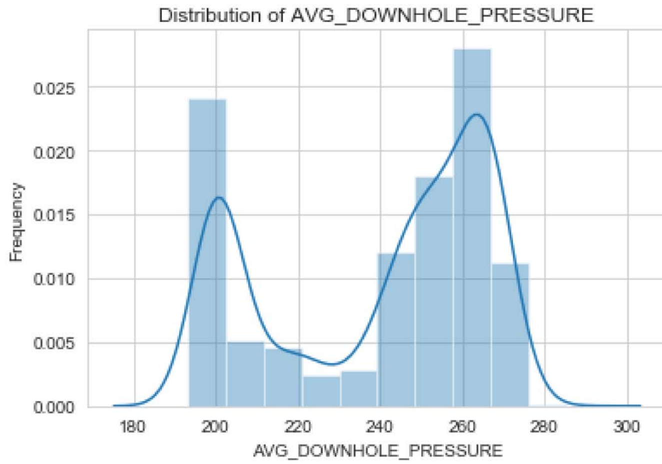

Figure 2b—Distribution Plot of Average Downhole Pressure

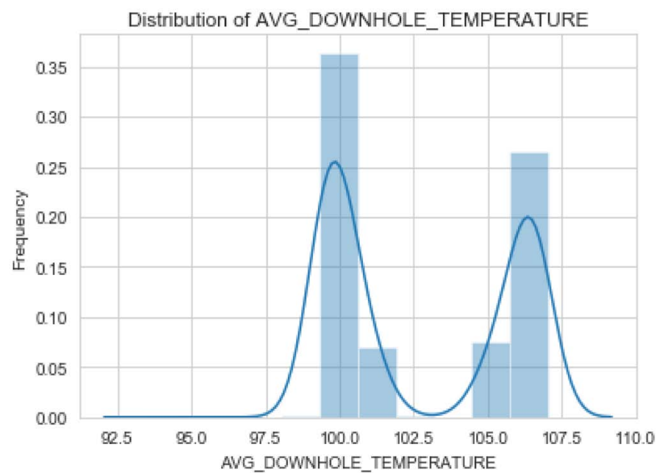

Figure 2c—Distribution Plot of Average Downhole Temperature

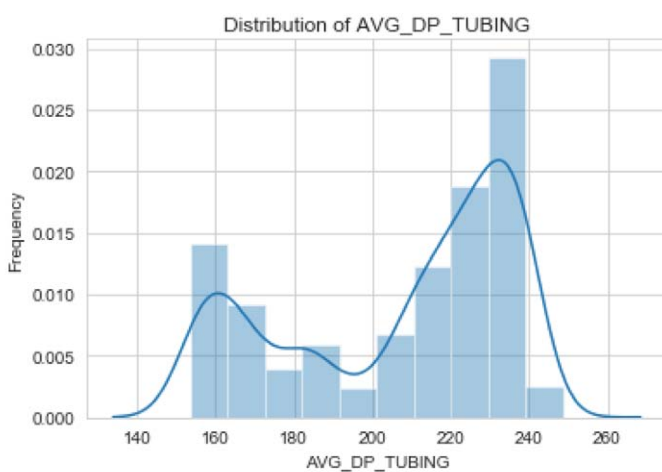

Figure 2d—Distribution Plot of Average DP Tubing

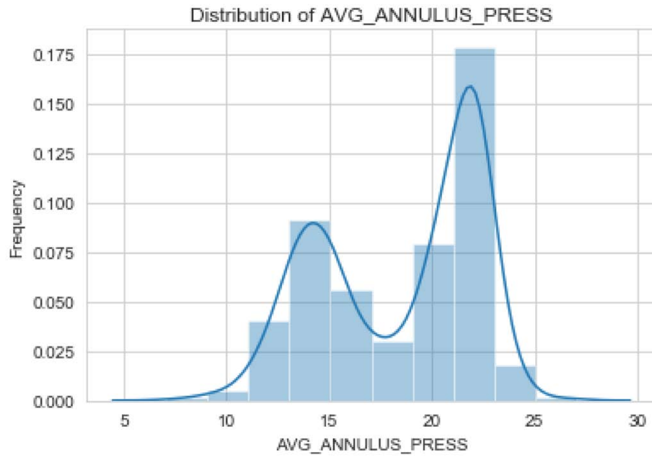

Figure 2e—Distribution of Average Annulus Pressure

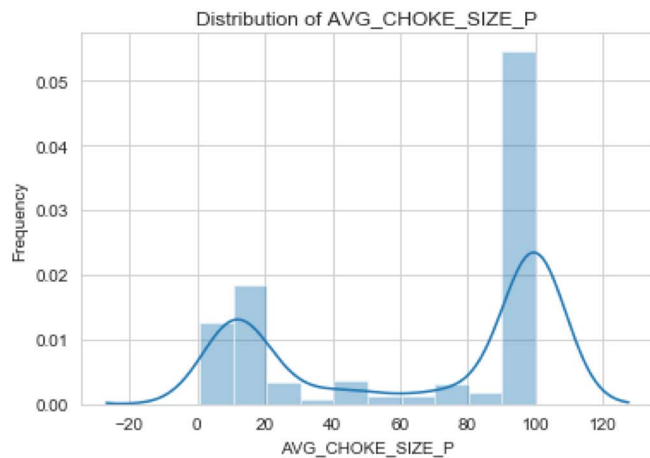

Figure 2f—Distribution of Average Choke Size

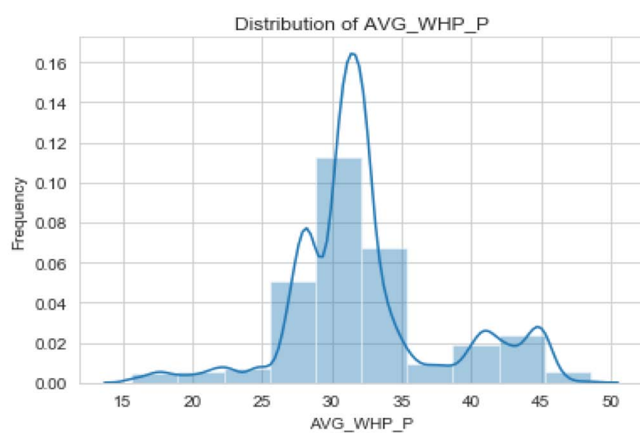

Figure 2g—Distribution Plot of Average Well-Head Pressure

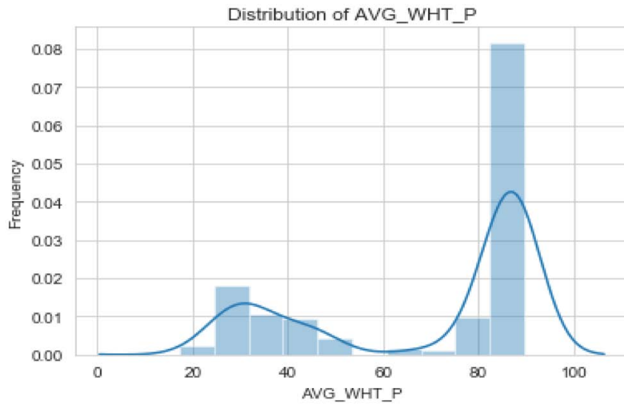

Figure 2h—Distribution Plot of Average Well-Head Temperature

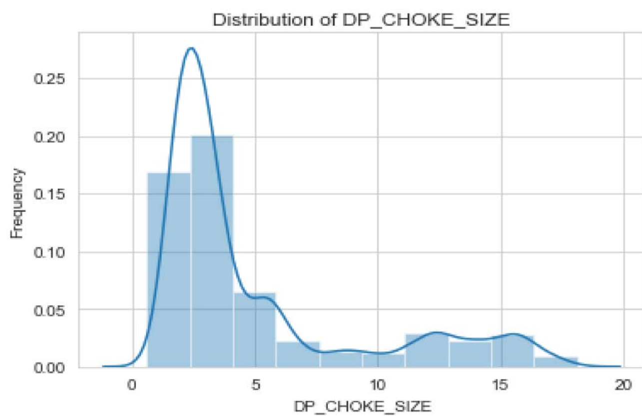

Figure 2i—Distribution Plot of DP Choke

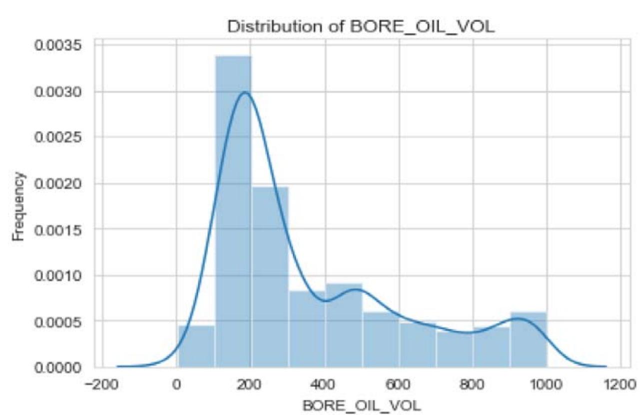

Figure 2j—Distribution Plot of Bore Oil Volume

Model Framework

The Multilayer Perceptron Regressor is a supervised learning algorithm that learns a function $f(.)$: $R^m \rightarrow R^{\circ}$ by training on the dataset, where m is the number of input dimensions and o is the number of output dimensions. For the features of the dataset (X = $X_{1,}$ $X_{2,}$ …, X_{10} and target variable y, which is a non-linear function approximator for the regression problem. (Scikit-Learn Documentation)

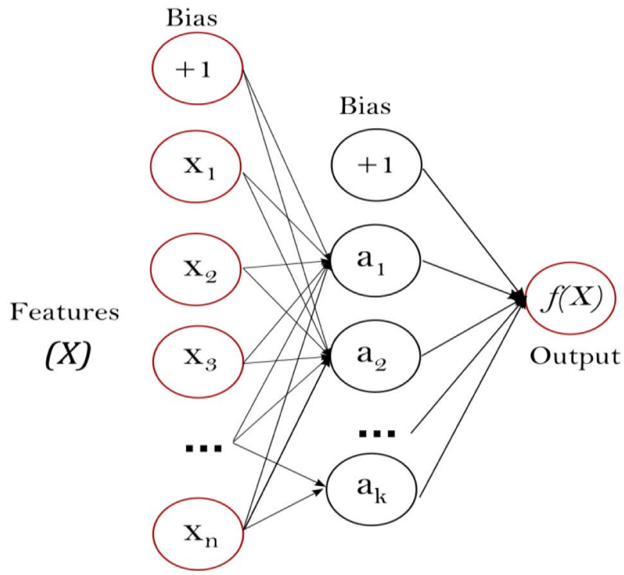

Figure 3—One Hidden Layer MLP

The input layer(X) is to the left side of Figure 4 above, it consists of a set of neurons $X_i \parallel \{X_1, X_2, \ldots, X_{10}\}$ representing the input features. The values from the previous layer is transformed by each neuron in the hidden layer with the weighted linear summation followed by a non-linear activation function

$$W_1X_1 + W_2X_2 + \ldots + W_{10}X_{10} \quad \text{(Eq. 1)}$$

$$g(.): R^m \rightarrow R^o \quad \text{(Eq. 2)}$$

Where:

$W_1, W_2 \ldots W_{10}$ are the weights of the input features.

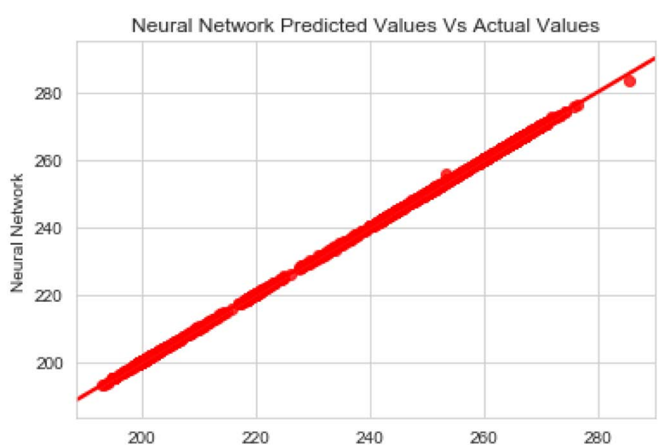

Figure 4a—ANN Predicted BHP Values VS Actual BHP values

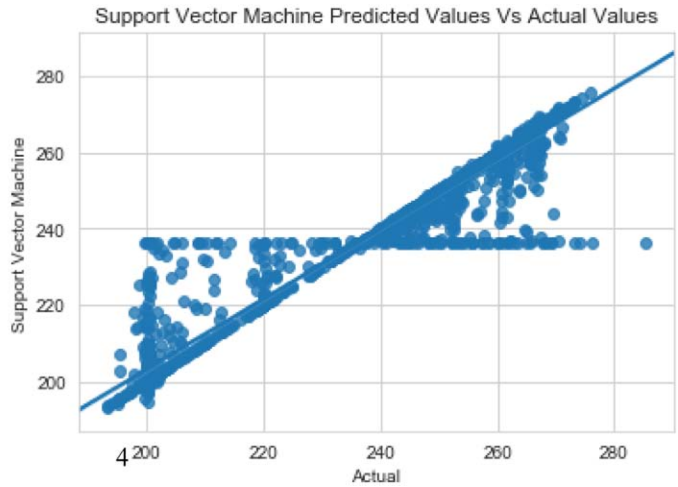

Figure 4b—SVM Predicted BHP values VS Actual BHP Values

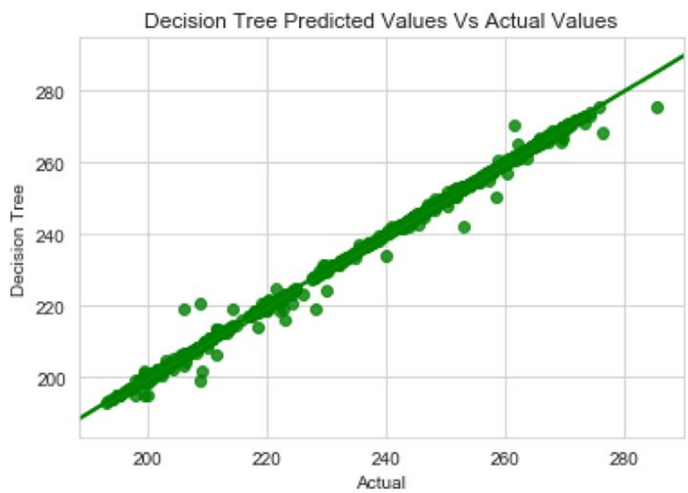

Figure 4c—Decision Trees Predicted BHP Values VS Actual BHP Values

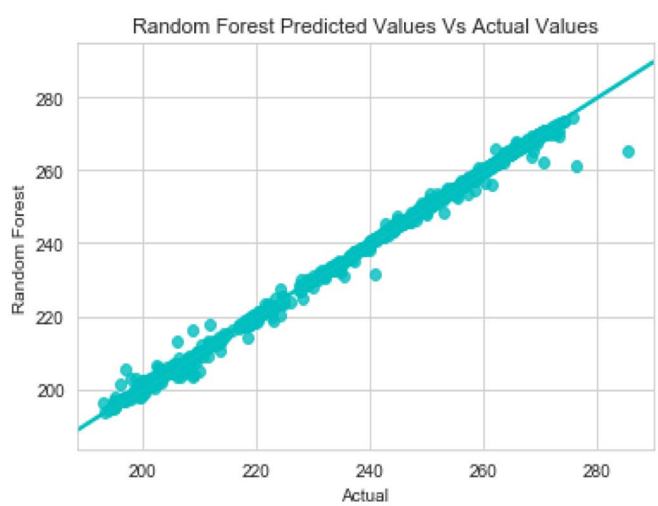

Figure 4d—Random Forest Predicted BHP Values VS Actual BHP Values

Finally, the last hidden layer gives values to the output layer to transform them into output values. The parameters used in the ANN model are explained below.

- Activation function – ‘relu’, the rectified linear unit function, which returns

$$f(x) = \max(0, x) \quad \text{(Eq. 3)}$$

- Alpha – 0.0001-is L2 penalty (regularization term) parameter.
- Hidden layer sizes (100) - it represents the number of neurons in the hidden layer.
- Learning rate – Adaptive-‘adaptive’ keeps the learning rate consistent to the underlying learning rate as long as training loss keeps reducing. Each time two consecutive epochs fail to reduce training loss by at least tolerance the present learning rate is divided by 5.
- Maximum Iteration (10,000) - Maximum number of iterations. The solver iterates until it converges.
- Solver – lbfgs - The solver for weight optimization, ‘lbfgs’(Limited memory Broyden-Fletcher-Goldfarb-Shanno), an optimizer in the family of quasi-Newton methods.

(Scikit-Learn Documentation)

However, for the data pre-processing, exploratory data analysis and model development, Python Programming language was used to achieve the processes above.

Results and Discussion

After completing the model training and testing with a certain portion of the datasets, the obtained predicted results were compared with the actual measured readings from the field. In a bid to validate the predicting power of the ANN model, the results were compared with other intelligent models, namely; Support Vector Machines (SVM), Decision Trees (DT), and Random Forest (RF). Using statistical metrics, each model is compared with each other. The statistical parameters used for comparison are the Coefficient of Determination (COD), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

The visualizations below (Figures 5a-d) are plots of predicted bottom-hole pressure against already measured pressure values (Appendix A). It can be observed that all the data points center on the line of best of fit, this verifies the predictability of each model.

Statistical Metrics

From the above plots (Figure 5a-d) and statistical metrics (Table 2), it can be observed that amongst all intelligent models used, ANN Model performed better with a Coefficient of Determination of 0.9999719, Mean Absolute Error of 0.0265742 and Root Mean Squared Error of 0.0740512. The plot for Artificial Neural Network (Figur 5a) also showed great prediction ability as all the data points lie on the line of best fit.

Table 2—Accuracy of ANN compared with other Intelligent Models

Machine Learning Algorithm	COD	MAE	RMSE
Artificial Neural Networks	0.9999719	0.0265742	0.0740512
Support Vector Machines	0.8021627	7.2280970	11.599087
Decision Trees	0.9966613	0.601207	1.5068196
Random Forest	0.9970530	0.5258371	1.4156647

Conclusion and Recommendation

This paper presented an Artificial Neural Network model for predicting Bottom-Hole Pressures of some production wells in Norway. The model was trained and tested using a certain portion of the obtained datasets and the predicted results were compared with the actual field measurements to validate the prediction ability of the model. Also, to test the performance of the ANN model, it was compared with other common intelligent models, and it was observed that the ANN model performed far better in comparison with other algorithms. Visualizations and statistical metrics were used to explain this comparison. For Industrial application, the model was trained on a general condition to further enable it useful for other production wells, meaning this model can be deployed for industrial usage by Production Engineers. For further research work, Artificial Neural Networks or other machine learning algorithms can be used to estimate other important features before passing the data for bottom-hole pressure estimation. These include flow regime recognition, liquid hold-up factor estimation, and multiphase flow pattern classification.

The adoption of big data analytics is in its infant stage in Oil and Gas Industry, the full-adoption can lead to improved efficiency in operations, lower operational costs and ability to make informed decisions. The major challenge is the availability of data for explorations, visualizations and modeling development. Lastly, since the Industry relies on informed decisions, data should be provided in a structured format for pattern recognition, insights derivation, and prediction.

Nomenclature

ANN	Artificial Neural Network
AVG	Average
BHP	Bottom-Hole Pressure
RMSE	Root mean square Error
MAE	Mean Absolute Square Error
COD	Coefficient of Correlation
RELU	the rectified linear unit function
SVM	Support Vector Machines
KDE	Kernel Density Estimate
NCS	Norwegian continental shelf
WHP	Well-Head Pressure
WHT	Well-Head Temperature
DP	Pressure Differential

References

Al Shammari, A. (2011). "Accurate Prediction of Pressure drop in two-phase Vertical Flow Systems using Artificial Intelligence". Society of Petroleum Engineers.

Baxendall, P.B. (1961), "The Calculation of Pressure Gradients in High rate Flowing wells", *Journal of Petroleum Technology*

Beggs, H. D., and Brill, J. P. (1973). A Study of Two-Phase Flow in Inclined Pipes, *Journal of Petroleum Technology*, Vol. **25**, No. 5, pp. 607-617.

Colebrook, C. F. (1939). Turbulent flow in pipes with particular reference to the transition region between the smooth and rough pipe laws. *Journal of the Institution of Civil Engineers (London)*, **11**(4), 133–156.

Duns, H., And Ros. N. C. J. (1963). Vertical flow of Gas and Liquid Mixtures in wells, Proc. of the Sixth World Congress, Vol.**10**, Section 2, Paper 22.PD6. 451465.

Fancher, G. H., and Brown, K. E. (1963). Prediction of Pressure gradients for Multiphase Flow in Tubing, Society of Petroleum Engineering Journal.

Hagedon, A. R and Brown, K. E. (1965). Experimental Study of Pressure Gradients occurring during continuous two-phase flow in small diameter vertical conduits. *Journal of Petroleum Technology*, p.475-484

Jahanandish, B. and Salimifard, H. J. (2011). Predicting Bottomhole Pressure in vertical multiphase flowing wells using artificial neural networks. *Journal of Petroleum Science and Engineering* Volume **75**, issues 3-4, Pages 336-342.

Li, X., Miskimins, J., and Hoffman, B. T. (2014). A combined bottom-hole pressure calculation procedure using multiphase correlations and artificial neural network models. Society of Petroleum Engineers. DOI: https://doi.org/10.2118/170683-MS.

Mehrotra, Kishan & Mohan, Chilukuri & Ranka Preface, Sanjay. (1997). Elements of Artificial Neural Nets.

Moody, L. F. (1944). "Friction factors for pipe flow." *Trans. ASME*, **66**:671-678.

Mukherjee, H. and Brill, J.P. (1985) Pressure Drop Correlation for Inclined Two-Phase Flow, *Journal of Energy Resources Technology*, Vol. **107**, pp. 549-554.

Orkiszewski, J. (1967). Predicting Two-Phase Pressure Drop in Vertical Pipes", *JPT*, Vol.**19** No 6, p.829-838.

Osman, E. A., Mohammed A. A., and Mohammed A. A. (2005). Artificial Neural Network Model for predicting bottom-hole flowing pressure in vertical multiphase flow, Society of Petroleum Engineers.

Poettmann, F. H. and Carpenter, P. G. (1952). The Multiphase flow of Gas, Oil, and Water through Vertical Flow Strings with Application to the Design of Gas Lift Installations", *Drilling and Production practice*.

Pedregosa et al. Scikit-learn: Machine Learning in Python, *JMLR* **12**, pp. 2825-2830, 2011.

Appendix A

Table 3—Measured and predicted BHP readings of all the intelligent models

Measured BHP	Intelligent Models Computed BHP			
	ANN	SVM	DT	RF
212.332	212.364	216.199	212.741	212.826
202.878	202.897	202.978	202.878	206.557
200.246	200.254	200.346	200.246	200.300
199.805	199.813	203.552	200.246	200.300
196.028	196.061	196.128	196.028	201.993
197.894	197.899	217.977	195.306	197.087
198.465	198.462	213.992	200.246	198.865
202.469	202.468	202.569	202.469	204.637
200.407	200.413	200.507	200.407	200.238
199.984	199.988	200.084	199.984	200.069
199.969	199.871	204.228	198.299	199.969
199.741	199.740	213.571	198.299	199.226
198.299	198.298	198.199	198.299	198.070
197.156	197.147	197.256	197.156	197.270
196.611	196.605	196.711	196.611	196.673
193.302	193.291	193.402	193.302	193.772
194.984	194.975	194.884	194.984	194.741
194.978	194.968	195.078	194.978	194.848
194.317	194.308	194.417	194.317	194.517
195.306	195.292	195.406	195.306	195.520

Appendix B

STATISTICAL PARAMETERS

Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_j - y^*_j\right|$$

Mean Squared Error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{actual} - y_{pred}\right)^2$$

Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{actual} - y_{pred}\right)^2}$$

Coefficient of Determination

$$R^2 = 1 - \frac{\Sigma_i^n\left(y_{actual} - y_{est}\right)^2}{\Sigma_i^n\left(y_{est} - y^*_{actual}\right)}$$

Where y* is the mean value