International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

Drilling Efficiency Improvement and Rate of Penetration Optimization by Machine Learning and Data Analytics

Sridharan Chandrasekaran
Department of Ocean Engineering,
Indian Institute Technology Madras, Chennai-600036, India.
Corresponding author: Oe14d033@smail.iitm.ac.in; sridharavc@gmail.com

G. Suresh Kumar
Department of Ocean Engineering,
Indian Institute Technology Madras, Chennai-600036, India.
E-mail: gskumar@iitm.ac.in

(Received November 3, 2019; Accepted February 20, 2020)

Abstract
Rate of Penetration (ROP) is one of the important factors influencing the drilling efficiency. Since cost recovery is an important bottom line in the drilling industry, optimizing ROP is essential to minimize the drilling operational cost and capital cost. Traditional the empirical models are not adaptive to new lithology changes and hence the predictive accuracy is low and subjective. With advancement in big data technologies, real- time data storage cost is lowered, and the availability of real-time data is enhanced. In this study, it is shown that optimization methods together with data models has immense potential in predicting ROP based on real time measurements on the rig. A machine learning based data model is developed by utilizing the offset vertical wells' real time operational parameters while drilling. Data pre-processing methods and feature engineering methods modify the raw data into a processed data so that the model learns effectively from the inputs. A multi – layer back propagation neural network is developed, cross-validated and compared with field measurements and empirical models.

Keywords- Neural network, Rate of penetration, Optimization, Drilling.

Nomenclature

ROP	Rate of Penetration (ft/hr)	h	Bit wear fraction
A_b	Area of bit (in^2)	q	Flow rate (gal/min)
D	Hole Depth (ft)	DOC	Depth of cut (ft/min)
WOB	Weight on Bit (klbs)	$b_o - b_6$	Regression constants
RPM	Rotation per minute	d	Diameter of bit (in)
SPP	Stand Pipe Pressure (psi)	s	Standard Deviation
g_p	Pore pressure gradient (lb/gal)	n	Number of samples
d_n	Diameter of nozzle (in)	μ	Apparent viscosity (Centipoise)
ρ_c	ECD at the bottom (lb/gal)	N	Number of rotation
$DIFF_{PRESS}$	Differential pressure (psi)	a_{1-8}	Constants of Bourgyne model

1. Introduction

The ever-increasing complexity of the wells escalate drilling cost and the focus of the drilling industry shifted towards reducing the NPT (Non-Productive time) and ILT (Invisible lost time). ILT is caused by under optimized drilling, poor rate of penetration and this inefficiency in the drilling operation is estimated to cost 15% of the total well cost (Zakariya et al., 2015). ROP is the related to the speed with which drilling takes place and maximizing it under rig and

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

operational environment has always been an objective for the drillers. In the gloomy oil market, it has become more imperative than ever to maximize the rate of penetration and reach the drilling target in the development wells at a faster pace.

Rate of penetration (ROP) prediction with mathematical models and optimization of the drilling variables based on these models has been active area of research. ROP depends on many drilling variables and its complex relationships, like the operational parameters, formation properties, compressive strength, well hydraulics, borehole shape and size, mud properties, type of bit, hole-cleaning etc. ROP is the key business performance indicator in the drilling industry and a measure of additional contractual incentives. Hence, accurately optimizing it in real time with mathematical model has a direct business benefit for the industry.

For rotary drilling process, the ROP model proposed (Bourgoyne and Young, 1974) is considered as a comprehensive model for the drilling optimization as it was based on physical behaviour and validated with field results.

$$ROP = \frac{dD}{dt} = \exp(x_{1+}x_2a_2 + x_3a_3 + x_4a_4 + x_5a_5 + x_6a_6 + x_7a_7 + x_8a_8) \quad (1)$$

where, x_{1-8} represents the effects on ROP and are evaluated by Table 1.

Table 1. Bourgoyne and Young Model

Equation	Effect
$x_{1=}a_1$	Formation Strength
$x_2 = (10000 - D)$	Formation Depth
$x_3 = D^{0.69}(g_p - 9)$	Formation Compaction / Pore pressure
$x_4 = D(g_p - \rho_c)$	Differential pressure
$x_5 = \ln(\frac{\frac{WOB}{d} - \left(\frac{WOB}{d}\right)_t}{4 - \left(\frac{WOB}{d}\right)_t})$	Bit Diameter and Bit weight
$x_6 = ln\,\frac{N}{60}$	Rotary Speed
$x_7 = -h$	Bit Wear
$x_8 = \frac{(\rho q)}{350\mu d_n}$	Bit Hydraulics

Warren and Armagost (1988) developed an imperfect cleaning model to predict the ROP in soft formation based on experimental data, dimensional analysis and generalized response curves. A case specific model developed for natural diamond bits (Hareland and Rampersad, 1994) is based on the assumption that when weight is applied on the drilling bit, it will penetrate a certain depth. The length of penetration depends on the cutter size, shape, rock strength and scraping action. This kind of bit model which was changed (Motahhari et al., 2010) to suit the PDC (polycrystalline diamond compact) bit and to include motor effects.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

The traditional empirical models discussed above involves a number of constants which needs to be evaluated based on field results or experimental methods This results in time consumption and cannot be generalized for the entire well resulting in inaccurate prediction (Eskandarian et al., 2017). The abundance of real time drilling data together with increased big data infrastructure and computing resulted in the development of machine learning and non-linear statistical models for predicting ROP. Neural network technique involving two parameters namely WOB and RPM (Dunlop et al., 2011) was used to optimize ROP. The data models suffer the drawback of understanding the physics behind the problem as they are mostly used as black box models which was overcome by enabling coupling conditions between neural network and regression analysis (Mantha and Samuel, 2016). Weak machine learning predictors are coupled together (ensemble) so that an integrated model produce accurate predictions than individual predictors. The data modelling by such ensemble methods of machine learning (Hegde and Gray, 2017) can infer parameters affecting ROP optimization rather than predicting them.

In the current work, a new model was developed to predict the ROP of adjacent wells by performing data analytics and developing an artificial neural network model from the offset well real time drilling data. In order to reduce the amount of uncertainties in the drilling scenario, only vertical wells drilled with similar BHAs from the same field is considered for the development of the model. In order to prove the effectiveness of the model from traditional approaches, the data model is compared with traditional models in the field namely the linear regression models and Bourgoyne and Young model. The workflow followed for the development of the model involves two phases

a) Data exploration
b) Model development

The data exploration is the process of transforming and converting the data into a form which could be easily handled by the mathematical algorithms for developing an accurate model. The second process is a computationally intensive process of choosing the best mathematical representations for the features in the dataset. In our work, we made use of the libraries in scikit learn – python for the model development process. The overall objective was to use every available information in real-time to perform ROP optimization overcoming static pre-drill configuration and allowing the use of data driven analytics in the drilling industry.

2. Data Exploration

The process of data exploration involves data selection, data quality analysis, exploratory data analysis and feature engineering.

2.1 Data Selection

In this study, data from vertical wells drilled in the same onsite field in Asia is used. The real time data is collected from drilling data acquisition units on the surface of the rig. It includes weight on bit (WOB), rotation (RPM), torque, stand pipe pressure (SPP), differential pressure, hookload measurement, hook height measurement, bit depth, flow measurement. These datasets are depth indexed sampled at 0.5 ft interval and the three wells used in this study spans a total footage of 6000 ft. In order to ensure consistency between wells, the wells are so chosen that their static parameters namely the drill string configuration, bottom- hole assemblies, mud properties, the nature of drill bits, blade counts, drill – bit cutter configurations are the same. These dataset span five different types of formation (named 1-5) (rock types) and these formations were drilled with 12.25-inch diameter bits. As a representation, the instantaneous ROP measurement from Well A

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

on the five continuous formation that it is shown in Figure 1. In general, ROP is a function of three overall characteristics.

ROP = function (*Operational Parameters, Environmental Parameters, Structural parameters*)

Operational Parameters involves the drilling parameters, environmental parameters involve formation parameters and structural parameters include the BHA and drill bit configuration. In this work, the wells are so chosen that the BHA configuration, drilling bit were the same. It ensures that the effect of one variable (structural parameter) is minimized. Moreover, the BHA configuration will be constant in a section and may not induce any variance which is essential for building an artificial neural network model section-wise.

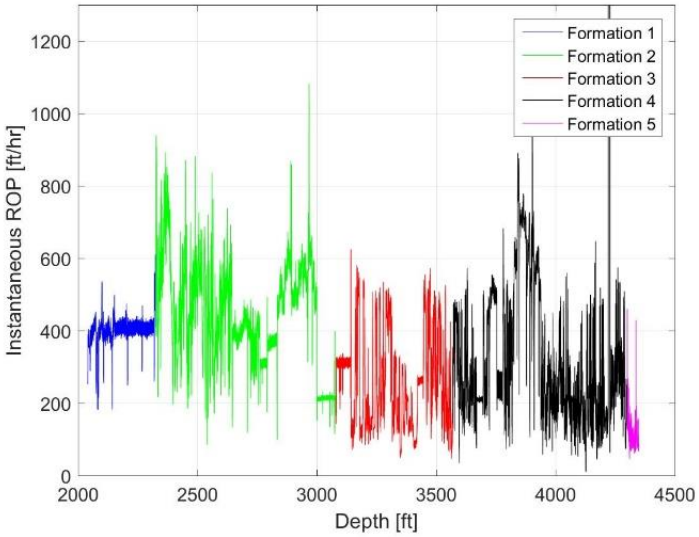

Figure 1. Well A – Instantaneous ROP measurement

2.2 Data Quality Analysis

Abnormal data can influence the model negatively and may restrict the model in its generalization. Suspect data is flagged in the range validation procedure which checks if the measurement can be within the limits of rig limitations and acceptable trends. For example, WOB can never be zero or negative during drilling in non- mud motor runs. The dataset is tested for anomaly by Z-score outlier detection algorithm where a threshold Z-Score of 3 is selected (Tripathy et al., 2013) and any data points above this threshold is marked as an outlier and excluded from the training data. The dataset is further processed and smoothened by a low pass second order Butterworth filter (Selesnick and Burrus, 1998) on high variance curves namely RPM curves and the flow curves as shown in Figure 2. These curves showed high magnitude in the statistic attribute (standard deviation / mean) as shown in Table 2.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

Table 2. High variance curves

Parameters	Standard deviation /mean
Torque	0.16
Stand pipe pressure (SPP)	1.79
RPM	4.62
Flow	6.48
Weight on Bit (WOB)	0.35
Differential Pressure	0.82
Hookload (HKLI)	0.06

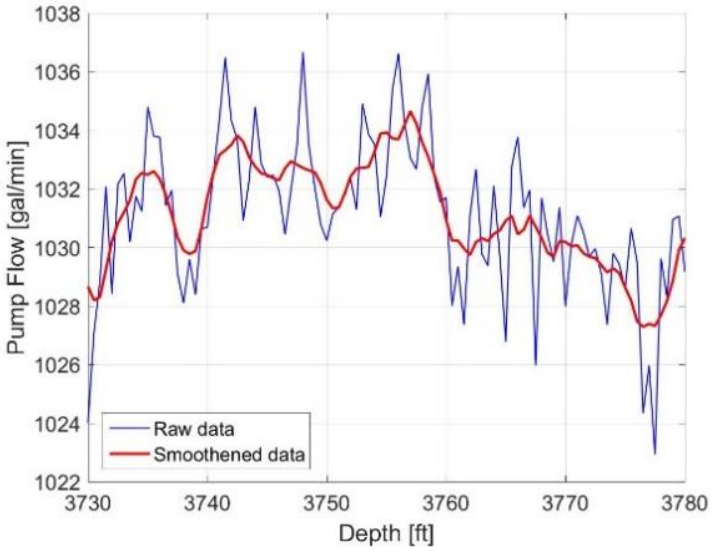

Figure 2. Smoothening of variables

2.3 Exploratory Data Analysis (EDA)

EDA is the aspect in data analysis that employs a variety of visualization and statistical techniques to obtain maximum information from the data and to uncover complex patterns in the data. In the well under consideration, the dataset contains five lithology types of varying hardness. It is observed from Figure 1 that the number of points for formation 5 is significantly lower than other formation types and in order to have a balanced training, formation type 5 was not taken into consideration during model building. The inter-relationships between the variables can be studied using cross plots as shown in Figure 3 where the regression coefficients between the two drilling variables are represented. A regression coefficient closer to 1 represents positive correlation and closer to -1 represents negative correlation between the drilling variables. It should be noted that when absolute value of the regression coefficient is closer to zero, it means that the variables are not linearly correlated and there are chances of higher order correlations or no correlation. Hence, these variables are retained in the subsequent model development. It is observed that DOC and ROP are perfectly correlated which is an expected behaviour. Similarly, the SPP is also correlated by 0.89 with the flow pumps which indicate there was no influx or abnormal pressure exhibited by the formation.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

2.4 Feature Engineering

Feature engineering is the process to generate better enhancements in the training dataset to uncover the underlying patterns in the data and its inter-relationship so that maximum accuracy is achieved from the machine learning model. In the current modelling scheme, feature augmentation is adopted where adjacent raw measurements are appended to the current measurement. For example, new features of flow measurement at the current depth is appended with the flow measurements at 0.5 ft, 1 ft, 1.5 ft and 2 ft less than the current depth. Similarly, other surface parameters (WOB, TORQUE and RPM) are also augmented to generate a feature space of 36 dimensions. Such a scheme is possible because considering the uncertainty scenario in the drilling process, the operating parameters at a measured depth will not be at exact depth but within a tolerance band. Moreover, the data acquisition system applies transformations functions like decimation or filtering on the raw data feed which can potentially induce numerical scaling, thereby validating the augmentation process.

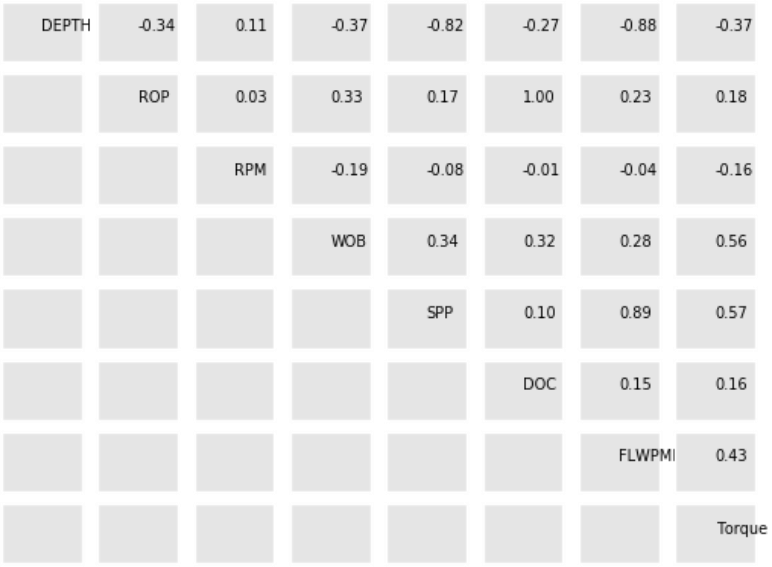

Figure 3. Pairs plot between the drilling variables for Well A

3. Model Development

3.1 Data Model

This section describes the implementation of artificial neural network type of machine learning for ROP modelling. Artificial neural network (ANN) are analogous to biological neural system consisting of dendrites that carry external impulses to the cell. The cell body processes the information and provides an output to the axon which is interconnected to other neurons through synapses. Figure 4 show the mechanism of artificial neuron where input channel is analogous to the dendrites, cell body to the activation function and axons to the output channel.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

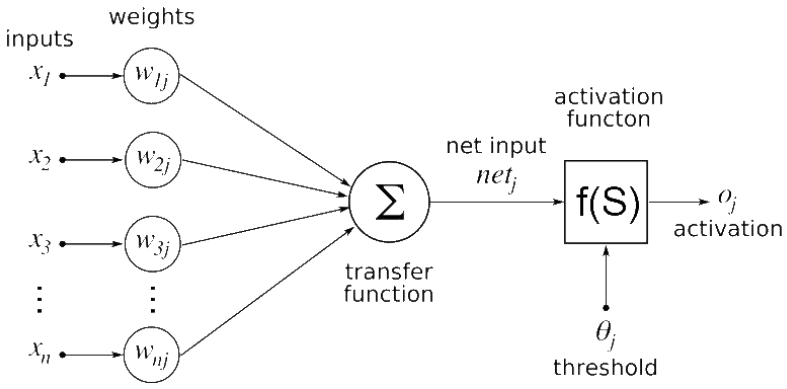

Figure 4. Mechanism of artificial neuron depicting biological neuron

In the modelling scheme, all the drilling parameters mentioned in Table 2 serve as input parameters to the neural network together with augmented features and derived features. The output of the network is the instantaneous rate of penetration. Hence, the problem is mathematically formulated as a classic supervised learning regression problem. The model was constructed with a back propagation three layered neural network using scikit learn libraries in Python language (Pedregosa et al., 2011).

The dataset of Well A was used with 80% of the input data taken for network training and 20% for validation. All the input parameters are normalized so that the input feature space and output ROP will fall in the range of zero and unity to avoid numerical scaling issues. The accuracy of the data model is demonstrated by root mean square which serves as a metric to score the predictions of the ANN model with the expected result. It is given by

$$\text{RMS Error} = \frac{\sqrt{\Sigma y_{predicted}^2 - y_{actual}^2}}{n} \quad (2)$$

The challenges in the design of neural network are to perfectly tune the hyper parameters and to make proper choice of activation function so that the network generalizes well and do not perform over fitting. The available activation functions with scikit-learn in python module is shown in Table 3 which can possibly cause the model to over fit or under fit.

Table 3. Activation function

Activation Function	Equation
No-Op Activation	$y = x$
Sigmoid	$y = \frac{1}{1+e^{-x}}$
Tanh	$y = tanh(x)$
Rectified Linear Unit	$y = \max(0, x)$

A sensitivity analysis on the activation yielded sigmoid logistic function to have the lowest training error as shown in Figure 5. After a number of simulations, the BP network utilized 50

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

neurons in the hidden layer to provide optimum accuracy in the training and validation dataset as shown in Figure 6. To avoid overfitting, the general approach is to perform cross validation where the structure is tested on a different data set other than the data set upon which it was trained. In the current study, a K-fold cross validation technique is employed where the data is split in 5 folds. This process is repeated for all the five folds in a completely randomized process until all the data is used for training and validation. This is shown in Figure 7 where the accuracy of the prediction for the different folds is represented and upon iterations, it is observed that the coefficient of determination is consistent across the folds thereby ensuring generalization of the model.

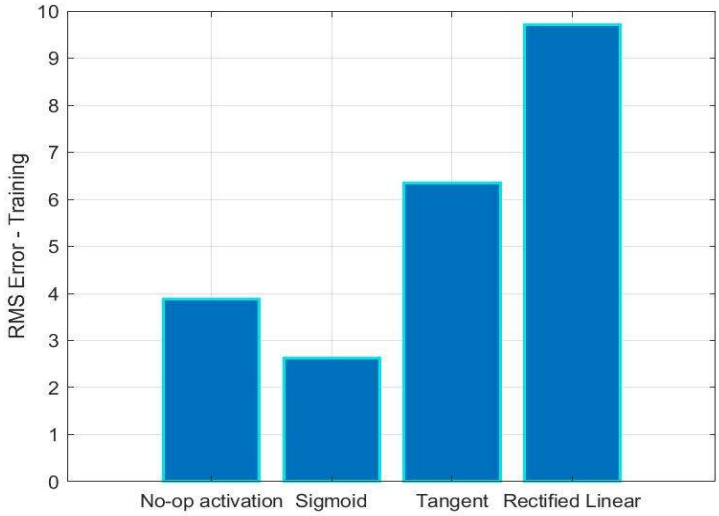

Figure 5. RMS error with activation functions

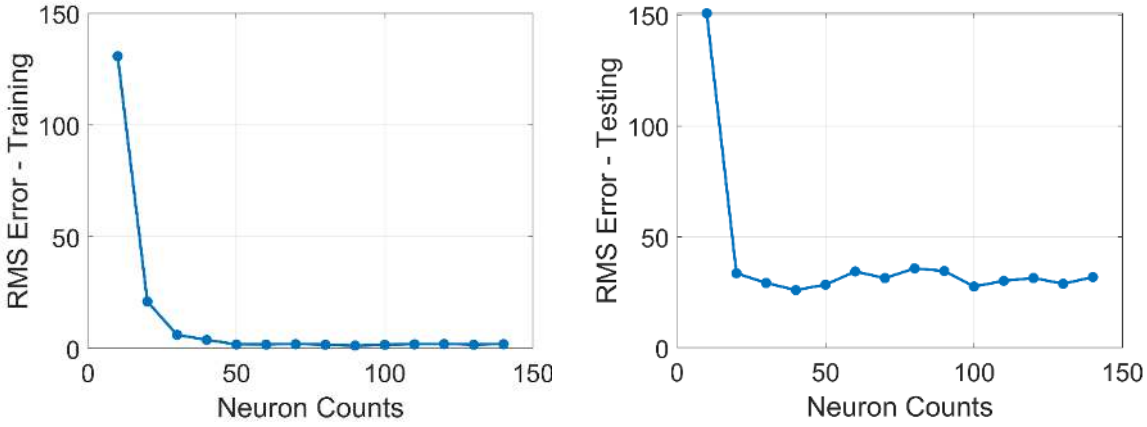

Figure 6. Optimum neurons in the hidden layer

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

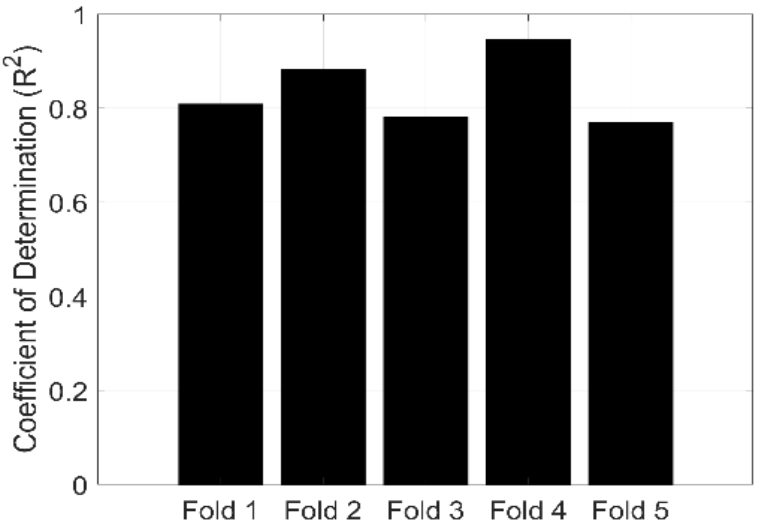

Figure 7. Comparison of model performance for different folds

3.2 Multivariate Regression

In order to compare the efficacy of the data model, two additional models namely multivariate regression and empirical models are taken into consideration. In the regression model, the outcome variable is rate of penetration (ROP) and the predictor variables are mentioned in Table 2. The regression modelling by least square approach is performed using Microsoft Excel spreadsheet program. ROP is chosen as the Y-ordinate and the rest of variables are chosen as the X-ordinate in the data analysis package of spreadsheet. The analysis on Well A computes coefficients (b_0 to b_6) as mentioned in equation 3.2 and this coefficient is applied on Well B drilling parameters to predict the ROP on that well.

$$ROP = b_0 + b_1 TORQUE + b_2 SPP + b_3 RPM + b_4 WOB + b_5 DIFF_{PRESS} + b_6 HKLI \quad (3)$$

3.3 Bourgyne Young Model

The ROP can be empirically calculated by the Bourgoyne and Young Model with the parameters from Table 1 which includes the drilling effects and the corresponding constants. These constants are evaluated from drilling measurements obtained from Well A. By adopting similar regression procedure similar to the multivariate regression, eight coefficients are computed with one constant representing the formation strength as intercept. Due to the non-availability of complete data, effect of bit wear and bit weight is neglected in this evaluation.

4. Results and Discussion

In the prediction phase of machine learning process, the neural network model is applied on the drilling parameters of Well B which was drilled with the same drill string configuration to predict ROP and compared with the empirical and regression model. It must be noted that the same feature engineering techniques applied on the Well A drilling parameters is also applied on Well B drilling parameters before model is evaluated with the new data. The plot of actual ROP and the predicted ROP for the formation 2 which was under consideration, evaluated by the three

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

models is shown in Figure 8. It is seen that the neural network model predictions are better than the other two models as confirmed by the error metrics in Table 4.

Table 4. Model performance comparison

Model	Root Mean Squared Error (ft/hr)
Bourgyne Young Model	307.18
Mutli-Variate Regression Model	251.59
Neural Network Model	88.36

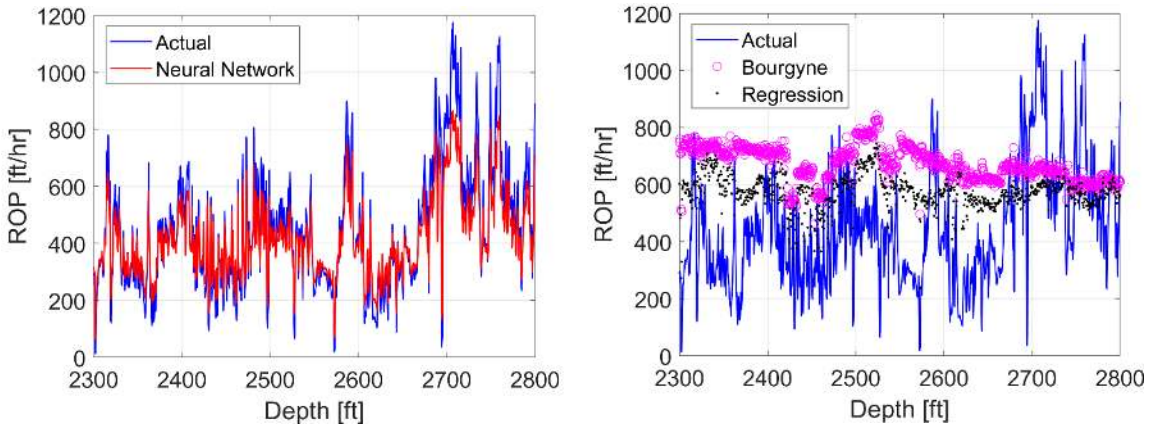

Figure 8. ROP prediction by neural network, Bourgyne and Young model and multivariate regression

The neural network was also compared with two other machine learning models namely
a) Support vector regression (SVR),
b) Random forest regression.

The SVR is based on fitting the observation points as close to the hyperplane and the random forest regression is based on fitting the classifying decision trees on the sub-samples and averaging to improve the predictive accuracy. It is observed from Figure 9 that prediction from these two machine learning models follows the trend of actual ROP. However, the root mean squared error from these two machine learning models are high than the neural network model as shown in Table 5. This validates the deployment of neural network-based machine learning model for estimating ROP from the drilling parameters.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

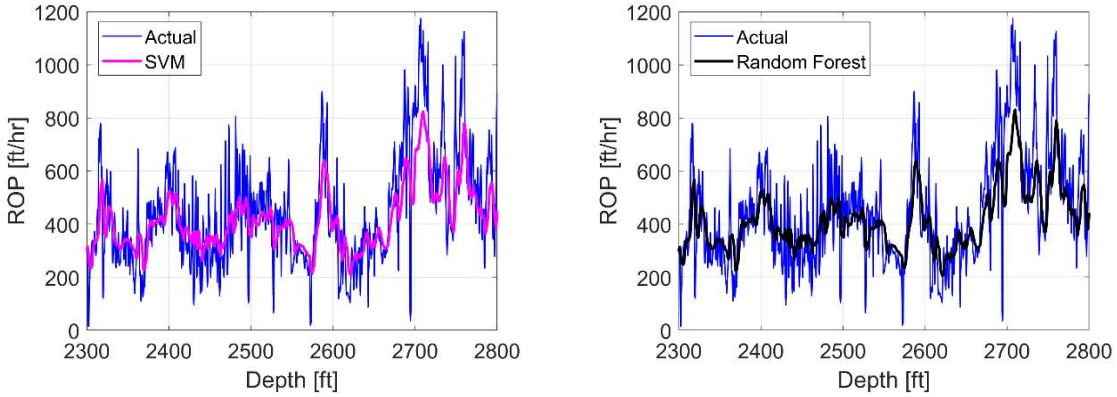

Figure 9. ROP prediction by support vector regressor and random forest regressor

Table 5. Model performance comparison

Model	Root Mean Squared Error (ft/hr)
Support Vector Regression	142,89
Random Forest Regression	145.33
Neural Network Model	88.36

4.1 Optimization

In this section, the drilling parameters are optimized to achieve the best ROP for a particular formation with the aid of neural network model and brute force algorithm. The traditional optimization methods tend to approach local maxima instead of the global maxima leading to sub-optimal solution. The brute force optimization evaluates every scenario on a sequential basis and this is because the variations in drilling parameters is step wise. Moreover, the computing facilities on the rig will be sufficient to support such iterative calculations.

As an example, the optimization is achieved by splitting formation type 2 which spans for 600 ft into 12 divisions of 50 ft each. The minimum and the maximum of WOB, RPM and flow pumps for every division is determined and used as reference limits. The brute force algorithm evaluates all the possible combinations of WOB, RPM and flow pumps between the limits with the help of a computer program and evaluates the maximum possible ROP. Mathematically, maximum ROP can be achieved with maximum WOB, RPM or flow but due to the manufacturing, downhole and rig constraints, it may not be feasible. Figure 10 demonstrates the three feature optimization where RPM, WOB and flow pumps are optimized for increased ROP. It is seen that there is an increase in the ROP for most of the section in consideration and reduction in the surface parameters correspondingly as shown in Table 6.

The increase in the ROP translates directly to reduction in the time required to drill the formation. Time saved to drill through the formation type 2 can be calculated by dividing the formation span by the optimized ROP and this is plotted in Figure 11. For this specific formation, the time save is approximately 45 minutes which in turn leads to a significant saving in the drilling cost.

In the practical sense, it is hard to change the surface parameters in seconds or sub-seconds unless a computerized and sophisticated control system was used. In order to provide guidance to the

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

driller to change the control parameters to achieve the best ROP, the ROP optimization technique is generally recommended over a fixed interval say 50ft, as opposed to the conventional way where a fixed operating parameter were used over one type of formation/section. Depending on the drilling limitation and the flexibility in the surface handling unit, this interval can be changed. Moreover, the best results are achieved in pad wells where there will be consistency in the drill string configurations and formation sequence. By providing such options with the help of simulations and optimization of this machine learning model, the driller can choose the right drilling parameters depending on the practical feasibility at the rig, lithology type and the thickness of the formation.

Table 6. Percentage decrease in operational parameters to achieve increased ROP

Parameter	Non-Optimized (Actual)	Optimized	% Change
WOB	20.20	18.37	-9.02
RPM	118.80	88.42	-25.57
FLWPMPS	943.39	822.65	-12.80
ROP	371.25	423.19	13.99

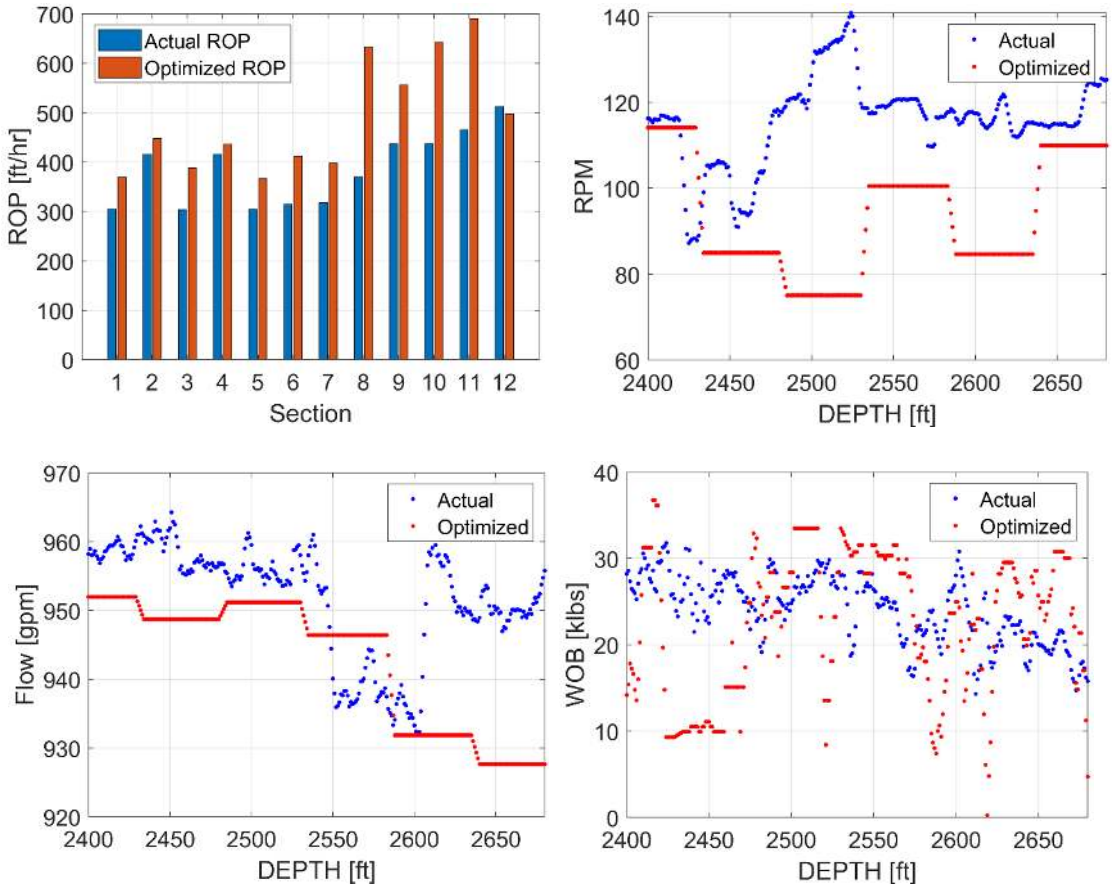

Figure 10. Actual ROP and Predicted ROP when optimizing for WOB, RPM, flow (top-left). Change in RPM (top-right), flow (bottom-left), WOB (bottom-right) for increase in ROP

International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 3, 381-394, 2020
https://doi.org/10.33889/IJMEMS.2020.5.3.032

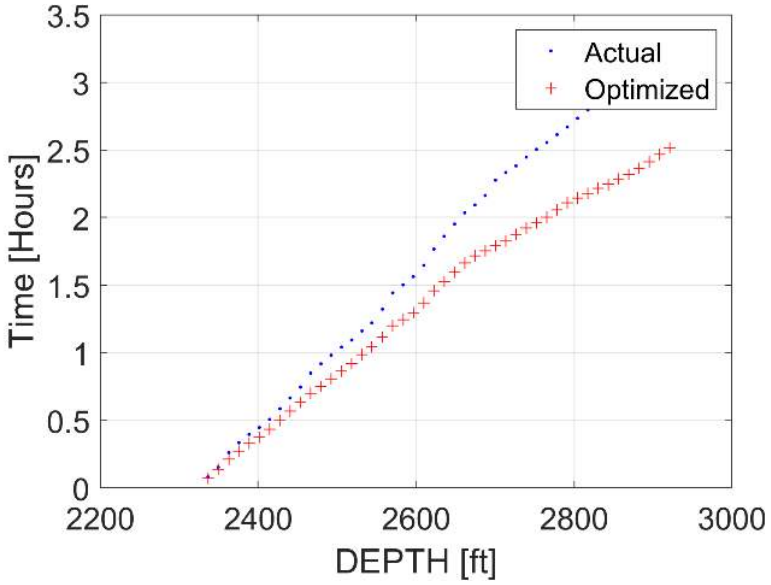

Figure 11. Time consumed to drill the section

5. Conclusion

This paper demonstrates the practical use of data model and learning methods in drilling engineering. Machine learning models are developed to predict ROP along the entire well taking into account the surface operational parameters namely RPM, WOB, Flow, Torque and stand pipe pressure. The feature engineering technique improved the prediction of the machine learning algorithm and accurately extracted the hidden patterns in the field data. This model was extended to perform three parameter (WOB, RPM, Flow) optimization to maximize the ROP over a specific formation type resulting in an increased ROP by 14 %. Since the implementation of such model is simple, this technique could be extended to real-time prediction and optimization at the surface of the rig without the need to rely on the downhole parameters.

Conflict of Interest

The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgments

The authors sincerely thank Mr. Jon Curtis (M/s Petrolink) and Mr. James Lazar (M/s Petrolink) for their support and encouragement in doing this research. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors also sincerely appreciate the editor and reviewers for their time and valuable comments.

References

Bourgoyne Jr, A.T., & Young Jr, F.S. (1974). A multiple regression approach to optimal drilling and abnormal pressure detection. *Society of Petroleum Engineers Journal*, *14*(04), 371-384.

Dunlop, J., Isangulov, R., Aldred, W.D., Sanchez, H.A., Flores, J.L.S., Herdoiza, J.A., Belaskie, J., & Luppens, C. (2011, January). Increased rate of penetration through automation. In *SPE/IADC Drilling Conference and Exhibition*. Society of Petroleum Engineers. Amsterdam, The Netherlands.

Eskandarian, S., Bahrami, P., & Kazemi, P. (2017). A comprehensive data mining approach to estimate the rate of penetration: application of neural network, rule based models and feature ranking. *Journal of Petroleum Science and Engineering*, *156*, 605-615.

Hareland, G., & Rampersad, P.R. (1994, April). Drag-bit model including wear. In *SPE Latin America/Caribbean Petroleum Engineering Conference*. Society of Petroleum Engineers. Buenos Aires, Argentina. https://doi.org/10.2118/26957-MS.

Hegde, C., & Gray, K.E. (2017). Use of machine learning and data analytics to increase drilling efficiency for nearby wells. *Journal of Natural Gas Science and Engineering*, *40*, 327-335.

Mantha, B., & Samuel, R. (2016, September). ROP optimization using artificial intelligence techniques with statistical regression coupling. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers. Dubai, UAE. https://doi.org/10.2118/181382-MS.

Motahhari, H.R., Hareland, G., & James, J.A. (2010). Improved drilling efficiency technique using integrated PDM and PDC bit parameters. *Journal of Canadian Petroleum Technology*, *49*(10), 45-52.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research, 12*(Oct), 2825-2830.

Selesnick, I.W., & Burrus, C.S. (1998). Generalized digital Butterworth filter design. *IEEE Transactions on Signal Processing*, *46*(6), 1688-1694.

Tripathy, S.S., Saxena, R.K., & Gupta, P.K. (2013). Comparison of statistical methods for outlier detection in proficiency testing data on analysis of lead in aqueous solution. *American Journal of Theoretical and Applied Statistics*, *2*(6), 233-242.

Warren, T.M., & Armagost, W.K. (1988). Laboratory drilling performance of PDC bits. *SPE Drilling Engineering*, *3*(02), 125-135.

Zakariya, R., Zein, A., Diab, E., Lotfy, A., Marland, C., Obaidli, Y.Y.A., Braik, H.A.A., Amin, M.S., & Attalah, M. (2015, September). A case study of real-time drilling optimization to improve well delivery through enhancing drilling rates and identifying invisible lost time to improve performance. In *SPE North Africa Technical Conference and Exhibition*. Society of Petroleum Engineers. Cairo, Egypt. https://doi.org/10.2118/175748-MS.