Journal of Natural Gas Science and Engineering 77 (2020) 103230

Contents lists available at ScienceDirect

Journal of Natural Gas Science and Engineering

journal homepage: http://www.elsevier.com/locate/jngse

ROP and TOB optimization using machine learning classification algorithms

Mayowa Oyedere *, Ken Gray

Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin, USA

A B S T R A C T

Drilling Optimization has consistently generated research interest over the years because the cost saving beneifts associated to improve drilling efficiency. Rate of penetration (ROP) and torque-on-bit (TOB) predictions have become critical to the successful drilling optimization efforts. Several physics-based and data-driven models have been developed for ROP and TOB prediction and majority of the data-driven models use regression-based approaches. This paper introduces a new approach to ROP and TOB prediction by modeling them as a classification problem consisting of two regions (low and high ROP and TOB respectively) based on a user-defined threshold. ROP and TOB are modeled as a function of with weight-on-bit (WOB), flow rate, rotray speed (RPM) and unconfined compressive strength (UCS). Five different machine learning classifcation algorithms - logistic regression, linear discriminant analysis (LDA), quadratic driscriminant analysis (QDA), support vector machines (SVM) and random forest were implemented in this paper to develop the classification model. Using the Area Under Curve (AUC) as the classification performance metric, results from the simulations showed that the best classifier should be chosen for each formation . Also, for the practical application of this approach to ROP and TOB prediction, a probability gradient heat map of RPM and WOB was developed as a tool to help the driller make informed decisions on the combinations of RPM and WOB values that would yield the desired regions of ROP and TOB.

1. Introduction

A large portion of the budget for an oil and gas exploration project is the drilling cost and the costs are calculated per day. As such, the longer it takes the drill a well due to drilling problems encountered, the higher the cost of the well. Therefore, any measure that helps reduce drilling problems and increase efficiency have direct cost saving implications (Hegde and Gray, 2017). Rate of penetration (ROP) measures the drilled depth for a given period of time (ft/hr) and hence is a worthy drilling parameter to be used for optimization. Even though other operations like tripping, bit change, hole problems affect the total drill time, the optimization of ROP has been established over the years as an important piece in drilling optimization puzzle. Also, ROP in conjunction with other parameters can be used to detect drilling problems like stick-slip and kicks.

Estimates of downhole torque also known as torque-on-bit (TOB) is another critical part of drilling optimization because it is one of the parameters used in the equation for Mechanical Specific Energy (MSE). MSE is a defined as the amount of energy required a unit volume of rock (Teale, 1965) hence accurate estimations of TOB will be help in improving the accuracy of the calculated MSE.

1.1. ROP models

1.1.1. Analytical models

Several ROP models have been developed over the years using on laboratory experiments, bit type and machine learning. Maurer (1962) explained that under perfect hole cleaning conditions, ROP is proportional to rate of penetration (RPM) and the weight-on-bit (WOB) squared. However, the author stated that under actual drilling conditions, ROP varies linearly with WOB until the hole cleaning problem outweighs the gains of increased WOB. This is often referred to as the floundering point beyond which increase in WOB causes a decrease in ROP because of the accumulation of the cuttings. An ROP model was developed for roller-cone bits as a function of nine parameters amongst which were WOB, bit diameter, RPM and formation strength (UCS). Bingham (1965) proposed a model that predicted ROP as a function of only RPM, WOB and bit diameter. The model predicts a power law relationship between ROP and WOB and a direct proportionality between ROP and RPM. Bourgoyne and Young (1974) expanded on the Bingham model and developed the most comprehensive model that eliminated the need for multiple individual models for the additional parameters. The additional parameters accounted for the effect of pore pressure, bit wear, bit hydraulics, rock drillability, rock compaction with depth and differential pressure. Eckel and Nolley (1949) added a

* Corresponding author.
E-mail addresses: mayowaoyedere@utexas.edu (M. Oyedere), ken_gray@mail.utexas.edu (K. Gray).

https://doi.org/10.1016/j.jngse.2020.103230
Received 21 November 2019; Received in revised form 24 February 2020; Accepted 25 February 2020

Reynolds number expression for hole cleaning in terms of bit hydraulics and mud properties to the ROP model and proved in Eckel (1968) that the Reynold's number expression was valid.

Hareland and Rampersad's (1994) ROP model for drag bit used conservation of mass to postulate that the rock volume removed by each bit is related to the contact area and the applied WOB. Alum and Egbon (2011) added annular friction loss and plastic viscosity capabilities to the Bourgoyne and Young model. Motahhari et al. (2010) also used the conservation of mass principle to develop a useful ROP model for directional and horizontal drilling operations with positive displacement motors (PDMs).

1.1.2. Data-driven models

The most common machine learning algorithm that has been used for ROP prediction is neural networks and has been implemented by several authors - Bilgesu et al. (1997), Jahanbakhshi et al. (2012), Gandelman (2012) and Amer et al. (2017) just to name a few. Hegde et al. (2015a, b) implemented other types of algorithms like, regression, boosting, random forest and support vector machine in the prediction of ROP. Wallace et al. (2015) developed a systematic workflow that applies statistical models to real time drilling operations. Hegde and Gray (2017) used machine learning and data analysis to develop a method for optimizing drilling parameters during post drilling analysis. Hegde et al. (2017) compared the ROP prediction using the conventional physics-based models and data driven models.

1.2. TOB models

The use of downhole sensors to measure TOB close to the bit is still not widely used in the industry. When used, they are added to the MWD tools and downhole data is transmitted to the surface using mud pulse telemetry and then used for MSE estimations. Trichel et al. (2016) developed a downhole automation system that used wired drill pipe to stream TOB data from the downhole tools to the surface at high speed. An alternative to using MWD tools is using the differential pressure across the mud motor because of its proximity to the bit (Logan, 2015). TOB is modeled to be directly proportional to the differential pressure however, the accuracy of this method highly depends on accuracy of the differential pressure values. Torque and drag models based on the motion (axial and rotational) of the drillstring in the well can also be used to estimate TOB. The motion of the drillstring is modeled using Coulomb's friction and a calibrated coefficient of friction. Given the surface torque, the model calculates the TOB by accounting for the torque losses that occur along the drillstring (Sheppard et al., 1987). In addition to these methods, Hegde and Gray (2018) used machine learning models to predict TOB using surface parameters.

1.3. ROP and TOB prediction as a classification problem

Most of the machine learning ROP and TOB prediction models developed so far use a regression-based approach as highlighted in the previous section. In this classical approach, the models are developed using training data and the trained model is used to make numeric predictions of ROP. Afterwards, an extra gradient-based or metaheuristic optimization scheme is run using the trained model to obtain the optimized values of input drilling parameters. Hegde et al. (2017) used a proprietary algorithm called Wider Windows Data Analytics Optimizer for optimization. Hegde et al. (2018) compared the performance of metaheuristic algorithms like simplex, differential evolution and particle-Swarm for real-time drilling optimization.

This paper presents a possible paradigm shift in the approach to ROP and TOB prediction where ROP and TOB predictions are modeled as a classification problem. Generally, in classification problems, the goal is to use the training data to build a classification model that divides the data into different classes. Afterwards, the model is used to make predictions of the class each test data point falls into. The classification in this paper is a binary classification and the aim is to divide the ROP and TOB values into two regions – low and high. With this classification approach, the classification algorithm can be used as an optimizer to generate a probability gradient plot for the high and low regions as shown in Figs. 11 and 14. This approach also can be potentially suited for business applications because the focus is no longer to the determine an optimal numerical value for ROP and TOB which can be variable but to determine the region of acceptable ROP and TOB. Interpretations are also easier since there is a clear delineation between the region of low and high ROP and TOB. The goal for example thus becomes to determine the drilling parameters that ensure drilling is in high ROP regions or low TOB regions.

To successfully make this classification, a user-defined threshold for ROP and TOB needs to be specified and there are a number of possible ways to define the threshold. It could be defined based on expert knowledge of the drilling conditions being experienced or based on the information from a previously drilled well in the field. Business goals could also guide the choice of the threshold - for example, a higher threshold for ROP might be desirable to reduce the time needed to drill the a well whereas a lower threshold for TOB may be better to reduce possible hole problems like increased vibrations and stick slip that could

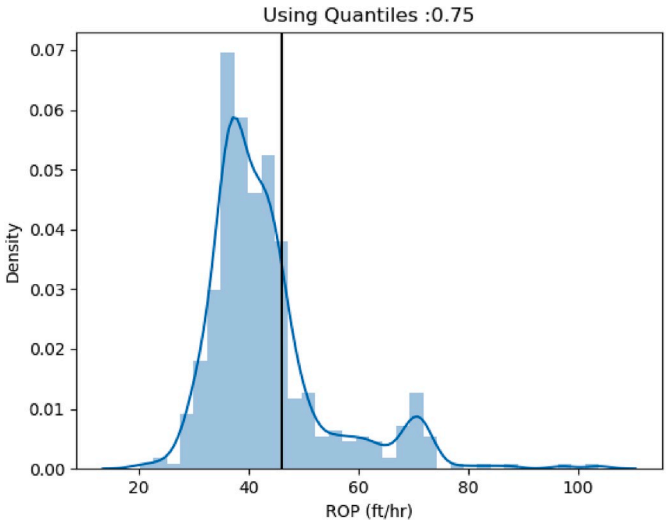

Fig. 1. Density plot of ROP in the Base Last Salt Formation with the 75th percentile threshold line.

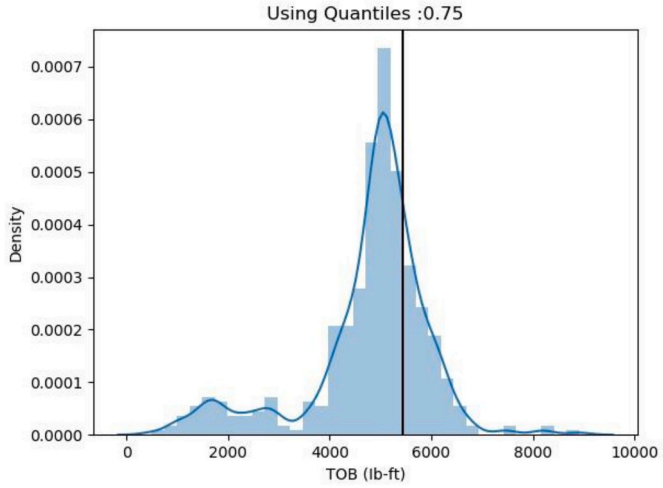

Fig. 2. Density plot of TOB in the Base Last Salt Formation with the 75th percentile threshold line.

arise from high TOB. Finally, the threshold can be defined using information from the training data and this is method employed in this paper. Fig. 1 shows the density plot for ROP with a 75th percentile threshold line while Fig. 2 shows the density plot for TOB with a 75th percentile threshold line.

2. Theory

The different types of machine learning classification algorithms have varying predictive performance based on the underlying principle of how the algorithms work. The aim of classification is to determine a boundary that is used to partition the data set into classes. This decision boundary can be linear or non-linear and classification algorithms perform differently based on the nature of the decision boundary. The classification algorithms used in this work are - logistic regression, support vector machines, linear and quadratic discriminant analysis and random forest. This section provides a brief overview of the underlying principles of each of the algorithms.

2.1. Logistic regression

Logistic regression is arguably the most commonly used classification algorithm because of its simplicity and diverse applicability. It is used to study the relationship between a categorical response variable, Y with a binary outcome (0,1) and a set of predictor variables, X. If the response variable has more than two unique values, then it is called multinomial logistic regression. In linear regression, the response variable is modeled as a linear function of all the predictor variables i.e.

$$Y = p(x) = \theta^T X \tag{1}$$

where θ is a vector containing the model parameters and X is a vector containing the predictor variables. However, logistic regression does not directly model Y but rather it models the probability that Y falls into a given category i.e. $Pr(Y = 1|\, X)$. Since probabilities lie between 0 and 1 and the linear regression function cannot ensure this condition, a sigmoid function is used to ensure the output is between 0 and 1 (Goodfellow et al., 2016). The logistic function is expressed as:

$$p(x) = Pr(Y = 1|\, X) = \frac{1}{1 + e^{-\theta^T x}} = \frac{e^{\theta^T X}}{1 + e^{\theta^T X}} \tag{2}$$

The sigmoid function is a convex function and a number of methods can used to find the optimum model parameters. However, for this application the maximum likelihood estimation is used to estimate the values of θ. The maximum likelihood function finds the values of θ that maximize the probability of observing the training data. Thus, Eqn. (2) can be rewritten as:

$$log\left(\frac{p(X)}{1 - p(X)}\right) = \theta^T X \tag{3}$$

The left-hand side of Eqn. (3) is called the log-odds or logit and the equation shows that the logistic regression model has a logit that is a linear function of X.

2.2. Linear discriminant analysis (LDA)

Logistic regression models the conditional distribution of the response variable Y given the predictor variables X. Some of the shortcomings of this method is that the parameters of the model become unstable when the classes are wide apart. Also, logistic regression is less stable than LDA when the number of observations is small. LDA addresses these shortcomings by modeling the distribution of the predictor variables X in each response class and then uses the Bayes theorem to perform prediction (James et al., 2013). The Bayes theorem is given by:

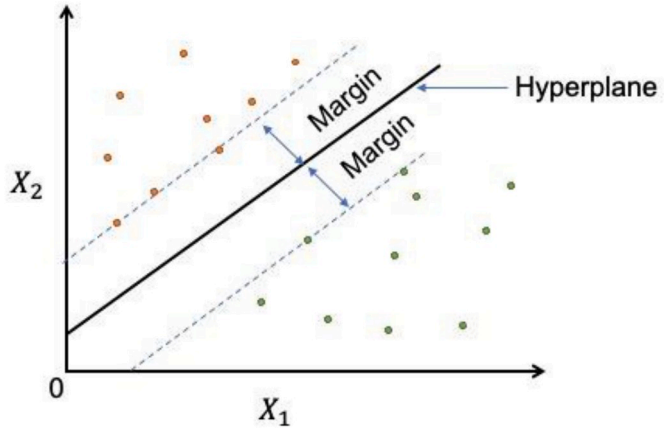

Fig. 3. A maximum margin hyperplane separating a data into two classes.

$$Pr(Y = k|\, X = x) = \frac{\pi_k f_k(x)}{\sum_{i=1}^{K} \pi_i f_i(x)} \tag{4}$$

π_k is the prior probability that an observation chosen at random belongs to the kth class. It is obtained by calculating the fraction of the training data that belong to the kth class. $f_k(x) \equiv \Pr(X = x|Y = k)$ and is referred to as the density function of X for an observation that belongs to the kth class. This implies that if the probability that an observation in the kth class has $X \approx x$ then $f_k(x)$ is relatively large and vice-versa. $f_k(x)$ is not easily determined so it is assumed to follow a normal or Gaussian distribution i.e.

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2\sigma_k^2}(x - \mu_k)^2\right) \tag{5}$$

where σ_k^2 and μ_k are the variance and mean values for the kth class. For simplicity, the classes are assumed to have the same variance σ . Therefore, Eqn (4) can be rewritten as (James et al., 2013):

$$Pr(Y = k|\, X = x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_k)^2\right)}{\sum_{i=1}^{K} \pi_i \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_i)^2\right)} \tag{6}$$

LDA approximates the Bayes classifier as explained above and this yields a linear decision boundary (James et al., 2013).

2.3. Quadratic discriminant analysis (QDA)

Just like LDA, QDA classifier assumes that the data from each class comes from a Gaussian distribution and then uses Bayes' theorem to classify the data. The major difference between LDA and QDA is that QDA assumes each class has its own unique covariance matrix so that a data point from the kth class is of the form $X \sim N(\mu_k, \Sigma_k)$ where Σ_k is a covariance matrix for the kth class. This makes the discriminant function quadratic and hence the name QDA.

The choice between LDA and QDA is determined by the bias-variance trade off. Since QDA calculates a different covariance matrix for each class, it means that more parameters will be computed making QDA more flexible than LDA which assumes the classes share a common covariance matrix. Practically, LDA yields better results when the training data is relatively small because it has a low variance. On the other hand, when the training data is large the assumption of the same covariance matrix becomes invalid so QDA makes better predictions (James et al., 2013).

2.4. Support vector machines (SVM)

Support vector machines is an algorithm that is used for binary classification when the response variable has two classes. In its simplest form for perfectly separable problems, SVM is called a maximal margin classifier. A maximal margin classifier uses a hyperplane to split the parameter space into two. For example, in a two-dimensional space, a hyperplane is a line while in a three-dimensional space, it is a plane. There are usually several possible hyperplanes however to determine the optimal separating hyperplane, the smallest perpendicular distance (margin) from the training data points to the hyperplanes is calculated. The hyperplane of choice is the separating hyperplane that has the largest margin as shown by the dashed lines in Fig. 3 and the data points that lie on the margin are called the support vectors. This decision boundary is then used to classify the test data based on the side of the boundary the data point falls. SVM is an extension of the maximal margin classifier because it can used for classifications problems where classes are not easily separable and also where the class boundaries are non-linear .

For non-linearly separable problems, SVM introduces a soft margin that allows some training data points to be on the incorrect side of the margin or the hyperplane (Cortes and Vapnik, 1995). These are the points that are misclassified by the classifier. Mathematically, this is equivalent to optimizing the cost function in Eqn. (7) below:

$$minimize_{\theta} \sum_{i=1}^{N} max\left(1 - y_i \theta^T x_i,\ 0\right) + \lambda \theta_2^2 \tag{7}$$

The first term is the equation is called the hinge loss, the second term is the regularization (L-2 norm) term and λ is the non-zero regularization parameter. For non-linear decision boundaries, SVM uses kernels to enlarge the feature space (Bosner et al., 1992). A kernel function estimates the similarity of two observations and some of the popular kernels are linear, sigmoid, polynomial and radial basis function respectively.

2.5. Random forest

Random forest is an ensemble machine learning algorithm that combines the predictive power of individual decision trees by combining several trees and averaging them out. Decision tress use a set of splitting rules to segment the predictor space into smaller regions as the tree grows until a stopping criterion is reached. For every data point that falls into a region, it is predicted as belonging to the most commonly occurring class in that region. A decision tree is non-linear in nature and is well suited for non-linear classification. Individual deep decision trees suffer from high variance where a little change in the tree structure causes a significant change in predictions. Random forest addresses this limitation using a process called bagging where bootstrapped samples of the training data are used to build the trees and then the predictions of the trees are averaged out (Breimann, 2001). Also, to ensure that the bagged trees are not correlated, the algorithm uses only a random sample of the predictor variables at each tree split (Ho, 1995).

3. Classification performance measure

Since machine learning algorithms are built based on training data and are used to make predictions on test data, it therefore imperative to have methods for evaluating the performance of the algorithms. Also, the models have error associated with them so in this section, the commonly used performance measures for classification machine learning algorithms are discussed.

3.1. Confusion matrix

This is not a performance measure in itself, but it contains important information that is used to calculate some performance metrics for classification problems with binary outputs. There are four possible outcomes for the binary output as shown in Table 1. The first possibility is cases with positive labels that are classified as positive and it is known as true positive. The second possibility, true negative, is when cases with negative labels are correctly classified as negative. The third possibility called false positive is when negative labels are incorrectly classified as positive. It is also referred to as type I error. Finally, the last possibility, false negative, is when positive labels are incorrectly classified as negative. False negatives are also referred to as type II error.

Table 1
Confusion matrix for classification models.

Confusion Matrix		Target	
		True	False
Model	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

3.2. Accuracy

This is the most basic performance metric for classification models that measures fraction of data points correctly classified by the model. For a two-class classifier accuracy is written as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}$$

Accuracy is used with some form of caution because it only works well when the classes are approximately balanced. However, it has been shown to fail when there is imbalance in the classes and one class is the majority in the data.

3.3. Precision

Precision measures the fraction of the correctly classified data points out of all the predicted data points with that label value hence it is sensitive the number of correctly classified data points. Precision is expressed as:

$$Precision = \frac{M_{i,i}}{\sum_j M_{i,j}} \quad \frac{TP}{TP + FP} \tag{9}$$

3.4. Recall

This is also called sensitivity and it measures the fraction of correctly classified data points for a label value out of all the data points that have that label value.

$$Recall = \frac{M_{i,i}}{\sum_i M_{i,j}} \quad \frac{TP}{TP + FP} \tag{10}$$

3.5. F-1 score

F1-score combines the recall and precision metrics to give a single metric that can be used to evaluate a classification algorithm. It calculates the harmonic mean of the precision and recall.

$$F1\ score = 2 \times \frac{Recall \times Precision}{Precision + Recall} \tag{11}$$

3.6. Receiver Operating Characteristics (ROC) curve

The ROC plot is a plot of true positive rate versus false positive rate and the ideal ROC curve will hug the top left corner (Fawcett, 2006). The area under the ROC curve (AUC) gives the overall performance summarized over all possible threshold as shown in Eqn. (12) below.

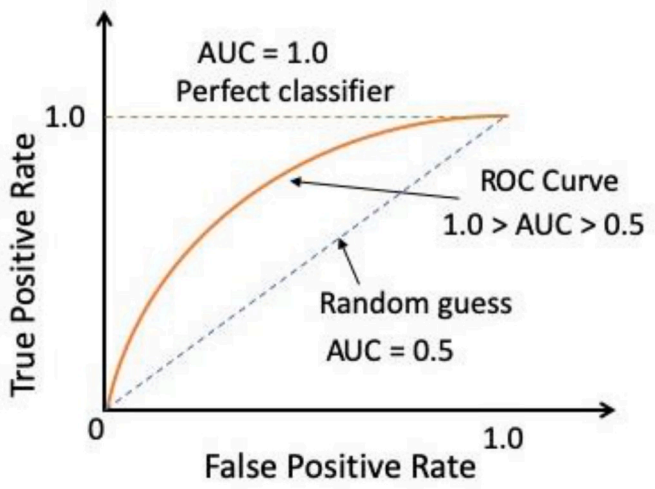

Fig. 4. ROC Curve with values of AUC for a balanced two-class problem.

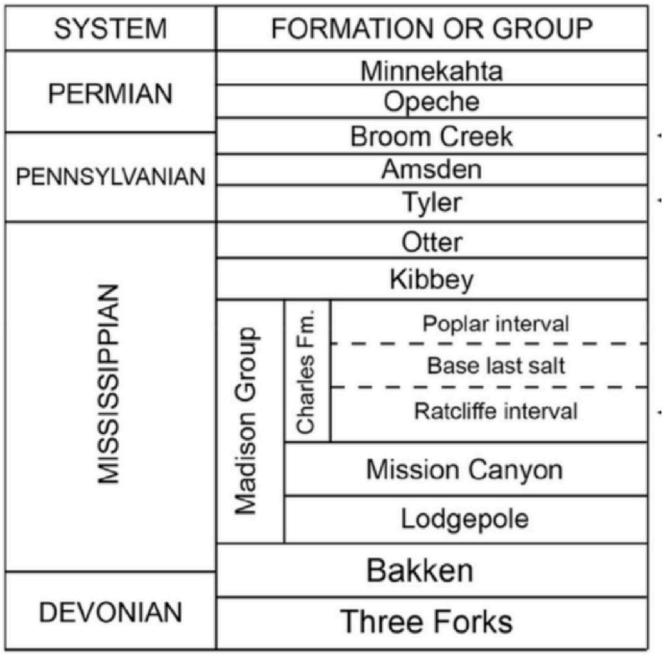

Fig. 5. Generalized stratigraphic column for the Williston Basin, North Dakota (Theloy, 2014).

$$AUC = \frac{S_p - n_p(n_n + 1)/2}{n_p n_n} \tag{12}$$

where, S_p is the sum of the all positive, $n_p$is the number of positives and n_n is the number of negatives.

The higher the AUC the lower the increase in false positive rate required to achieve a required true positive rate and for an ideal classifier, the AUC is 1 as shown in Fig. 4. Therefore, AUC is useful for comparing classifiers so that the classifier with the higher AUC is generally the better one.

4. Methodology

4.1. Dataset

The dataset used in this work is from a Bakken shale horizontal well drilled in North Dakota. The stratigraphic section of the Bakken shale is shown in Fig. 5.

The vertical section of the well spans 20 different formations from 4,871 ft true vertical depth (TVD) to 9128 ft TVD. The dataset contains depth-based (every 0.25 ft) measured readings of surface and downhole drilling parameters. Depth-based data is usually suited for ROP analysis because it is easier to locate formation boundaries which will facilitate lithology-based analysis. (Wallace et al., 2015). Only 12 out of the 20 formations were used for this study. The entire vertical section of the well was drilled with the same Smith 616 PDC bit and without a downhole motor or rotary steerable. Figs. 6 and 7 show the range of ROP and TOB values for each of the formations.

4.2. Feature engineering

The fundamental problem statement of a machine learning algorithm is the estimation of a functional relationship between input variables or features (in this case surface drilling parameters) and an output variable (in this case ROP and TOB). From the foregoing, it is clear that the choice of features used in the machine learning algorithm will impact the performance of the algorithm when used for predictions. It may desirable to use all the input features including the irrelevant ones to approximate the functional relationship but in practice this potentially creates two problems. Firstly, including irrelevant features increase the computational cost of the predictions and secondly, it may also lead to overfitting.

Feature engineering is the process of using domain knowledge to select the features that will ensure that maximum accuracy is obtained from the algorithm. Some of the possible input features include travelling block height, differential pressure, RPM, WOB, flowrate and unconfined compressive strength (UCS). However, knowledge of the parameters used in the traditional physics based ROP and TOB models provide good indicators of the input features to be used as inputs. For real-time applications during drilling, the most desirable features are surface parameters that the driller has control over and can be adjusted. Dunlop et al. (2011) developed a real-time automated system that monitors drilling performance and constantly adjusts WOB and RPM to maximize ROP. Jahanbakhshi et al., 2012 used additional features like rock mechanical properties, hydraulics and bit type in their artificial neural network model. Hegde et al. (2015a) in their work showed that using RPM, WOB, flowrate and UCS as the input parameters for bagging and random forests algorithms produced higher computational accuracy and efficiency. This formed the basis for further studies carried out (Hegde et al. (2017), Hegde and Gray (2017), Hegde and Gray (2018). These are also the input features used in this work.

4.3. Model development and selection

From the data, WOB and flowrate were surface measurements while the RPM was measured downhole using MWD tools. To simplify and ensure consistency in prediction, parameters like BHA, drilling mud and bit are assumed to be constant. The models are lithology dependent and so are retrained for each formation (Hegde and Gray, 2018). The data in each formation is split into training and test sets, the classification algorithms are built on the training set and then evaluated on the test set. In this work the train/test split ratio in 60/40 which means that the first 60% of the data is used to train the model and while the remaining unseen 40% of the data is used to test the model.

Some classification algorithms in addition to model parameters also have hyperparameters which determine the characteristics of the model. SVM has number of hyperparameters that can be used to develop the model. For example, the kernel function choice – (linear, polynomial or radial basis), C (the penalty term for misclassification) and gamma (the goodness of fit of the model to the training set). Random forest also has hyperparameters like the number of trees, number of features to be considered at each split and how deep the trees grow. Cross validation is a technique used to determine the best hyperparameters for optimal

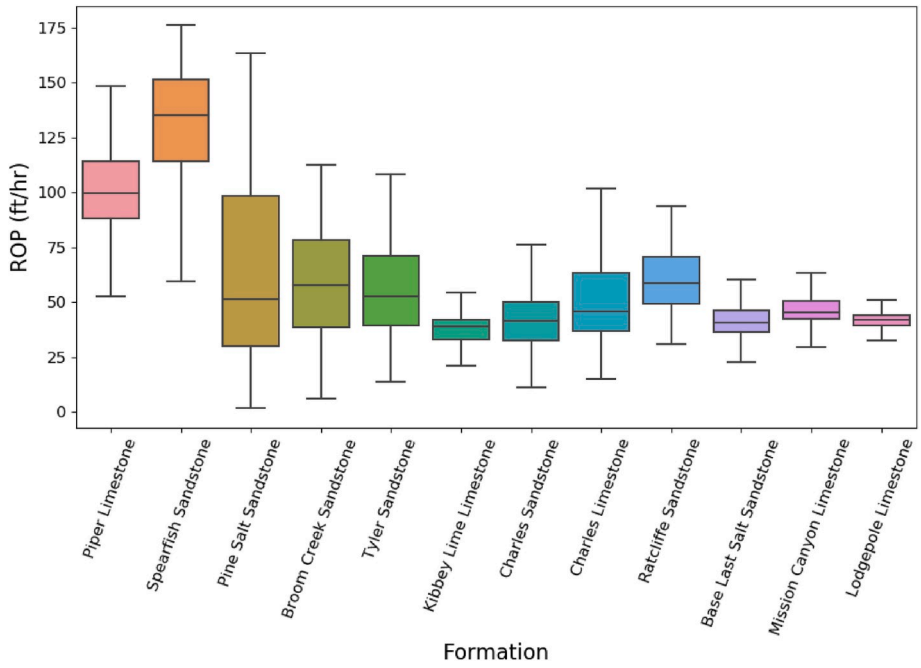

Fig. 6. Boxplot showing the range of ROP values for each of the 12 formations.

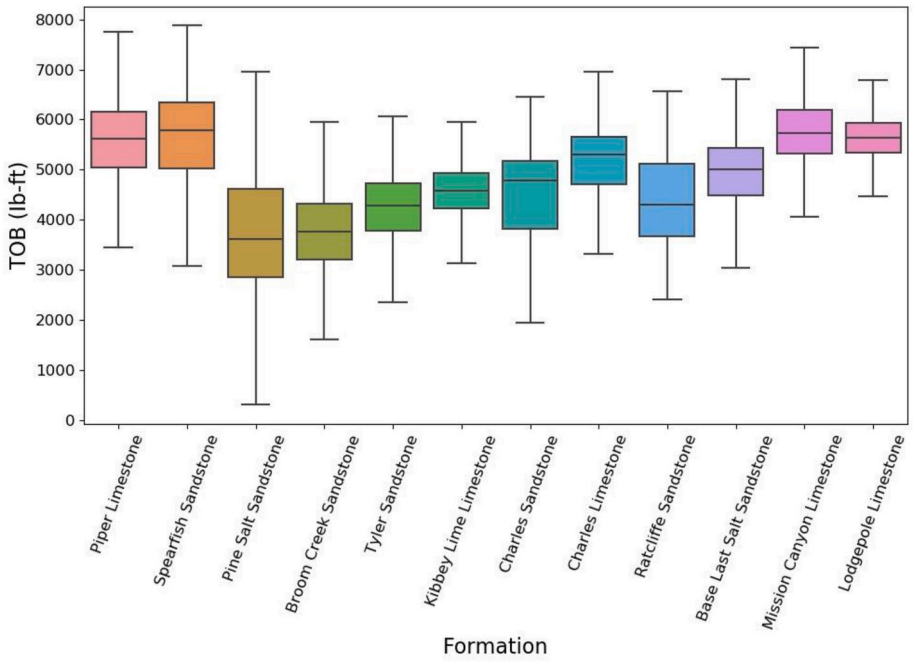

Fig. 7. Boxplot showing the range of ROP values for each of the 12 formations.

model performance. However, cross validation still suffers from some bias and so nested cross validation is used to address this bias by using an inner loop for parameter tuning while the outer loop is used for error estimation. Python's scikit learn package was used for the implementation of the machine learning algorithms.

Class imbalance occurs in binary classification problems when the classes are not approximately equally represented with one class having more data points than the other. Class imbalance can result in significant error in the predictions for the minority class. There are several techniques for handling class imbalance – undersampling, upsampling, negative downsampling and weighting (Hegde et al., 2019). For this application, imbalance was introduced by the selection of the 75th percentile as the threshold for splitting the data into two classes. Synthetic Minority Over-sampling Technique (SMOTE) is used to address imbalance in the paper (Chawla et al., 2011). It is more robust than the techniques listed above because it combines oversampling of the minority class and undersampling of the majority class.

5. Results and discussion

5.1. Application to ROP prediction

The different classification algorithms implemented in this paper classify ROP into two classes – low and high ROP. The pre-determined

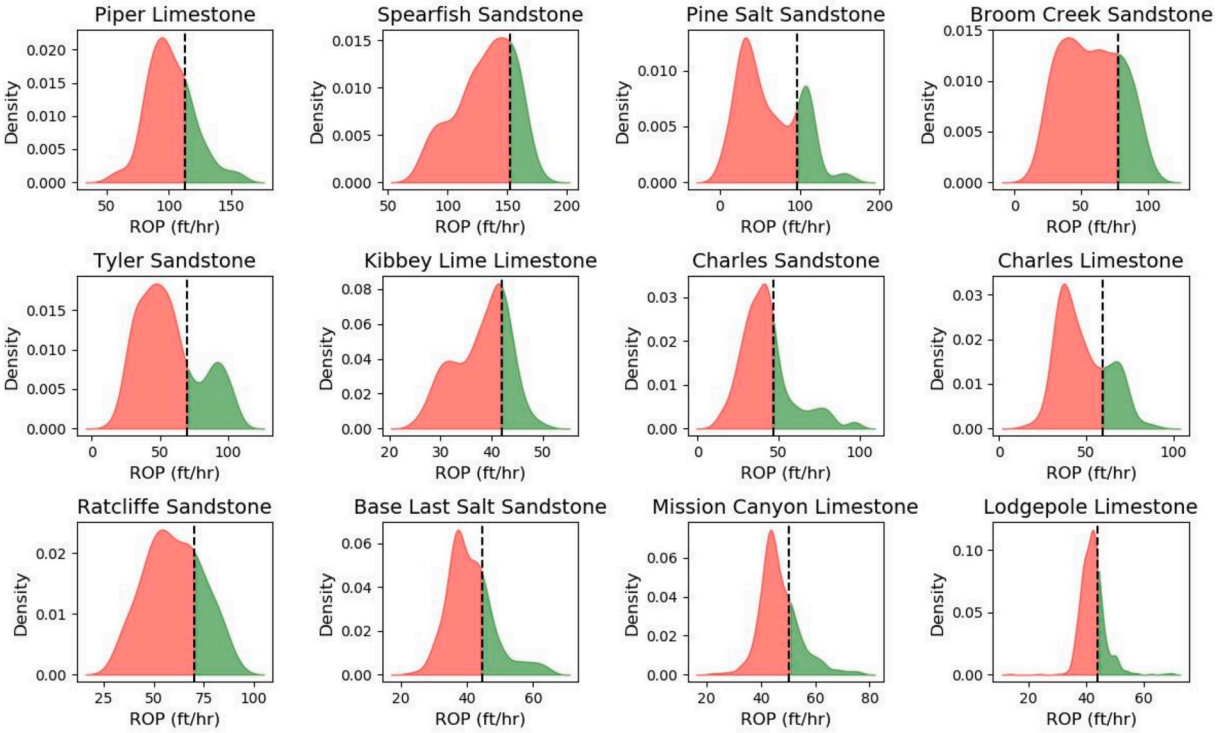

Fig. 8. Density plot of ROP for the 12 formations. The red region is the region of low ROP and the green (desired) region is the region of high ROP.

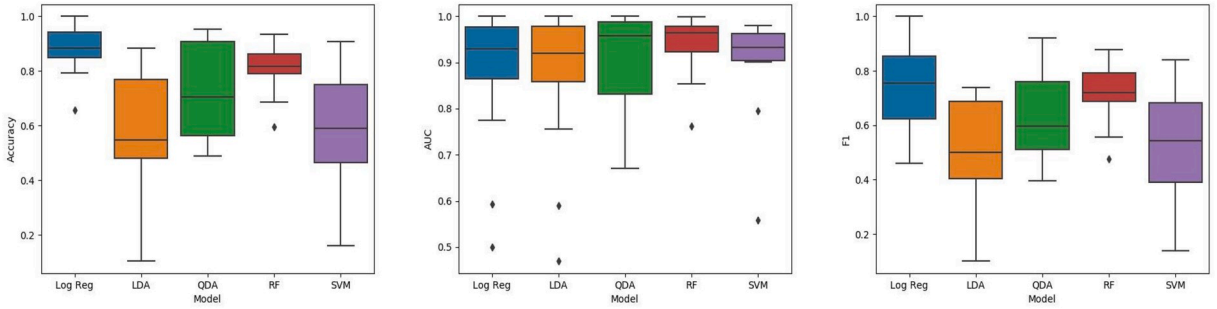

Fig. 9. Boxplot showing the performance of the five classification algorithms used for ROP prediction. Left figure (accuracy), middle figure (AUC), right figure (F1).

threshold (black line in Fig. 8) used for ROP was based on the data from each formation and was chosen as the 75th percentile of the ROP values. As explained previously, this threshold is user-defined and is guided by the desirable drilling outcomes (it need not be the 75th percentile). For applications in this paper, the desirable outcome is obtaining high ROP and it is assumed that ROP values higher than the 75th percentile is the region of high ROP. If an ROP value is higher than the 75th percentile, the data point is classified as high (green region in Fig. 8) while if it is lower than the 75th percentile, the data point is classified as low (red region in Fig. 8).

The evalaution metrics for the five algorithms were calculated and Fig. 9 below shows a boxplot of the metrics for all the formations. The line in the box is the median value, the bottom and top edges of the box are the first and third quartiles respetively while the lower and the upper whiskers represent the minimum and maximum values respectively. For the accuracy and F1 metrics, logistic regression was the overall best performing algorithm followed by random forest, QDA, LDA and finally SVM. However as mentioned in section 3.6, AUC is the best performance metric for comparing different classifiers. From the boxplot for AUC, all the classifiers performed reasonably well however, overall random forest algorithm was the best performing classifier across most of the formations followed by logistic regression.

Table 2
Tabulation showing the AUC scores from the five ROP classifiers for each of the formations.

FORMATION	AUC SCORES				
	Log Reg	LDA	QDA	RF	SVM
Piper Limestone	0.94	0.93	0.97	0.96	0.98
Spearfish Sandstone	0.94	0.92	0.97	0.97	0.93
Pine Salt Sandstone	0.77	0.76	0.98	0.98	0.96
Broom Creek Sandstone	0.90	0.89	0.84	0.76	0.90
Tyler Sandstone	0.99	0.99	1.00	1.00	0.90
Kibbey Lime Limestone	0.59	0.59	0.67	0.85	0.56
Charles Sandstone	0.90	0.90	0.90	0.96	0.97
Charles Limestone	1.00	1.00	1.00	1.00	0.95
Ratcliffe Sandstone	1.00	1.00	1.00	0.98	0.97
Base Last Salt Sandstone	0.50	0.47	0.71	0.86	0.80
Mission Canyon Limestone	0.92	0.92	0.82	0.94	0.93
Lodgepole Limestone	0.97	0.97	0.95	0.97	0.93

Table 2 shows the AUC scores for the five classifiers across all the formations. Four of the five algorithms (except random forest) consistently performed poorly in the Kimberly Lime Limstone formation while

Table 3
Tabulation showing the AUC scores of the best ROP classifiers for each of the formations.

FORMATION	AUC	Best Classifier
Piper Limestone	0.98	QDA
Spearfish Sandstone	0.97	RF
Pine Salt Sandstone	0.98	RF
Broom Creek Sandstone	0.90	Log Reg
Tyler Sandstone	1.00	RF
Kibbey Lime Limestone	0.85	RF
Charles Sandstone	0.97	SVM
Charles Limestone	1.00	RF
Ratcliffe Sandstone	1.00	Log Reg
Base Last Salt Sandstone	0.86	RF
Mission Canyon Limestone	0.94	RF
Lodgepole Limestone	0.97	RF

logistic regression and LDA were the low performers in Base Last Salt Sandstone as shown in Table 2. Hypothesis testing, a statistical method, is used to understand this observed trend in the Kimberly Lime Limestone formation. Given the null hypothesis that each of the input feature is not related to ROP, individual t-tests were run for each of the input features and the corresponding p-values obtained were compared to the critical p-value (0.05) for rejecting the null hypothesis. RPM and WOB had p-values less than 0.05 showing that they were statistically signifcant features while flow rate and UCS had p-values greater than 0.05 indicating that they were not statistically significant features for predicting ROP. This provides a possible explanation why the algorithms performed poorly in the Kimberly Lime Limestone formation. Table 3 shows that the best classifier for each formation based on the AUC score varies from one formation to the other. This means that instead of trying to identify an overall best classifier for all the formations, a better approach is to use to the best classifier for each individual formation.

5.1.1. ROP optimization

In order to carryout optimzation of ROP for each formation the best classifier for each formation was used. The random forest, logistic regression and LDA classifiers allows the input features to be ranked in order of their importance as shown in Fig. 10 below. However, the radial basis function used as the kernel for SVM does not allow features to be ranked because it is non-linear and the data set is mapped into a multi-

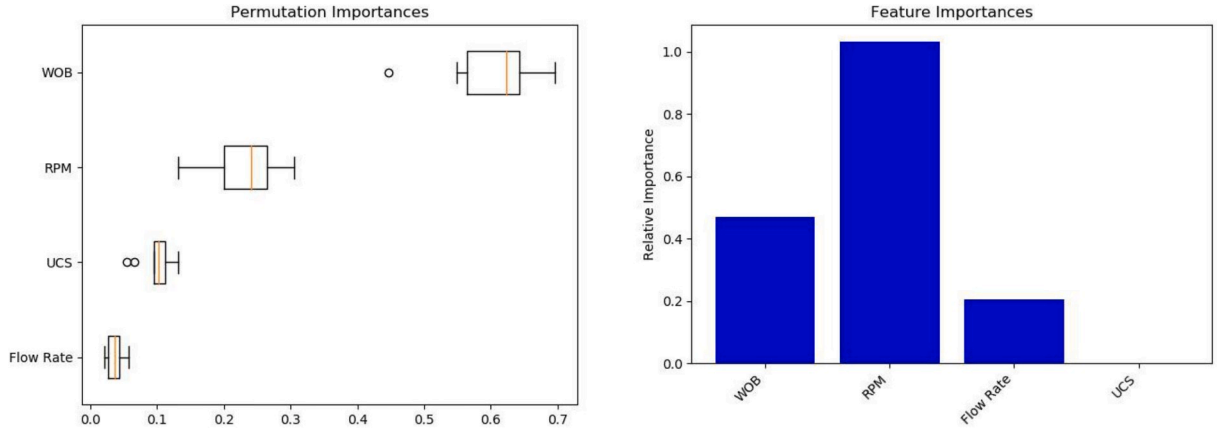

Fig. 10. Permutation importance of random forest for base salt sandstone (left); feature importance of logistic regression for broom creek sandstone (right).

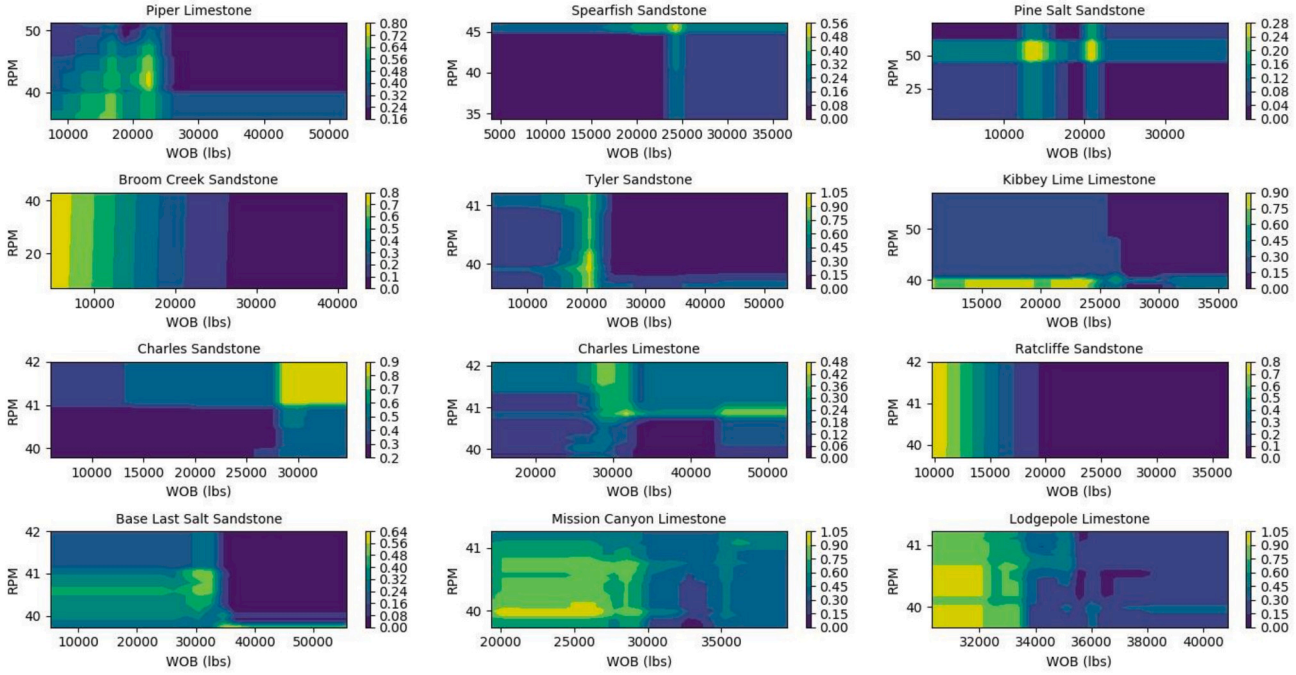

Fig. 11. Optimized probability gradient plot of RPM and WOB for the 12 formations with increasing probability of high ROP from the blue to the yellow region.

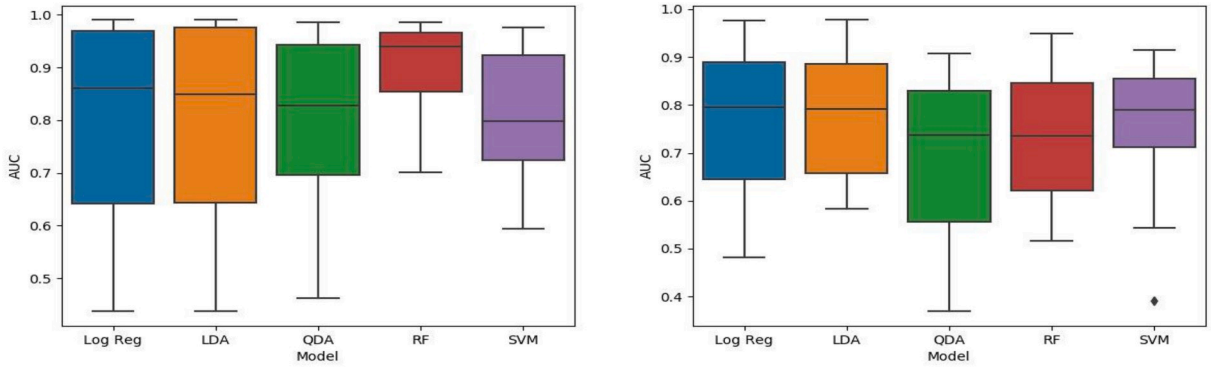

Fig. 12. Boxplot showing the performance of the five classifiers used for ROP prediction. Left figure (25th percentile threshold), right figure (50th percentile threshold).

Table 4
Tabulation showing the AUC scores of the best ROP classifiers using the 25th and 50th percentile threshold.

FORMATION	25th Percentile		50th Percentile	
	AUC	Best Classifier	AUC	Best Classifier
Piper Limestone	0.97	LDA	0.92	Log Reg
Spearfish Sandstone	0.98	RF	0.71	QDA
Pine Salt Sandstone	0.99	RF	0.78	SVM
Broom Creek Sandstone	0.99	Log Reg	0.98	LDA
Tyler Sandstone	0.71	RF	0.73	RF
Kibbey Lime Limestone	0.90	RF	0.66	Log Reg
Charles Sandstone	0.94	QDA	0.83	Log Reg
Charles Limestone	0.80	RF	0.92	Log Reg
Ratcliffe Sandstone	0.97	RF	0.92	RF
Base Last Salt Sandstone	0.99	LDA	0.83	LDA
Mission Canyon Limestone	0.87	RF	0.77	Log Reg
Lodgepole Limestone	0.70	RF	0.60	LDA

dimensional space. QDA also has a similar challenge with feature ranking. For applications in this paper, the next best classifer (random forest) was chosen as the classifier for optimzation in formations (Piper Limestone and Charles Sandstone) where SVM and QDA were the best classifiers.

For most of the formations WOB and RPM were the top two features of interest since the goal is to optimize using paramters that can be adjusted by the driller. Once a classifier has been chosen, the next step is to use the algorithm to create a gradient map of the combination of WOB and RPM values and their corresponding probabilites of yielding high ROP. Fig. 11 below shows a probability gradient plot of RPM and WOB. The color bar shows the gradation of the probablility of obtaining high ROP based on the pre-defined threshold while changing RPM and WOB. The probability increases from the its lowest value in the blue region to its highest value in the yellow region. Broom Creek and Ratcliffe Sandstone have different patterns from the other formations because the classifier used in the formations is logistic regression The aim for the ROP optimization scheme is to increase the probability that a data falls in the yellow region. This is done by adjusting the WOB and RPM

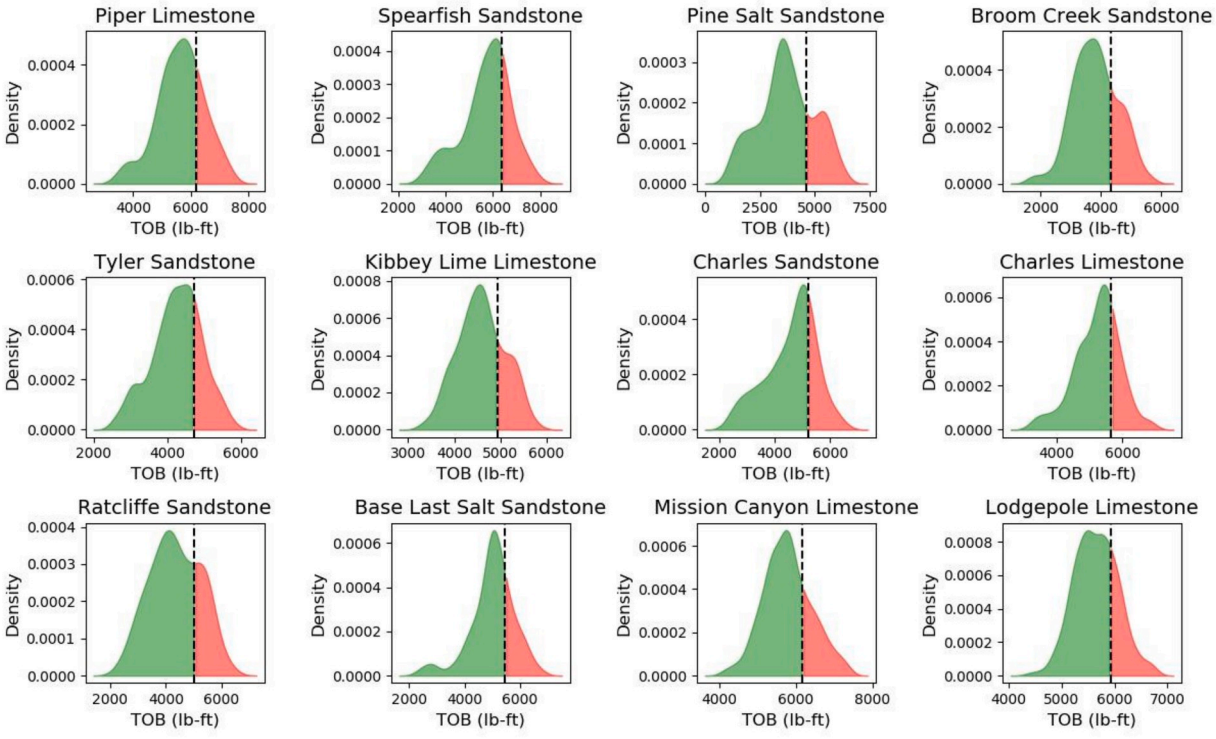

Fig. 13. Density plot of TOB for the 12 formations. The red region is the region of low TOB and the green (desired) region is the region of high TOB.

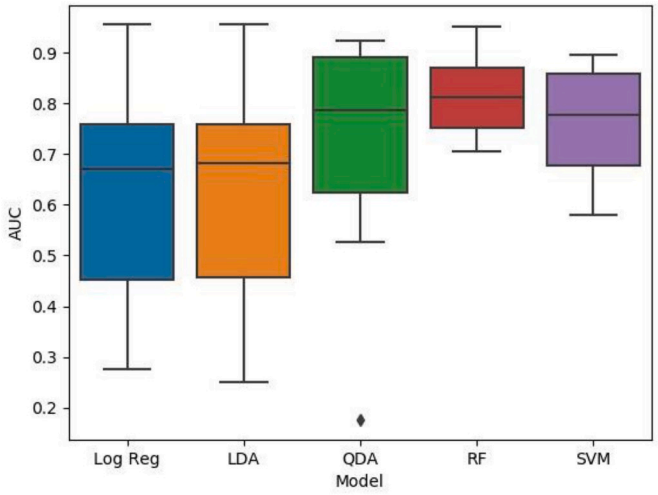

Fig. 14. Boxplot showing the AUC scores of the five classifiers used for TOB prediction.

Table 5
Tabulation showing the AUC scores from the five TOB classifiers for each of the formations.

FORMATION	AUC SCORES				
	Log Reg	LDA	QDA	RF	SVM
Piper Limestone	0.28	0.28	0.76	0.70	0.77
Spearfish Sandstone	0.69	0.71	0.62	0.75	0.70
Pine Salt Sandstone	0.49	0.50	0.91	0.95	0.87
Broom Creek Sandstone	0.95	0.95	0.92	0.76	0.78
Tyler Sandstone	0.28	0.25	0.87	0.91	0.90
Kibbey Lime Limestone	0.64	0.64	0.53	0.93	0.88
Charles Sandstone	0.33	0.33	0.18	0.78	0.62
Charles Limestone	0.86	0.86	0.82	0.73	0.58
Ratcliffe Sandstone	0.72	0.72	0.92	0.86	0.86
Base Last Salt Sandstone	0.77	0.77	0.89	0.85	0.78
Mission Canyon Limestone	0.65	0.65	0.63	0.75	0.61
Lodgepole Limestone	0.76	0.76	0.69	0.85	0.82

simultaneously or individually as needed with the goal of staying within the yellow region. The plots show that the yellow region varies in size for the different formation. Formations like Spearfish Sandstone, Pine Salt Sandstone and Piper Limestione would require extreme caution and care when drilling through in order to achieve the highest possible ROP values.

5.1.2. Parametric study on varying the threshold for ROP prediction

This section presents an experiment on the impact of using different thresholds on the performance of the classifiers. In previous analysis, the 75th percentile threshold was used however in this section the 25th and 50th percentiles are used as thresholds. Fig. 12 shows the boxplot of the AUC scores for using the 25th and 50th percentile. Comparing Figs. 9 and 12, it is evident that the performance of the classifiers changes as the threshold changes and this is because of the difference in the data points that will lie on the either side of the threshold. There were no outliers for the 25th percentile while the only outlier for the 50th percentile was from the poor performance of the SVM algorithm in the Kibbey Lime Limestone formation. Overall the random forest algorithm was the best performer across 8 out of the 12 formations for the 25th percentile threshold while for the 50th percentile logistic regression was the best classifier in 5 out of the 12 formations. Table 4 shows the best classifiers for the formations using the two percentiles and it is clear that as the percentiles change the best classifiers for each formation potentially changes. This change will also impact the nature of the optimization space as seen in Fig. 11.

Table 6
Tabulation showing the AUC scores of the best TOB classifiers for each of the formations.

FORMATION	AUC	Best Classifier
Piper Limestone	0.76	SVM
Spearfish Sandstone	0.75	RF
Pine Salt Sandstone	0.95	RF
Broom Creek Sandstone	0.95	Log Reg
Tyler Sandstone	0.91	RF
Kibbey Lime Limestone	0.93	RF
Charles Sandstone	0.78	RF
Charles Limestone	0.86	Log Reg
Ratcliffe Sandstone	0.92	QDA
Base Last Salt Sandstone	0.89	QDA
Mission Canyon Limestone	0.75	RF
Lodgepole Limestone	0.85	RF

5.2. Applications to TOB

The classification of TOB into low and high regions is slighty different from that of ROP. For ROP the goal was to obtain as high ROP as possible hence the desirable region was the region of high ROP. However for TOB, the deisred region is the region of low TOB because from the equation of MSE low TOB results in less energy used up during drilling. Just as in ROP, the predetermined threshold for each formation is the 75th percentile of the TOB values. From Fig. 13 the green (desirable) region is the region with low TOB values while the red region is the region with high TOB values.

Fig. 14 below shows the boxplot of the AUC scores for the five classifiers run for all the formations while Table 5 shows the individual values. The outlier from the QDA was due to its poor performance in the Charles Sandstone formation. On average, the random forest classifier performed the best in most of the formations as shown in Table 6.

5.2.1. TOB optimization

The TOB optimization process is similar to that of ROP. Formations with SVM (Piper Limestone) and QDA (Ratcliffe Sandstone and Base Last Salt Sandstone) classifiers were replaced with the next best performing classifier (random forest) inorder to be able to rank the features. Fig. 15 shows the permutation importance for random forest and feature importance from logistic regression and it can be seen again that RPM and WOB are the two higly ranked features. Fig. 16 shows the probability gradient map for obtaining a low TOB based on the defined threshold. The yellow (desired) region gives the combination of WOB and RPM values that maximize the probability of obtaining a low TOB. Formations like Spearfish Sandstone, Piper Limestone, Base Last Sanstone and Ratcliffe Sanstone have wider regions of high probability giving the driller more leaway to adjust the RPM and WOB to obtain low TOB.

This probability heat map potentially provides several applications. Firstly, it gives the driller a great tool to understand and make quick informed decisions on RPM and WOB values. Secondly, it shows the driller the currrent drilling state and the possible combinations of RPM and WOB that can be used to increase the probability of obtaining high ROP or low TOB as the case may be. The driller would then choose values which are attainable and have the highest probability of the desired drilling performance.

5.2.2. Parametric study on varying the threshold for TOB prediction

Fig. 17 shows the AUC scores using the 25th and 50th percentile threshold and comparing that with Fig. 14, it is seen that the performance of the classifiers again changes as the threshold changes. The best classifier for each formation also changed as the threshold changed as seen in Table 7 and this will also impact the optimization space during optimization. For the 50th percentile threshold, no classifier outperformed the others and they performed best in Ratcliffe Sandstone which is the outlier seen in Fig. 17 (right). For the 25th percentile,

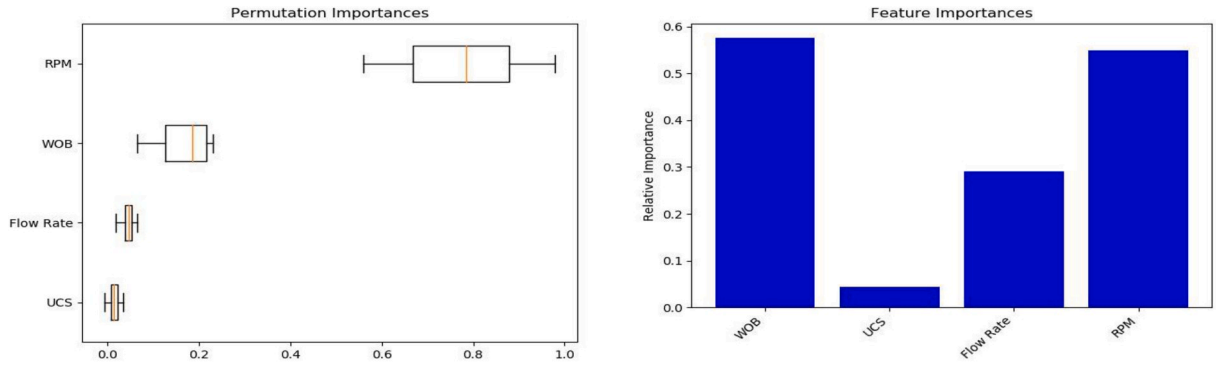

Fig. 15. Permutation importance of random forest for ratcliffe sandstone (left); feature importance of logistic regression for charles limestone (right).

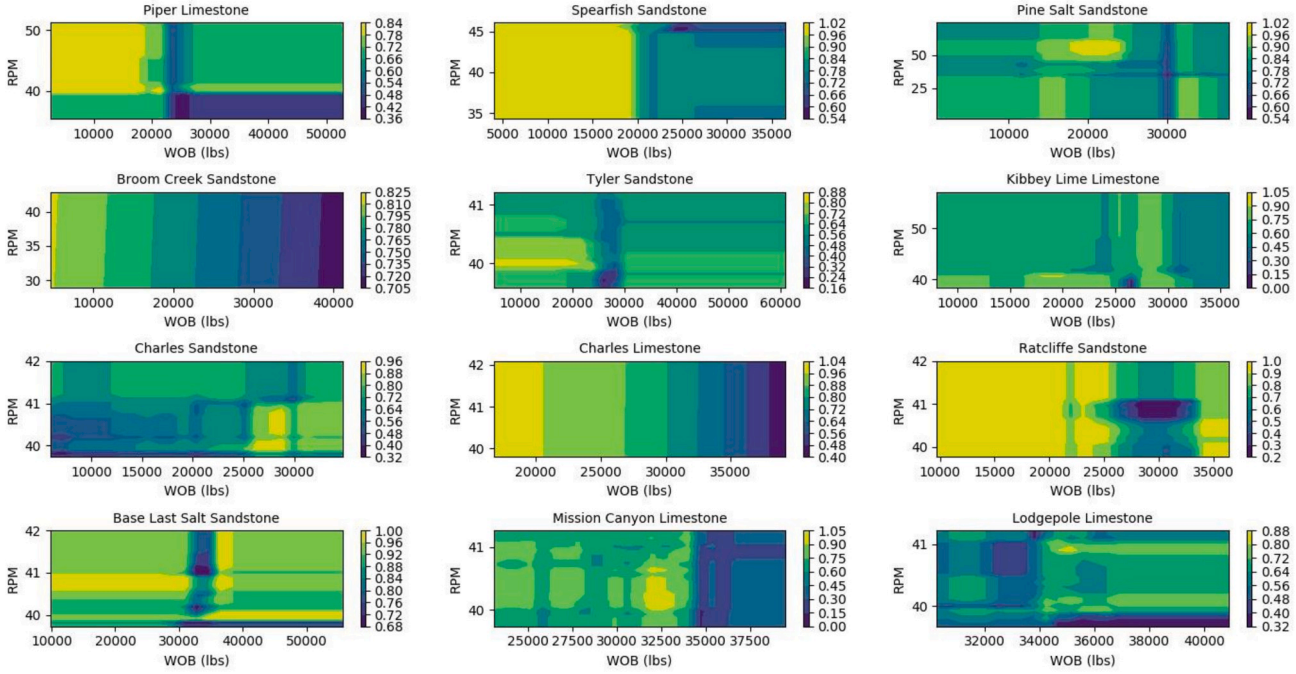

Fig. 16. Optimized probability gradient plot of RPM and WOB for the 12 formations with increasing probability of low TOB from the blue to the yellow region.

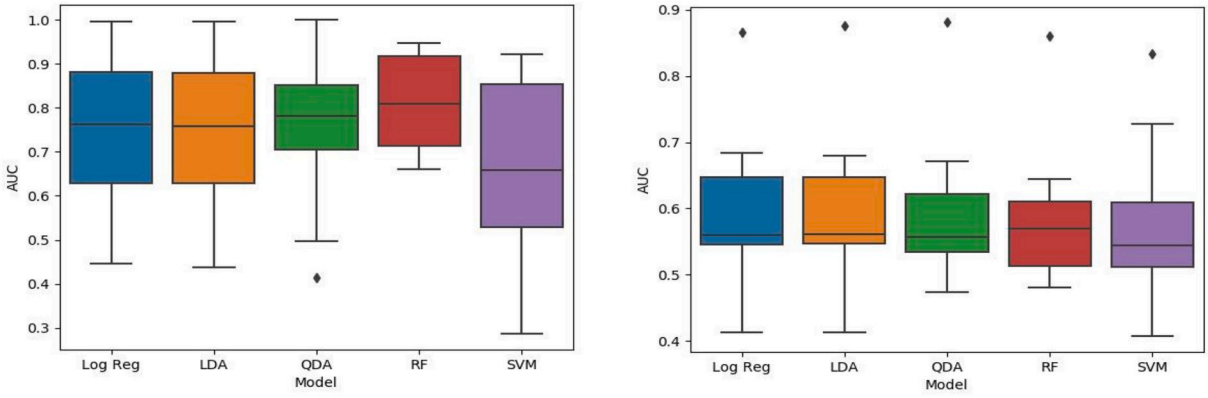

Fig. 17. Boxplot showing the performance of the five classifiers used for TOB prediction. Left figure (25th percentile threshold)), right figure (50th percentile threshold).

Table 7
Tabulation showing the AUC scores of the best TOB classifiers using the 25th and 50th percentile threshold.

FORMATION	25th Percentile		50th Percentile	
	AUC	Best Classifier	AUC	Best Classifier
Piper Limestone	0.73	LDA	0.55	QDA
Spearfish Sandstone	0.92	RF	0.68	Log Reg
Pine Salt Sandstone	0.88	RF	0.73	SVM
Broom Creek Sandstone	0.99	QDA	0.55	LDA
Tyler Sandstone	1.00	QDA	0.60	RF
Kibbey Lime Limestone	0.77	RF	0.55	Log Reg
Charles Sandstone	0.80	QDA	0.51	RF
Charles Limestone	0.95	RF	0.62	QDA
Ratcliffe Sandstone	0.92	SVM	0.88	QDA
Base Last Salt Sandstone	0.67	RF	0.64	Log Reg
Mission Canyon Limestone	0.85	RF	0.59	SVM
Lodgepole Limestone	0.72	RF	0.60	SVM

random forest was the best classifier for 7 out of the 12 formations.

6. Conclusion

This paper introduced a novel approach to ROP and TOB prediction using machine learning classification algorithms. ROP and TOB were modeled as a function of drilling paramters – RPM, WOB, flowrate and UCS. Using a user-defined threshold, the algorithms classify the data set into regions of high and low ROP and TOB respectively. Five classification algorithms namely-logistic regression, LDA, QDA, SVM and random forest were used to develop the models. The performance of the algorithms was evaluated using the AUC performance metric and the best algorithm was the one that had the highest AUC score.

Results showed that the appropriate approach for choosing the best classifier was to determine the best classifier for each formation rather than an overall best classifier for all the formations. Random forest classifier which is good for modeling non-linear problems outperformed the other classifiers in a number cases. For practical applications, varying the threshold was shown to impact the performance of the classifiers for each formation indicating that the threshold can be adjusted as the user deems appropriate. This ensures that results are unique for each threshold and dataset. Probablity gradient heat maps that show the probability that a given RPM and WOB setting will be in the region of high ROP and low TOB were also shown. Using this heat map, a driller can sucessfully adjust RPM and WOB (within operational limits) to fall in the region of high probability.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Mayowa Oyedere: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - original draft, Writing - review & editing. **Ken Gray:** Supervision.

Acknowledgements

The authors would like to thank sponsors of the Wider Windows Industrial Affiliate Program at The University of Texas at Austin: British Petroleum, Chevron, ConocoPhillips, Halliburton, Marathon Oil Company, National Oilwell Varco, Occidental Oil and Gas, and Shell for their financial support and technical inputs in this work.

Acronyms

RPM	Rotation per minute
WOB	Weight on bit
UCS	Unconfined Compressive Strength
PDC	Polycrystalline diamond compact
LDA	Linear Discriminant Analysis
QDA	Quadratic Discriminant Analysis
SVM	Support Vector Machines
ROP	Rate of penetration
TOB	Torque on bit
MSE	Mechanical Specific Energy
ROC	Receiver operating characteristics
AUC	Area under curve

References

Alum, M.A., Egbon, F., 2011. Semi-analytical models on the effect of drilling fluid properties on rate of penetration (ROP). In: Nigerian Annual International Conference and Exhibition. Abuja, Nigeria, July 30 – August 3.
Amer, M.M., Sattar, A., El-Sayed, A.H., 2017. An ROP predictive model in nile delta area using artificial neural networks. In: SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition. Dammam, Saudi Arabia, April 24-27.
Bilgesu, H.I., et al., 1997. A new approach for the prediction of rate of penetration (ROP) values. In: SPE Eastern Regional Meeting. Society of Petroleum Engineers.
Bingham, M.G., 1965. A new approach to interpreting rock drillability. Technical Manual Reprint. Oil Gas J. 93.
Bosner, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. Pittsburg, PA, USA, July 27-29.
Bourgoyne Jr., A.T., Young Jr., F.S., 1974. A multiple regression approach to optimal drilling and abnormal pressure detection. Soc. Petrol. Eng. J. 14 (4), 371–384.
Breimann, L., 2001. Random Forests. Machine Learning 45 (1), 5–32.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2011. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357, 2002.
Cortes, C., Vapnik, V., 1995. Support vector networks. Mach. Learn. 20 (3), 273–297.
Dunlop, J., Isangulov, R., Aldred, W., Arismendi Sanchez, H., Sanchez Flores, J.L., Alarcon Herdoiza, J., Belaskie, J., Luppens, J.C., 2011. Increased rate of penetration through automation. SPE/IADC 139897.
Eckel, J.R., 1968. Microbit studies of the effect of fluid properties and hydraulics on drilling rate, II. In: 43rd Annual Fall Meeting of the Society of Petroleum Engineers of AIME. Houston, TX, USA, September 29- October 2.
Eckel, J.R., Nolley, J.P., 1949. An Analysis of Hydraulic Factors Affecting the Rate of Penetration of Drag-type Rotary Bits. American Petroleum Institute.
Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27 (8), 861–874. ISSN 0167-8655.
Gandelman, R.A., 2012. Prediction of ROP and Real Time Optimization of Operational Parameters while Drilling Offshore Oil-Wells. M.S. Thesis. Universidade Federal do Rio de Janeiro.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT press.
Hareland, G., Rampersad, P.R., 1994. Drag-bit model including wear. In: SPE Latin America/Caribbean Petroleum Engineering Conference. Society of Petroleum Engineers.
Hegde, C.M., Daigle, H., Millwater, H., Gray, K.E., 2017. Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. J. Petrol. Sci. Eng. 159, 295–306. ISSN 0920-4105.
Hegde, C.M., Gray, K.E., 2018. Evaluation of coupled machine learning models for drilling optimization. J. Nat. Gas Sci. Eng. 56, 397–407, 1875–5100-2018.
Hegde, C.M., Gray, K.E., 2017. Use of machine learning and data Analytics to increase drilling efficiency for nearby wells. J. Nat. Gas Sci. Eng. 40, 327–335, 1875–5100.
Hegde, C.M., Millwater, H., Gray, K.E., 2019. Classification of drilling stick slip severity using machine learning. J. Petrol. Sci. Eng. 179, 1023–1036, 2019-0920–4105.
Hegde, C.M., Wallace, S.P., Gray, K.E., 2015a. Using trees, bagging and random forests to predict rate of penetration during drilling. In: Presented at SPE Middle East Intelligent Oil & Gas Conference & Exhibition, Abu Dhabi, United Arab Emirates, 15-16 September. SPE-176792.
Hegde, C.M., Wallace, S.P., Gray, K.E., 2015b. Use of regression and bootstrapping in drilling: inference and prediction. In: Presented at SPE Middle East Intelligent Oil & Gas Conference & Exhibition, Abu Dhabi, United Arab Emirates, 15-16 September. SPE-176791.
Ho, T.K., 1995. Random decision forestes. In: IEEE Proceedings of the Third International Conference on Document Analysis and Recognition, Montreal, Canada, August 14-16.
Jahanbakhshi, R., Keshavarzi, R., Jafarnezhad, A., 2012. Real-time prediction of rate of penetration during drilling operation in oil and gas wells. In: 46th US Rock Mechanics/Geomechanics Symposium. American Rock Mechanics Association.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, vol. 112. Springer.
Logan, W.D., 2015. Engineered Shale Completions Based on Common Drilling Data. Society of Petroleum Engineers.

Maurer, W.C., 1962. The “perfect-cleaning” theory of rotary drilling. J. Petrol. Technol. 1270–1274.
Motahhari, H.R., Hareland, G., James, J.A., 2010. Improved drilling efficiency technique using integrated PDM and PDC bit parameters. J. Can. Petrol. Technol. 49 (10), 45–52.
Sheppard, M.C., Wick, C., Burgess, T., 1987. Designing Well Paths to Reduce Drag and Torque. Society of Petroleum Engineers.
Teale, R., 1965. The concept of specific energy in rock drilling. Int. J. Rock Mech. Min. Sci. 2 (1), 57–73.
Theloy, C., 2014. Integration of Geological and Technological Factors Influencing Production in the Bakken Play, Williston Basin. Colorado School of Mines.
Trichel, D.K., Isbell, M., Brown, B., Flash, M., McRay, M., Nieto, J., Fonseca, I., 2016. Using Wired Drill Pipe, High-Speed Downhole Data, and Closed Loop Drilling Automation Technology to Drive Performance Improvement across Multiple Wells in the Bakken. Society of Petroleum Engineers.
Wallace, S.P., Hegde, C.M., Gray, K.E., 2015. System for real time drilling performance optimization and automation based on statistical learning methods. In: Presented at SPE Middle East Intelligent Oil & Gas Conference & Exhibition, Abu Dhabi, United Arab Emirates, 15-16 September. Society of Petroleum Engineers. SPE 176804.