



Data-driven approaches tests on a laboratory drilling system

Erik Andreas Løken¹ · Jens Løkkevik¹ · Dan Sui¹

Received: 20 November 2019 / Accepted: 16 March 2020
© The Author(s) 2020

Abstract

In recent years, considerable resources have been invested to exploit vast amounts of data that get collected during exploration, drilling and production of oil and gas. Data-related digital technologies potentially become a game changer for the industry in terms of reduced costs through increasing operational efficiency and avoiding accidents, improved health, safety and environment through strengthening situational awareness and so on. Machine learning, an application of artificial intelligence to offer systems/processes self-learning and self-driving ability, has been around for recent decades. In the last five to ten years, the increased computational powers along with heavily digitized control and monitoring systems have made machine learning algorithms more available, powerful and accurate. Considering the state-of-art technologies that exist today and the significant resources that are being invested into the technologies of tomorrow, the idea of intelligent and automated drilling systems to select best decisions or provide good recommendations based on the information available becomes closer to a reality. This study shows the results of our research activity carried out on the topic of drilling automation and digitalization. The main objective is to test the developed machine learning algorithms of formation classification and drilling operations identification on a laboratory drilling system. In this paper, an algorithm to develop data-driven models based on the laboratory data collocated in many scenarios (for instance, drilling different formation samples with varying drilling operational parameters and running different operations) is presented. Moreover, a testing algorithm based on data-driven models for new formation detection and confirmation is proposed. In the case study, results on multiple experiments conducted to test and validate the developed machine learning methods have been illustrated and discussed.

Keywords Drilling rig · Drilling automation · Drilling data · Machine learning · Classification

Introduction

Background

Recently, the concept of drilling digitalization and automation has advanced from primarily being automation of rig floor equipment to novel solutions that rapidly can be deployed to the rig environment and assist drillers in a variety of operations. Aside from providing an early warning to drillers, intelligent systems aim to improve efficiency and reduce financial costs through continuous monitoring and interaction with drillers. Smart drilling systems could also be anticipated to suggest operating parameters to drillers through correlating real-time drilling data with vast amounts

of historic data stored in a virtual environment. Digital systems target solutions and new technologies to even exert full control of all rig equipment if permissible (top drive, draw works, mud pumps, elevator, rough neck and so on), leaving only major decision points to be determined by drillers. The later automation level described above is most likely still several years away from being deployable to fields. A timeline that highlights artificial intelligence applications in drilling practices is given in Bello et al. (2015). Short-term advances in drilling automation and digitalization lie in developing simple, yet robust tools for drillers to strengthen the understanding of operations during critical phases.

Related research problems

The past decade has seen rapid growth in the ability of networked and mobile computing systems to gather and transport vast amounts of data, or Big Data (Mayer-Schneider and Cukier 2013). Machine learning (Shalev-Shwartz and

✉ Dan Sui
dan.sui@uis.no

¹ Energy and Petroleum Engineering Department, University of Stavanger, Stavanger, Norway

Ben-David 2014; Kelleher et al. 2015) becomes a more and more powerful tool to solve problems of obtaining useful insights, predictions and decisions (Jordan and Mitchell 2015). Given big amounts of wells that get drilled every year, approaches for data interpretation, performance prediction and optimization, and decisions making based on historical data through machine learning approaches are important areas of research.

Literature review

In recent years, many research works and studies have proposed to develop and implement machine learning approaches in different drilling applications, aiming to aid drilling engineers to detect drilling incidents, predict drilling parameters, analyze drilling behaviors and advise drilling actions. For instance, several works to use machine learning classification approaches to identify drilling-related parameters and drilling incidents have been proposed. In Sun et al. (2019), a machine learning approach was proposed to identify the lithology while drilling that provides valuable information for drilling geosteering of oilfield development. In Klyuchnikov et al. (2019), machine learning classification methods were used to identify rock types around the drill bit. Hegde et al. (2019) have proposed to use machine learning to identify the drilling stick slip severity to help vibration mitigation during rate of penetration optimization. The most recent work (Zaytsev et al. 2020) used machine learning to detect drilling incidents for directional drilling.

Besides classification approaches, machine learning has huge capacity for predictions and regressions. In Hegde et al. (2017), different rate of penetration (ROP) models developed via physics-based and machine learning approaches have been evaluated through uncertainty analysis. Similar comparison work has been done by Soares and Gray (2019), where machine learning models were observed to reduce test errors much more effectively than analytical models with incremental data availability. A detailed literature review on machine learning methods for ROP prediction and optimization has been given in Barbosa et al. (2019). In Spesivtsev et al. (2018), a bottom hole pressure prediction model has been used for multi-phase wellbore flows via the machine learning approach. In Kanin et al. (2019), laboratory data has been used to develop machine learning model for pressure prediction. Artificial neural network model for predicting the density of oil-based drilling fluids in high-temperature and high-pressure wells has been presented in Agwu et al. (2019). In AlAzani et al. (2019), cuttings concentration for horizontal and deviated wells was predicted using machine learning. In addition, machine learning approaches were used in many other applications, for instance, mud loss estimation during lost circulation (DunnNorman et al. 2018), permeability prediction (Arigbe et al. 2018), titration-based

asphaltene precipitation (Gholami et al. 2015), oil/gas ratio for volatile oil and gas condensate reservoirs (Fattah and Khamis 2018) and hydraulic fracturing prediction (Makhotin et al. 2019). In Al-Mudhafar (2017), both machine learning classification and regression approaches were used for lithofacies classification and permeability prediction.

Our novelty and contributions

In this paper, the data-driven models developed to classify different rock formations are presented. The models have been developed, got trained and validated using time-based experimental data collected in a laboratory environment on a test bench. Furthermore, unsupervised machine learning models (DeepAI 2019; Roman 2019; Michael 2019) have been developed to classify drilling operations such as tripping and rotating on bottom. Learning outcome from the study is to show how to develop machine learning algorithms from the data collection phase to real-time algorithm implementation phase. Laboratory testing and evaluation is an essential part of promoting the adaptation of digital technologies. Such study is a useful and cost-effective solution for testing data-driven approaches before expensive full-scale testing and development.

Drilling rig

Figures 1 and 2 show the laboratory drilling rig and its sketch, respectively. The detailed information about the rig structure, its software and control system was given in Løken and Løkkevik (2019), Løken et al. (2018, 2019). The top drive is controlled by a driver to set the rotary speed (RPM) and maximum torque. The construction is equipped with a complete hoisting system consisting of actuators, stepper motors and brakes. The top plate is where the top drive and other components are mounted. It is positioned between three tri-axial load cells connected to the actuators to provide enough lifting force and for proper stabilization. The circulation system is a simple system consisting of two pumps. Each pump has a maximum flow rate of 19 L/min and the maximum working pressure at 3.1 bar.

The rig includes the following sophisticated functions and capabilities (Khadisov et al. 2019):

- Conducting vertical/deviated well drilling tests in manual/autonomous mode;
- Having a data management system for data processing, analysis, visualization and storage;
- Being instrumented with high-speed and reliable down-hole and surface sensors;
- Having an adaptive advisory system for optimization.

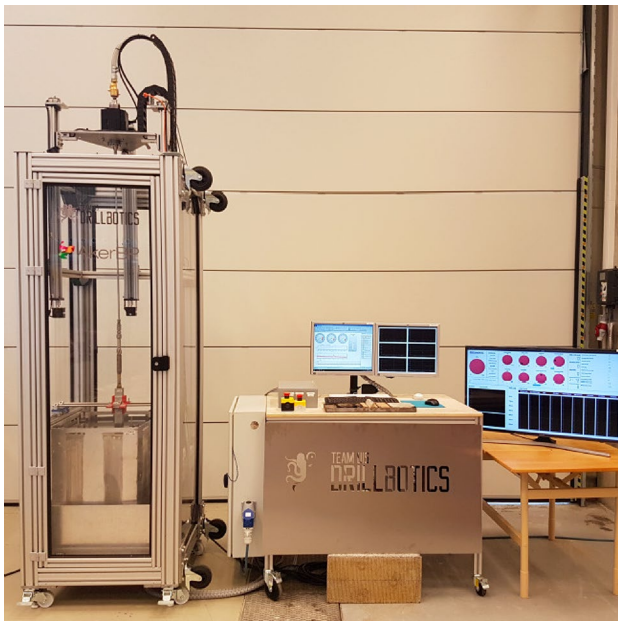


Fig. 1 Laboratory rig test platform

Having such drilling system allows us to conduct multiple experiments in a laboratory scale and create possibilities to test and validate the developed data-driven approaches.

Model development

Drilling data

Data pre-processing

In order to develop accurate models, the major importance lies in ensuring that the data with high quality. According to Good and Hardin (2006), the following steps should be carried out in order to improve the data quality:

- Review quality assurance reports,
- Describe the dataset with statistics,
- Remove duplicate values,
- Verify physical units of measured data,
- Remove missing data,
- Remove outliers.

Before a data-driven model is developed, cleaning the dataset is essential (van der Aalst 2016; James 2016). Data cleaning includes several steps, but not limited to: outlier removal, removing invalid data, removing missing data, duplicates and so on.

Invalid data If a significant part of the dataset falls outside of a validity range, one approach is to replace the values with NaN (not a number) and later remove the complete row of observations. Measurements are kept from the other variables (sensors) in the dataset, but discard measurements in a single variable where invalid data is present. Invalid data can cause issues when developing data-driven algorithms. For the drilling data captured using the laboratory drilling system, invalid data would typically be data measured outside of the specific sensors measurement range, Table 1.

Missing data A number of reasons lead the data missing in a dataset. One common cause is when different sensors get sampled with varying sampling frequencies, for instance, 10 Hz for one sensor and 20 Hz for another. Second common cause could be hardware (electrical) failure, where the signal is lost for a short duration of time. Third cause could be that the data is held up in the buffer where the computer stores the data short-term before it gets used.

To handle missing data, the common interpolation techniques (linear, quadratic, cubic or polynomial) can be used, see (Al Bakri et al 2014).

Outlier removal Outliers are ones that are situated away from the main observation window. An important factor to consider before removing outliers is to find out whether they consist of relevant information or are the result of noises. In some datasets, for example, when dealing with kick detection or stuck pipe detection, the important information could be apparent in the outlying points. In our research, several techniques have been evaluated for optimal outlier removal. The interquartile range (IQR) method has been identified as the most optimal when dealing with outliers, see the detailed discussions in Holdaway (2014).

Normalization and standardization Considering drilling data where the variables or features originate from different sources or sensors, an important task is to scale all data to a common unit range. Ideally data that is normal distributed gets represented as values from 0 to 1. This can be achieved through performing a linear feature scaling (LFS), by considering the minimum and maximum value of each variable (James 2016). For a dataset, $X = \{x_1, x_2, \dots, x_n\}$, the normalized data point becomes

$$x_i^N = \frac{x_i - \min(X)}{\max(X) - \min(X)}. \quad (1)$$

While the LFS provides a sensible method to scale data that has no predefined range, this technique could still cause a challenge if a significant outlier is present. The outlier, which could be either very large or very small, would then cause the rest of data to be skewed either toward 0 or 1,

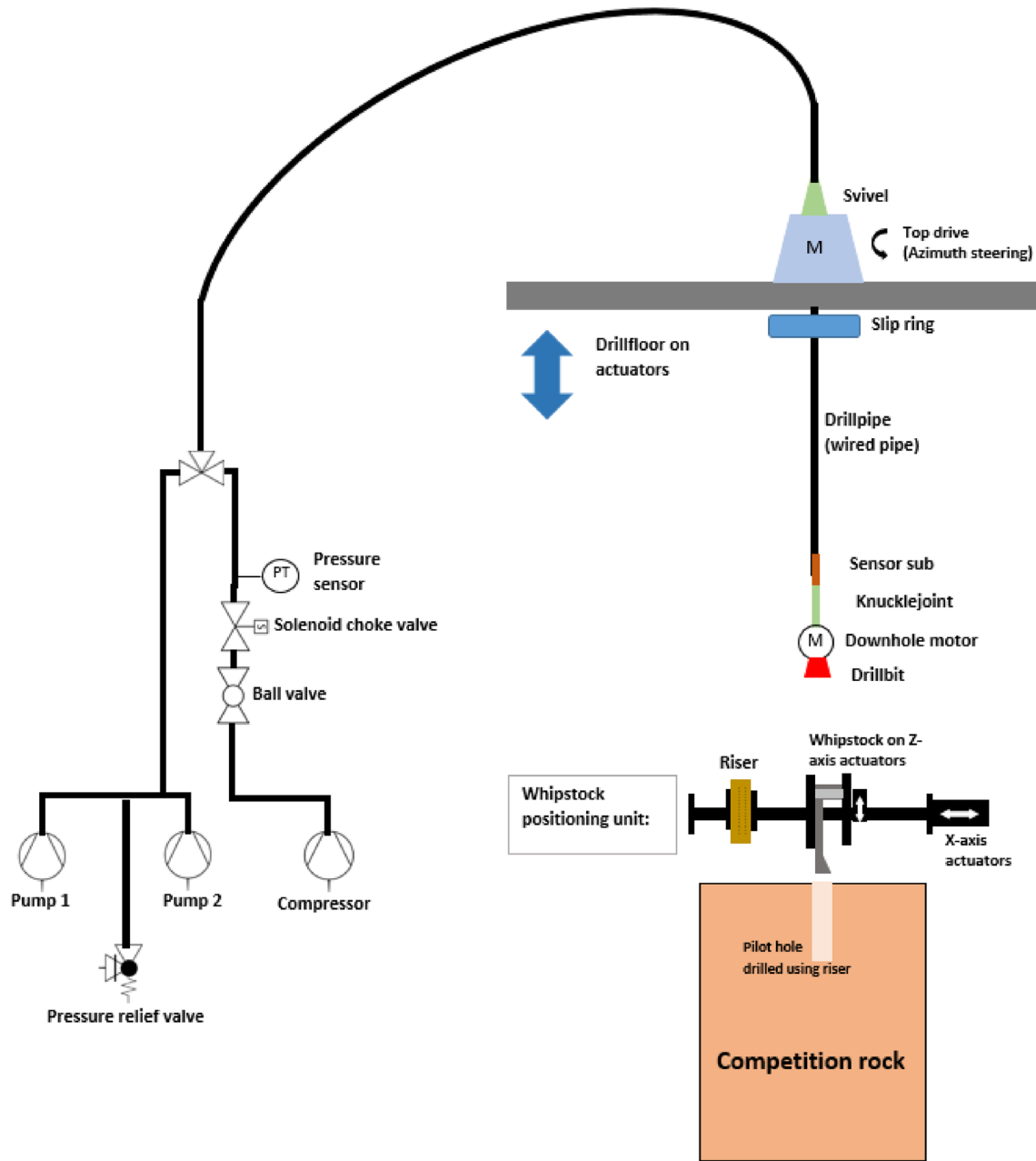


Fig. 2 Schematics of the rig construction

Table 1 Sensor measurement range

Measurement	Unit	Min range	Max range
LC (load cell)	(g)	- 10000	10000
RPM	(rev/min)	0	1500
Torque	(Nm)	0	8.59
Depth	(mm)	0	1000
WOB (weight on bit)	(kg)	- 30	30
Pressure	(bar)	0	10

see (James 2016). Standardization is the other commonly used technique. It refers to a process of subtracting the mean value of the set of values for a variable from each measurement and dividing by the standard deviation of the set of values, see (James 2016). The standardized data point is calculated as

$$x_i^S = \frac{x_i - \mu}{\sigma} = \frac{x_i - \frac{\sum_{j=1}^n x_j}{n}}{\sqrt{\frac{\sum_{j=1}^n (x_j - \mu)^2}{n-1}}}, \quad (2)$$

Table 2 Data is collected from 6 formations

Type	RPM = 400	RPM = 600	RPM = 800	RPM = 1000
Cement	WOB = 3,6	WOB = 3,6,10	WOB = 10	–
Salt	WOB = 3,6	WOB = 3,6,10	WOB = 8,10	–
Sandstone	WOB = 3,6	WOB = 3,6,10	WOB = 3,6,10	WOB = 10
Chalk	WOB = 5	WOB = 5	WOB = 10	WOB = 10
Shale	WOB = 3,6	WOB = 3,6,10	WOB = 3,6,10	WOB = 10
Granite	WOB = 3,6	WOB = 3,6,10	WOB = 3,6,10	WOB = 10

WOB = 3,6 means that WOB is set 3 kg and 6 kg, respectively

**Fig. 3** Collection of different rock specimen drilled to gather experimental drilling data

where σ represents the standard deviation and μ is the true mean value of the set. For our case, the measurements for each variable relative to the threshold are considered. For instance, for the weight on bit data, the load cells are configured to measure - 300 N (compression) to 300 N (tension) of force, Table 1. Therefore, the first step of processing the WOB data would remove all measurements where the data is invalid, leaving only those measurements within the (- 300 N, 300 N) range. In terms of normalizing data, Eq. (1) is used to calculate normalized data based on the range of sensor measurements.

Laboratory data

Formation classification Six different rocks were drilled with different drilling parameter combinations given in Table 2. The rock samples are shown in Fig. 3.

The process of concatenating all experiments and labeling them is repeated for rock formations 1 through 6. The pool of data consists of the relatively big number of observations for each rock formation specimen, Table 3.

The difference in number of observations per rock specimen is based on the availability of different rock specimen

Table 3 Data concatenation for rock classification

Rock specimen	Label	Number of samples
Cement	1	8.263×10^6
Chalk	2	1.864×10^6
Granite	3	1.473×10^7
Sandstone	4	1.020×10^7
Salt	5	7.149×10^6
Shale	6	1.375×10^7

Table 4 Data concatenation for rig operation classification

Rig operation	Label	Number of samples
POOH	1	8.153×10^5
RIH	2	9.633×10^5
ROnB	3	1.77×10^6

to drill, as well as the drilling speed. (A 150-mm-thick chalk specimen is drilled in less than a minute for reference; however, a well drilled in granite rock would require several hours to drill.)

Rig operations classification A total of nine experiments were conducted to collect data on three rig operations in an attempt to develop models to distinguish between drilling and non-production time (NPT) activities such as tripping. These three operations are tripping up (POOH), tripping down (RIH) and rotating on bottom (ROnB). The experiments contain data for each operation, either with or without bit rotation, circulation or a combination of both. The data is labeled so that each operation is represented by Table 4.

Feature engineering

Feature selection

Natural features when classifying rock formations and rig operations are: LC(1/2/3) denote the hook load strain gauge measurements from load cells; RPM (rotary speed of drill string); torque (surface torque); depth (measured depth); WOB and pump pressure.

Several drilling-related features have been created from the above natural features, as shown in Table 5 (More information of drilling parameters in Table 5 is given in “Appendix.”)

Several statistical features have also been created for the cases of rock classification. They describe the average value, standard deviation, median, maximum, minimum, P25, P50 and P75 value for each natural feature, like pressure, weight on bit or torque.

Similar to the DOC and BA, some features considering the data transformation are created to add additional interactions of the drilling parameters, Table 6. The basis for calculating the interactions of natural features is a data analysis experiment conducted to investigate whether the feature importance of these natural interactions is higher than the natural features.

Feature extraction

Principal component analysis (PCA) (Otterbach 2019) is a method of analyzing small or large datasets. It extracts the numerical values from the variables and calculates a set of new orthogonal variables called principal components. The benefit of using this method is to extract only the required information to explain the variance in the data and thus reduce the size of the dataset by keeping only the valuable information required for prediction and classification. After creating the principal components, the quality of the model can be evaluated by cross-validation (Herv and Williams 2010). The following workflow as shown in Fig. 4 is to extract the features that have provided the highest score in the feature importance evaluation.

Table 5 Engineering features, where MSE is mechanical specific energy; DOC is depth of cut and BA is bit aggressiveness

Features	Description (“Appendix”)
Mean depth	Average height sensor measurement
ROP	Change of bit depth per time unit
MSE	Mechanical energy used to remove a unit rock volume
DOC	Distance that drill bit cuts into formation per revolution
BA	How high the DOC is
D-exponent	Being used to correct drilling rate

Table 6 Artificial features

Feature	Parameters	Feature	Parameters
TF1	RPM * WOB	TF2	RPM * WOB/ROP
TF3	RPM ²	TF4	WOB ²
TF5	WOB/RPM		

Feature extraction methods give a good indication of the importance of features from data science perspective. When working with drilling data, manual feature selection and optimization should be performed, besides these standard methods. Some features that are considered important from drilling engineers’ perspective to describe a particular phenomenon (such as bit-rock interaction for rock formation classification) should get selected rather than blindly trusting the score from an algorithm. A high accuracy score does not guarantee that the model can correctly classify the observations in a new dataset if the selected features are not directly applicable.

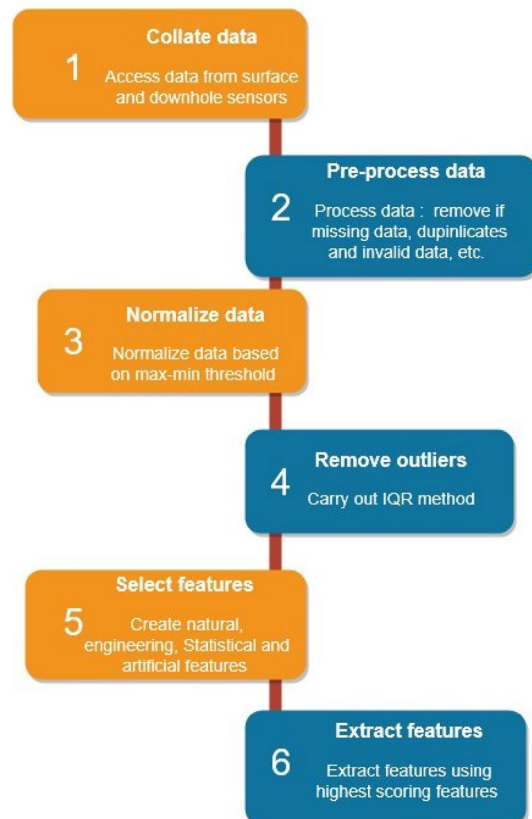


Fig. 4 Flowchart of data flow and processes performed for real-time classification

Machine learning models

The different classifiers used to develop the models in “Discussions” section have all been taken from the Scikit Learn library (Scikit 2019). These are: multilayer perceptron (MLP) classifier (Wilson 1994), decision tree (DT) classifier (Kamiski et al. 2017), support vector machine (SVM) classifier (Christiani and have-Taylor 2000), random forest (RF) classifier (Ho 1998), gradient boosting (GB) classifier (Elith 2018), K-neighbors (K-NN) classifier (Altman 1992), K-means (Hartigan and Hartigan 1979), density-based spatial clustering of applications with noise (DBSCAN) (Fan et al 2011) and tree-based pipeline optimization tool (TPOT) classifier (Randal 2019). The flowchart for model development is shown in Fig. 5.

Results

Cases

A sensitivity study is conducted to evaluate which features result in the most optimal models for the drilling cases given below. Since the most optimal features have only been presented for rock formation classification, each classification task will also be presented with the recommended feature priority. Regardless of the feature priority from the algorithm, manual selection is performed to ensure that only those features that are regarded as applicable are used. The cases in this study are shown below:

- Laboratory rock formation classification—4 cases (Case 1–Case 4)
- Laboratory rig operation classification—3 cases (Case 5–Case 7)

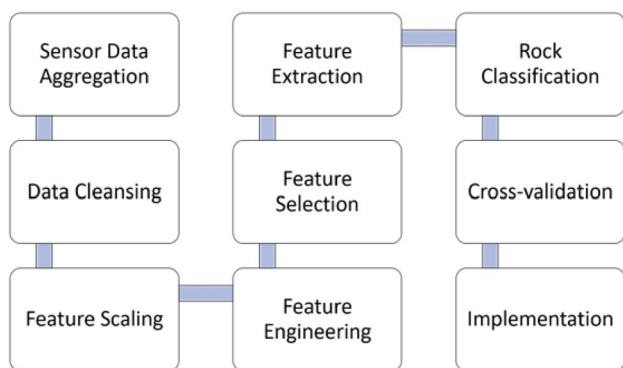


Fig. 5 Flowchart of model development

Table 7 shows the cases with different machine learning methods. Table 8 shows the cases with different features used to the models.

Evaluation (Cases 1–4)

For the support vector machine, the ability to extract linear combinations of features is high, but the model is both weak with regard to computational scalability and natural handling of mixed-type data. For Cases 3 and 4 however, when the number of rock types has been reduced to three, an increase by approximately 10% can be noted. The same applies to the multilayer perceptron model, which appears to perform much better when the type of samples has been reduced to three. With regard to K-NN, the model appears to score better when the number of features is low.

Figure 6 shows the output from Case 1, where the best predictions are achieved with the decision tree, gradient boosting and random forest models. Figure 7 shows the output from Case 2. While the decision tree, gradient boosting and random forest models continue to deliver the best predictions, all models except the multilayer perceptron and support vector machine now deliver almost identical predictions.

While the above experiments were conducted for six different formations, several of the formations are similar in drillability such as sandstone and cement. For this reason, the models from Case 3 have been trained on class 3: granite, 4: sandstone and 5: salt, respectively, representing a hard-drilling formation, a medium to hard-drilling formation and a soft formation. From Table 9 and Fig. 8, it is seen that except from K-NN model, all other models perform well. Finally, the same dataset is run through the models in Case 4 that have been developed with the six highest scoring features. From the results, Fig. 9, all models except for MLP and K-NN perform well.

Considering all models, it is our recommendation to use decision tree classifiers for rock formation classification on the laboratory drilling rig. It can be observed that the

Table 7 Cases with methods

Case	Classification	Models
1	6 formations	MLP, DT, SVM, GB, RF, K-NN, TPOT
2	6 formations	MLP, DT, SVM, GB, RF, K-NN, TPOT
3	3 formations	MLP, DT, SVM, GB, RF, K-NN, TPOT
4	3 formations	MLP, DT, SVM, GB, RF, K-NN, TPOT
5	Rig operations	K-NN
6	Rig operations	K-means, DBSCAN
7	Rig operations	K-means, DBSCAN

Table 8 Cases with features

Case	Features	Labels
1	LC1, LC2, LC3, RPM, Torque, WOB, ROP, TF1, TF2, TF3, TF4, TF5, MSE, DOC, BA, D-exponent	1, 2, ..., 6
2	RPM, Torque, ROP, TF3, DOC, BA (6 highest scores)	1, 2, ..., 6
3	LC1, LC2, LC3, RPM, Torque, WOB, ROP, TF1, TF2, TF3, TF4, TF5, MSE, DOC, BA, D-exponent	3, 4, 5
4	RPM, Torque, ROP, TF3, DOC, BA (6 highest scores)	3, 4, 5
5	LC1, LC2, LC3, RPM, Torque, WOB, ROP, TF1, TF2, TF3, TF4, TF5, MSE	1, 2, 3
6	ROP, WOB	1, 2, 3
7	ROP _{min} , ROP _{median}	1, 2, 3

Table 9 Model accuracy

Case	MLP (%)	DT (%)	SVM (%)	RF (%)	GB (%)	K-NN (%)	TPOT (%)
1	95.53	99.66	89.66	98.50	99.26	85.37	98.29
2	93.10	99.71	85.12	98.68	99.18	92.34	98.02
3	99.36	99.93	98.83	99.53	99.85	89.92	99.32
4	98.60	99.94	97.50	99.66	99.93	97.74	99.83

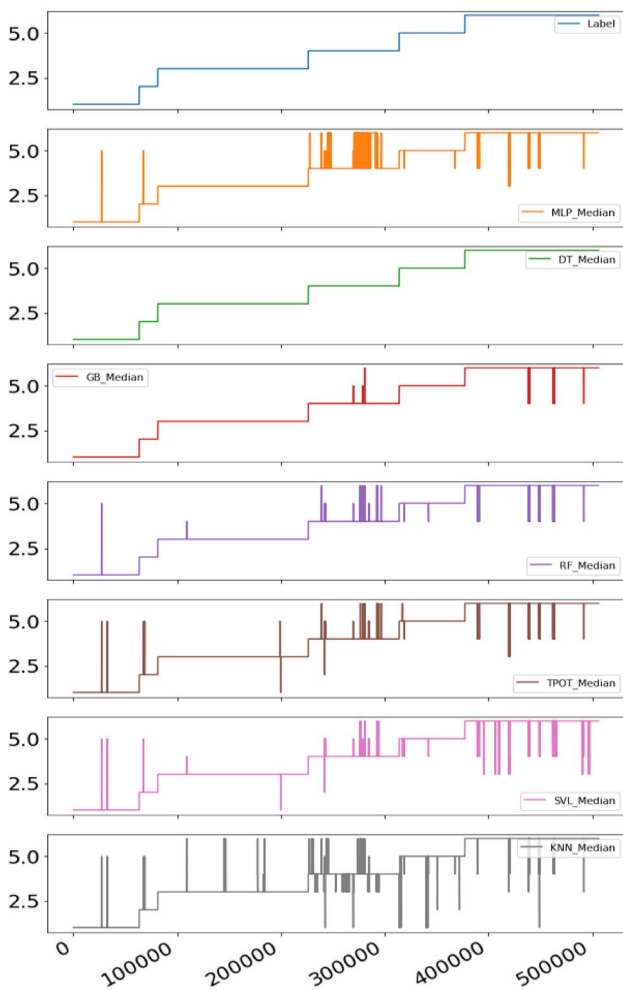


Fig. 6 Prediction from Case 1

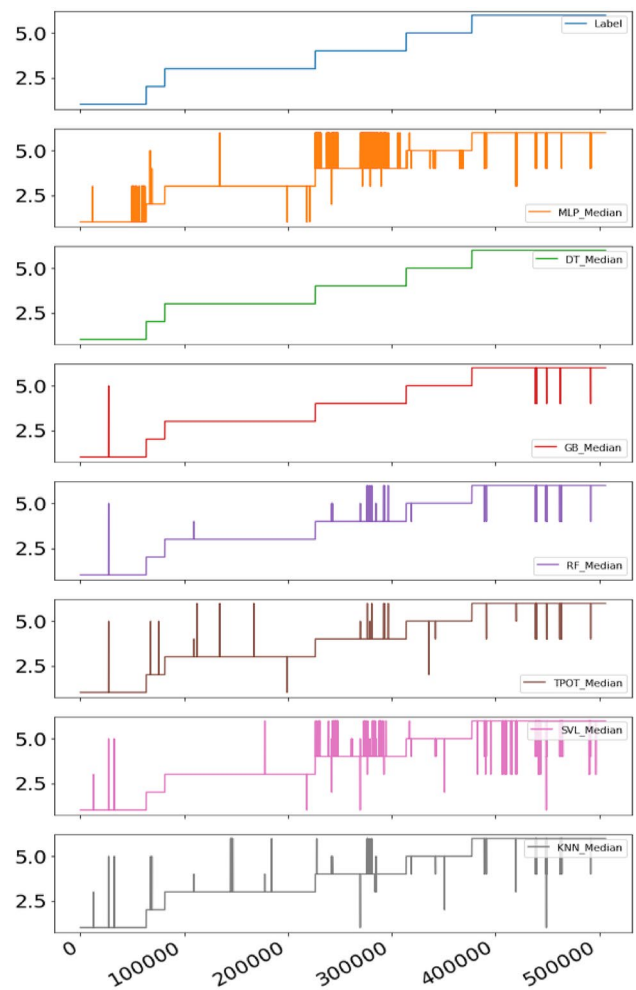


Fig. 7 Prediction from Case 2

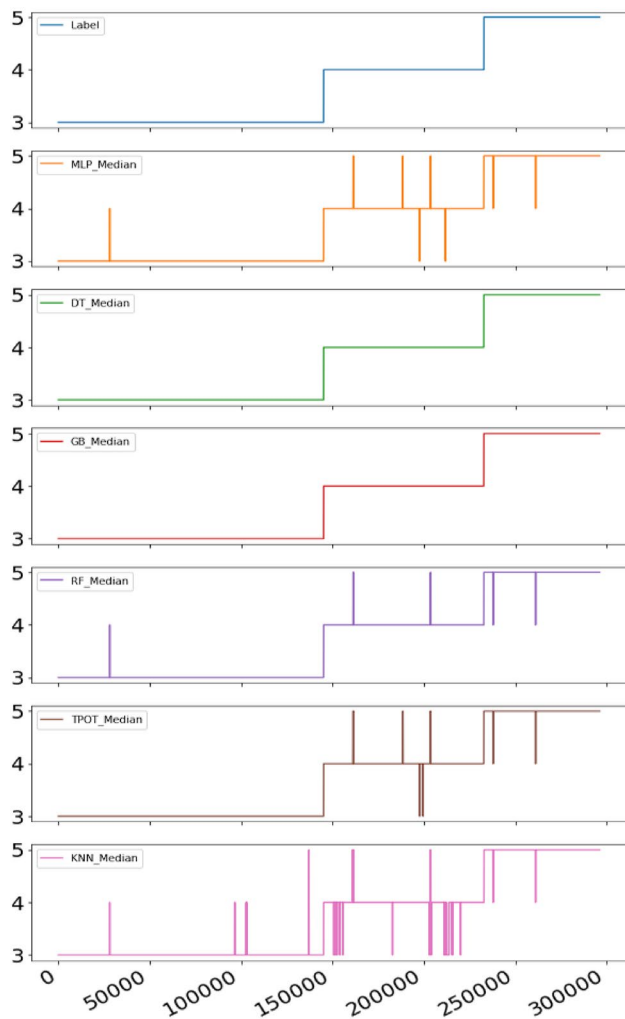


Fig. 8 Prediction from Case 3

number of features to train and classify formations can be reduced from 16 to 6 without losing the accuracy.

Evaluation (Cases 5–7)

The three rig operations POOH, RIH and ROnB can be predicted, as shown in Fig. 10 and Table 10. The dataset is run on the K-NN model in Case 5 that is built using a combination of natural features and engineered features. From Case 6, it shows that the unsupervised K-means model is capable of identifying the three different rig operations using the WOB and the ROP. The unsupervised DBSCAN model, however, interprets that four clusters are present, suggesting that only natural features are not robust enough. Considering Case 7, the two engineered features ROP median and ROP maximum—ROP minimum, both K-means and DBSCAN models are capable of identifying the different rig operations by their correct classes. Considering the results from Cases 5–7, there is no challenge in classifying the rig operations using the K-NN model developed.

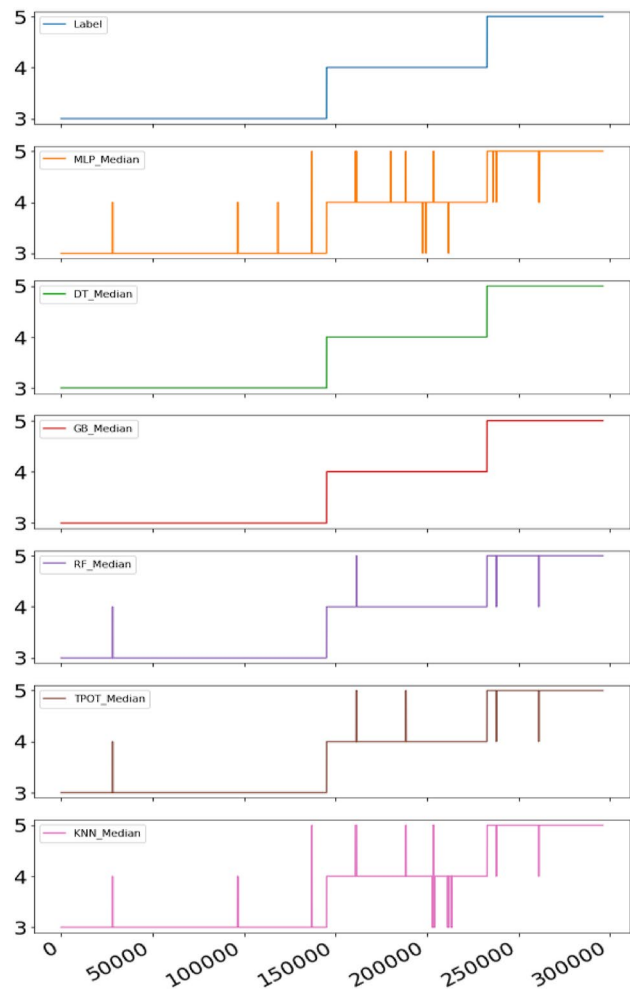


Fig. 9 Prediction from Case 4

Table 10 Results for rig operation classification, where ARI is adjusted rand index (Alexander 2017)

Case	K-NN	K-Means	DBSCAN
5	100%	–	–
6	–	ARI: 1.0000	ARI: 0.9989
7	–	ARI: 0.9970	ARI: 0.9970

It appears that high accuracy can be achieved using only a few selected features being either natural or engineered. For Cases 6 and 7, the models appear to more easily be capable of separating the engineered features from each other.

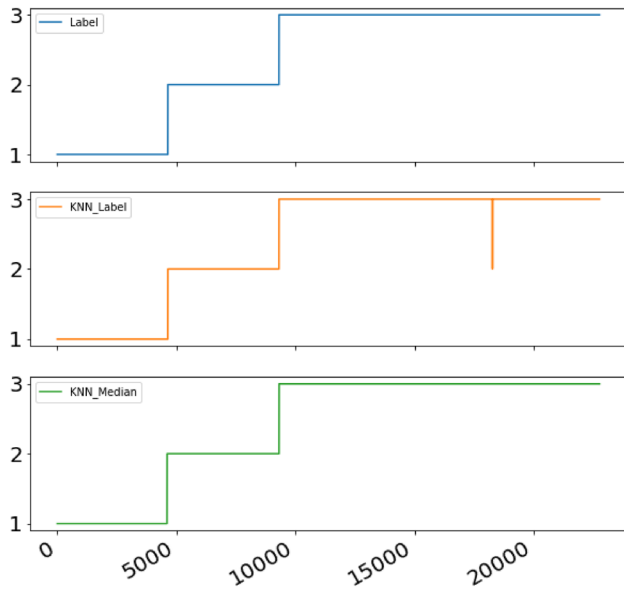


Fig. 10 Prediction from Case 5 (laboratory rig operations classification with raw- and median-filtered prediction)

Implementation

Voting system

A voting system has been developed to combine the predictions from the seven models into one formation class prediction with a confidence level score. The voting system could be further used to signal that a new formation is possibly detected, as well as to confirm a new formation that indeed has been encountered. From analyzing the performance of the models and checking the model performance on a test set separately, the weights in Table 11 are given to the different models.

The control system is configured to operate at 60 Hz, i.e., 60 predictions per second per model. The voting system can be illustrated by considering a case, as shown in Table 12. Each weight given in the table is counted as a separate problem. For instance, a model is given weight 2; the prediction from that model is equal to the prediction of two models that each have weight 1.

Then, a count is performed of the six classes. They represent six different rock formations, and a percentage score is calculated that represents the number of times that the class is predicted divided by 11 (total amount of all predictions multiplied by the weight), Table 13.

Table 12 Example for voting system

Model	Label prediction	Model weight
MLP	3	1
DT	3	3
SVM	4	1
RF	3	2
GB	5	2
TPOT	4	1
K-NN	3	1

Table 13 Example for confidence level calculation

Class (rock type)	Occurrences	Confidence level (%)
1 (cement)	0	0
2 (chalk)	0	0
3 (granite)	7	63.64
4 (sandstone)	2	18.18
5 (salt)	2	18.18
6 (shale)	0	0

This suggests that the machine should recognize a granite is being drilled with a 63.64% confidence, a 18.18% chance that the formation is a sandstone and a 18.18% chance that salt is being drilled. The prediction and confidence level is performed once every second.

Confirmation

New formation detection is handled by evaluating whether a class (formation type) gets predicted with a higher confidence level than 60% that is different from the previously confirmed formation class. Then, new formation confirmed is handled by considering the predictions over the last 10 s. For example, if 70% of the predictions in the last 10 s are of the same class (all with a higher confidence level than 60%) the machine could now replace granite with sandstone as the formation being drilled.

Table 14 shows how it works in real-time operation in terms of formation detection and confirmation. In such example, a new formation is not yet confirmed in the second last row, since even though a new formation gets detected, this formation class has not occurred in 70% of the last 10 seconds worth of predictions. The highest class is only filled into the array

Table 11 Weights added for real-time voting system

MLP	DT	SVM	RF	GB	TPOT	K-NN
1	3	1	2	2	1	1

Table 14 Example for real-time rock classification

Time interval (s)	Highest label	New formation detected?	Formation confirmed	Current label
t:[0,9]	Sandstone	Yes	No	Shale
t:[1,10]	Sandstone	Yes	No	Shale
t:[2,11]	Shale	No	No	Shale
t:[3,12]	Sandstone	Yes	No	Shale
t:[4,13]	Sandstone	Yes	No	Shale
t:[5,14]	Sandstone	Yes	No	Shale
t:[6,15]	Sandstone	Yes	No	Shale
t:[7,16]	Sandstone	Yes	No	Shale
t:[8,17]	Salt	Yes	No	Shale
t:[9,18]	Sandstone	Yes	Yes	Sandstone

if the confidence score from the voting output is higher than 60%.

Discussions

It shows that the machine learning models achieve high accuracy to detect different rock formations and rig operations. There are, however, several limitations and challenges of machine learning:

- First and foremost, it should be emphasized that the model accuracy heavily depends on the quality of the data used to train the models. It means that while a good model can be created for one objective, there is no guarantee that a good model can be used for another, unless the data can accurately describe the phenomena.
- Secondly, the models depend heavily on the environment that they have been trained. An example of this is a model that has been trained on data acquired in the laboratory environment, but when used in the field is not able to make the correct prediction, even if the trend might be the same. More scaling issues shall be considered in terms of model development phase.
- Thirdly, another limitation is to understand which features that must be selected in order to correctly detect the phenomena that the model gets developed for, and to blindly trust different importance evaluation techniques.
- When compared to physical models, it is our perception that it can both be difficult to detect and correct if the machine learning model makes a mistake. Machine learning approaches look like a black box to be difficult to be interpreted, translated or understanding. This is related to the complexity of fully understanding the processes that go into each decision that the machine makes.

- Finally, a major limitation lies in computational power available to train a model on large sets of data. If for instance a deep learning model gets developed from an immense number of observations, the required hardware to train such model can be both expensive and inaccessible. There has, however, been a big shift in recent years toward cloud computing, where one can upload the data and use the computational power of a data center to build the model. This also applies to the time that it takes to train a model. If either the time available to train the model or to make a prediction is limited, it is absolutely necessary to understand which models are computationally expensive to build, and which are not.

Conclusion

In our experimental tests, a total of six different rock formations can successfully get classified on the laboratory drilling rig by using machine learning approaches. Moreover, the predictions from the machine learning models for formation classification can be combined through the proposed voting system to present the output prediction along with a confidence level. Specifically, a new formation can be confirmed by voting if it has been detected successfully over a number of consecutive iterations. Having a new formation detected, it allows the control system to initiate either a new search for an optimal ROP or to use pre-determined drilling parameters for the WOB and rotational speed, based on analysis of previous runs. Different drilling scenarios have been introduced to test, evaluate and validate our approaches on the rig while drilling different formations. Model calibrations regarding data processing, feature selection, hyper-parameter tuning and machine learning architecture choice and model validations to validate model results with the real system can be easily conducted by running different tests.

The developed approach of pre-processing the data, selecting the most optimal features and developing multiple models along with a voting system has resulted in reliable results. Future recommendations are:

- Integration of reinforcement learning on the rig, in which the models constantly get improved by correction of the prediction outputs from models,
- Developing a larger database containing both different rock formations drilled while varying drilling parameters,
- Develop models and perform PCA based on downhole measurements or surface measurements that accurately describe the bit interaction with the formations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Mechanical specific energy (MSE) is a measure of the mechanical energy required to remove a unit volume of rock. The MSE calculation is given in Fear and Pessier (1992):

$$\text{MSE} = \frac{\text{WOB}}{A_{\text{bit}}} + \frac{120 \times \pi \times \text{RPM} \times \tau}{A_{\text{bit}} \times \text{ROP}} \quad (3)$$

Depth of cut (DOC) can be described as the axial distance that the drill bit cuts into the formation per revolution. The formula to calculate the DOC is given by Kenneth and Rusel (2016):

$$\text{DOC} = \frac{\text{ROP}}{5 \times \text{RPM}} \quad (4)$$

Bit aggressiveness (BA) is determined by how high the DOC of a drill bit is, and depends on the bits backrake angle (cutting angle) and the exposure of cutters. The equation for bit aggressiveness calculation (Karadzhova 2014) is

$$\mu = \frac{36 \times \tau}{\text{WOB} \times D_{\text{bit}}} \quad (5)$$

D-exponent describes the so-called drillability of different formations, see (Akisanmi 2016). It is calculated by:

$$d = \frac{\log\left(\frac{\text{ROP}}{60 \times \text{RPM}}\right)}{\log\left(\frac{12 \times \text{WOB}}{10^3 \times D_{\text{bit}}}\right)} \quad (6)$$

References

- Agwu OE, Akpabio JU, Dosunmu A (2019) Artificial neural network model for predicting the density of oilbased muds in high-temperature, high-pressure wells. *J Pet Explor Prod Technol* 10:1081–1095
- Akisanmi OA (2016) Automatic management of rate of penetration in heterogeneous formation rocks. Master's thesis, University of Stavanger
- AlAzani K, Elkhatny S, Ali A, Ramadan E, Abdulraheem A (2019) Cutting concentration prediction in horizontal and deviated wells

- using artificial intelligence techniques. *J Pet Explor Prod Technol* 9(4):2769–2779
- Alexander J (2017) Gates and Yong-Yeol Ahn. The impact of random models on clustering similarity. *J Mach Learn Res* 18(1):3049–3076
- AlHameedi ATT, Alkinani HH, DunnNorman S, Flori RE, Hilgedick SA, Amer AS, Alsaba M (2018) Mud loss estimation using machine learning approach. *J Pet Explor Prod Technol* 9(2):1339–1354
- Al-Mudhafar WJ (2017) Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms. *J Pet Explor Prod Technol* 7(4):1023–1033
- Altman NS (1992) An introduction to kernel and nearest-neighbor non-parametric regression. *Am Stat* 46(3):175–185
- Arigbe OD, Oyeyeyin MB, Arana I, Ghazi MD (2018) Real-time relative permeability prediction using deep learning. *J Pet Explor Prod Technol* 9(2):1271–1284
- Barbosa LFFM, Nascimento A, Mathias MH, Carvalho JA Jr (2019) Machine learning methods applied to drilling rate of penetration prediction and optimization—a review. *J Pet Sci Eng* 183:106332
- Bello O et al (2015) Application of artificial intelligence methods in drilling system design and operations: a review of the state of the art. *J Artif Intell Soft Comput Res* 5(2):121–139
- Christiani N, Shawe-Taylor J (2000) Support vector machines: data analysis machine learning and applications. Cambridge University Press, Cambridge
- DeepAI (2019) Unsupervised Learning, viewed 25.03.2019
- Elith J (2008) A working guide to boosted regression trees. *J Anim Ecol* 77(4):802–813
- Fan et al (2011) Mr-dbscan: an efficient parallel density-based clustering algorithm using mapreduce. In: 2011 IEEE 17th international conference on parallel and distributed systems, pp 473–480
- Fattah KA, Khamis MA (2018) Estimating oilgas ratio for volatile oil and gas condensate reservoirs: artificial neural network, support vector machines and functional network approach. *J Pet Explor Prod Technol* 9(1):573–582
- Fear MJ, Pessier RC (1992) Quantifying common drilling problems with mechanical specific energy and a bit-specific coefficient of sliding friction. In: SPE-24584
- Gholami A, Mohammadzadeh O, Kord S, Moradi S, Dabir B (2015) Improving the estimation accuracy of titration-based asphaltene precipitation through power-law committee machine (PLCM) model with alternating conditional expectation (ACE) and support vector regression (SVR) elements. *J Pet Explor Prod Technol* 6(2):265–277
- Good PI, Hardin JW (2006) Common errors in statistics (and how to avoid them). Wiley, New York
- Hartigan JA, Hartigan MA (1979) A k-means clustering algorithm. *J R Stat Soc Ser C* 28(1):100–108
- Hegde C, Daigle H, Millwater H, Gray K (2017) Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. *J Pet Sci Eng* 15:295–306
- Hegde C, Millwater H, Gray K (2019) Classification of drilling stick slip severity using machine learning. *J Pet Sci Eng* 179:1023–1036
- Herv A, Williams LJ (2010) Principal component analysis. Wiley Interdiscip Rev Comput Stat 2(4):433–459
- Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20(8):832–844
- Holdaway K (2014) Harness oil and gas big data with analytics: optimize exploration and production with data driven models, 1st edn. Wiley, New York
- James S (2016) An introduction to data analysis using aggregation functions in R. Springer, New York

- Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. *Science* 349(6245):255–260
- Kamiski B, Jakubczyk M, Szufel P (2017) A framework for sensitivity analysis of decision trees. *Central Eur J Oper Res* 26(1):135–159
- Kanin EA, Osiptsov AA, Vainshtein AL, Burnaev EV (2019) A predictive model for steady-state multiphase pipe flow: machine learning on lab data. *J Pet Sci Eng* 180:727–746
- Karadzhova GN (2014) Drilling efficiency and stability comparison between tricone, PDC and Kymera Drill Bits. Master's thesis, University of Stavanger
- Kelleher JD, Namee BM, D'Arcy A (2015) Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. The MIT Press, Cambridge
- Kenneth E, Russel SC (2016) Innovative ability to change drilling responses of a PDC bit at the rigsite using interchangeable depth-of-cut control features. In: SPE-178808-MS. Society of Petroleum Engineers
- Khadisov M, Petersen H, Jakobsen A, Sui D (2019) Developments and experimental tests on a laboratory-scale drilling automation system. *J Pet Explor Prod Technol* 10:605–621
- Klyuchnikov N, Zaytsev A, Gruzdev A, Ovchinnikov G, Antipova K, Ismailova L, Muravleva E, Burnaev E, Semenikhin A, Cherepanov A, Koryabkin V, Simon I, Tsurgan A, Krasnov F, Koroteev D (2019) Data-driven model for the identification of the rock type at a drilling bit. *J Pet Sci Eng* 178:506–516
- Løken E, Trulsen A, Holsaeter AM, Wiktorski E, Sui D, Ewald R (2018) Design principles behind the construction of an autonomous laboratory-scale drilling rig. *IFAC-OOGP* 51(8):62–69
- Løken E, Geekiyanage S, Sui D (2019) Small-scale autonomous drilling development for drilling digitalization. In: *Oil Gas European Magazine*
- Løken E, Løkkevik J (2019) Optimization of an intelligent autonomous drilling rig: testing and implementation of machine learning and control algorithms for formation classification, downhole vibrations management and directional drilling. Master thesis of University of Stavanger
- Makhotin I, Koroteev D, Burnaev E (2019) Gradient boosting to boost the efficiency of hydraulic fracturing. *J Pet Explor Prod Technol* 9(3):1919–1925
- Mayer-Schnberger V, Cukier K (2013) *Big data: a revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, Boston
- Michael J (2019) Understanding k-means clustering in machine learning. Retrieved 2019-10-31
- Mustafa Al Bakri et al (2014) Filling missing data using interpolation methods: study on the effect of fitting distribution. *Trans Tech Publications, Zurich*, pp 889–895
- Otterbach J (2019) Principal Component Analysis (PCA) for Feature Selection and some of its Pitfalls, viewed 20.04.2019
- Randal S (2019) Olson. TPOT, viewed 12.03.2019
- Roman V (2019) Unsupervised machine learning: Clustering analysis. Retrieved 2019-10-01
- Scikit Learn (2019) Scikit-learn machine learning in python, viewed 05.02.2019
- Shalev-Shwartz S, Ben-David S (2014) *Understanding machine learning: from theory to algorithms*. Cambridge University Press, Cambridge
- Soares C, Gray K (2019) Real-time predictive capabilities of analytical and machine learning rate of penetration (rop) models. *J Pet Sci Eng* 172:934–959
- Spesivtsev P, Sinkov K, Sofronov I, Zimina A, Umnov A, Yarullin R, Vetrov D (2018) Predictive model for bottomhole pressure based on machine learning. *J Pet Sci Eng* 166:825–841
- Sun J, Lia Q, Chen MQ, Ren L, Huang GH, Li CY, Zhang ZX (2019) Optimization of models for a rapid identification of lithology while drilling—a win-win strategy based on machine learning. *J Pet Sci Eng* 176:321–341
- van der Aalst W (2016) *Process mining: data science in action*, 2nd edn. Springer, New York
- Wilson E, Tufts DW (1994) Multilayer perceptron design algorithm. In: *Proceedings of IEEE workshop on neural networks for signal processing*
- Zaytsev A, Romanenkova E, Antipova K, Simon I, Makarov V, Koroteev D, Gurina E, Klyuchnikov N (2020) Application of machine learning to accidents detection at directional drilling. *J Pet Sci Eng* 184:106519

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.