OTC-32447-MS

Real-Time Machine Learning Application for Formation Tops and Lithology Prediction

Wael Ziadat and Hany Gamal, Weatherford; Salaheldin Elkatatny, King Fahd University of Petroleum and Minerals

This paper was prepared for presentation at the Offshore Technology Conference held in Houston, TX, USA, 1 – 4 May, 2023.

Abstract

During the drilling operation, the drill string is subjected to different geological formations which have distinct lithological characteristics that greatly affect the drilling performance and may ultimately result in increased costs of the project. The lithology of a formation can vary significantly, thus it is of paramount importance to accurately detect lithology changes and formation tops while drilling. In order to do so, geologic data and logs are often utilized by experts and operators to identify lithological variations. Machine learning algorithms and random forest have been employed in recent years to improve the process of lithology prediction, enabling more accurate results at faster rates. Machine learning-based systems incorporate a wide range of indicators such as rock types, mineral composition, sedimentary structures and microfossils for efficient lithology prediction. Additionally, random forest classifiers are beneficial due to their robustness with respect to outliers as well as their ability to capture complex relationships between variables from multivariate input datasets. With this approach, an effective operational strategy can be formulated based on the identified formation lithology in order to reduce incident costs associated with unexpected wellbore issues or instability caused by lithological changes. This technique also provides valuable insight into understanding subsurface conditions for more efficient resource exploration and production operations. limitations and drawbacks of this approach as cost and lag time. The current study proposed an intelligent machine learning solution for auto-detecting drilled formation tops and lithology types while drilling in real-time utilizing drilling surface data. Machine learning techniques are technically employed for developing real-time prediction models for the formation tops and lithology type from the surface drilling data as weight on bit, drill string speed, torque, pumping pressure and rate, and drilling penetration rate. This study implemented random forest and decision trees as two machine learning classifiers to develop real-time models using a data set of composite lithology schemes of five drilled formations. The methodology approach presents a comprehensive layout for data collection, preprocessing, data statistics and analytics, feature engineering, model development, parameters optimization, and prediction performance evaluation. The results showed a high prediction performance for the models for training and testing with overall accuracy higher than 95 through detecting complex lithology schemes. Predicting the drilled formation's tops and lithology while drilling in real-time through the developed

solution will provide a technical guide for optimizing the drilling parameters for better drilling performance and optimized mechanical-specific energy to have a safe operation and cost savings.

Keywords: Machine learning, random forest, decision tree, real-time, formation tops, lithology

Introduction

Machine learning has been used for lithology and formation top prediction with high accuracy by different studies (Mahmoud et al., 2021; Yoon et al., 2019). The random forest (RF) algorithm was found to provide a good combination between speed, accuracy, complexity reduction capability, and interpretability (Breiman 2001). RF was applied by many researchers for lithology and formations tops predictions from petrophysical log data with good results. One of the most important elements in devising a successful drilling program is to accurately predict the type of drilled formation and lithology. This is because drilling parameters must be optimized according to each formation's road map for optimal performance, as well as for designing an effective mud program and casing design. During the drilling operation, knowing the exact formational top and lithology type is essential for ensuring that drilling parameters are adjusted for maximum efficiency. Furthermore, chemical analysis of both the mud used during the drilling process and the drilled formation by its lithology is imperative in order to safely drill with optimum functionality while saving costs (Caenn et al., 2011; Gamal et al., 2021). The depth at which casings are set also relies heavily on being able to detect the correct formation tops and their corresponding lithology type (Mahmoud et al., 2021). Machine learning algorithms can be utilized to accurately anticipate these formations and lithologies.

Random Forest Machine Learning algorithms have been found to be particularly effective in predicting formations and lithologies based on certain data points such as pressure and temperature readings from downhole logs (Kumar & Razdan, 2017). Random Forest algorithms can identify patterns amongst parameters such as rock composition or grain size, which would otherwise require significant manual analysis from experienced professionals (Liu et al., 2016). In addition, Machine learning algorithms are not only precise but also timesaving when compared to traditional methods of prediction. Additionally, Machine learning applications enable engineers to use more extensive datasets than ever before when making decisions on how best to drill a given well (Uddin & Nassar, 2020). Machine learning also allows engineers to better assess risk management issues within a more thorough context (Mukherjee & Chatterjee, 2018).

The application of Machine Learning techniques in predicting top formations and lithologies provides an invaluable tool with regards to obtaining accurate results while saving resources. Machine Learning technology has opened new opportunities for predictive analytics that allow engineers to make informed decisions rapidly whilst reducing operating costs. Machine Learning algorithms can provide predictions regarding various geological properties such as permeability or porosity which will help engineers efficiently estimate resource reserves without having to conduct additional tests or experiments. These advancements are essential for improving safety standards in oilfield operations whilst simultaneously allowing for increased economic savings through improved efficiency. Detect lithology and formation tops in drilling operations using the available petrophysical logs data such as gamma ray, resistivity, density, and neutron. Machine learning (ML) algorithms are one of the most useful artificial intelligence techniques used in oil & gas applications.

Formations Tops and Lithology Type Detection

The determination of drilled formations tops is a task that requires specific techniques and approaches and can vary depending on the experience gained by the drilling and petrophysics teams from the offset drillings. Technically, the most commonly used approach for detecting the formation tops is logging tools such as gamma-ray (Z. Losoya et al., 2021). However, this tool has some limitations, such as being expensive and having difficulty in obtaining data and interpretation in real-time. Moreover, monitoring the drillability rate (rate of penetration) during drilling operations is also used to discern different profiles from one formation to another based on their interaction with drilling parameters and strength characteristics. But this method might not be highly accurate due to other factors influencing in the drilling environment like mud characteristics and parameters that interact in a complex manner (Elkatatny et al., 2019). Additionally, technical analysis of cuttings obtained from mud circulation system and mud logging process is also employed, though this technique has a delay time as well as requiring advanced experimental analyses to obtain results. In conclusion, Machine Learning (ML) using random forest algorithms could help improve formation tops prediction by providing more accurate results with shorter times.

Conclusions from the existing technical approaches for detecting the tops of drilled formations are the lag time, cost, and measurement errors. These limitations directed the research and development teams to leverage the new technology-based solutions and machine learning applications for detecting the formation top and lithology from the big data collected from the sensors either downhole or at the surface. Logging data is considered the common tool for modeling as input features to detect the formation type by machine learning applications. Sonic data, gamma ray, photoelectric effect, resistivity, porosity, and neutron-density porosity are utilized for lithology detection (Wang and Zhang, 2008; Xie et al., 2018; Nanjo and Tanaka, 2019). The rig sensor data provide much more information about the drilled formation characteristics due to the interaction with the drill string and bottom hole assembly. Surface drilling data such as rate of penetration (ROP) or drillability rate, weight on bit (WOB), drilling speed (DS), drilling torque (T), pumping rate (Q), and pressure at the standpipe (SPP) are commonly utilized to reveal information about the drilled rock by generating complex relationships by the high computational capabilities and machine learning algorithms. Much research already studied such applications for detecting the formation lithology type (Mohammad Ali, 2015; Elkatatny et al., 2019; Mahmoud et al., 2021) and extra parameters might be included as bite features and mud hydraulics.

Through the literature, many techniques were implemented for such purposes as support vector (SVM) machine, artificial neural network (ANN), random forest (RF), gradient tree boosting, convolution neural network (CNN), backpropagation neural network (BPNN), functional neural networks (FNN), adaptive neuro-fuzzy inference system (ANFIS) (Al-AbdulJabbar et al., 2018; Mohamed et al., 2019; Nanjo and Tanaka, 2020). It is worth mentioning that each technique has its power and computational way, but the important factor for modeling this problem is the data for the field of interest and the model input features, and how the interrelation with the drilled formation characteristics. That is why different tools might be used, and results show that the application is relative to the data set.

The current research presents a technical contribution to the literature through the developed intelligent solution for auto-detecting the drilled formation top and lithology type by machine learning application. The study implemented two classifiers named random forest (RF) and decision tree (DT) to develop the formation top prediction model for five different lithology schemes during the drilling operation. Three wells from the field of interest were utilized to develop the model to train and test the developed classifiers. The model input features are only the surface drilling mechanical parameters as the rate of penetration (ROP), weight on bit (WOB), drilling speed (DS), and drilling torque (T). The study added to the drilling automation and digitalization by auto-detecting the formation top and lithology type while drilling in real-time.

Machine learning algorithms such as random forest, support vector machine (SVM) and artificial neural network (ANN) have been applied to the process of lithology prediction. Machine learning methods provide an accurate way to predict lithology type by capturing the information from Drilling data with high precision. Machine Learning algorithms are capable of creating a model that can automatically classify the formation and lithology types using all available input features for reducing the cost, time, and errors associated with manual inspection of drilling data. The Machine Learning algorithms showed excellent capabilities & accuracy in predicting formation tops and lithology compared with traditional statistical & analytical approaches. Machine Learning Applications offer more efficient ways of analyzing big datasets from different sources like logs, sensors and surfaces drills, which could lead to better accuracy and repeatability continuously.

Data Description and Machine Learning Application

Research Approach

The study followed a technical approach that considered the full process for developing the ML models starting from collecting rig sensor data, data preprocessing, building the models, evaluating the accuracy, and finally saving the best results for detecting the formation tops and lithology type. The data was gathered from the rig sensors and then forwarded to the data wrangling process for cleaning and quality enhancement by removing the outliers and performing advanced statistical data analysis to study the data range, distribution, and correlation coefficient. Building the ML models started after the full data preprocessing phase and the model building was studied with deep sensitivity for optimizing the model parameters to provide the best results for the prediction. The models were evaluated with statistical metrics to evaluate the performance and compare the predicted versus actual values. The models would be retrained again with more training data or algorithms in case the prediction performance was not accepted till reached the best accuracy level to report the best results and model parameters. The best prediction results from the optimized ML models were reported to show the models' prediction capabilities.

Data Description

The current study utilized real field data during drilling operation that comprises the drilling surface data with the actual formation tops and lithology type as per the geological reports. Mainly rig sensors data that record the drilling mechanical and hydraulic data are the basic inputs for the models' ROP, WOB, DS, and T. Figure 1 represents the drilled formation tops and types across the field of interest for the four wells data. Formation (1) is mainly dolomitic lithology, followed by an anhydrite sequence (Formation 2), Formation (3) is a dolomitic limestone composition, followed by limestone formation (4), and finally Formation (5) which is mainly a share formation. As shown from the schematic cross-section that all drilled wells penetrated the same lithology scheme with different thicknesses. The figure shows also the data points distribution along the penetrated formation as formation 1 has 10% of the data points, 9% for formation 2, while formation 3, 4, and 5 have 31%, 24%, and 26% respectively.

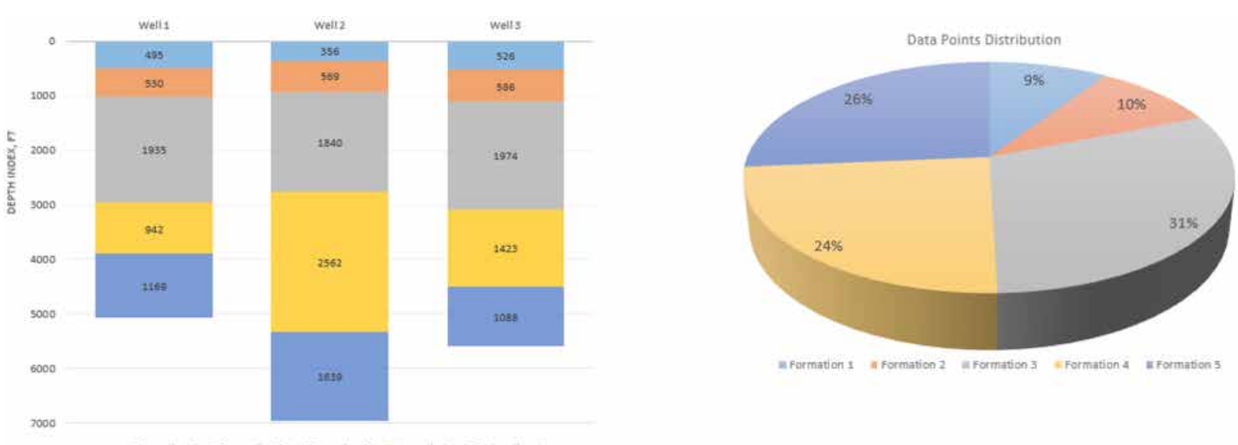

Figure 1—Wells and formation data across the field

Data Wrangling and Analytics

The collected drilling data was preprocessed by diverse pipelines to remove the illogic values within the data set as zeros and negative values, remove the parameters outliers, and smooth the drilling data to remove the noise and enhance the data quality. The cleaned data set was then analyzed to explore the interrelationships and data distributions.

The surface drilling parameters are commonly observed at the surface while drilling and these surface parameters are significantly affected by the interaction between the drillstring and wellbore wall that reveals indications and information about the drilled zones. The application of statistical analysis to the data gathered demonstrates good variety and representation of the data as it encompasses a wide range of input parameters (Table 1). The drillstring rotation had a range from 24 up to 104 rpm, the drilling torque from 1.0 up to 13.1 klbf.ft, the weight on bit from 2.2 to 52.9 klbf., and the drillability rate from 8.1 to 111.4 ft/hr. along the different dilled formations.

Table 1—Data statistical analysis

Parameter	ROP	DS	T	WOB
Min	8.1	24.0	1.0	2.2
Max	111.4	104.0	13.1	52.9
Mean	52.5	42.6	7.6	24.3
Standard Deviation	21.0	10.2	3.2	10.9

Studying the data analytics through pair plots for the data set features to represent the correlation of the parameters through visualization is shown in Figure 2. It is clear from the plots that it is a complex type of correlation between the drilling parameters and the drilled formation types for the problem that is highly candidate for the applications of machine learning to leverage the high computational capabilities to capture the existing patterns between the surface drilling data and the drilled formation tope and lithology type.

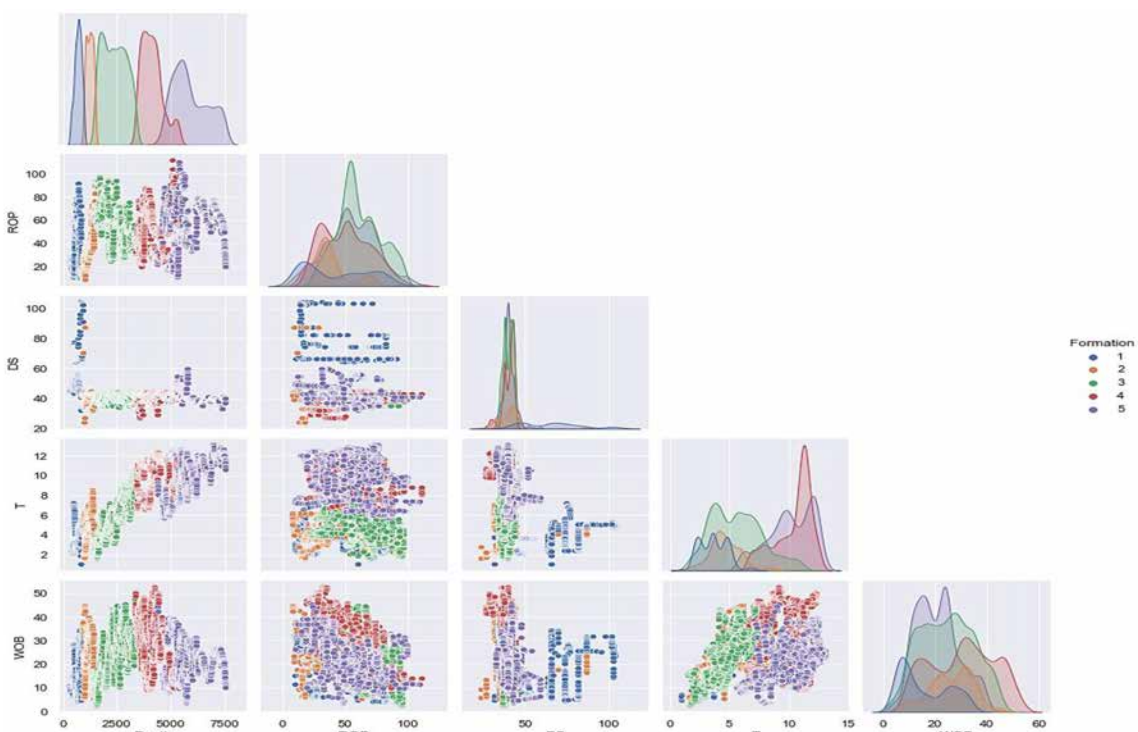

Figure 2—Pair plots for the data set features.

In addition, the feature importance between the drilling data and the formation type is analyzed and results are shown in Figure 3 which shows a direct linear relationship between the feature and formation type for all parameters except the indirect relationship for the drilling speed. Drilling torque had stronger feature relative importance (0.75) for the formation type followed by ROP parameter (0.18) and WOB (0.08) with indirect (-0.46) for the DS. The data heatmap (Figure 4) showed complex relationship types and degrees for the drilling data that added more challenges for the machine learning application and the problem complexity.

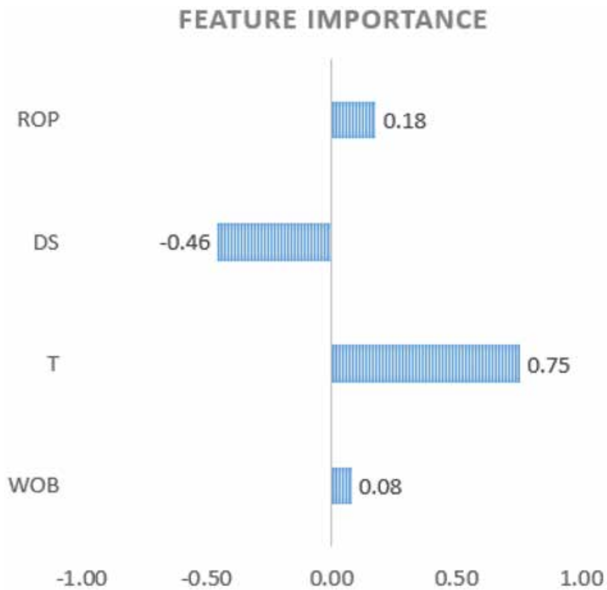

Figure 3—Feature importance chart

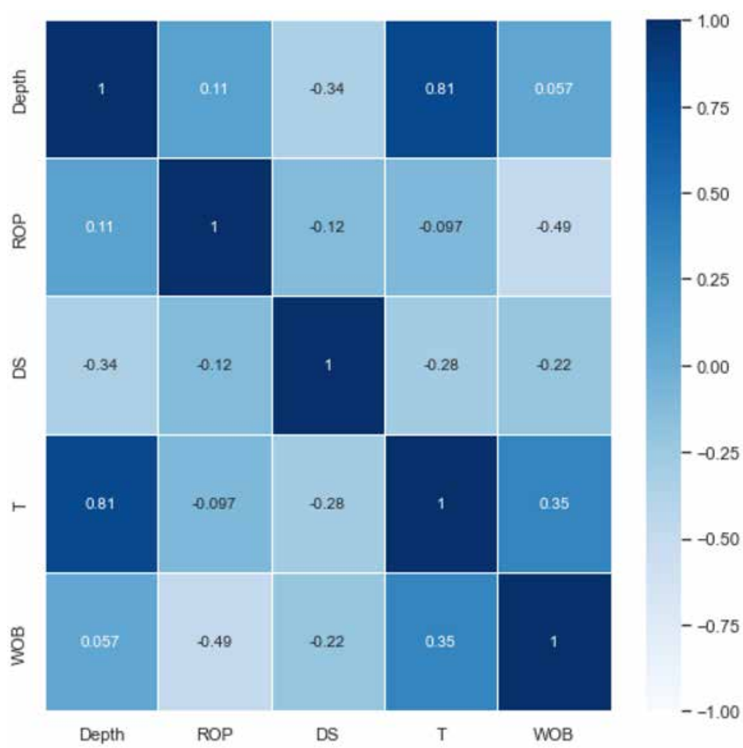

Figure 4—Data heat map

Machine Learning Models

The study implemented four techniques/classifiers to evaluate the capabilities of the algorithms to capture the parameter patterns for autodetecting the formation top and type during the drilling (Cover and Hart, 1967) operation. Neural network (NN), k-nearest neighbor (KNN), random forest (RF), decision tree

(DT), logistic regression (LR), and support vector machine (SVM) were utilized and compared during the training and testing for the models. These techniques are now widely applied in ML research applications for petroleum big data (Zhang et al., 2022). The current study employed two ML tree-based techniques (Decision tree and random forest) for developing the formation tops auto-detection models as these techniques have a wide applications within the petroleum industry for regression and classification (Alkinani et al., 2019).

Decision tree (DT) is considered one of the ML algorithms and it is a simple approach for application (Li and Chan, 2010). The technique employed straightforward rules for decision-making that is based on inferred decision instructions (Reddy et al., 2020). The technique has a hierarchical construction that contains root node, decision nodes, leaf nodes, and branches. Optimizing these parameters will lead to enhancing the model performance for better prediction (Sun et al., 2020).

Random forest is another supervised machine learning algorithm for classification and regression purposes. RF was introduced in 1995 and had several modifications over time (Breiman, 2001; Ho, 1995; Kleinberg, 2000, 1996), the technique is designed to overcome the overfitting that usually happens in classical discussion trees (Hastie et al., 2009). Similar to the other machine learning algorithms, several applications of random forest in the oil industry have been reported for classification (Kim et al., 2018) and regression (Hegde et al., 2015; Nasir and Rickabaugh, 2018; Sun et al., 2019). Figure 5 represents the schematic layout for the random forest as it includes a number N of decision trees in the model structure.

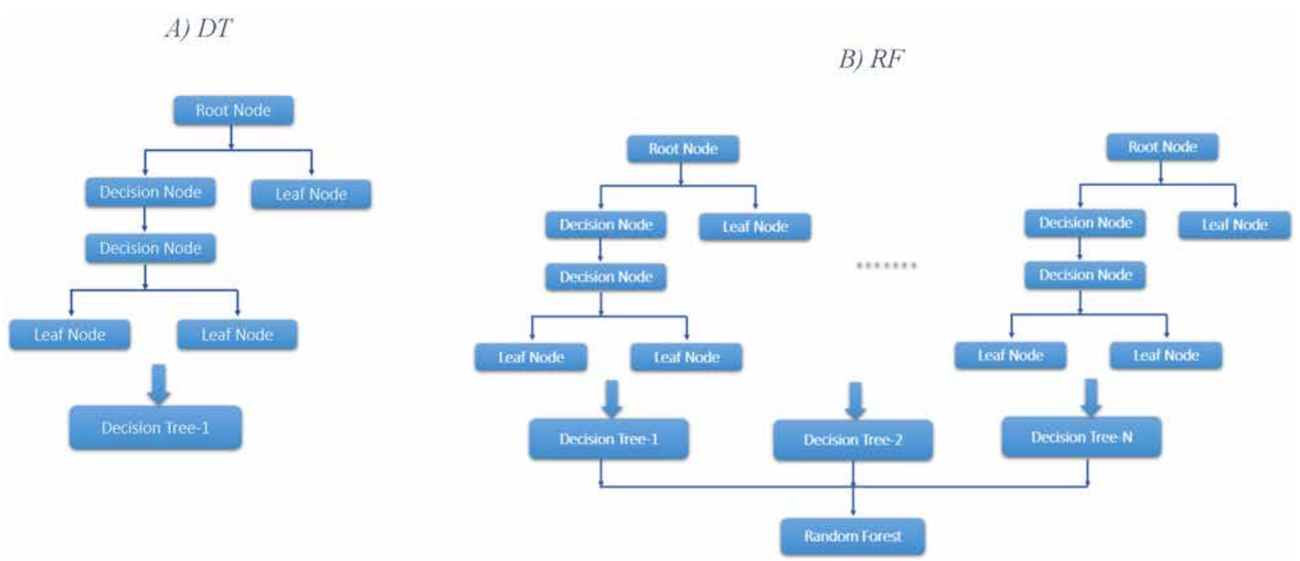

Figure 5—Random Forest and decision tree schematic. A) DT and B) RF (Gamal et al., 2022)

Model Performance

The developed models were evaluated during the sensitivity analysis for each model to get the best results for predicting the three models of downhole vibrations. The performance of each model is evaluated using the following metrics: accuracy, recall, precision, and F1-score.

According to the combination of actual data labels and predicted classes, the classification results can be divided into four cases: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The accuracy, defined by the following equation, measures the percentage of correctly classified samples.

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

The recall is indicating the percentage of real positive samples that are classified as positive.

$$Recall = \frac{TP}{TP + FN}$$

The precision is measuring the proportion of actual positive samples within the samples that are predicted to be positive.

$$Precision = \frac{TP}{TP+FP}$$

The F1 score is the harmonic mean of Recall and Precision and can be used to evaluate the model thoroughly. It is calculated as:

$$F1 = \frac{2*\text{Pecision}*\text{Recall}}{\text{Pecision} + \text{Recall}} = \frac{2*TP}{2*TP+FP+FN}$$

The confusion matrix is a 2D array of predicted and actual target labels. The entries of confusion matrix C[i][j] are equal to the number of observations predicted to have facies j, but are known to have facies i.

Results and Discussion

Each model was tested during sensitivity analytics for a wide range of every parameter to get the optimized model parameters for the best accuracy and prediction results. This step is exactly performed by evaluating the model prediction performance in terms of accuracy, recall, precision, and F1-score through each case/run for the model.

Model Optimization

Optimizing the model parameters is a crucial step for getting the best prediction result through the model training and testing phases, and this step is commonly achieved through deep sensitivity trials for all the model parameters that affect the prediction performance. Automated trials and error runs were done to evaluate the impact of every model parameter on the predicted values of vibrations.

Models Accuracy

The optimized models were evaluated through the model training and testing data sets and Table 3 represents the RF-Model accuracy results across each drilled formation type. Table 4 represents the accuracy results for DT-Model during the model training and testing. Figures 6 and 7 present the confusion matrix for two model results. The results highlight the overall outstanding performance of the RF and DT classifiers over the training as well as testing data sets. The models overall accuracy showed 0.99 and 0.98 for RF and DT respectively.

Table 3—RF-Model accuracy for the training and testing phases

RF-Model	Training				Testing			
	Precision	Recall	F1-score	Points	Precision	Recall	F1-score	Points
Formation 1	1.00	1.00	1.00	474	1.00	0.99	1.00	204
Formation 2	1.00	1.00	1.00	474	0.98	1.00	0.99	229
Formation 3	1.00	1.00	1.00	1563	1.00	0.99	0.99	668
Formation 4	1.00	1.00	1.00	1249	0.99	0.99	0.99	507
Formation 5	1.00	1.00	1.00	1343	0.98	0.99	0.99	579
Accuracy		1.00		5103		0.99		2187

Table 4—DT-Model accuracy for the training and testing phases

DT-Model	Training				Testing			
	Precision	Recall	F1-score	Points	Precision	Recall	F1-score	Points
Formation 1	1.00	1.00	1.00	474	0.99	0.99	0.99	204
Formation 2	0.99	1.00	1.00	474	0.97	0.96	0.97	229
Formation 3	1.00	1.00	1.00	1563	0.99	0.99	0.99	668
Formation 4	1.00	1.00	1.00	1249	0.99	0.98	0.98	507
Formation 5	1.00	1.00	1.00	1343	0.98	0.99	0.99	579
Accuracy		0.98		5103		0.99		2187

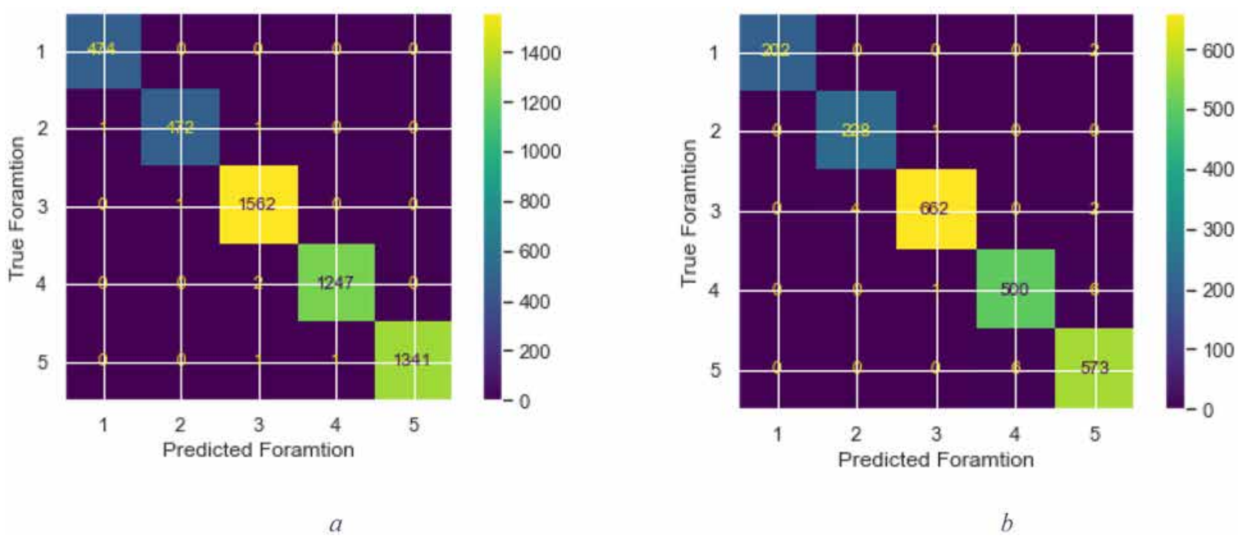

Figure 6—Confusion matrix for RF-Model. A) Training, b) Testing.

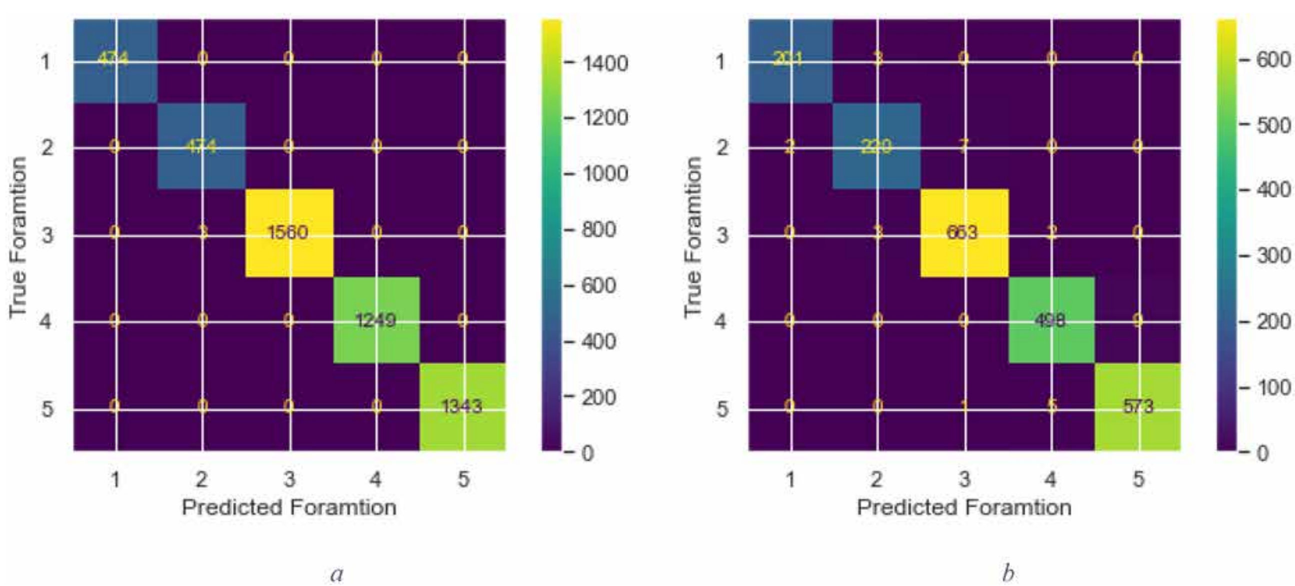

Figure 7—Confusion matrix for DT-Model. A) Training, b) Testing.

Conclusions

Accurate detection of the drilled formation top and lithology type is an essential task during drilling operations. Not only does it enable the optimization of technical designs and parameters, such as casing setting and mud rheology, but it also ensures safe wellbore stability. This study proposed a machine learning application to be implemented on the rig site as an intelligent solution to detect the formation top and type in real-time from the surface rig sensor drilling data. To develop this Machine Learning (ML) model, three wells were studied in detail. The workflow involved collecting data, wrangling it, and running analytics to create Machine Learning models using Decision Trees (DT) and Random Forests (RF). In addition, hyperparameter optimization was conducted for both classifiers to ensure the best possible accuracy for formation tops and lithology type prediction. The results showed that both DT and RF had high capabilities for predicting formation tops with an impressive accuracy rate of 0.99 and 0.98 respectively. It was also observed that RF had greater prediction power than DT in some cases due to its growing accuracy levels over time when compared against a traditional ML approach. Furthermore, model performance could be further improved by proper tuning of hyperparameters along with ongoing data collection.

The Machine Learning (ML) models presented in this study can be employed as a smart solution on drilling rigs to provide real-time recognition of drilled formation tops and lithology types. Additionally, the research sets out to generalize the model's application for large datasets regarding the field of focus, incorporating more wells data and leveraging Machine Learning capabilities for such a complex problem. With the help of this solution, drilling and petrophysics teams would be able to make more informed decisions while having safer drilling operations at an overall lower cost. This Machine Learning application would eliminate the need for downhole equipment or extra logging tools and also bypass delays due to analysis of drilled cuttings. The model's ability to detect real-time formation tops will be immensely beneficial in helping operators anticipate subsurface lithological changes that could potentially lead to safety hazards. Furthermore, it would reduce the risk associated with having limited knowledge of underground formations by providing accurate predictions of subsurface rocks and structure based on collected data from adjacent wells. To ensure accurate predictions, Machine Learning algorithms like random forest will be used in conjunction with geological information from reference sections generated through well logs, core samples, cuttings sample analysis, seismic surveys etc. This approach can help construct more precise models that are capable of learning from small datasets efficiently. In addition to yielding higher accuracy rates for lithology prediction tasks than traditional Machine Learning approaches, random forest has demonstrated an improved capacity for recognizing patterns within data collected from multiple sources. As a result, these models can be utilized not only for accurate drilling operations but also for more precise reservoir characterizations.

References

1. Al-AbdulJabbar, A., Elkatatny, S., Mahmoud, M., Abdulraheem, A., 2018. Predicting Formation Tops While Drilling Using Artificial Intelligence. Presented at the SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Society of Petroleum Engineers. https://doi.org/10.2118/192345-MS
2. Alkinani, H.H., Al-Hameedi, A.T.T., Dunn-Norman, S., Alsaba, M.T., Amer, A.S., 2019. Review of the applications of decision tree analysis in petroleum engineering with a rigorous analysis, in: SPE Middle East Oil and Gas Show and Conference, MEOS, Proceedings. Society of Petroleum Engineers (SPE). https://doi.org/10.2118/194795-ms
3. Alsaihati, A., Elkatatny, S., Gamal, H., 2022. Rate of penetration prediction while drilling vertical complex lithology using an ensemble learning model. *Journal of Petroleum Science and Engineering* **208**, 109335. https://doi.org/10.1016/J.PETROL.2021.109335
4. Breiman, L., 2001. Random forests. *Machine Learning* **45**, 5–32. https://doi.org/10.1023/A:1010933404324
5. Caenn, R., Chillingar, G.V., 1996. Drilling fluids: State of the art. *Journal of Petroleum Science and Engineering* **14**, 221–230. https://doi.org/10.1016/0920-4105(95)00051-8

6. Caenn, R., Darley, H.C.H., Gray, G.R., 2011. Composition and Properties of Drilling and Completion Fluids, *Composition and Properties of Drilling and Completion Fluids*. https://doi.org/10.1016/C2009-0-64504-9
7. Elkatatny, S., Al-Abduljabbar, A., Mahmoud, A.A., 2019. New Robust Model to Estimate Formation Tops in Real Time Using Artificial Neural Networks (ANN). *Petrophysics* **60**, 825–837. https://doi.org/10.30632/pjv60n6-2019a7
8. Gamal, H., Alsaihati, A., Elkatatny, S., 2022. Predicting the Rock Sonic Logs While Drilling by Random Forest and Decision Tree-Based Algorithms. *Journal of Energy Resources Technology* 144. https://doi.org/10.1115/1.4051670
9. Gamal, H., Elkatatny, S., Abdulraheem, A., 2020. Rock drillability intelligent prediction for a complex lithology using artificial neural network. Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference 2020, ADIP 2020.
10. Gamal, H., Elkatatny, S., Adebayo, A., 2021. Influence of mud filtrate on the pore system of different sandstone rocks. *Journal of Petroleum Science and Engineering* **202**, 108595. https://doi.org/10.1016/j.petrol.2021.108595
11. Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, *Springer Series in Statistics*. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-84858-7
12. Hegde, C., Wallace, S., Gray, K., 2015. Using trees, bagging, and random forests to predict rate of penetration during drilling, in: Society of Petroleum Engineers - SPE Middle East Intelligent Oil and Gas Conference and Exhibition. Society of Petroleum Engineers. https://doi.org/10.2118/176792-ms
13. Ho, T.K., 1995. Random decision forests, in: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR. IEEE Computer Society, pp. 278–282. https://doi.org/10.1109/ICDAR.1995.598994
14. Hossain, M.E., Al-Majed, A.A., 2015. Fundamentals of Sustainable Drilling Engineering, Fundamentals of Sustainable Drilling Engineering. John Wiley & Sons. https://doi.org/10.1002/9781119100300
15. Kim, Y., Hardisty, R., Torres, E., Marfurt, K.J., 2018. Seismic facies classification using random forest algorithm. 2018 SEG International Exposition and Annual Meeting.
16. Kleinberg, E.M., 2000. On the algorithmic implementation of stochastic discrimination. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 473–490. https://doi.org/10.1109/34.857004
17. Kleinberg, E.M., 1996. An overtraining-resistant stochastic modeling method for pattern recognition. *The Annals of Statistics* **24**, 2319–2349. https://doi.org/10.1214/aos/1032181157
18. Li, X., Chan, C.W., 2010. Application of an enhanced decision tree learning approach for prediction of petroleum production. *Engineering Applications of Artificial Intelligence* **23**, 102–109. https://doi.org/10.1016/j.engappai.2009.06.003
19. Mahmoud, A.A., Elkatatny, S., Al-AbdulJabbar, A., 2021. Application of machine learning models for real-time prediction of the formation lithology and tops from the drilling parameters. *Journal of Petroleum Science and Engineering* **203**, 108574. https://doi.org/10.1016/j.petrol.2021.108574
20. Mohamed, I.M., Mohamed, S., Mazher, I., Chester, P., 2019. Formation Lithology Classification: Insights into Machine Learning Methods. Presented at the SPE Annual Technical Conference and Exhibition, OnePetro. https://doi.org/10.2118/196096-MS
21. Mohammad Ali, A.M., 2015. Artificial Intelligence for Lithology Identification through Real-Time Drilling Data. *J Earth Sci Clim Change* 06. https://doi.org/10.4172/2157-7617.1000265

22. Nanjo, T., Tanaka, S., 2020. Carbonate Lithology Identification with Generative Adversarial Networks. Presented at the International Petroleum Technology Conference, OnePetro. https://doi.org/10.2523/IPTC-20226-ABSTRACT
23. Nanjo, T., Tanaka, S., 2019. Carbonate Lithology Identification with Machine Learning. Presented at the Abu Dhabi International Petroleum Exhibition & Conference, OnePetro. https://doi.org/10.2118/197255-MS
24. Nasir, E., Rickabaugh, C., 2018. Optimizing Drilling Parameters Using a Random Forests ROP Model in the Permian Basin, SPE Liquids-Rich Basins Conference - North America. Society of Petroleum Engineers, Midland, Texas, USA. https://doi.org/10.2118/191796-MS
25. Reddy, G.T., Reddy, M.P.K., Lakshmanna, K., Kaluri, R., Rajput, D.S., Srivastava, G., Baker, T., 2020. Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access* **8**, 54776–54788. https://doi.org/10.1109/ACCESS.2020.2980942
26. Sun, D., Lonbani, M., Askarian, B., Jahed Armaghani, D., Tarinejad, R., Thai Pham, B., Van Huynh, V., 2020. Investigating the Applications of Machine Learning Techniques to Predict the Rock Brittleness Index. *Applied Sciences* **10**, 1691. https://doi.org/10.3390/app10051691
27. Sun, Y., Ma, G., Wang, L., Zhang, F., Li, X., Mei, L., 2019. Organic-matter content prediction based on the random forest algorithm: Application to a Lower Silurian shale-gas reservoir. SEG International Exposition and Annual Meeting.
28. Wang, K., Zhang, L., 2008. Predicting formation lithology from log data by using a neural network. *Petroleum Science* **5**, 242–246. https://doi.org/10.1007/s12182-008-0038-9
29. Xie, Y., Zhu, C., Zhou, W., Li, Z., Liu, X., Tu, M., 2018. Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. *Journal of Petroleum Science and Engineering* **160**, 182–193. https://doi.org/10.1016/J.PETROL.2017.10.028
30. Z. Losoya, E., Vishnumolakala, N., Noynaert, S.F., Medina-Cetina, Z., Bukkapatnam, S., Gildin, E., 2021. Automatic Identification of Rock Formation Type While Drilling Using Machine Learning Based Data-Driven Models. Presented at the IADC/SPE Asia Pacific Drilling Technology Conference, OnePetro. https://doi.org/10.2118/201020-MS
31. Zhang, Junlong, He, Y., Zhang, Y., Li, W., Zhang, Junjie, 2022. Well-Logging-Based Lithology Classification Using Machine Learning Methods for High-Quality Reservoir Identification: A Case Study of Baikouquan Formation in Mahu Area of Junggar Basin, NW China. *Energies* **15**, 3675. https://doi.org/10.3390/en15103675