Article

Applications of Different Classification Machine Learning Techniques to Predict Formation Tops and Lithology While Drilling

Ahmed Farid Ibrahim, Ashraf Ahmed, and Salaheldin Elkatatny*

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Accurate prediction of formation tops and lithology plays a critical role in optimizing drilling processes, cost reduction, and risk mitigation in hydrocarbon operations. Although several techniques like well logging, core sampling, cuttings analysis, seismic surveys, and mud logging are available for identifying formation tops, they have limitations such as high costs, lower accuracy, manpower-intensive processes, and time or depth lags that impede real-time estimation. Consequently, this study aims to leverage machine learning models based on easily accessible drilling parameters to predict formation tops and lithologies, overcoming the limitations associated with traditional methods. Data from two wells (A and B) in the Middle East, encompassing drilling mechanical parameters such as rate of penetration (ROP), drill string rotation (DSR), pumping rate (Q), standpipe pressure (SPP), weight on bit (WOB), and torque, were collected for real-field analysis. Machine learning models including Gaussian naive Bayes (GNB), logistic regression (LR), and linear discriminant analysis (LDA) were trained and tested on the data set from well A, while the data set from well B was utilized for model validation as unseen data. The formations of wells A and B consist of four lithologies, namely, sandstone, anhydrite, carbonate/shale, and carbonates, necessitating the development of multiclass classification models. The drilling parameters, specifically the WOB and ROP, exhibited a strong influence on lithology identification. Among the models, GNB demonstrated exceptional performance in predicting formation lithology from the drilling parameters, achieving accuracy and nearly perfect precision, recall, and F1 score for the different classes. LDA and LR models accurately predicted sandstone and carbonate lithologies, although some misclassifications occurred in approximately 5% of points for anhydrite and around 20% in carbonate/shale formations. During validation, the models demonstrated accuracies of around 0.96, 0.95, and 0.92 for the GNB, LR, and LDA, respectively. The study highlights the efficacy of the developed machine learning models in accurately predicting the formation lithology and tops in real time. This is achieved by utilizing readily available drilling parameters, making the approach highly accurate and cost effective by leveraging existing real-time drilling data.

1. INTRODUCTION

One of the most crucial elements to ensuring effective and secure drilling operations is the instantaneous (in real time) and accurate assessment of the formation tops and the lithology of the section presently being drilled.[1] While building the casing program, knowing the tops of formations is important because it is necessary when selecting the suitable casing setting depths for effective zonal isolation and efficiently designing the suitable mud weight for controlling the wellbore conditions. The ability to accurately forecast the location and characteristics of different rock layers or formations is essential for understanding the geology of an area and making informed drilling decisions. Furthermore, an accurate prediction of formation tops and lithology can help reduce drilling costs, increase efficiency, and minimize risks. Knowing the formation tops and lithology can help predict potential drilling hazards and optimize drilling mechanical parameters, such as the WOB and rate of penetration, to minimize risk and increase efficiency. For example, if a formation is known to be prone to instability or to contain high-pressure fluids, drilling parameters can be adjusted to minimize the risk of a blowout or other incident.[2,3]

Reservoir characterization is of key importance for formation tops and lithology prediction. Understanding the lithology of a formation can present key information about the type and quality of hydrocarbons and the potential for fractures, porosity, and other reservoir characteristics.[4,5] This information can be used to optimize production and minimize environmental impact. For example, knowing the lithology of a formation can help in determining the best completion method or in identifying potential areas for hydraulic fracturing.

There are several methods that can be used to define formation tops and lithology including well logging, core sampling, cuttings analysis, seismic surveys, and mud logging.[6−8] The logging tool takes measurements of the rock's various
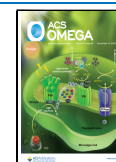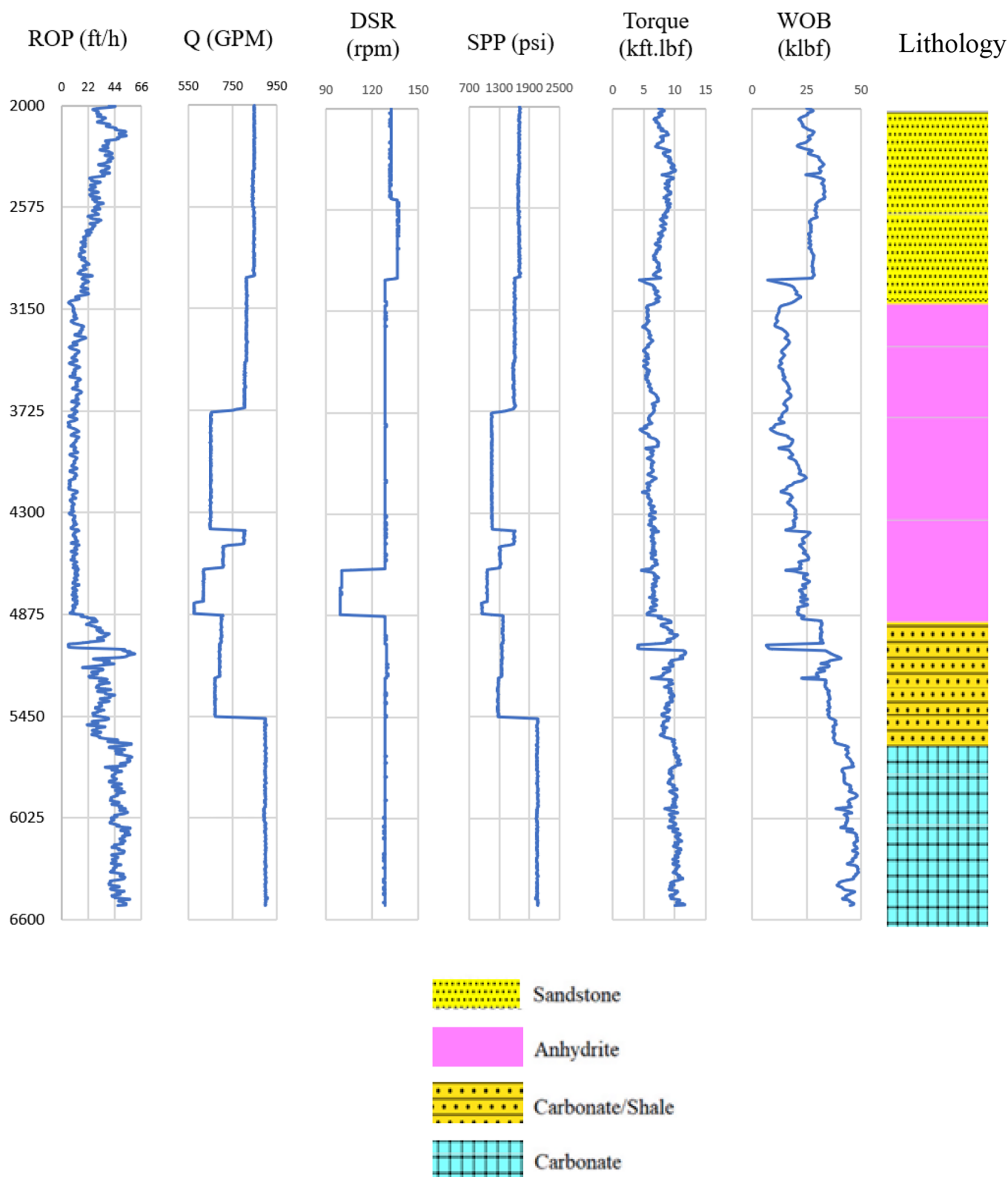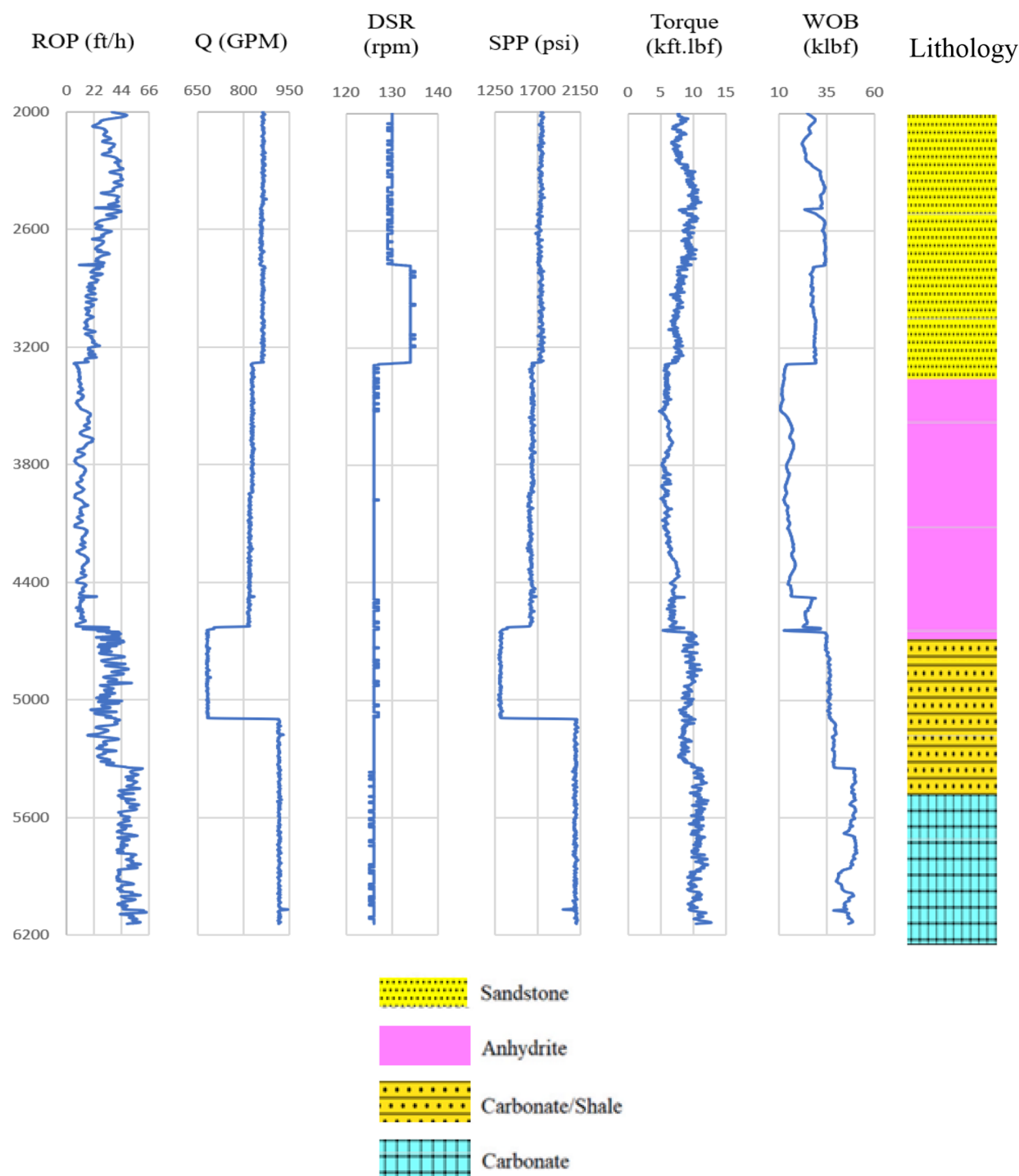
Figure 1. Lithology and graphical presentation of the drilling parameters of well A.

physical and chemical properties, such as gamma ray (GR), density, porosity, and resistivity, which can be used to infer the formation tops and lithology.[9−11] Core sampling involves physically removing a cylindrical sample of rock, called a "core", from the formation. The core can be analyzed in a laboratory to determine the formation tops and lithology, as well

as other information about the rock's physical and chemical properties.[12,13] Cuttings analysis for the rock cuttings that are brought to the surface during drilling can be performed. The cuttings can be analyzed to determine the formation tops and lithology, as well as other information about the rock's physical and chemical properties.[14,15] Seismic surveys involve the use of

Figure 2. Lithology and graphical presentation of the drilling parameters of well B.

sound waves to create images of the subsurface. Seismic data can be used to map the location and thickness of different rock layers, and this information can be used to infer the formation tops and lithology.[16,17] Mud logging involves analyzing samples of drilling mud. Mud logging can provide information about the rock types and formation tops that the drill has encountered.[18] In addition, in the situation of mud losses, mud logging and other techniques will not be available for formation tops determination.

Despite the benefits of various techniques used to determine formation tops, there are inherent limitations associated with

each method. These limitations include high costs, relatively low accuracy, substantial personnel requirements, and delays in obtaining measurements, in terms of time or depth. Consequently, these constraints hinder the applicability and lessen the competence of the existing methods for determining formation tops.[6,19]

When drilled cuttings are considered, there is a time delay in reaching the surface. This lag increases substantially with deeper boreholes, further impeding the real-time prediction of the currently drilled formation. Employing techniques such as logging while drilling (LWD) or GR for every formation is

economically impractical and fails to deliver instantaneous information as needed. Additionally, the LWD and GR sensors are not positioned at the drill bit, meaning that the formations predicted by these logs do not precisely match the present formation being drilled. Furthermore, mud logging exhibits its own delay time, varying from minutes to hours depending on factors, such as well depth and drilling fluid characteristics. This is due to the necessity of bottom-up circulation of the wellbore to remove the mud or cuttings. To address these challenges, many companies combine these techniques to enhance the formation tops estimation, taking into account the formation criticality and its proximity to the reservoir.

This study aims to estimate formation tops by utilizing machine learning (ML) models that compile six drilling parameters: rate of penetration (ROP), drill string rotation (DSR), pumping rate (Q), standpipe pressure (SPP), weight on the bit (WOB), and torque. By leveraging these models, the real-time detection of formation tops becomes faster compared to other methods as it eliminates the need for log processing and waiting for drilled cuttings to reach the surface. This approach enables accurate and real-time determination of formations at a significantly lower cost given the availability of real-time drilling data.

2. METHODOLOGY

2.1. Data Analysis. Real-field data for drilling mechanical parameters, including the ROP, DSR, Q, SPP, WOB, and torque, were gathered from two vertical onshore gas wells (well A and well B) located in the Middle East. The data sets consisted of 5535 data points from well A and 3128 data points from well B. These data were recorded on a footage basis using surface real-time data transmitter sensors, capturing information about various operations conducted in the borehole sections (such as drilling, tripping, and running the casing). The data points recorded while drilling new footage were considered for this study at the initial step of data preparation, while the remaining data was disregarded.

Figure 1 and Figure 2 depict the drilling parameters of well A and well B and the corresponding lithologies. Among these parameters, the weight on bit and torque exhibit the dynamic conduct of the drilling operation, whereas Q, DSR, and SPP remain relatively fixed on the surface with infrequent changes. The ROP primarily varies due to differences in formation characteristics, but it is also influenced by fluctuations in other drilling parameters. Therefore, the combination of drilling parameters with ROP is deemed to improve the accuracy of predicting changes in the formation lithology, particularly where the drilling parameters display significant fluctuations. This approach is crucial, as ROP in these depths is not solely determined by the type of formation.

The data sets collected from both wells underwent a thorough analysis to ensure their representativeness, reliability, and quality. The data were filtered and cleaned for outliers or missing data or zero value data. The data outliers were removed using the standard deviation method, where the data were standardized using their mean and standard deviation, and then the data outside −3 to 3 were excluded. This step was essential to ensuring that the accuracy of the model is not compromised, should it be used for further and different data sets. The statistical characteristics of the well A and well B data sets are summarized in Table 1. These statistics play a crucial role, as machine learning practitioners need to verify that the input data falls within the specified ranges. The statistics indicate that the

Table 1. Statistics of the Dataset

parameter	ROP	Q	DSR	SPP	torque	WOB
mean	26.6	811.7	127.7	1660.9	8.0	28.7
standard deviation	14.8	89.7	5.7	303.2	1.8	11.4
minimum	5.3	575.4	99.0	946.5	4.1	6.5
25% percentiles	11.7	691.6	128.0	1327.0	6.4	11.7
median	24.9	838.8	128.0	1672.7	7.9	27.9
75% percentiles	39.8	864.7	129.0	1741.9	9.4	39.8
maximum	63.4	942.8	137.0	2126.0	12.8	50.8

ROP ranges from 5.3 to 63.4 ft/h, the drilling fluid Q varies from 575.4 to 942.8 GPM, DSR falls between 99 and 137, SPP ranges from 946.5 to 2126 psi, the torque values range from 4.1 to 12.8 klb$_f$-ft, and the WOB varies between 6.5 and 50.8 klb$_f$.

Well A and B formation consists of four lithologies; sandstone, anhydrite, carbonate/shale, and carbonates. Hence, multiclass classification techniques will be used and each formation was represented by a certain class as described in Table 2.

Table 2. Representation of the Four Lithologies

lithology	class number
sandstone	0
anhydrite	1
carbonate/shale	2
carbonates	3

The study examined the importance of drilling mechanical parameters in predicting formation tops by assessing their correlation coefficient and linear relationship strength for the obtained data sets. The different formation lithologies were labeled as 0, 1, 2, and 3, and then the correlation coefficient was calculated between the drilling mechanical parameters and the lithology label using Pearson and Spearman methods. The correlation coefficient is valid physically; for example, a negative relation of DSR indicates possible sticking and slipping while drilling with high DSR. Figure 3 presents the correlation coefficient using the Pearson method (A) and the Spearman method (B). Figure 3A demonstrates that WOB plays a significant role in formation top prediction, showing a comparatively high correlation coefficient of 0.68. The torque and ROP exhibit correlation coefficients of 0.56 and 0.57, respectively, indicating a moderate relationship with formation tops. DSR and SPP exhibit lower correlation coefficients of −0.28 and 0.42, respectively. On the other hand, Q shows the weakest correlation coefficient of 0.2, indicating an absence of a linear relationship; however, Q cannot be excluded from the input feature as Q highly impacts the other parameters such as SPP and ROP, and WOB. Comparing the Pearson correlation coefficient to the Spearman correlation coefficient reflects the nonlinear relationship between the parameters. The correlation coefficient between the lithology with other input parameters almost did not change with using Pearson and Spearman methods, except for the DSR parameter. The correlation coefficient increased from (−0.28) to (−0.61) by using Pearson and Spearman methods, respectively, which reflect a nonlinear relationship between the lithology and DSR. In addition, Figure 4 presents a cross plot between the different parameters with formation tops in addition to the data distribution for each.[20,21] The cross plots between the lithology (Y) and the different parameters almost show a linear relation except with DSR (×3), which shows a nonlinear relation. In addition, the weakest

Figure 3. Correlation coefficient of the parameters with formation tops: (A) Pearson correlation coefficient. (B) Spearman correlation coefficient.

relationship between the inputs and lithology was found to be in the case of Q where the data are scattered and do not follow a certain trend, which confirms the results from Figure 3.

2.2. Model Development. The collected data were utilized to apply various ML techniques for predicting formation lithology and tops based on drilling parameters. The formation, as depicted in Figures 1 and 2, comprises four distinct lithologies. Consequently, different ML tools were employed to classify these lithologies into four classes: sandstone (class 0), anhydrite (class 1), carbonate/shale (class 2), and carbonates (class 3). The procedures for developing the ML models are presented in Figure 5. Following data collection and preprocessing, the model was constructed using the data from the first well, which was divided into training and testing data sets in a 70:30 ratio.[22] The data from well B served as unseen data for model validation.

This study involved the development of several machine learning methods, namely, Gaussian naive Bayes (GNB), logistic regression (LR), and linear discriminant analysis (LDA). For each ML model, the hyperparameters were optimized to improve the model performance. The grid search method was applied for hyperparameter selection. Grid search is a methodical approach to tweaking hyperparameters. For each hyperparameter, a grid of potential values is defined and the model's performance is assessed for each conceivable combination. The best performance-producing hyperparameter settings are then selected. Initially, a range of values for each hyperparameter to be tuned was selected for each ML method. All possible combinations of hyperparameter values were generated from the defined ranges. The ML model was trained for each combination of hyperparameters using the training data and the model's performance was evaluated using a chosen evaluation metric (e.g., accuracy, F1 score[22]) on a validation set or through cross-validation. The hyperparameter combination that resulted in the highest performance based on the evaluation metric was identified.

GNB is a probabilistic classifier that operates based on Bayes' theorem and assumes conditional independence among features. It calculates the probabilities of a data point belonging to each class and predicts the class with the highest probability. GNB is commonly used for classification tasks involving

continuous features and has demonstrated successful performance in various real-world applications. It is a straightforward and computationally efficient algorithm capable of handling large data sets, suitable for both binary and multiclass classification problems.

LR is a popular classification method used in machine learning for binary and multiclass classification tasks. It models the relationship between input features and the probability of the target variable belonging to a specific class using the sigmoid function. By estimating optimal coefficients during training, LR calculates the probabilities and applies a threshold to make predictions. It offers advantages such as interpretability, simplicity, and efficiency, making it suitable for both small and large data sets. However, its assumption of a linear relationship between features and the log-odds may limit its performance in complex classification scenarios.

LDA is a supervised classification method used for dimensionality reduction and separating classes in ML. It aims to find a linear projection of input features that maximizes the separation between classes. By estimating class means and covariances, LDA transforms the data into a lower-dimensional space and assigns new data points to classes based on discriminant scores. LDA is computationally efficient, interpretable, and applicable to multiclass problems. However, it assumes a multivariate normal distribution and equal class covariances, which may affect its performance in certain scenarios.

To assess the performance of the classification models, various performance indicators were employed. One of these indicators is the confusion matrix, which provides a concise summary of the classification model's performance by providing information on true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs). Furthermore, accuracy, precision, recall, and F1 score were calculated using the specific equations listed below to accurately evaluate the effectiveness of the classification model.

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{N} \tag{1}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$
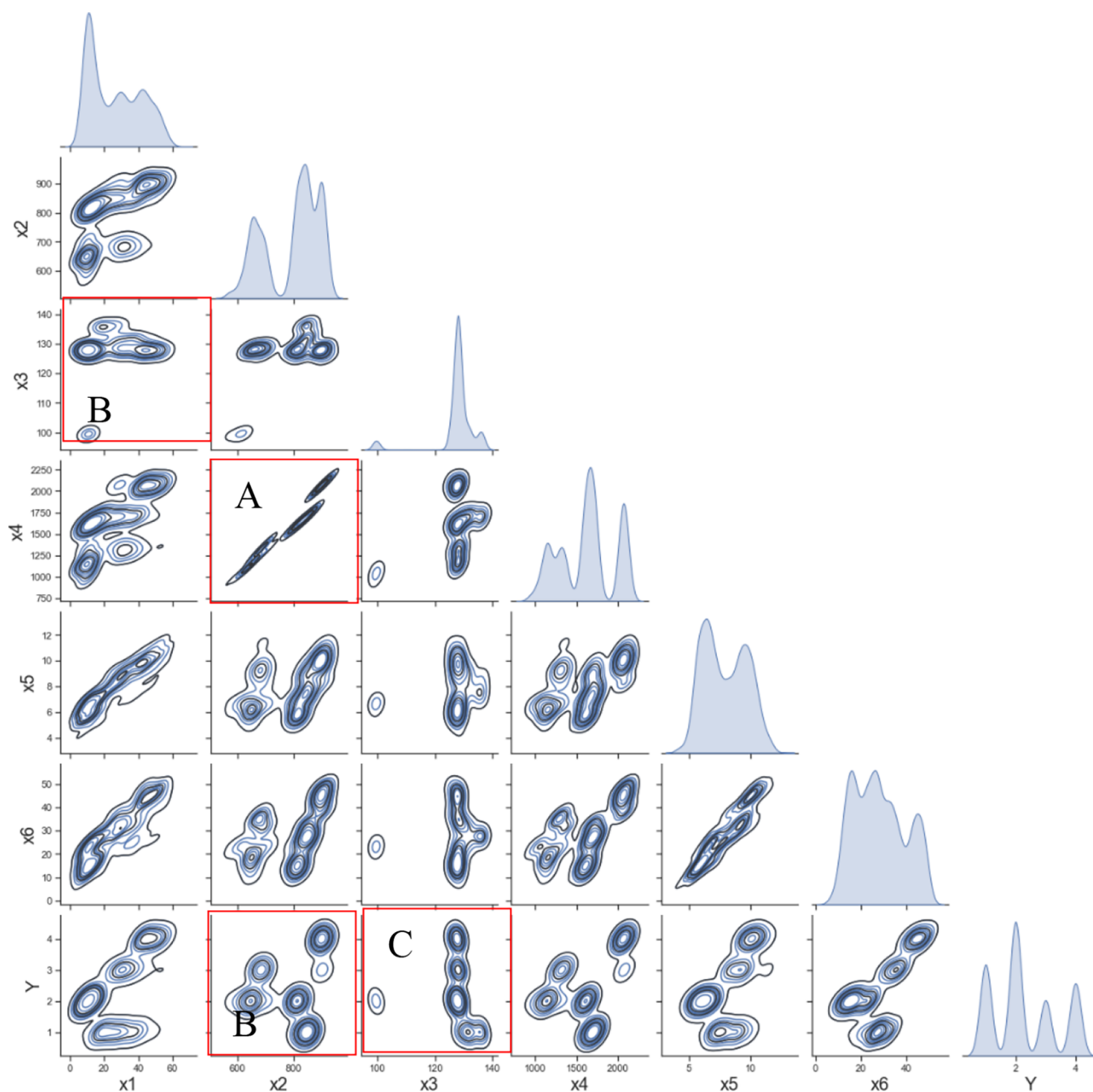
Figure 4. Cross plot for the different parameters with formation tops and the data distribution for each parameter where Y = lithology, $\times 1$ = ROP, $\times 2$ = Q, $\times 3$ = DSR, $\times 4$ = SPP, $\times 5$ = torque, and $\times 6$ = WOB. (A) reflects a strong positive linear relationship, (B) reflects a weak relationship, and (C) reflects a negative relationship.

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{F1 score} = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \tag{4}$$

TP and TN represent the numbers of instances accurately recognized by the model as positive and negative, respectively. FP signifies the number of instances mistakenly identified as positive when they are truly negative. FN denotes the number of instances erroneously identified as negative when they are really positive. Equation 1, eq 2, eq 3, and eq 4 describe the definition of each parameter in the confusion matrix. Accuracy measures the proportion of correctly classified instances out of the total instances. While high accuracy is desirable, it may not be suitable for imbalanced data sets where one class dominates. Precision quantifies the ratio of TP predictions to the total predicted positives. High precision indicates that when the model predicts a positive outcome, it is likely to be correct, and precision is crucial when FPs are costly. Recall calculates the ratio of TP predictions to the actual positives in the data set. It is important in scenarios where missing positive instances are a concern. The F1 score is the harmonic mean of precision and recall. It provides a balanced assessment of a model's performance.
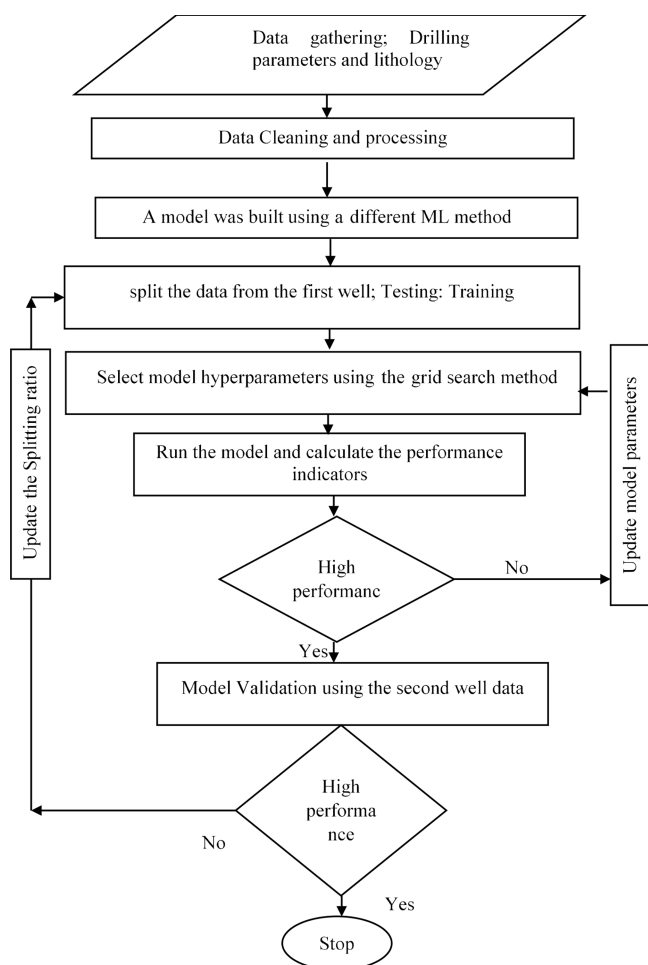
Figure 5. Models' development processes.

3. RESULTS AND DISCUSSION

3.1. GNB Model. GNB machine learning model was trained using data from well A, which was divided into training and testing data sets in a 70:30 ratio. The selected optimal hyperparameters were priors = 4 and var_smoothing = 1e−09. Figure 6 illustrates the confusion matrix for the training and testing data sets, demonstrating the accurate prediction of different formation lithologies by the GNB model, with only a

few instances being inaccurately predicted. Tables 3 and 4 provide a summary of the performance indicators for the

Table 3. Summary of Performance Indicators for Training Dataset Using the GNB Model

class	precision	recall	F1 score	total points
0	1	0.89	0.94	1006
1	0.93	0.99	0.96	1468
2	0.95	0.9	0.92	608
3	0.96	1	0.98	792
accuracy = 0.95				3874

Table 4. Summary of Performance Indicators for the Testing Dataset Using the GNB Model

class	precision	recall	F1 score	total points
0	0.99	0.89	0.94	424
1	0.94	0.98	0.96	633
2	0.94	0.91	0.93	262
3	0.95	1	0.97	342
accuracy = 0.95				1661

training and testing data sets, showing an overall accuracy of 0.95, along with precision, recall, and F1-score values exceeding 0.91 for the various classes. These results highlight the GNB model's ability to predict formation lithology based on drilling parameters.

3.2. LR Model. In a similar manner, the LR technique was employed using the data from well A, with the following optimal hyperparameters selected: solver = lbfgs, dual = false, tol = 0.0001, C = 1.0, fit_intercept = true, intercept_scaling = 1, class_weight = none. Figure 7 presents the confusion matrix for both the training and testing data sets, illustrating the LR model's accurate predictions in classifying formation lithology. Similar to the GNB results, the LR model exhibited a few inaccurately predicted points in the testing data set. The evaluation of performance indicators for the testing data set is summarized in Tables 5 and 6. Notably, the LR model demonstrated an impressive overall accuracy of 0.96, with precision, recall, and F1-score values surpassing 0.95 for different classes. These excellent outcomes underscore the LR model's remarkable capability in accurately predicting formation lithology based on drilling parameters.
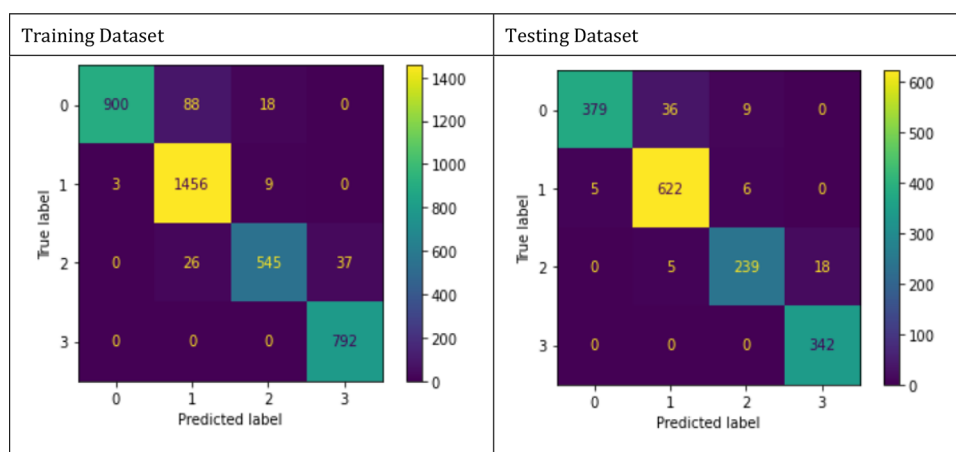

Figure 6. Confusion matrix for the training and testing data sets using the GNB model.
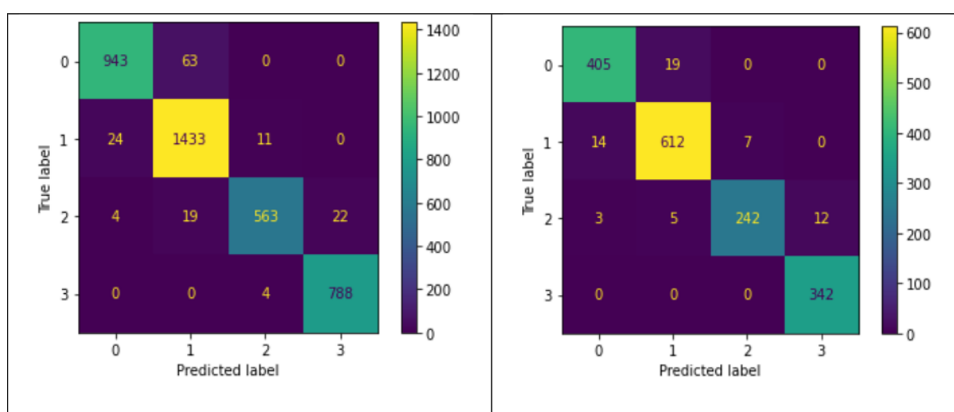
Figure 7. Confusion matrix for the training and testing data sets using the LR model.

Table 5. Summary of Performance Indicators for the Training Dataset Using the LR Model

class	precision	recall	F1 score	total points
0	0.97	0.94	0.95	1006
1	0.95	0.98	0.96	1468
2	0.97	0.93	0.95	608
3	0.97	0.99	0.98	792
accuracy = 0.96				3874

Table 6. Summary of Performance Indicators for the Testing Dataset Using the LR Model

class	precision	recall	F1 score	total points
0	0.99	0.89	0.94	424
1	0.94	0.98	0.96	633
2	0.94	0.91	0.93	262
3	0.95	1	0.97	342
accuracy = 0.95				1661

3.3. LDA Model. The data collected from well A served as the foundation for building the LDA model. The optimal hyperparameters for the LDA model were selected as follows: n_estimators = 100, max_depth = none, min_samples_split = 2, min_samples_leaf = 1, and max_features = 'sqrt'. The LDA model demonstrated a fair level of accuracy in predicting lithology for a majority of the points in both the training and testing data sets, as depicted in Figure 8. The confusion matrix for these data sets highlights the precise predictions achieved by

the LDA model in classifying formation lithology. In Tables 7 and 8, the high performance of the LDA model is evident, with an overall accuracy of 0.93 and precision, recall, and F1-score values close to 0.9 for different classes.

Table 7. Summary of Performance Indicators for the Training Dataset using the LDA Model

class	precision	recall	F1 score	total points
0	0.99	0.9	0.94	1006
1	0.92	0.99	0.95	1468
2	0.98	0.74	0.84	608
3	0.85	1	0.92	792
accuracy = 0.93				3874

Table 8. Summary of Performance Indicators for Testing Dataset Using the LDA Model

class	precision	recall	F1 score	total points
0	0.99	0.89	0.94	424
1	0.93	0.98	0.95	633
2	0.97	0.77	0.86	262
3	0.86	1	0.92	342
accuracy = 0.93				1661

3.4. Validation. The unseen data from well B were utilized to validate the developed ML models, namely, LDA, LR, and GNB, in predicting formation lithology based on drilling parameters. In Figure 9, the lithology predictions for well B
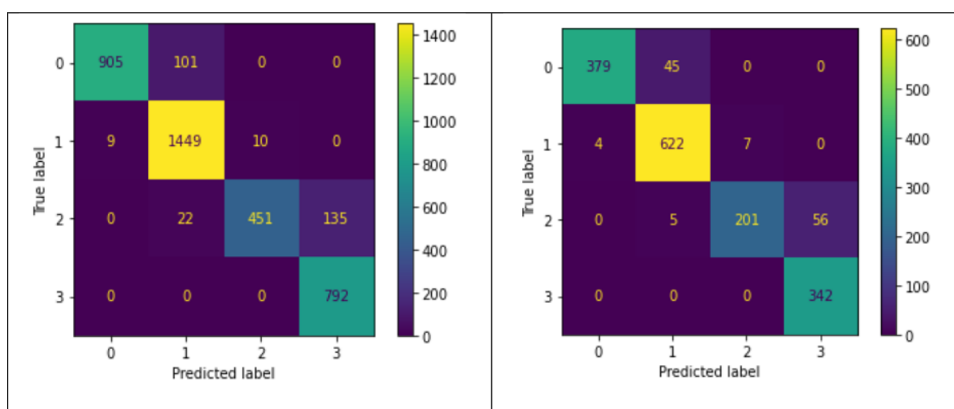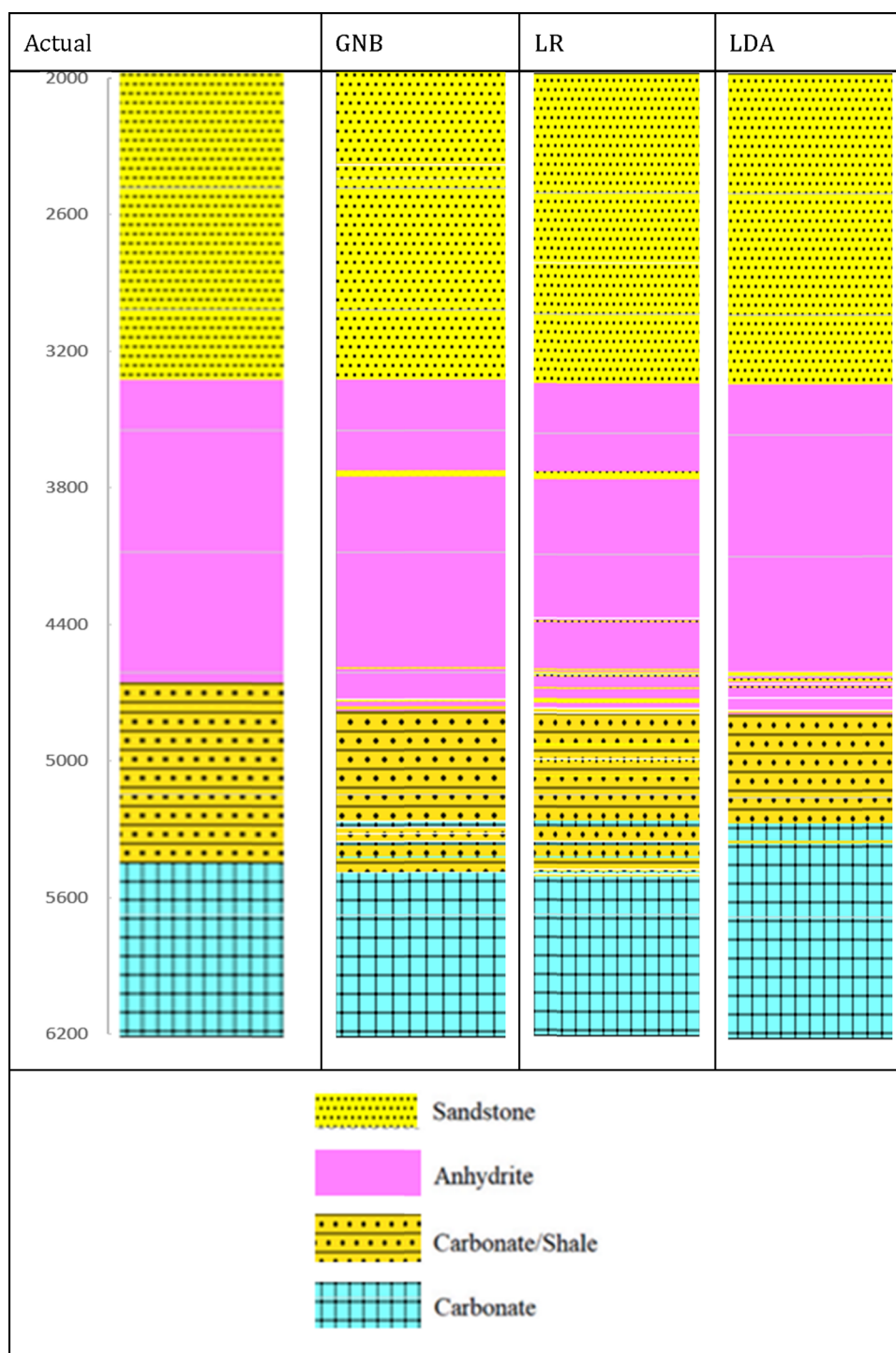

Figure 8. Confusion matrix for the training and testing data sets using the LDA model.

Figure 9. Different lithology prediction using ML models for the validation data set from well B.

generated by these models are compared to the actual lithology. All three models exhibited accurate predictions of sandstone and carbonate formations for all assigned data points based on the drilling parameters. However, when it came to predicting anhydrite and carbonate/shale formations, GNB demonstrated superior performance. GNB accurately predicted the lithology for most intervals, with a few intervals being misclassified. Notably, GNB occasionally predicted some anhydrite formations as sandstone and some points of carbonate/shale as pure carbonate formations. Although LDA and LR also performed well, they displayed a slightly increased tendency to

misclassify intervals, including misinterpreting some carbonate/shale formations as sandstone formations.

To assess the performance of the three models, a confusion matrix was generated for each model, as shown in Figure 10. The results demonstrated exceptional accuracy in predicting sandstone and carbonate formations across all three models. When it came to predicting anhydrite formations, GNB misclassified 34 points out of 1021, whereas LDA and LR misclassified 43 and 64 points, respectively.

In the case of the carbonate/shale formation prediction, GNB provided reasonably accurate lithology predictions with only a
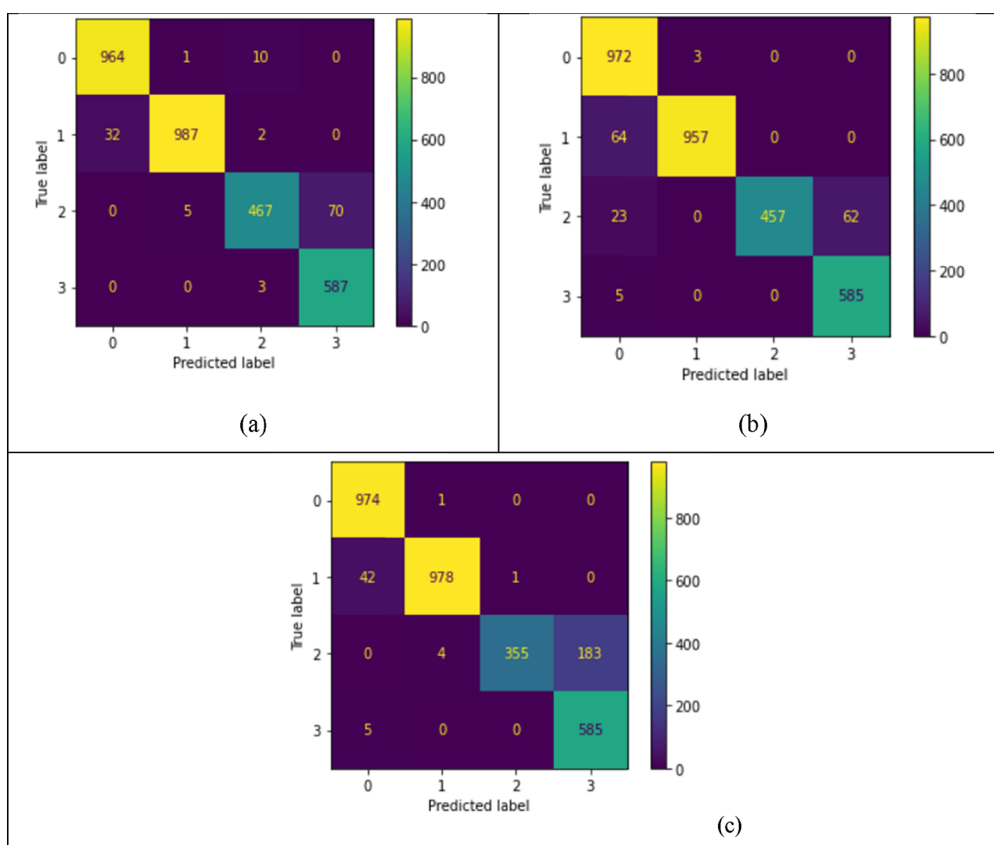
Figure 10. Confusion matrix for the validation data sets using (a) GNB, (b) LR, and (c) LDA models.

few instances misinterpreted as pure carbonate formations. However, LDA and LR encountered some difficulty in distinguishing between pure carbonate formations and shale/carbonate formations, with misclassification around 20% of the data. These findings were further supported by various performance indicators from the confusion matrices as summarized in Table 9, Table 10, and Table 11.

Table 9. Summary of Performance Indicators for Validation Dataset Using the GNB Model

class	precision	recall	F1 score	total points
0	0.97	0.99	0.98	975
1	0.99	0.97	0.98	1021
2	0.97	0.86	0.91	542
3	0.89	0.99	0.94	590
accuracy = 0.96				3128

Table 10. Summary of Performance Indicators for Validation Dataset Using the LR Model

class	precision	recall	F1 score	total points
0	0.91	1	0.95	975
1	1	0.94	0.97	1021
2	1	0.84	0.91	542
3	0.9	0.99	0.95	590
accuracy = 0.95				3128

Table 9 displays the confusion matrix for the GNB model. For the identification of sandstone lithology (class 0), the precision stands at 0.97, denoting an accuracy of about 97% in predicting this class. The recall rate of 0.99 signifies the model's adeptness

Table 11. Summary of Performance Indicators for the Validation Dataset Using the LDA Model

class	precision	recall	F1 score	total points
0	0.95	1	0.98	975
1	0.99	0.96	0.98	1021
2	1	0.65	0.79	542
3	0.76	0.99	0.86	590
accuracy = 0.92				3128

at capturing almost all instances of this class. A robust equilibrium between precision and recall is evident from the F1 score of 0.98. Moving on to anhydrite lithology (class 1) detection, the model showcases high precision at 0.99, effectively reflecting its accuracy in predicting this class. The recall value of 0.97 underlines the model's capability to capture a substantial portion of actual instances. The F1 score, mirroring the preceding classes, reaches 0.98, showcasing a harmonious trade-off between precision and recall. Likewise, in the prediction of carbonates/shale formations (class 2), the model's precision of 0.97 indicates a 97% accuracy rate when predicting this class. While the recall value of 0.86 suggests that the model captures most instances, it might miss a few. An F1 score of 0.91 portrays a balanced alignment between precision and recall for this class. Shifting focus to class 3, representing carbonate formation, the model attains a precision of 0.89, indicating relatively accurate predictions for this class. The high recall rate of 0.99 implies the model's proficiency in capturing nearly all instances of this class. An F1 score of 0.94 attests to a sturdy equilibrium between precision and recall. The model's overall accuracy is 0.96, signifying its capacity to accurately predict 96% of total points. This high accuracy, coupled with consistently

elevated precision and recall values for most classes, underscores the model's proficiency across diverse classes. The F1 scores for each class emphasize the harmony between precision and recall, substantiating a robust performance. Additionally, the well-balanced trade-offs between precision and recall validate the model's effectiveness in both accurate classification and capturing true class instances.

In contrast, Table 11 presents results indicating lower model performance in the case of LDA. Examining the prediction for class 0, the precision stands at 0.95, signifying that when the model predicts this class, it is accurate around 95% of the time. The recall value of 1 reflects the model's adeptness in identifying all instances of this class. A balanced trade-off between precision and recall is evident from the F1 score of 0.98. Transitioning to class 1, the precision impressively reaches 0.99, highlighting the model's precision in predicting this class. A recall rate of 0.96 indicates the model's ability to capture a substantial portion of actual instances. Similarly, for class 0, the F1 score remains 0.98, indicating a strong overall performance. However, in the context of class 2, precision stands at 1, suggesting consistent accuracy when the model predicts this class. Yet, the lower recall rate of 0.65 suggests that the model misses some instances of this class. The F1 score, at 0.79, reveals a balance between precision and recall for this class. In the scenario of class 3, the precision reaches 0.76, indicating a moderate accuracy in predicting this class. A high recall rate of 0.99 suggests the model's effectiveness in capturing the majority of class instances. An F1 score of 0.86 demonstrates a reasonable equilibrium between precision and recall. The overall model accuracy is 0.92, illustrating its capability to accurately predict 92% of the total points. However, it is important to note that accuracy alone might not provide a comprehensive understanding due to class imbalance. Hence, considering precision and recall is essential to discern where the model excels and where it encounters challenges. Notably, the model demonstrates particular strength in predicting class 1, while class 2 showcases a trade-off between precision and recall.

The current study showed the applicability of using machine learning to predict the formation tops and lithology. Leveraging machine learning for formation tops and lithology predictions while drilling enhances real-time insights, accuracy, cost efficiency, and risk mitigation. These advantages collectively drive efficient drilling operations, elevate decision-making, and augment wellbore positioning precision and geological understanding. It provides real-time insights by analyzing large volumes of data as drilling progresses. This immediate understanding of formation tops and lithology changes empowers drillers to make informed decisions promptly, enhancing wellbore accuracy and minimizing drilling risks. Machine learning models excel in accuracy due to their ability to learn from historical data, reducing errors associated with manual interpretation. This leads to reliable predictions and improved wellbore positioning.

The main limitation to applying these methods is data availability. The ML methods were used to predict four different lithologies, namely, sandstone, anhydrite, shale, and carbonates. Hence, low-accuracy results may be found if another lithology is introduced to the models, in addition, the input data should be within the data ranges summarized in Table 1.

4. CONCLUSIONS

Accurate prediction of formation tops and lithology plays a crucial role in optimizing drilling operations, reducing costs, and mitigating risks in hydrocarbon exploration and production.

However, traditional methods for identifying formation lithology suffer from high costs, lower accuracy, extensive manpower requirements, and time or depth lags, hindering real-time estimation. To address these challenges, this study employed various classification ML techniques to predict formation lithology and tops using readily available drilling parameters. The key findings of this research are summarized below:

- Three machine learning models, namely, GNB, LR, and LDA, were developed and evaluated using real-field data from two wells.
- The GNB model exhibited exceptional performance in accurately predicting formation lithology, achieving an accuracy along with high precision, recall, and F1-score values for different lithology classes.
- The models' accuracy in the validation data set was roughly 0.96, 0.95, and 0.92 for GNB, LR, and LDA, respectively.
- The LDA and LR models showcased precise predictions for sandstone and carbonate lithologies, although there were some misclassifications, approximately 5% for anhydrite and around 20% for carbonate/shale formations.
- The results emphasized the significant influence of drilling parameters such as WOB and ROP on lithology identification.

Overall, this study highlights the efficacy of the developed ML models in accurately predicting real-time formation tops and lithologies using readily available drilling parameters. By leveraging existing instantaneous drilling data, this approach offers high accuracy and cost-effectiveness. The implementation of ML techniques in drilling operations enables improved lithology prediction, leading to more informed decision-making and enhanced operational efficiency.

■ AUTHOR INFORMATION

Corresponding Author

Salaheldin Elkatatny − *College of Petroleum Engineering and Geosciences and Center for Integrative Petroleum Research, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;* ⦿ orcid.org/0000-0002-7209-3715; Phone: +966-594663692; Email: elkatatny@kfupm.edu.sa

Authors

Ahmed Farid Ibrahim − *College of Petroleum Engineering and Geosciences and Center for Integrative Petroleum Research, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;* ⦿ orcid.org/0000-0001-7258-8542

Ashraf Ahmed − *College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;* ⦿ orcid.org/0000-0003-1680-0179

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.3c03725

Author Contributions

Conceptualization, S.E. and A.F.I.; methodology, A.F.I.; formal analysis, A.F.I. and A.A.; investigation, A.F.I. and A.A.; resources, S.E.; data curation, A.F.I.; writing—original draft preparation, A.F.I. and A.A.; writing—review and editing, S.E. and A.A.; visualization, A.F.I.; supervision, S.E. All authors have read and agreed to the published version of the manuscript.

Notes

The authors declare no competing financial interest.

NOMENCLATURE

DSR: drill string rotation
GNB: Gaussian naive Bayes
GR: gamma ray
LDA: linear discriminant analysis
LR: logistic regression
LWD: logging while drilling
ML: machine learning models
Q: pumping rate
ROP: rate of penetration
SPP: standpipe pressure
WOB: weight on bit

REFERENCES

(1) Al-AbdulJabbar, A.; Elkatatny, S.; Mahmoud, M.; Abdelgawad, K.; Al-Majed, A. A Robust Rate of Penetration Model for Carbonate Formation. *J. Energy Resour. Technol.* **2019**, *141* (4), No. 042903.

(2) Bourgoyne, A. T. Jr.; Millheim, K. K.; Chenevert, M. E.; Young, F. S. Jr. *Applied Drilling Engineering*; Society of Petroleum Engineers: TX, 1986.

(3) Hossain, M. E.; Al-Majed, A. A. *Fundamentals of Sustainable Drilling Engineering*; Wiley-Scrivener: USA, 2015 DOI: 10.1002/9781119100300.

(4) Al-Baldawi, B. Reservoir Characterization and Identification of Formation Lithology from Well Log Data of Nahr Umr Formation in Luhais Oil Field Southern Iraq. *Iraqi J. Sci.* **2016**, *57* (1B), 436−445.

(5) Mishra, A.; Sharma, A.; Patidar, A. K. Evaluation and Development of a Predictive Model for Geophysical Well Log Data Analysis and Reservoir Characterization: Machine Learning Applications to Lithology Prediction. *Nat. Resour. Res.* **2022**, *31* (6), 3195−3222.

(6) Cannon, S. Introduction. In *Petrophysics: A Practical Guide*; Wiley-Blackwell, 2015; pp 8−10 DOI: 10.1002/9781119117636.

(7) Holstein, E. D.; Warner, H. R. Overview of Water Saturation Determination For the Ivishak (Sadlerochit) Reservoir, Prudhoe Bay Field. In *SPE Annual Technical Conference and Exhibition*; Society of Petroleum Engineers, 1994. DOI: 10.2118/28573-MS.

(8) Zhu, L.; Li, H.; Yang, Z.; Li, C.; Ao, T. Intelligent Logging Lithological Interpretation With Convolution Neural Networks. *Petrophysics* **2018**, *59* (6), 799−810.

(9) Losoya, Z. E.; Vishnumolakala, N.; Noynaert, S. F.; Medina-Cetina, Z.; Bukkapatnam, S.; Gildin, E. Automatic Identification of Rock Formation Type While Drilling Using Machine Learning Based Data-Driven Models 2021 DOI: 10.2118/201020-MS.

(10) Nanjo, T.; Tanaka, S. Carbonate Lithology Identification with Machine Learning. *Abu Dhabi Int. Pet. Exhib Conf.* 2019, D021S060R001, DOI: 10.2118/197255-MS.

(11) Wang, K.; Zhang, L. Predicting Formation Lithology from Log Data by Using a Neural Network. *Pet Sci.* **2008**, *5* (3), 242−246.

(12) Thiercelin, M. J.; Plumb, R. A. Core-Based Prediction of Lithologic Stress Contrasts in East Texas Formations. *SPE Formation Evaluation* **1994**, *9* (04), 251−258.

(13) Thomas, A.; Rider, M.; Curtis, A.; Macarthur, A. A Novel Approach for Automated Lithology Extraction from Core Photographs. In *72nd EAGE Conference and Exhibition incorporating SPE EUROPEC 2010*; European Association of Geoscientists & Engineers, 2010. DOI: 10.3997/2214-4609.201400891.

(14) Arabjamaloei, R.; Edalatkha, S.; Jamshidi, E.; Nabaei, M.; Beidokhti, M.; Azad, M. Exact Lithologic Boundary Detection Based on Wavelet Transform Analysis and Real-Time Investigation of Facies Discontinuities Using Drilling Data. *Pet Sci. Technol.* **2011**, *29* (6), 569−578.

(15) Lentz, N.; Carr, R.; Yarbrough, L.; Neset, K.; Lucero, B.; Kirst, T. On-Site XRF Analysis of Drill Cuttings in the Williston Basin. In *Proceedings of the 2nd Unconventional Resources Technology Conference*; American Association of Petroleum Geologists: Tulsa, OK, USA, 2014. DOI: 10.15530/urtec-2014-1934308.

(16) Christeson, G. L.; Morgan, J. V.; Gulick, S. P. S. Mapping the Chicxulub Impact Stratigraphy and Peak Ring Using Drilling and Seismic Data. *J. Geophys. Res. Planets* **2021**, *126* (8), No. e2021JE006938, DOI: 10.1029/2021JE006938.

(17) Rao, V. V.; Raju, J. S.; Rao, B. P.; Rao, P. K. Bed Rock Investigation by Seismic Refraction Method - A Case Study. *J. Indian Geophys. Union* **2004**, *8* (3), 223−228.

(18) Pixler, B. O. Mud Analysis Logging. *Journal of Petroleum Technology* **1961**, *13* (04), 323−326.

(19) Elkatatny, S. New Approach to Optimize the Rate of Penetration Using Artificial Neural Network. *Arab J. Sci. Eng.* **2018**, *43* (11), 6297−6304.

(20) Joshi, D.; Patidar, A. K.; Mishra, A.; Mishra, A.; Agarwal, S.; Pandey, A.; Dewangan, B. K.; Choudhury, T. Prediction of Sonic Log and Correlation of Lithology by Comparing Geophysical Well Log Data Using Machine Learning Principles. *GeoJournal* **2021**, Springer Nature DOI: 10.1007/s10708-021-10502-6.

(21) Khilrani, N.; Prajapati, P.; Patidar, A. K. Contrasting Machine Learning Regression Algorithms Used for the Estimation of Permeability from Well Log Data. *Arabian Journal of Geosciences* **2021**, *14* (20), 2070.

(22) Patidar, A. K.; Singh, S.; Anand, S. Subsurface Lithology Classification Using Well Log Data, an Application of Supervised Machine Learning; 2023; pp 227−240 Springer: Singapore. DOI: 10.1007/978-981-99-1620-7_18.