Article

Data-Driven Framework for Real-time Rheological Properties Prediction of Flat Rheology Synthetic Oil-Based Drilling Fluids

Ahmed Abdelaal, Ahmed Farid Ibrahim, and Salaheldin Elkatatny*

Read Online

ACCESS | 📊 Metrics & More | 📖 Article Recommendations

ABSTRACT: Appropriate mud properties enhance drilling efficiency and decision quality to avoid incidents. The detailed mud properties are mainly measured in laboratories and are usually measured twice a day in the field and take a long time. This prevents real-time mud performance optimization and adversely affects proactive actions. As a result, it is critical to evaluate mud properties while drilling to capture mud flow dynamics. Unlike other mud properties, mud density (MD) and Marsh funnel viscosity (MFV) are frequently evaluated every 15−20 min in the field. The goal of this study is to predict the rheological properties of flat rheology synthetic oil-based mud (SOBM) in real time using machine learning (ML) techniques such as random forest (RF) and decision tree (DT). A proposed approach is followed to first predict the viscometer readings at 300 and 600 RPM (R_{600} and R_{300}) and then estimate the other mud properties using the existing equations in the literature. A set of data contained MD, MFV, and viscometer readings (R_{300} and R_{600}) for different samples from the same mud type. The mud samples were collected after going through a shale shaker. MD and MFV are measured by a mud balance and a Marsh funnel, respectively, while rheology is evaluated using a viscometer. The data were randomly split into training, testing, and validation data sets. The ML models' performance was evaluated through average absolute percentage error (AAPE) and correlation coefficient (R). The proposed models predicted the viscometer readings as a middle stage with a low AAPE that did not exceed 4.5% for both models. The suggested models forecasted the rheological properties with a good degree of accuracy, with an AAPE being less than 7% for most of the parameters. The proposed models can save costs and time since there is no need to include additional tools in the rig location. Furthermore, these models will significantly aid in avoiding serious problems and achieving better rig hydraulics and hole cleaning, which in turn will technically and economically enhance drilling operations.
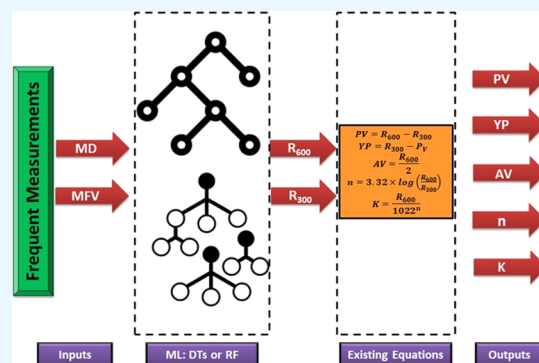
1. INTRODUCTION

Drilling fluids are a mixture of a base fluid and additional ingredients in specific ratios used while drilling. Several materials are added to adjust the mud properties such as, but are not limited to, weighting agents for density control, viscosifiers for rheology control (e.g., plastic viscosity (PV), yield point (YP), and gel strength), and lost circulation materials. The cost of drilling fluids is relatively low, but suitable mud programs and modification of mud properties while drilling have an impact on the entire well cost.[1] Although mud represents 5−15% of total drilling expenditures, it may address most of the drilling problems. Drilling fluids are mainly classified as water-based mud (WBM) and oil-based mud (OBM). OBM typically contains a base oil representing the external continuous phase, a saline aqueous solution representing the internal phase, emulsifiers at the interface, additives for suspension, weighting materials, and control additives for oil-wetting, fluid loss, and rheology. OBM has two main categories named invert-emulsion and all-oil drilling fluids.[2] An inverted emulsion mud contains approximately 50:50 to 95:5 volume oil-to-water ratio. An all-oil mud contains 100% oil and does not have an aqueous internal phase. The inverted emulsion drilling fluid is used to tackle some drilling problems, such as shale instability, minimize damage to water zones, and protect casing and tubing against corrosion.[3,4] Invert emulsion mud is characterized by its low toxicity, and brine is added to control salinity to prevent water molecules from invading formations.[5] It is mainly used to drill high-pressure high-temperature wells owing to its thermal stability that outperforms WBM and can be used in drilling up to 400 °F.[6]

Flat rheology SOBM is an inverted emulsion mud possessing a flat rheology profile that maintains consistent fluid properties that are independent of temperature and pressure. This flat rheology profile is considered a major improvement over conventional invert systems, which are known to have higher

viscosities at low temperatures and lower rheology at high bottom-hole temperatures. In comparison to conventional invert systems, flat rheology SOBM systems provide lower ECD, leading to fewer drilling problems and minimizing the loss of whole drilling fluids. The advantage of a flat rheology profile is that it allows pumping at higher flow rates compared to conventional mud systems, thus increasing ROP while providing lower ECD values. It can also be specially formulated and customized to achieve the flatness of the gel strength structure. The flat rheology SOBM develops a gel strength structure that will reduce cutting slip velocities when pumps are off and prevents sagging especially in deviated and extended reach wellbores. The flat rheology SOBM system is different from standard invert systems in two main areas. The first is that the emulsifier and wetting agent package has a greater solid tolerance as compared to conventional systems. The second is that the content of organophilic clay has been significantly reduced using a mixture of high-performance clays.

Most drilling muds show behaviors that can be described as non-Newtonian fluids. Dilatant (shear-thickening) and pseudo-plastic (shear-thinning) fluids are both non-Newtonian in which the relationship of viscosity versus shear rate is not linear. A typical mud has a yield point and exhibits shear-thinning behavior. At elevated shear rates, most models simulate a typical mud well. Differences among the models are mostly observed at small shear rates that are most important for hole cleaning and weighting material suspension. The Bingham plastic (BP) model has a yield point; however, it does not accurately describe the fluid behavior at low shear rates. The power law (PL) model more accurately describes the behavior at low shear rates; however, it does not contain a yield point that gives inaccurate results at extremely low shear rates. A typical mud has a behavior between the BP model and the PL model. This type of behavior can be approximated by Herschel Bulkley's (HB) model.[7] The fluid consistency index (K) and flow behavior index (n) are the parameters of the PL model. Generally, K describes the fluid thickness and is somehow comparable to effective viscosity. If mud viscosity increases, then K should increase to describe the stress/rate relationship. n represents the degree of non-Newtonian behavior. When n equals 1, the PL model turns out to be the Newtonian model. If n is higher than 1, the fluid is described as a dilatant, where effective viscosity goes up with an increasing shear rate. For mud, pseudoplastic behavior is common and characterized by an n-value between 0 and 1.[7] Rheological models are mathematical equations between shear rate and shear stress that are used to predict fluid flow behavior.[8] Mud rheology in the oil and gas industry can be described by different models such as the BP model (eq 1), PL model (eq 2), and HB (yield power law, YPL) model (eq 3).[9,10] The HB model more precisely describes the mud rheological behavior than other models and gives acceptable estimations of measured rheological properties for both WBM and nonaqueous mud. The HB model collapses to the BP model when n equals 1 and the PL model when $\tau_0 = 0$. The HB model is complex and requires three measurements as a minimum to solve for its unknowns.[9]

$$\tau = YP + PV \times (\gamma) \tag{1}$$

$$\tau = K(\gamma)^n \tag{2}$$

$$\tau = \tau_0 + K \times (\gamma)^n \tag{3}$$

where τ_0 is the fluid yield stress, lbf/100 ft^2 or Pa, and γ is the shear rate, 1/s.

The advancement of technology has made it possible to measure, collect, and store increasingly vast quantities of information. The term drilling data analytics refers to the process of utilizing drilling data to gain valuable insights and improve the performance of drilling operations. The fundamental idea behind data-driven models is to identify connections between the input and output variables of a system, without relying on an explicit understanding of the system's physical behavior. Data-driven models were introduced in the petroleum industry for different purposes, such as sonic velocities and times prediction,[11−13] static Young's modulus estimation,[14] and oil production and dynamic risk profile prediction in production systems.[15] The use of machine learning techniques in data-driven models can provide significant benefits compared to conventional analytical or numerical models. These benefits include adaptable model inputs, improved accuracy in predicting outcomes, and the capability to identify hidden patterns.[16]

ML is a discipline that straddles the boundaries of computer science, statistics, artificial intelligence, and mathematical optimization. The structure of statistical and differential equation-based models is determined by assumptions about the system to be modeled. The strategy in ML is to let data govern the model. Rather than establishing assumptions about the model's shape, an algorithm is used to grow up the model structure.[17] ML has different techniques, such as ANN, functional networks, support vector machines, ANFIS, RF, and DT, which show a good performance for prediction and classification.[18] Different ML models were developed for various purposes, such as fracture pressure prediction,[19] density log generation,[20] resistivity prediction,[21−23] pore pressure gradient prediction while drilling,[24−27] formation lithology prediction,[28] coal pay zones identification,[29] rock geomechanical parameters prediction,[30] stuck pipe prediction,[31] equivalent circulation density prediction,[32,33] rheological properties prediction,[2,34] carbon dioxide solubility estimation in ionic liquids,[35] compressibility factor estimation,[36] ROP prediction,[37] water saturation prediction,[38] early kick detection and estimation in managed pressure drilling,[39] and lithology classification.[40] ML is widely used in the petroleum industry at large and drilling engineering in different areas, such as drilling fluids, drilling problems, and well control.[16,41] In this study, DTs and RF were employed to provide two predictive models for viscometer readings that can be used to estimate different rheological properties, such as PV, YP, apparent viscosity (AV), n, and K, using the existing equations in the literature.

Drilling fluid literature studies have been flooded with many publications over the last two decades that presented models for mud rheology. Empirical, theoretical, and ML models, as well as ensembles of models, have all been used in the development of these models. Mud rheology modeling has the advantage of predicting mud behavior in extreme conditions such as high stress, high temperature, and high pressure when conducting experiments is difficult.[42] Table 1 summarizes some efforts done to predict the rheological properties of various mud types using ML.

Real-time/dynamic/operational data feed into operations can enhance the economics, safety, and integrity of the systems in the oil and gas industry with ML method incorporation. Mamudu et al.[15] introduced prediction models for oil production and dynamic risk profile in production systems.

Table 1. Summary of the Related Research on Mud Rheological Property Prediction

reference	mud type	technique	inputs	outputs
Razi et al.[43]	WBM	artificial neural networks (ANN)	shear stress at 600 rpm, shear rate, temperature, and concentration	PV, YP, and AV
Elkatatny et al.[44]	invert emulsion mud	ANN	MFV, solid content, and MD	PV, YP, and AV
Elkatatny[45]	KCl polymer mud	ANN	MF viscosity, solid content, and MD	PV, YP, and AV
Elkatatny and Mahmoud[46]	NaCl WBM	ANN	MF viscosity, solid content, and MD	PV, YP, and AV
Bispo et al.[47]	WBM	ANN	temperature, concentrations of barite, bentonite, and xanthan gum.	AV
Al-Azani et al.[48]	OBM	ANN	MFV, MD, solid percent	PV and AV
Avcı[49]	WBM	ANN	shear rate, and temperature	Shear stress
Abdelgawad et al.[50]	bentonite spud mud	ANN	MFV, MD, and solid percent	PV, YP, and AV
Elzenary[51]	invert emulsion mud	ANN and adaptive neuro fuzzy inference system (ANFIS)	MFV, MD, and solid percent	PV, YP, and AV
Gowida et al.[52]	CaCl$_2$ brine	ANN	MFV and MD	PV, YP, and AV
Gomaa et al.[53]	high-overbalanced bridging mud	ANN	MFV and MD	PV, YP, and AV
Oguntade et al.[54]	mud with modified biopolymer	ANN	water volume, the concentration of bentonite, and the concentration of biopolymer	PV, YP, and AV
Gowida et al.[55]	high-bentonite mud	ANN	MFV and MD	PV, YP, and AV
Alsabaa et al.[2]	invert emulsion mud	ANFIS	MFV and MD	PV, YP, and AV
Elkatatny[56]	NaCl mud	ANN	MFV and MD	PV, YP, and AV
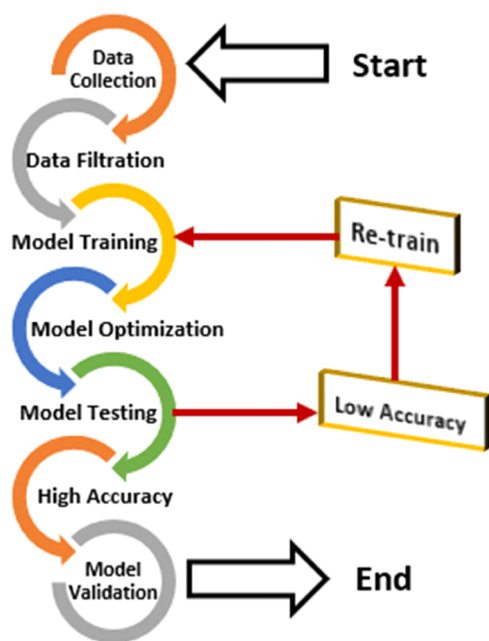
Figure 1. Methodology flow chart for the prediction models.

The authors mentioned that the suggested approach may provide field operators with an improved chance to receive current predictions regarding the probability of any potential production decline at any given moment during production activities. The proposed method is cost-effective that reduces the computational workload involved in history-matching processes. It also fills the gaps in current oilfield development systems by providing dynamic risk forecasts and production predictions. Onalo et al.[11] provided a data-driven model for sonic wave prediction as these logs are not available in all wells.
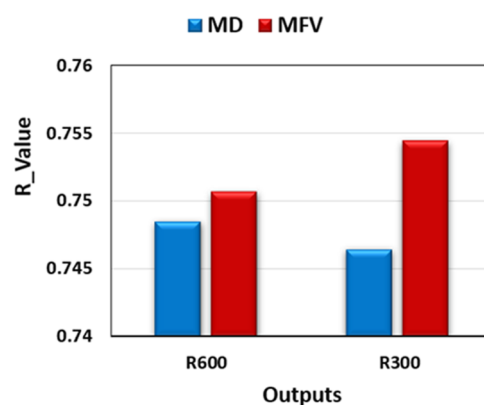
Figure 2. Relationship of outputs (R_{600} and R_{300}) with each input (MD and MFV) for the SOBM in terms of R.

The authors described their model as a cost-effective and robust tool for formation evaluation and can provide log data if the data have been corrupted for any reason.[11] The previous two examples show the importance of data-driven models in terms of assessing and reducing risks, cost reduction, safety enhancement, and nonproductive time reduction.

For proper hydraulic control, accurate and frequent measurements of mud rheological parameters are required. It is particularly vital for intelligent drilling since it provides mud data for establishing the ROP optimization model. Proper mud properties can enhance drilling efficiency and minimize drilling problems. The mud properties are mostly measured in the lab, making real-time mud optimization and decision-making challenging. If mud properties cannot be monitored and decisions cannot be quickly taken, ROP may slow down, potentially resulting in accidents and significant financial losses. Consequently, it is significant to find a way to estimate mud
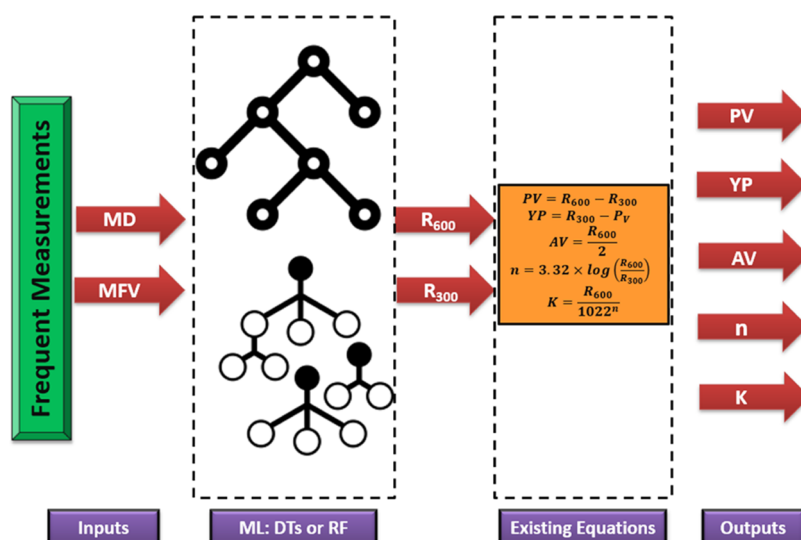
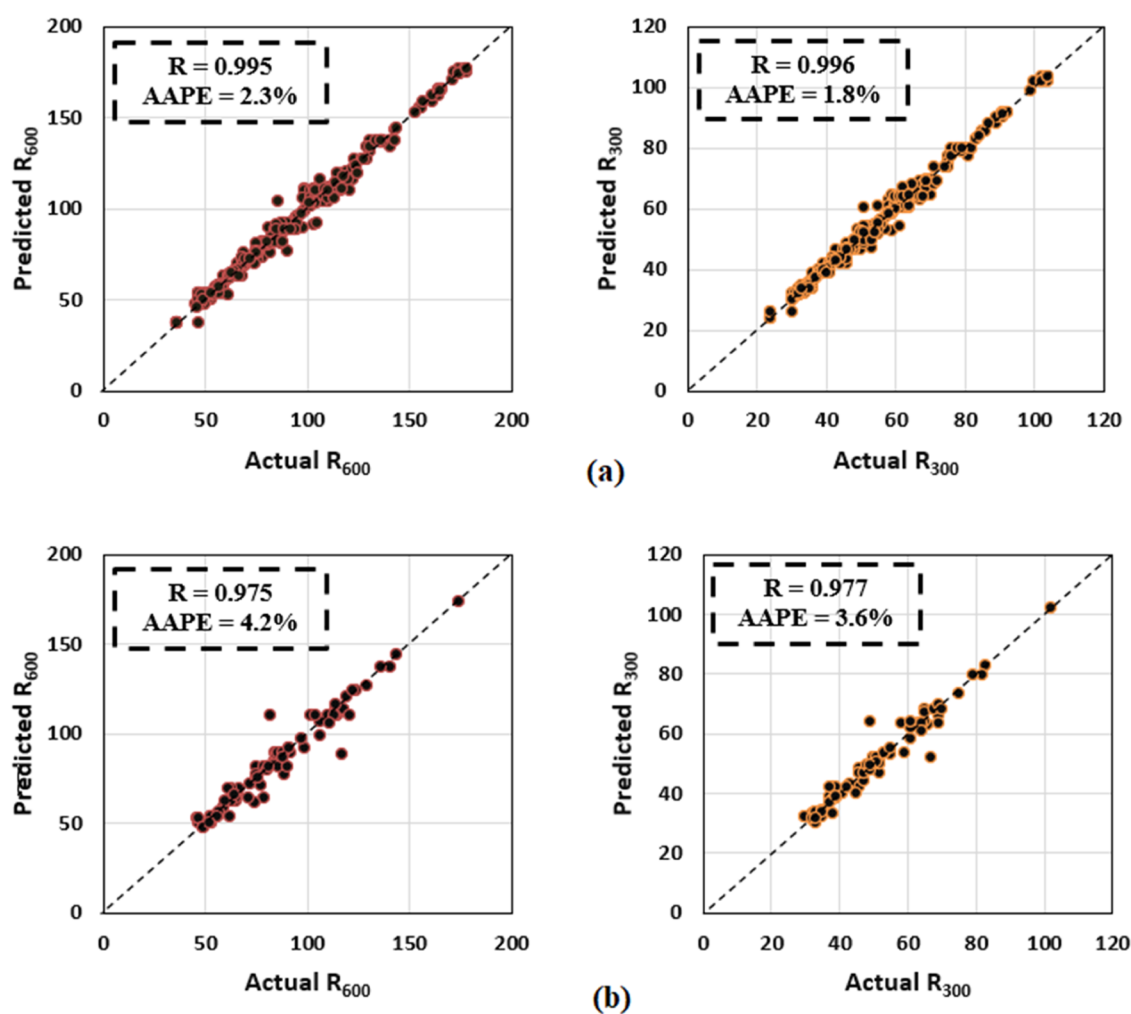**Figure 3.** Summary of the prediction framework in this study.

Figure 4. Cross-plots of the predicted versus actual viscometer readings using the DT-based model for training (a) and testing (b).

properties while drilling. Rheological properties are estimated using a manually controlled rotational viscometer. Given the complexity of indirect determination of drilling fluid rheology in the field, various approaches have been introduced in the literature. Recently, some researchers have investigated the real-

time estimation of mud rheology. The real-time estimation of mud rheology includes four main techniques: (1) online rotational Couette viscometer, (2) pipe viscometer, (3) Marsh funnel-based models (ML or mathematical models), and (4) tuning fork technology.[57] The goal of this work is to predict the
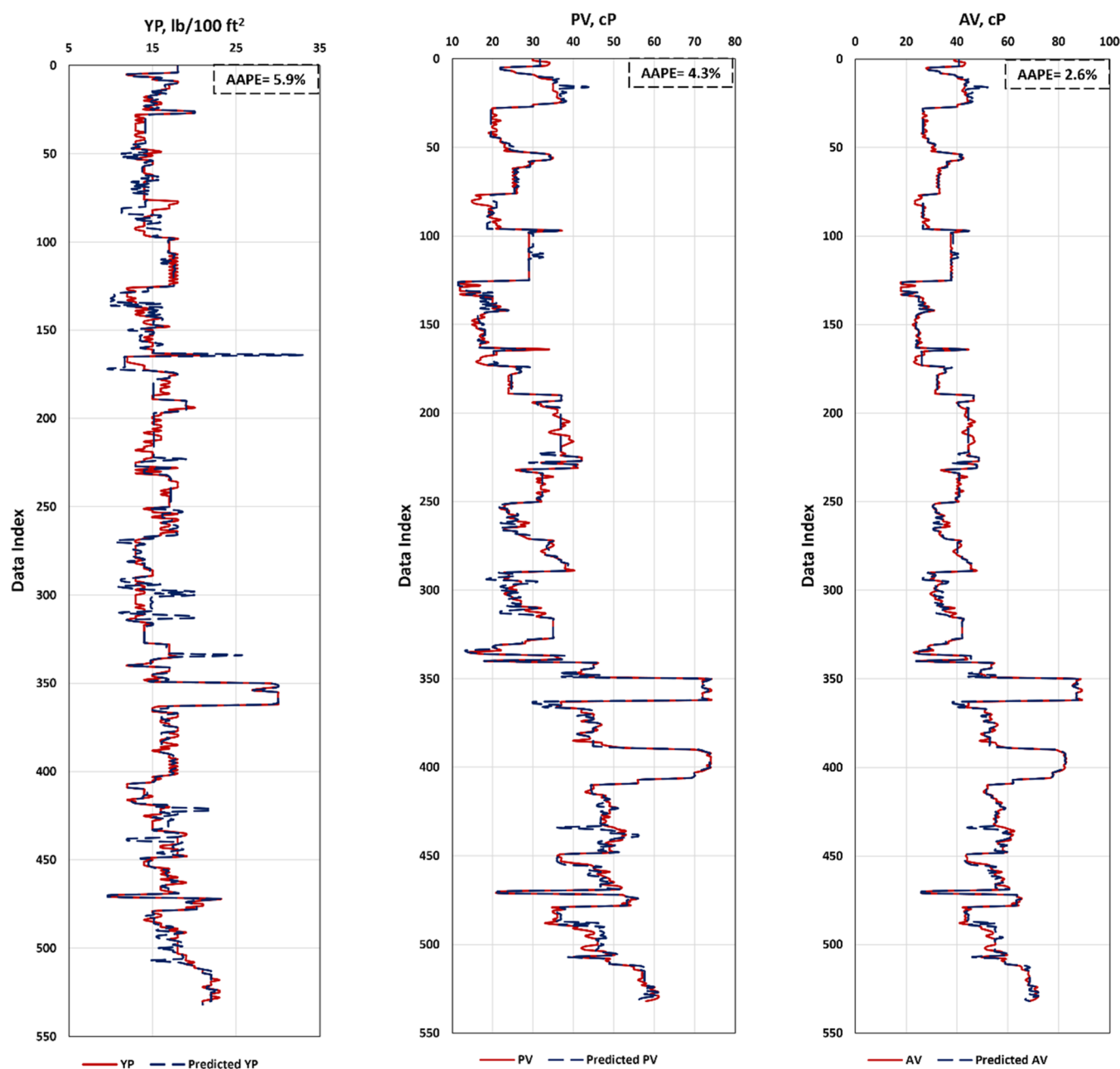
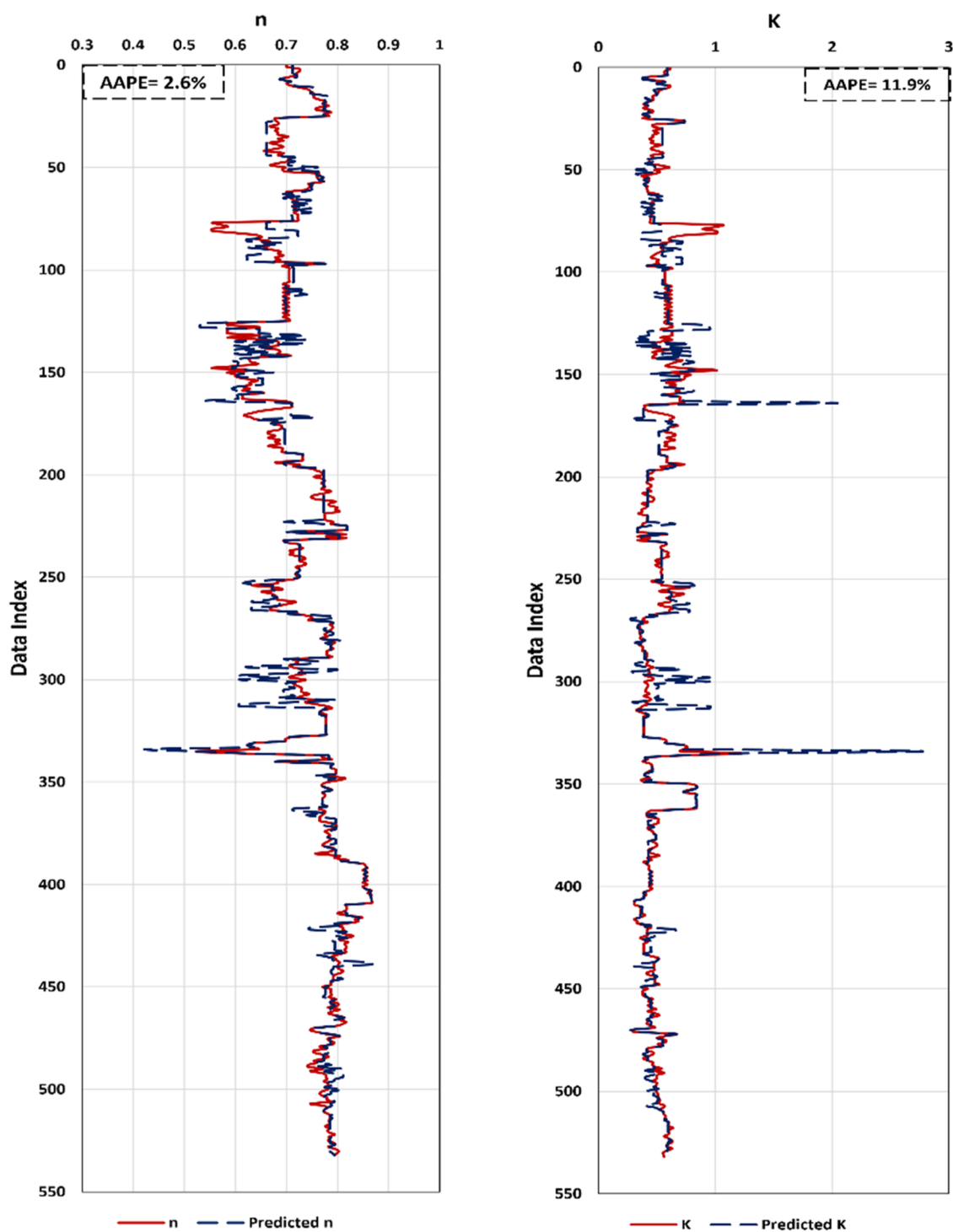**Figure 5.** Profiles of the predicted versus actual values of YP, PV, and AV using the developed DT-based model.

rheological properties of flat rheology SOBM in real-time by ML using frequent measurements such as MD and MFV, usually, measured every 15−20 minutes. A proposed approach is introduced to first predict the viscometer readings (R_{600} and R_{300}) as an intermediate stage and then estimate the other mud properties using the existing equations in the literature. The ML models can save costs and time as there is no need to use additional tools in the rig location. In addition, these models will help to avoid serious problems and achieve better rig hydraulics and hole cleaning, which in turn will technically and economically enhance drilling operations. Section 2 shows the methodology, data description along with their statistical analysis, the ML approaches, and their optimization scheme, followed by Section 3, which introduces the obtained results of the models with an in-depth discussion, Section 4 is designed for

the validation stage of the proposed models, and lastly, Section 5 summarizes the work outcomes.

2. MATERIALS AND METHODOLOGY

This work used field data that were measured while drilling using various flat rheology SOBM samples. Figure 1 briefly shows the workflow followed to develop two ML models (RF and DT) for rheological property prediction in real time. The workflow started with data acquisition, data cleaning, and filtration to get high-quality data, the training stage for ML models and optimizing the model hyperparameters, testing model accuracy, re-training the model in case of low accuracy, and finally validating the proposed models using a blind holdout data set that was not involved in training and testing stages.

2.1. Data Description and Preparation. Any learning algorithm is as good as its input parameters, and usually, input
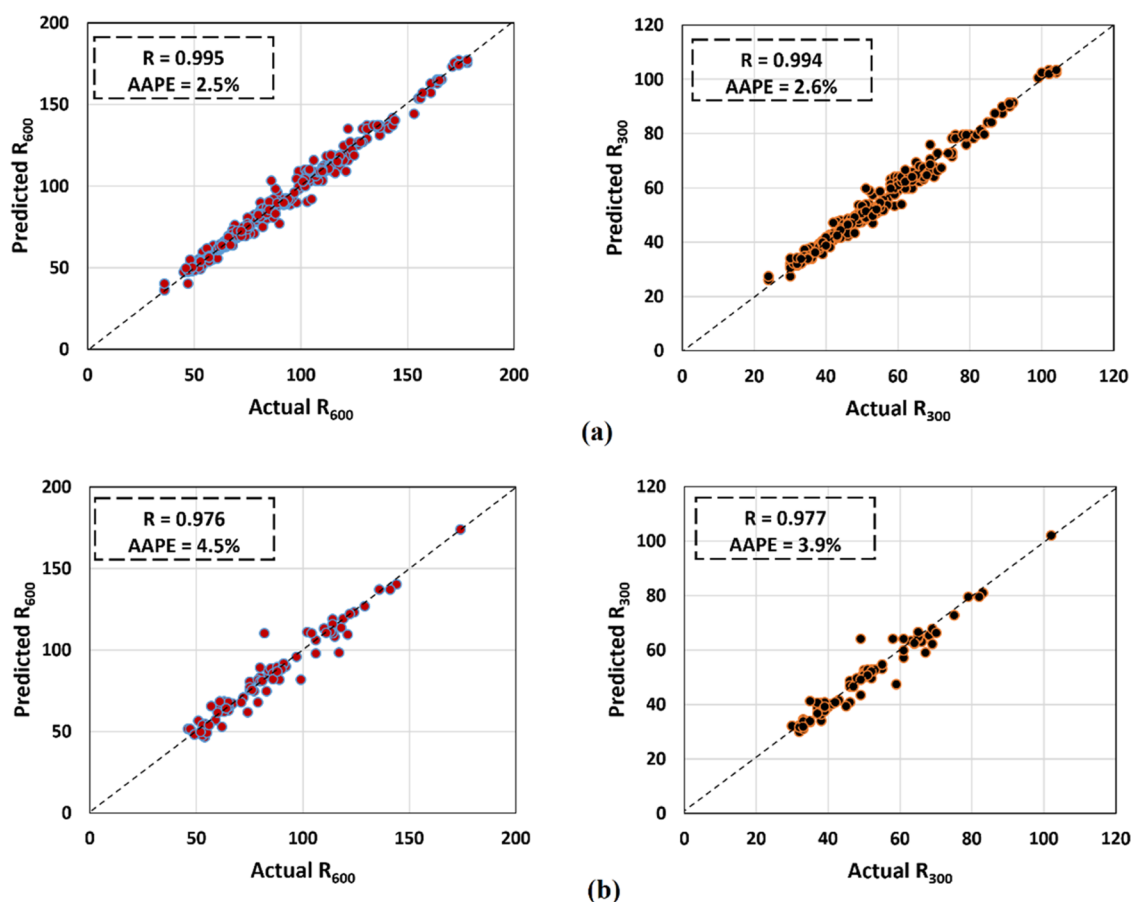
Figure 6. Profiles of the predicted versus actual values of n and K using the developed DT-based model.

data are complex and are collected from different sources with various formats. Due to this complexity, the biggest part of the efforts put into ML projects is dedicated to data preparation and exploration stage.[58] The collected data set (533 data points) contained MD, MFV, and viscometer readings at 300 (R_{300}) and 600 (R_{600}) RPM for different samples from the same mud type. PV, YP, AV, n, and K were estimated using R_{300} and R_{600} through eqs 4−8.[7] The mud samples were collected after going through a shale shaker. MD and MFV are measured by a mud balance and a Marsh funnel, respectively, while the rheology is evaluated using a viscometer. The data set covered broad ranges, which is

Table 2. Statistical Analysis of the SOBM Data Set

parameter	MD	MFV	R_{600}	R_{300}
min.	70	27	36	24
max.	120	120	178	104
mean	95.8	67	87	52
standard deviation	13.98	12.63	30.76	16.49
range	50	93	142	80
skewness	−0.35	0.50	0.90	0.99

Figure 7. Cross-plots of the predicted versus actual viscometer readings using the RF model for training (a) and testing (b).

essential to have reliable general models that could confidently capture the problem's nature as follows: MD ranged from 70 to 120 pcf with a mean of 95.8 and standard deviation (σ) of 13.98, MVF from 27 to 120 s/quart with a mean of 67 and σ of 12.63, R_{600} from 36 to 178 with a mean of 87 and σ of 30.76, and R_{300} from 24 to 104 with a mean of 52 and σ of 16.49. Table 2 summarizes the statistical analysis of the data used for training and testing in this study. Table 3 presents a typical sample of the flat rheology SOBM data.

$$PV = R_{600} - R_{300} \tag{4}$$

$$YP = R_{300} - PV \tag{5}$$

$$AV = \frac{R_{600}}{2} \tag{6}$$

$$n = 3.32 \times \log\left(\frac{R_{600}}{R_{300}}\right) \tag{7}$$

$$K = \frac{R_{600}}{1022^n} \tag{8}$$

The data should be filtered and analyzed since it came from the field where there may be human or/and instrumental errors during measurements. Moreover, data filtration is a crucial step for enhancing data quality, which in turn provides a better machine learning prediction.[59] First, the data set was investigated searching for any missing value, repeated data points, negative values, and unreasonable magnitudes which miss the engineering sense. Then, the outliers can be eliminated

Table 3. Sample of the SOBM Field Data

MD (pcf)	MFV (s/quart)	R_{600}	R_{300}
104	65	124	68
98	82	91	53
77	56	49	32
115	72	106	62
86	55	66	41
120	65	120	69
73	48	59	38

Table 4. Summary of the Different Metrics for the Viscometer Reading Prediction Using the DT Model

metric	R_{600} train	R_{600} test	R_{300} train	R_{300} test
AAPE (%)	2.3	4.2	1.8	3.6
MAE	1.88	3.36	0.95	1.74
RMSE	3.09	5.85	1.57	3.00
R	0.995	0.975	0.996	0.977
R^2	0.990	0.950	0.991	0.955

in various ways such as a box and whisker plot. The relationship of each output with each input was checked using the R-value. The R-value may take any value between −1 and 1 inclusive. A high R-value reflects a strong linear relationship. The positive R-value means the existence of a direct relationship, while a negative value means an inverse relationship between the two variables. Almost no linear relationship exists when the R-value is around zero.[60] As shown in Figure 2, R_{600} had R-values of
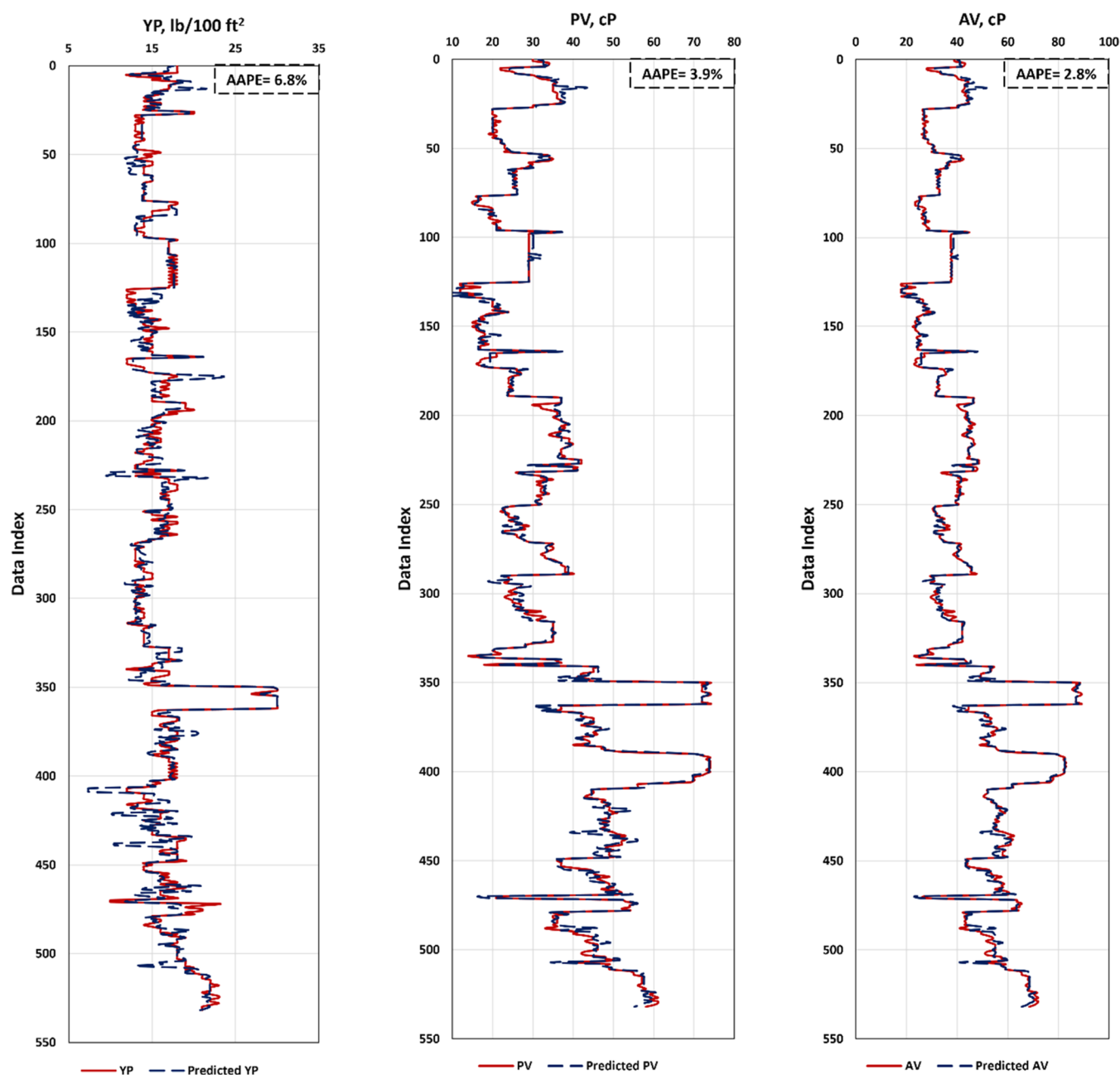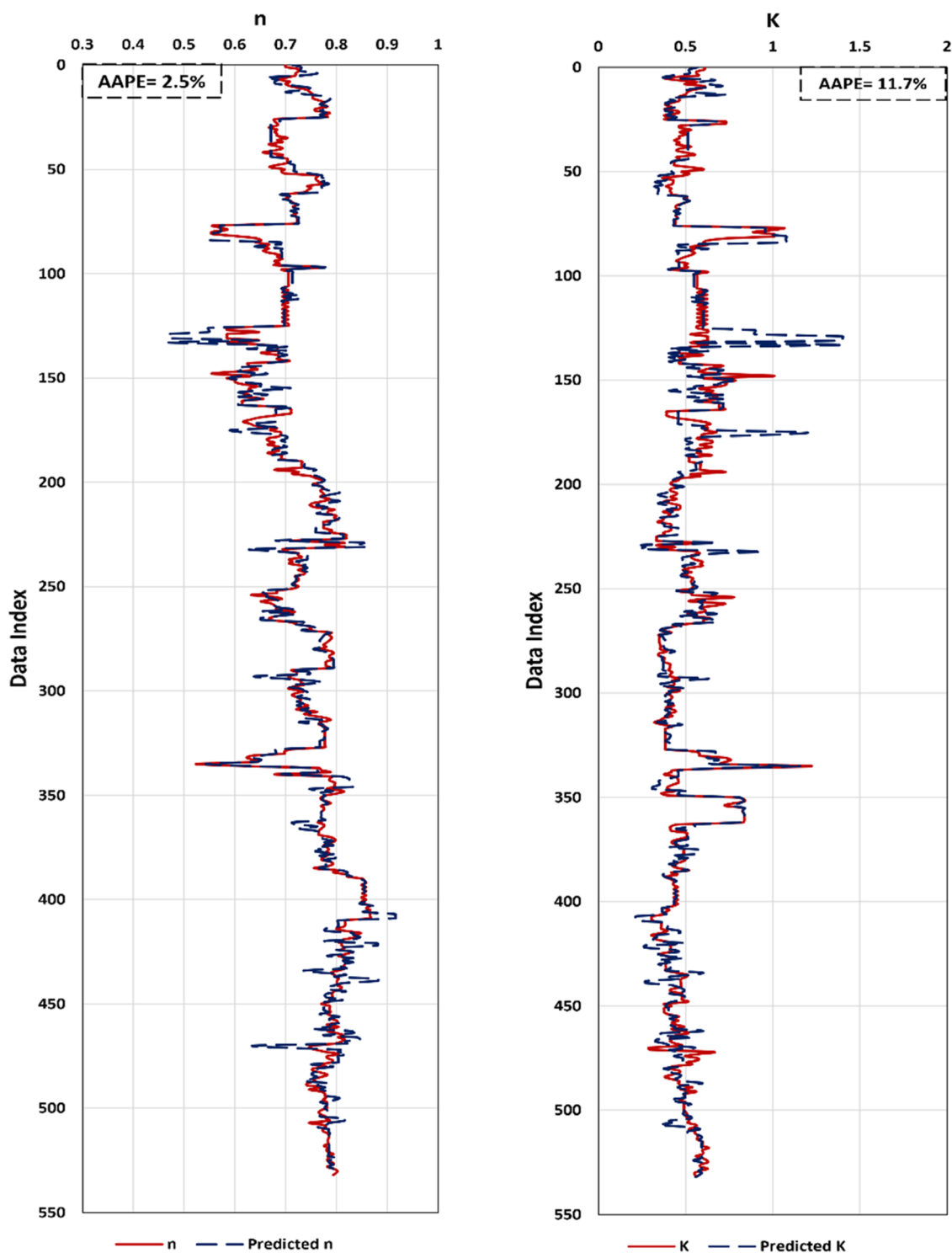
Figure 8. Profiles of the predicted versus actual values of YP, PV, and AV using the developed RF-based model.

0.748 and 0.75 with MD and MFV, respectively. Similarly, R_{300} data had R-values higher than 0.74 with all inputs. This proves a remarkable dependence of viscometer readings on the chosen input variables. ML can discover the direct and indirect relations between inputs and target variables. As mud weight increases which can be translated into more solid content, viscosity and viscometer readings will most probably go up, which agrees with the R-values presented in Figure 2. The frequent measurements of the flat rheology SOBM such as MD and MFV were used as input parameters to predict the viscometer readings (R_{600} and R_{300}) as a middle stage to be substituted in the existing equations of rheological properties to get PV, YP, AV, n, and K. The SOBM rheological property prediction framework introduced in this work can be summarized in Figure 3.

2.2. Machine Learning Algorithms. *2.2.1. DT Model.* DT learning or induction of DT is a predictive modeling approach

utilized in statistics, data mining, and ML. A DT is a nonparametric supervised learning technique for classification and regression problems. It utilizes a decision tree to transfer from observations of an element (represented by branches) to conclusions about the element's target value (represented by leaves). A classification tree is a tree model in which the output variable is the class (discrete) to which the data belong; in this tree structure, the leave represents the class label and the branch represents the feature conjunction that leads to the class label. A regression tree is a decision tree in which the output variable can be continuous values (usually real numbers). DTs are among the most popular ML approaches owing to their intelligibility and simplicity.[61] The DT technique is simple to be understood and interpreted and trees can be graphically displayed in a way that is easy for nonexperts to interpret.[62] Statistical tests can be used to validate the predictive model, which makes it possible to check

Figure 9. Profiles of the predicted versus actual values of n and K using the developed RF-based model.

the model's reliability. It is a nonparametric technique that makes no assumptions about the training data. It closely reflects human decision-making more than other techniques.

After checking data quality, 533 data points were selected to build the predictive models. The obtained data had been divided into two groups with approximately 80:20 ratio for training and testing. The DT hyperparameters should be tuned to get the optimum prediction performance. The DT model hyperparameters, including different combinations of various available options for DT hyperparameters, were optimized by running

many scenarios for each hyperparameter. The R-value and different types of errors (root mean squared error (RMSE), mean absolute error (MAE), and AAPE) were calculated. The hyperparameters giving the highest R and the lowest errors between the actual and the predicted outputs were chosen. The max_features hyperparameter has different options, such as "auto", "sqrt", and "log 2," which indicates the maximum number of features DT is allowed to try when looking for the best split. In case the "sqrt" option is selected, then max_features = sqrt (n_features), and so on. The max_depth shows the
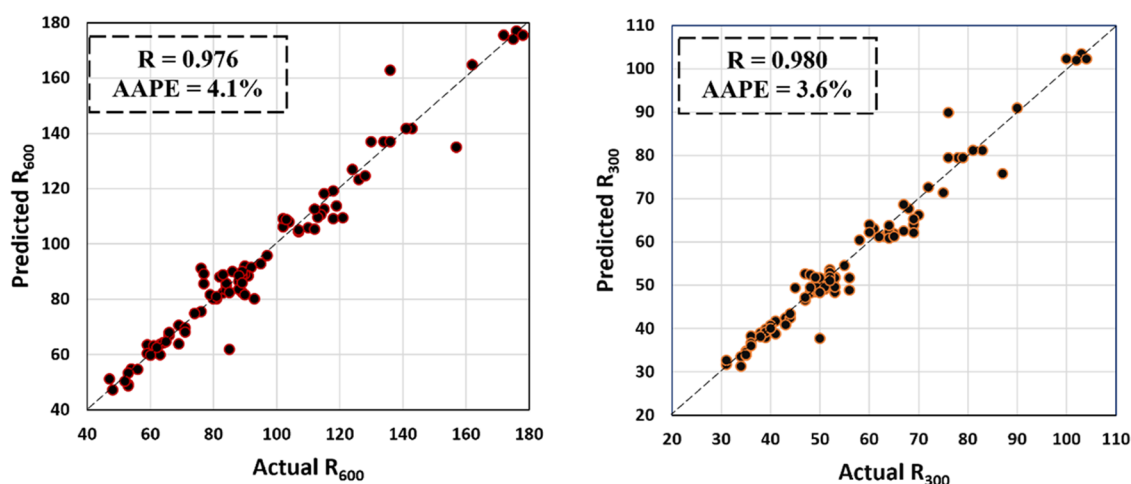
Figure 10. Cross-plots of the predicted versus actual viscometer readings using the DT-based model for the validation.
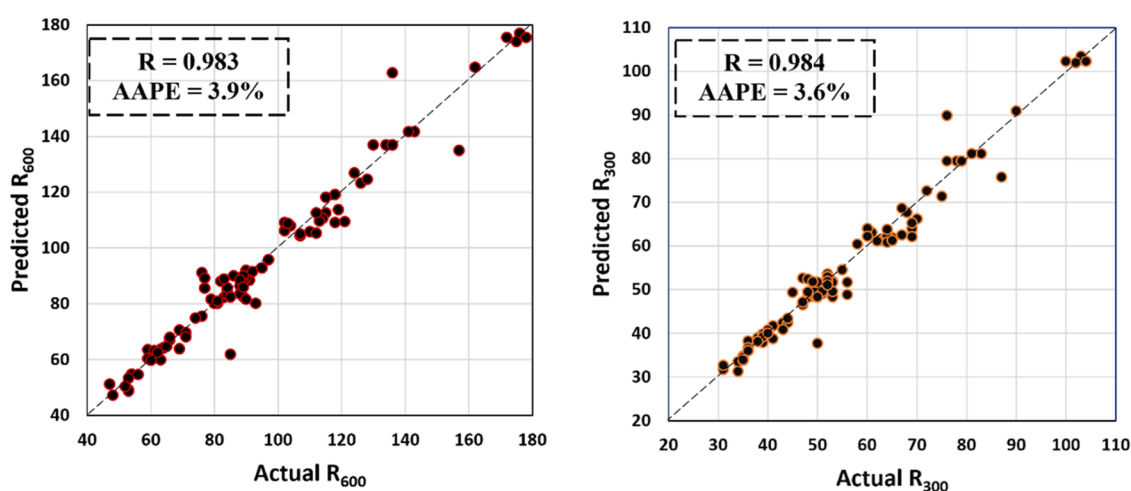

Figure 11. Cross-plots of the predicted versus actual viscometer readings using the RF-based model for the validation.

maximum depth of the tree or the maximum number of levels in the tree. The min_samples_leaf indicates the minimum number of samples required to be at a leaf node. When the algorithm splits the data, the goal is to reduce impurity as much as possible. The lower the impurity is, the greater the informative power of the split is. The quantity of impurities in the tree reduces as it grows deeper. The hyperparameter called min_impurity_decrease is used to prevent the tree from splitting further.

2.2.2. RF Model. RF is an ensemble learning approach for regression and classification problems.[63] It contains hundreds or thousands of trees in which individual tree training is performed using a slightly different observation set.[64] It employs a technique known as bootstrapping, which is an iterative resampling technique for estimating population statistics by sampling the data with replacement. In a procedure known as aggregation, the forecasts of every tree are averaged to get the final prediction. Bootstrapping excludes approximately one-third of the original data set, and this sample is used to internally test the RF model accuracy.[65] RF is better than a single DT owing to the capability of limiting overfitting without increasing the error margin.[66] The same procedure for data preparation and division is followed for the RF model. The RF hyperparameters should be tuned to get the optimum prediction performance. RF shares many hyperparameters with DTs in addition to some

Table 5. RF Model Optimum Hyperparameters

| hyperparameter	options	selection
max_features	["auto", "sqrt", "log$_2$"]	Sqrt
max_depth	[3, 4, 5, ⋯, 30]	8
N_estimators	[3, 4, 5, ⋯, 150]	100

Table 6. Summary of the Different Metrics for the Viscometer Reading Prediction Using the RF Model

metric	R_{600} train	R_{600} test	R_{300} train	R_{300} test
AAPE (%)	2.5	4.5	2.6	3.9
MAE	2.12	3.48	1.32	1.88
RMSE	3.23	5.66	1.87	3.02
R	0.995	0.976	0.994	0.977
R^2	0.990	0.953	0.988	0.954

specific ones. For instance, the n_estimator parameter indicates the number of trees in the forest.

3. RESULTS AND DISCUSSION

3.1. DT Model. After the optimization process, the DT model managed to predict R_{600} and R_{300} with good R-values of
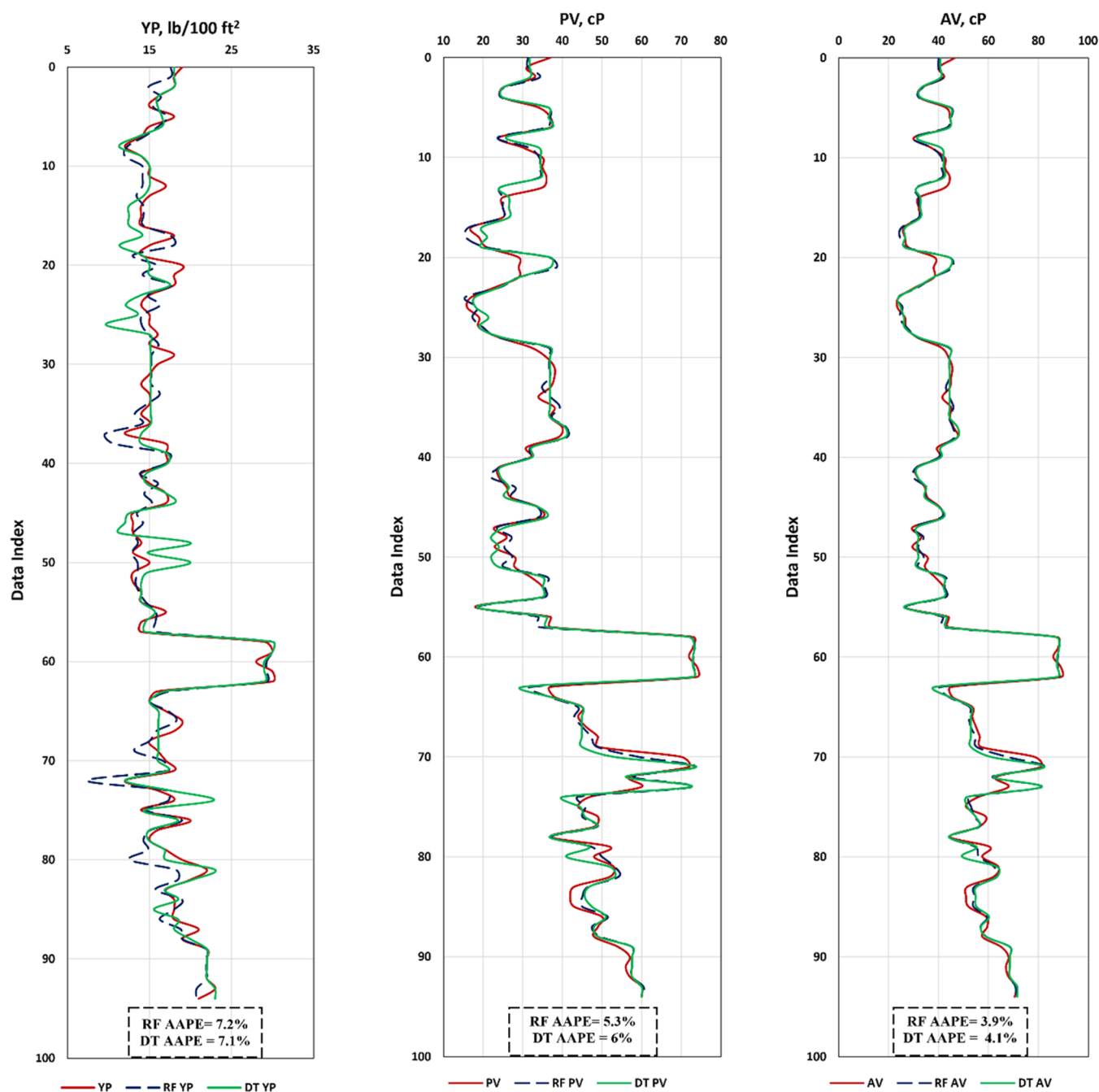
Figure 12. Profiles of the predicted versus actual YP, PV, and AV using the developed models for the validation data set.

around 0.995 and 0.98 and with a coefficient of determination (R^2) of 0.99 and 0.95 for training and testing, respectively. The different types of errors for viscometer reading prediction are listed in Table 4 showing a high prediction accuracy. The optimum hyperparameters of the DT model were max_depth of 10 and max_features of log 2. A remarkable matching was observed between the estimated and recorded outputs as indicated in Figure 4, proved by high R-values and low errors for both the training and testing stages. Figures 5 and 6 show the profiles of the predicted versus actual rheological properties showing a reasonable prediction accuracy.

3.2. RF Model. After the optimization process, the optimum RF hyperparameters are listed in Table 5. The RF model managed to estimate R_{600} and R_{300} with good R-values of around 0.99 and 0.98 and with an R^2 of 0.99 and 0.95 for training and

testing, respectively. The different types of errors for viscometer reading prediction are listed in Table 6 showing a high prediction accuracy. Figure 7 presents the cross-plots of the estimated versus the recorded viscometer readings in which the points remarkably coincide with the 45° line, which is confirmed

Table 7. Statistical Analysis of the SOBM Data Det

parameter	MD	MFV	R_{600}	R_{300}
min.	70	27	36	24
max.	120	120	178	104
mean	95.8	67	87	52
standard deviation	13.98	12.63	30.76	16.49
range	50	93	142	80
skewness	−0.35	0.50	0.90	0.99
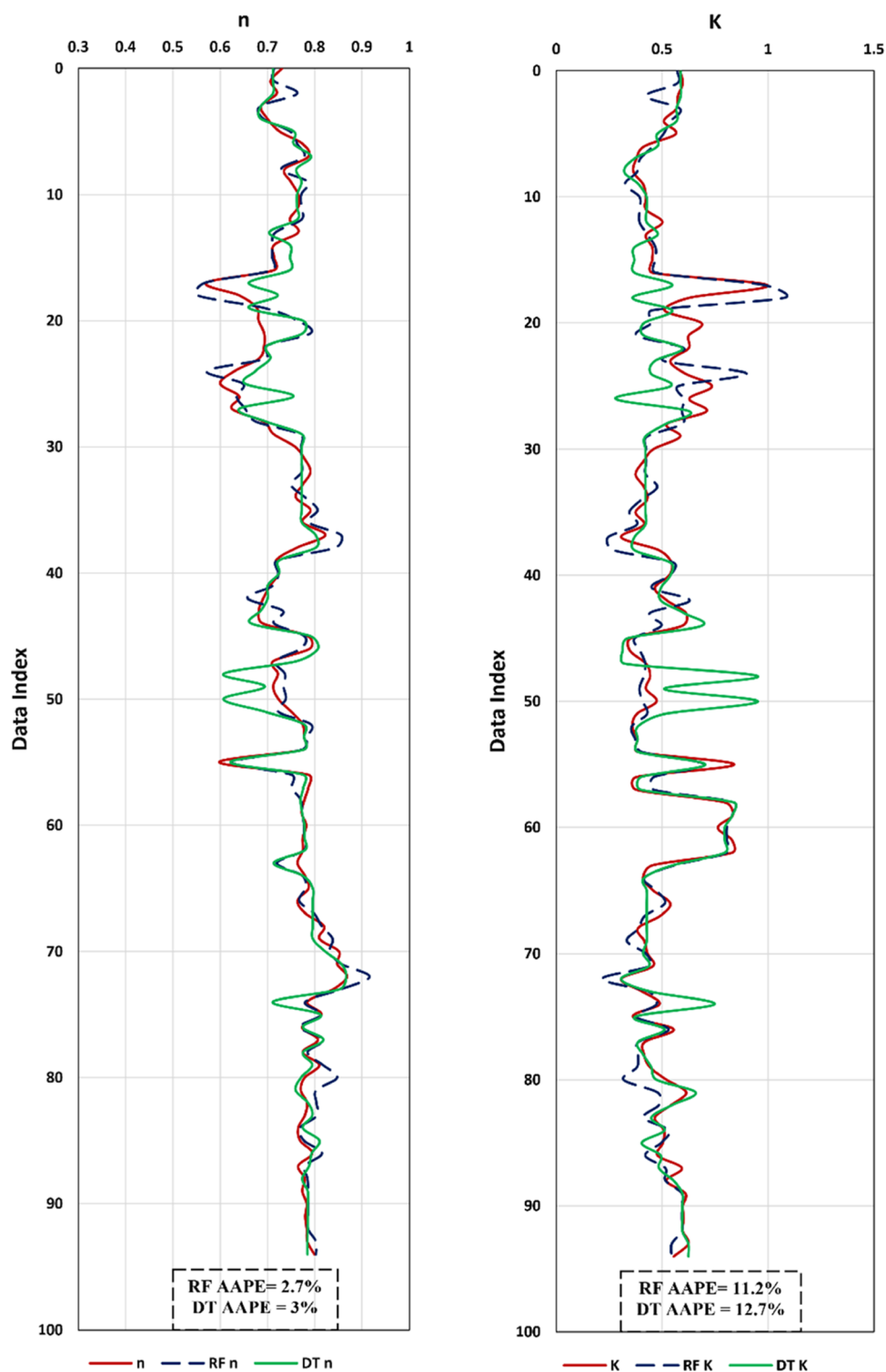
Figure 13. Profiles of the predicted versus actual *n* and *K* using the developed models for the validation data set.

by high *R*-values and low errors for both training and testing stages. Figures 8 and 9 show the profiles of predicted versus actual rheological properties showing a reasonable estimation accuracy.

4. MODEL VALIDATION

To check the performance of the proposed models, another blind holdout data set (95 examples) was used for validation

purposes. The data included the frequent mud measurements (MD and MFV) with their respective viscometer readings in addition to their different rheological properties. The MD and MFV were used as inputs to feed the proposed models and the results were compared with the actual outputs values. Table 7 lists the statistical parameters of the validation data set used for prediction evaluation. The validation data fall within the same ranges of the data used for developing ML models. A good matching was observed between the estimated and the recorded
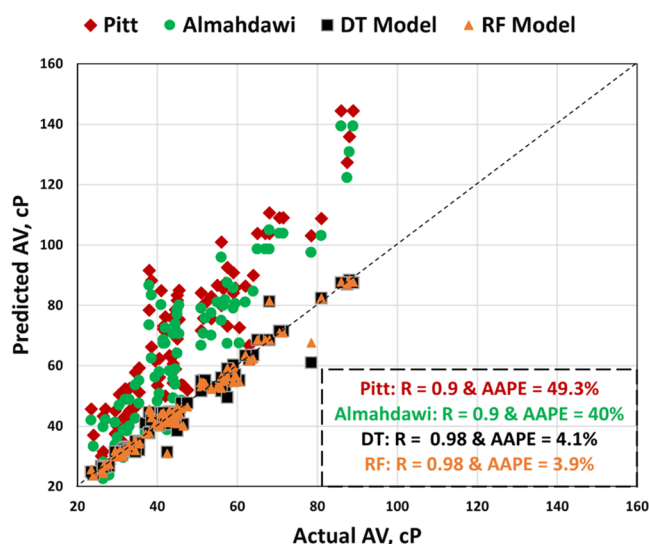
Figure 14. Cross-plot of the predicted versus actual AV values using different models.
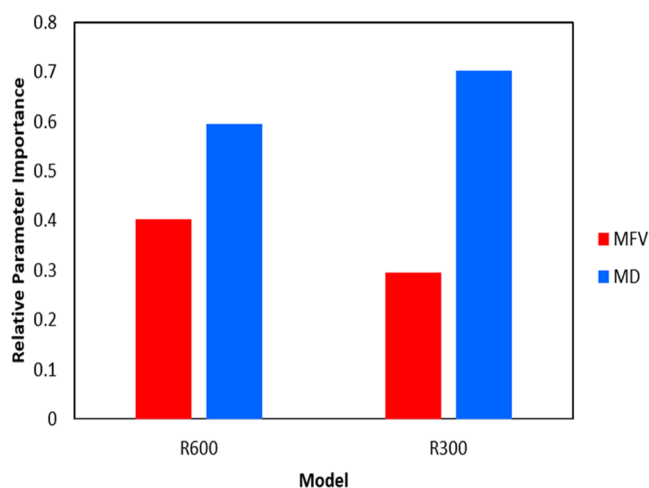

Figure 15. Relative importance of the input parameters on R_{600}, and R_{300} ML models.

R_{600} and R_{300} readings as shown in Figures 10 and 11. The R-value of the RF model was slightly higher than that of the DT model for both R_{600} and R_{300}. Moreover, the AAPE of the RF model was lower than that of the DT model that did not exceed 4.1% for both models. Figures 12 and 13 show the profiles of the

Table 8. Summary of the Different Metrics for AV Prediction Using Different Models

parameter	Pitt correlation[67]	Almahdawi correlation[68]	DT model	RF model
AAPE (%)	49.26	40.03	4.14	3.93
RMSE	27.39	23.34	3.42	2.98
R	0.90	0.90	0.98	0.98
R^2	0.82	0.81	0.95	0.97

predicted rheological properties versus the actual values with their respective AAPE. The results showed that the RF model outperformed the DT model, which gave the lowest errors in all properties except YP.

Recently, ML algorithms have shown promising results in predicting drilling fluid rheology. Research into new types of drilling fluids, including those with novel rheological properties, can expand the understanding of fluid behavior and lead to improved models. Integrating drilling data with rheology predictions can lead to a more comprehensive understanding of drilling fluid behavior and its effect on drilling performance. The prediction models should be tested and validated in different drilling environments, such as offshore, onshore, and unconventional drilling, to understand their performance and accuracy under different conditions.

5. COMPARISON BETWEEN THE DEVELOPED AV MODELS AND EXISTING MODELS

Two models were introduced in the literature to estimate AV. MD and MFV were used as input parameters to predict AV as introduced by Pitt (2000) in eq 9. Then, Almahdawi et al. modified Pitt's equation to include a constant equal to 28 instead of 25 as presented in eq 10 because the authors found that the new constant gave better results.

$$AV = D(T - 25) \qquad (9)$$

$$AV = D(T - 28) \qquad (10)$$

where AV in cP, D is the mud density in g/cc, and T is the MF time in s/quart.

For a fair comparison, the validation data set that was not included in building the ML models was used. The AV values were calculated using the different models and the metrics such as AAPE, RMSE, R, and R^2 were compared as listed in Table 8. ML models outperformed the other models by AAPE not exceeding 4.1% and an R of 0.98. The other models in the literature gave AAPE in the range of 40−49% and a smaller R-
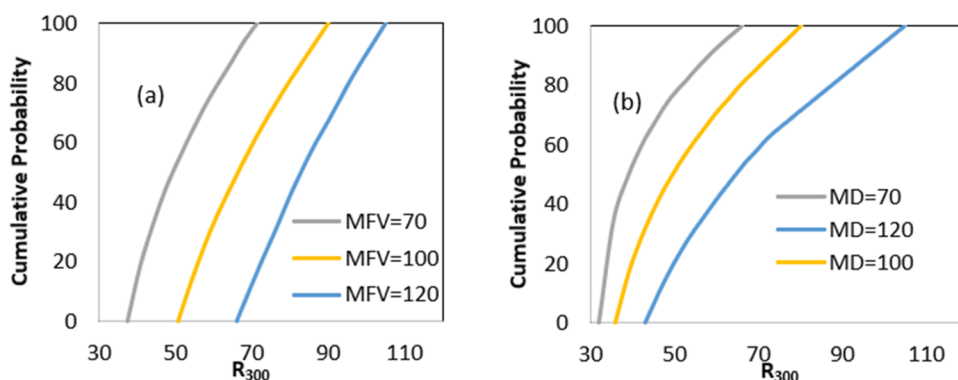

Figure 16. Sensitivity analysis distribution for (a) constant MFV and varying MD values and (b) constant MD and varying MFV.

value of around 0.81. Figure 14 shows a comparison cross-plot for AV prediction using different models. It shows that the Pitt and Almahdawi correlations overestimate the AV values.

6. INPUT IMPORTANCE AND SENSITIVITY ANALYSIS

To examine the relative importance of the input parameters and their effect on the output parameters, the relative importance method in the RF model was used. This method measures the reduction in impurity that occurs when splitting a feature in the decision trees. The higher the impurity reduction is, the more important the feature is. Figure 15 presents the relative importance of the input parameters for the different ML models to predict R_{600} and R_{300}. Mud density has higher importance than funnel viscosity to predict the output parameters, especially in the case of the R_{300} model. These results agreed with the correlation coefficient of the input data as presented in Figure 2.

To examine the sensitivity analysis, the model was used to predict R_{300} and R_{600} with varying input parameters while fixing others. Figure 16 shows the cumulative distribution of the predicted R_{300} with assuming constant MFV and randomly varying MD in Figure 16a and then predicting R_{300} with assuming constant MD and randomly varying MFV as shown in Figure 16b. At constant MFV, with variable values of MD, R_{300} varied from 37 to 71 (range = 34) at an MFV value of 70. With increasing MFV, the whole curve shifted to higher R_{300} values with a minimum of 66 and a maximum of 105 (range = 39) at an MFV value of 120. Similarly, at constant MD with variable values of MFV, R_{300} varied from 32 to 66 (range = 34) at an MD value of 70. With increasing MD, the whole curve shifted to higher R_{300} values with a minimum of 43 and a maximum of 105 (range = 62) at an MD value of 120. The variability of MD has more influence than MFV, which is in agreement with the results from Figure 15.

7. CONCLUSIONS

This study introduced a machine learning (ML) framework for predicting rheological properties of drilling fluids while drilling, using frequent mud measurements such as MD and MFV as inputs to predict viscometer readings, and then estimating other mud properties using existing equations in the literature. The results demonstrated the potential for promoting mud system automation and enabling proactive decisions to avoid drilling problems and enhance the safety and integrity of drilling operations. The main outcomes of this study include the following:

(1) The proposed models accurately predicted viscometer readings with a low average absolute percentage error (AAPE) of less than 4.5% for both models.

(2) The models predicted rheological properties with reasonable accuracy, with an AAPE less than 7% for most parameters.

(3) The models were validated using a blind holdout data set from the same mud type, achieving high correlation coefficients (R-values) of 0.98.

(4) The ML models outperformed empirical correlations for apparent viscosity prediction, demonstrating their potential for improving drilling operations.

The proposed framework could also save time and costs associated with time-consuming lab work and nonproductive time. However, it is important to note that the proposed models and hyperparameters were specifically designed for flat rheology SOBM, and different mud types may have varying responses in viscometer readings and Marsh funnel viscosity. Input parameters should also be within the same training ranges and units for reasonable prediction.

AUTHOR INFORMATION

Corresponding Author

Salaheldin Elkatatny − *Department of Petroleum Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;* ⊙ orcid.org/0000-0002-7209-3715; Email: elkatatny@kfupm.edu.sa

Authors

Ahmed Abdelaal − *Department of Petroleum Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*

Ahmed Farid Ibrahim − *Department of Petroleum Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;* ⊙ orcid.org/0000-0001-7258-8542

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c06656

ABBREVIATIONS

AAPE average absolute percentage error
ANFIS adaptive neuro fuzzy inference system
ANN artificial neural network
AV apparent viscosity
BP Bingham plastic
DTs decision trees
ECD equivalent circulation density
K consistency factor or index
MAE mean absolute error
MD mud density
MFV Marsh funnel viscosity
ML machine learning
n flow behavior index
pcf pound per cubic feet
PL power law
PV plastic viscosity
R correlation coefficient
R^2 coefficient of determination
R_{300} viscometer reading at 300 RPM
R_{600} viscometer reading at 600 RPM
RF random forest
RMSE root mean squared error
ROP rate of penetration
RPM revolution per minute
SOBM synthetic oil-based mud
YP yield point

REFERENCES

(1) Hansen, E. Automatic Evaluation of Drilling Fluid Properties, Master's Thesis; University of Stavanger: Norway, 2012.

(2) Alsabaa, A.; Gamal, H.; Elkatatny, S.; Abdulraheem, A. Real-Time Prediction of Rheological Properties of Invert Emulsion Mud Using Adaptive Neuro-Fuzzy Inference System. *Sensors* **2020**, *20*, No. 1669.

(3) Gray, G. R.; Grioni, S. Varied Applications of Invert Emulsion Muds. *J. Pet. Technol.* **1969**, *21*, 261−266.

(4) Growcock, F. B.; Ellis, C. F.; Schmidt, D. D. Electrical Stability, Emulsion Stability, and Wettability of Invert Oil-Based Muds. *SPE Drill. Completion* **1994**, *9*, 39−46.

(5) Hossain, M. E.; Al-Majed, A. A. Fundamentals of Sustainable Drilling Engineering. *Fundam. Sustainable Drill. Eng.* **2015**, 1−754.

(6) Lee, J.; Shadravan, A.; Young, S. In *Rheological Properties of Invert Emulsion Drilling Fluid under Extreme HPHT Conditions*, SPE/IADC Drilling Conference, Proceedings, 2012; pp 946−954. DOI: 10.2118/151413-MS.

(7) Hughes, B. Drilling Fluids Reference Manual, 2006.

(8) Wiśniowski, R.; Skrzypaszek, K.; Małachowski, T. Selection of a Suitable Rheological Model for Drilling Fluid Using Applied Numerical Methods. *Energies* **2020**, *13*, No. 3192.

(9) Rabia, H. Well Engineering & Construction Hussain Rabia. 2002.

(10) Abegunrin, O. A.; Okhuevbie, S. O.; Animasaun, I. L. Comparison between the Flow of Two Non-Newtonian Fluids over an Upper Horizontal Surface of Paraboloid of Revolution: Boundary Layer Analysis. *Alexandria Eng. J.* **2016**, *55*, 1915−1929.

(11) Onalo, D.; Adedigba, S.; Oloruntobi, O.; Khan, F.; James, L. A.; Butt, S. Data-Driven Model for Shear Wave Transit Time Prediction for Formation Evaluation. *J. Pet. Explor. Prod. Technol.* **2020**, *10*, 1429−1447.

(12) Oloruntobi, O.; Onalo, D.; Adedigba, S.; James, L.; Chunduru, R.; Butt, S. Data-Driven Shear Wave Velocity Prediction Model for Siliciclastic Rocks. *J. Pet. Sci. Eng.* **2019**, *183*, No. 106293.

(13) Onalo, D.; Oloruntobi, O.; Adedigba, S.; Khan, F.; James, L.; Butt, S. Dynamic Data Driven Sonic Well Log Model for Formation Evaluation. *J. Pet. Sci. Eng.* **2019**, *175*, 1049−1062.

(14) Onalo, D.; Oloruntobi, O.; Adedigba, S.; Khan, F.; James, L.; Butt, S. Static Young's Modulus Model Prediction for Formation Evaluation. *J. Pet. Sci. Eng.* **2018**, *171*, 394−402.

(15) Mamudu, A.; Khan, F.; Zendehboudi, S.; Adedigba, S. A Hybrid Intelligent Model for Reservoir Production and Associated Dynamic Risks. *J. Nat. Gas Sci. Eng.* **2020**, *83*, No. 103512.

(16) Zhong, R.; Salehi, C.; Johnson, R. Machine Learning for Drilling Applications: A Review. *J. Nat. Gas Sci. Eng.* **2022**, *108*, No. 104807.

(17) Hansson, K.; Yella, S.; Dougherty, M.; Fleyeh, H. Machine Learning Algorithms in Heavy Process Manufacturing. *Am. J. Intell. Syst.* **2016**, *6*, 1−13.

(18) Mohaghegh, S. Virtual-Intelligence Applications in Petroleum Engineering: Part I - Artificial Neural Networks. *J. Pet. Technol.* **2000**, *52*, 64−73.

(19) Ahmed, S. A.; Elkatatny, S.; Abdulraheem, A.; Mahmoud, M.; Ali, A. Z. In *New Approach to Predict Fracture Pressure Using Functional Networks*, SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, SATS 2018; Society of Petroleum Engineers, 2018. DOI: 10.2118/192317-ms.

(20) Zhong, R.; Johnson, R.; Chen, Z. Generating Pseudo Density Log from Drilling and Logging-While-Drilling Data Using Extreme Gradient Boosting (XGBoost). *Int. J. Coal Geol.* **2020**, *220*, No. 103416.

(21) Abdelaal, A.; Ibrahim, A. F.; Elkatatny, S. Data-Driven Approach for Resistivity Prediction Using Artificial Intelligence. *J. Energy Resour. Technol.* **2022**, *144*, No. 103003.

(22) Ibrahim, A. F.; Abdelaal, A.; Elkatatny, S. Formation Resistivity Prediction Using Decision Tree and Random Forest. *Arabian J. Sci. Eng.* **2022**, *47*, 12183−12191.

(23) Abdelaal, A.; Ibrahim, A. F.; Elkatatny, S. Empirical Correlation for Formation Resistivity Prediction Using Machine Learning. *Arabian J. Geosci.* **2022**, *15*, 1−12.

(24) Abdelaal, A.; Elkatatny, S.; Abdulraheem, A. Data-Driven Modeling Approach for Pore Pressure Gradient Prediction While Drilling from Drilling Parameters. *ACS Omega* **2021**, *6*, 13807−13816.

(25) Abdelaal, A.; Elkatatny, S.; Abdulraheem, A. Real-Time Prediction of Formation Pressure Gradient While Drilling. *Sci. Rep.* **2022**, *12*, No. 11318.

(26) Abdelaal, A.; Elkatatny, S.; Abdulraheem, A. In *Pore Pressure Estimation While Drilling Using Machine Learning*, ARMA/DGS/SEG 2nd International Geomechanics Symposium; OnePetro, 2021.

(27) Abdelaal, A.; Elkatatny, S.; bin Waheed, U.; Hanafy, S. M. In *Formation Pressure Abnormality Identification Using Artificial Neural Networks: A Classification Model*, 56th U.S. Rock Mechanics/Geomechanics Symposium, 2022. DOI: 10.56952/ARMA-2022-0824.

(28) Ren, X.; Hou, J.; Song, S.; Liu, Y.; Chen, D.; Wang, X.; Dou, L. Lithology Identification Using Well Logs: A Method by Integrating Artificial Neural Networks and Sedimentary Patterns. *J. Pet. Sci. Eng.* **2019**, *182*, No. 106336.

(29) Zhong, R.; Johnson, R. L.; Chen, Z. Using Machine Learning Methods To Identify Coal Pay Zones from Drilling and Logging-While-Drilling (LWD) Data. *SPE J.* **2020**, *25*, 1241−1258.

(30) Siddig, O.; Gamal, H.; Elkatatny, S.; Abdulraheem, A. Applying Different Artificial Intelligence Techniques in Dynamic Poisson's Ratio Prediction Using Drilling Parameters. *J. Energy Resour. Technol.* **2022**, *144*, No. 073006.

(31) Rostami, H.; Manshad, A. K. A New Support Vector Machine and Artificial Neural Networks for Prediction of Stuck Pipe in Drilling of Oil Fields. *J. Energy Resour. Technol.* **2014**, *136*, No. 024502.

(32) Gamal, H.; Abdelaal, A.; Elkatatny, S. Machine Learning Models for Equivalent Circulating Density Prediction from Drilling Data. *ACS Omega* **2021**, *6*, 27430−27442.

(33) Gamal, H.; Abdelaal, A.; Alsaihati, A.; Elkatatny, S.; Abdulraheem, A. In *Artificial Neural Network Model for Predicting the Equivalent Circulating Density from Drilling Parameters*; OnePetro, June 18, 2021.

(34) Alsabaa, A.; Elkatatny, S. Improved Tracking of the Rheological Properties of Max-Bridge Oil-Based Mud Using Artificial Neural Networks. *ACS Omega* **2021**, *6*, 15816−15826.

(35) Aghaie, M.; Zendehboudi, S. Estimation of CO2 Solubility in Ionic Liquids Using Connectionist Tools Based on Thermodynamic and Structural Characteristics. *Fuel* **2020**, *279*, No. 117984.

(36) Salem, A. M.; Attia, M.; Alsabaa, A.; Abdelaal, A.; Tariq, Z. Machine Learning Approaches for Compressibility Factor Prediction at High- and Low-Pressure Ranges. *Arabian J. Sci. Eng.* **2022**, *47*, 1−12.

(37) Soares, C.; Gray, K. Real-Time Predictive Capabilities of Analytical and Machine Learning Rate of Penetration (ROP) Models. *J. Pet. Sci. Eng.* **2019**, *172*, 934−959.

(38) Miah, M. I.; Zendehboudi, S.; Ahmed, S. Log Data-Driven Model and Feature Ranking for Water Saturation Prediction Using Machine Learning Approach. *J. Pet. Sci. Eng.* **2020**, *194*, No. 107291.

(39) Habib, M. M.; Imtiaz, S.; Khan, F.; Ahmed, S.; Baker, J. Early Detection and Estimation of Kick in Managed Pressure Drilling. *SPE Drill. Completion* **2021**, *36*, 245−262.

(40) Moazzeni, A.; Haffar, M. A. Artificial Intelligence for Lithology Identification through Real-Time Drilling Data. *J. Earth Sci. Clim. Change* **2015**, *06*, 1−4.

(41) Zendehboudi, S.; Rezaei, N.; Lohi, A. Applications of Hybrid Models in Chemical, Petroleum, and Energy Systems: A Systematic Review. *Appl. Energy* **2018**, *228*, 2539−2566.

(42) Renpu, W. *Advanced Well Completion Engineering*, 3rd ed.; Elsevier, 2011; pp 1−715. DOI: 10.1016/C2010-0-66820-9.

(43) Razi, M. M.; Mazidi, M.; Razi, F. M.; Aligolzadeh, H.; Niazi, S. Artificial Neural Network Modeling of Plastic Viscosity, Yield Point, and Apparent Viscosity for Water-Based Drilling Fluids. *J. Dispersion Sci. Technol.* **2013**, *34*, 822−827.

(44) Elkatatny, S.; Tariq, Z.; Mahmoud, M. Real Time Prediction of Drilling Fluid Rheological Properties Using Artificial Neural Networks Visible Mathematical Model (White Box). *J. Pet. Sci. Eng.* **2016**, *146*, 1202−1210.

(45) Elkatatny, S. Real-Time Prediction of Rheological Parameters of KCl Water-Based Drilling Fluid Using Artificial Neural Networks. *Arabian J. Sci. Eng.* **2017**, *42*, 1655−1665.

(46) Elkatatny, S.; Mahmoud, M. In *Real Time Prediction of the Rheological Parameters of NaCl Water-Based Drilling Fluid Using Artificial Neural Networks*, Society of Petroleum Engineers - SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, 2017; 1726−1740.

(47) da Silva Bispo, V. D.; Scheid, C. M.; Calçada, L. A.; da Cruz Meleiro, L. A. Development of an ANN-Based Soft-Sensor to Estimate the Apparent Viscosity of Water-Based Drilling Fluids. *J. Pet. Sci. Eng.* **2017**, *150*, 69−73.

(48) Al-Azani, K.; Elkatatny, S.; Abdulraheem, A.; Mahmoud, M.; Al-Shehri, D. In *Real Time Prediction of the Rheological Properties of Oil-Based Drilling Fluids Using Artificial Neural Networks*, Society of Petroleum Engineers - SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition 2018, SATS 2018, 2018 DOI: 10.2118/192199-MS.

(49) Avcı, E. An Artificial Neural Network Approach for the Prediction of Water-Based Drilling Fluid Rheological Behaviour. *Int. Adv. Res. Eng. J.* **2018**, *2*, 124−131.

(50) Abdelgawad, K.; Elkatatny, S.; Moussa, T.; Mahmoud, M.; Patil, S. Real-Time Determination of Rheological Properties of Spud Drilling Fluids Using a Hybrid Artificial Intelligence Technique. *J. Energy Resour. Technol.* **2019**, *141*, No. 032908.

(51) Elzenary, M. Estimation of Hole Cleaning Condition in Real Time While Drilling (Operating Point of View). 2019.

(52) Gowida, A.; Elkatatny, S.; Ramadan, E.; Abdulraheem, A. Data-Driven Framework to Predict the Rheological Properties of $CaCl_2$ Brine-Based Drill-in Fluid Using Artificial Neural Network. *Energies* **2019**, *12*, No. 1880.

(53) Gomaa, I.; Elkatatny, S.; Abdulraheem, A. Real-Time Determination of Rheological Properties of High over-Balanced Drilling Fluid Used for Drilling Ultra-Deep Gas Wells Using Artificial Neural Network. *J. Nat. Gas Sci. Eng.* **2020**, *77*, No. 103224.

(54) Oguntade, T.; Ojo, T.; Efajemue, E.; Oni, B.; Idaka, J. In *Application of ANN in Predicting Water Based Mud Rheology and Filtration Properties*, Society of Petroleum Engineers - SPE Nigeria Annual International Conference and Exhibition20, NAIC 2020, 2020 DOI: 10.2118/203720-MS.

(55) Gowida, A.; Elkatatny, S.; Abdelgawad, K.; Gajbhiye, R. Newly Developed Correlations to Predict the Rheological Parameters of High-Bentonite Drilling Fluid Using Neural Networks. *Sensors* **2020**, *20*, No. 2787.

(56) Elkatatny, S. Artificial Neural Network Models for Real-Time Prediction of the Rheological Properties of NaCl Mud. *Arabian J. Geosci.* **2020**, *13*, 1−10.

(57) Liu, N.; Zhang, D.; Gao, H.; Hu, Y.; Duan, L. Real-Time Measurement of Drilling Fluid Rheological Properties: A Review. *Sensors* **2021**, *21*, No. 3592.

(58) Lantz, B. Machine Learning with R, 3rd ed.; 2019, p 458.

(59) Kotsiantis, S. B. Use of Machine Learning Techniques for Educational Proposes: A Decision Support System for Forecasting Students' Grades. *Artif. Intell. Rev.* **2011**, *37*, 331−344.

(60) Nettleton, D.Selection of Variables and Factor Derivation. In *Commercial Data Mining*; Elsevier, 2014; pp 79−104. DOI: 10.1016/B978-0-12-416602-8.00006-6.

(61) Wu, X.; Kumar, V.; Ross, Q. J.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G. J.; Ng, A.; Liu, B.; Yu, P. S.; Zhou, Z. H.; Steinbach, M.; Hand, D. J.; Steinberg, D. Top 10 Algorithms in Data Mining. *Knowl. Inf. Syst.* **2007**, *14*, 1−37.

(62) James, G.; Witten, D.; Hastie, T.; Tibshirani, R.Tree-Based Methods. In *An Introduction to Statistical Learning*, Springer Texts in Statistics; Springer, 2013; pp 303−335. DOI: 10.1007/978-1-4614-7138-7_8.

(63) Zhang, C.; Ma, Y. *Ensemble Machine Learning: Methods and Applications*; Springer, 2012.

(64) Blakely, L.; Reno, M.; Broderick, R. *Evaluation and Comparison of Machine Learning Techniques for Rapid QSTS Simulations*; Sandia National Lab, 2018.

(65) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.

(66) Hegde, C.; Gray, K. E. Use of Machine Learning and Data Analytics to Increase Drilling Efficiency for Nearby Wells. *J. Nat. Gas Sci. Eng.* **2017**, *40*, 327−335.

(67) Pitt, M. J. The Marsh Funnel and Drilling Fluid Viscosity: A New Equation for Field Use. *SPE Drill. Completion* **2000**, *15*, 3−6.

(68) Almahdawi, F. H. M.; Al-Yaseri, A. Z.; Jasim, N. Apparent Viscosity Direct from Marsh Funnel Test. *Iraqi J. Chem. Pet. Eng.* **2014**, *15*, 51−57.