



# Characterization based machine learning modeling for the prediction of the rheological properties of water-based drilling mud: an experimental study on grass as an environmental friendly additive

Atif Ismail<sup>1</sup> · Hafiz Muhammad Awais Rashid<sup>1</sup> · Raouf Gholami<sup>2</sup> · Arshad Raza<sup>3</sup>

Received: 28 September 2021 / Accepted: 14 December 2021  
© The Author(s) 2021

## Abstract

The successful drilling operation depends upon the achievement of target drilling attributes within the environmental and economic constraints but this is not possible only on the basis of laboratory testing due to the limitation of time and resources. The chemistry of the mud decides its rheological potential and selection of the techniques required for recycling operations. Conductivity, pH, and photometer testing were performed for the physio-chemical characterization of the grass to be used as an environmental friendly drilling mud additive. In this study, different particle sizes (75, 150, and 300  $\mu\text{m}$ ) of grass powder were mixed in mud density of 8.5, 8.6, and 8.7 ppg in the measurement of gel strength and viscosity of drilling mud. The grass additive was added in different weight conditions considering no additive, 0.25, 0.5, and 1 g to assess the contribution of grass on the gel strength and viscosity of the drilling mud. The machine learning techniques (Multivariate Linear Regression Analysis, Artificial Neural Network, Support Vector Machine Regression, k-Nearest Neighbor, Decision Stump, Random Forest, and Random Tree approaches) were applied to the generated rheological data. The results of the study show that grass can be used for the improvement of the gel strength and viscosity of the drilling mud. The highest improvement of the viscosity was seen when grass powder of 150  $\mu\text{m}$  was added in the 8.7 ppg drilling mud in 0.25, 0.5, and 1 g weights. The gel strength of the drilling mud was improved when the grass additive was added to the drilling mud 8.7 ppg. Random forest and Artificial Neural Network had the same results of 0.72 regression coefficient ( $R^2$ ) for the estimation of viscosity of the drilling mud. The random tree was found as the most effective technique for the modeling of gel strength at 10 min ( $GS_{10_{\text{min}}}$ ) of the drilling mud. The predictions of Artificial Neural Network had 0.92  $R^2$  against the measured gel strength at 10 s ( $GS_{10_{\text{sec}}}$ ) of the drilling mud. On average, Artificial Neural Network predicted the rheological properties of the mud with the highest accuracy as compared to other machine learning approaches. The work may serve as a key source to estimate the net effect of grass additives for the improvement of the gel strength and viscosity of the drilling mud without the performance of any large number of laboratory tests.

**Keywords** Drilling engineering · Natural additive · Machine learning · Supervised learning · Neural network · Support vector machine regression

---

✉ Atif Ismail  
atifismail.uet@gmail.com

<sup>1</sup> Department of Geological Engineering, University of Engineering and Technology (UET), Lahore, Pakistan

<sup>2</sup> Department of Energy Resources, University of Stavanger, Stavanger, Norway

<sup>3</sup> Department of Petroleum Engineering, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

## Introduction

Machine learning (ML), the subclass of artificial intelligence (AI), is a technique in which models are formulated based on data, and the models are forced to recognize the pattern in the data by training processes (Davenport and Kalakota 2019). ML is based on the principles of computer science, statistics, and all other fields of study which can model the behavior of decision making in doubtful conditions. ML has been used efficiently in the field of automation, speech recognition, computer vision, neuroscience, and many other domains of life for software development concerning predictions (Jordan and Mitchell 2015; Kersting 2018). Data scientists consider machine learning as an important tool to handle large sources of data for the development of accurate data-driven predictive models (Provost and Fawcett 2013). ML is used to model the nonlinear relationships by observation of the available patterns in the data. The performance of machine learning depends upon the selection of the algorithm used in the training phase of the model (Kotsiantis et al. 2007). The popularity of ML has increased because of recent developments in algorithms and improved computing facilities (Dimiduk et al. 2018; LeCun et al. 2015; Schmidt and Lipson 2009).

Recently, the application of AI has advanced the reservoir characterization in the petroleum engineering domain (Al-Bulushi et al. 2009; Al-Marhoun et al. 2012; Anifowose et al. 2014; Asadisaghandi and Tahmasebi 2011; Barros and Andrade 2013; Dutta and Gupta 2010; Hegde et al. 2019; Ismail et al. 2017a, b; Waqas et al. 2020). The reported influential performance of AI has enhanced the application of machine learning in the petroleum industry. Machine learning has been used in the prediction of various rock and fluid properties that are significant in the execution of different petroleum operations. Since 1970, artificial neural network (ANN), regression modeling, and support vector machine (SVM) have been used successfully in the field of geophysics, well logging, and production engineering. ML has become an important part of data analysis in all kinds of industries, but relatively this ML is underutilized in geoscience (Maniar et al. 2018). The application of ML has improved the precision and efficiency of the drilling operation that saved the industry from excessive financial and technical constraints. In the oil and gas industry, machine learning models have improved the relationship between input and output variables without significantly dependent upon the structure of the system (Noshi and Schubert 2018). The prediction of accidents in directional drilling operations can be done by using ML techniques. The designed models work by comparison of real time-data with the accidents that happened in the past (Gurina et al. 2020). Drilling optimization plays an important role in the economical and technical

goals during drilling operations. The drilling optimization is done by identification of the optimum parameters using different empirical and machine learning techniques (Hegde and Gray 2018). The identification of the lithology and formation horizons is successfully done by using the neural network models (Mahmoud et al. 2021).

Drilling mud is a viscous and heavy fluid used in the drilling operations to carry the cuttings from the borehole, maintenance of borehole pressure, and lubrication of the drilling bit. Drilling mud is considered as one of the most important parts of the drilling activity. A wide range of chemicals and polymers are used for the achievement of the required properties of the drilling fluid. The main properties of the drilling mud include viscosity, yield point, gel strength, mud density, fluid loss control property, rate of penetration, and filtration control agents (Amanullah et al. 2016; Barbosa et al. 2019; Gul and van Oort 2020). The control of the physical properties of the drilling mud is very important to get its optimum performance. Viscosity and gel strength are two of the most important properties to control the drilling fluid (Abdou and El-Sayed Ahmed 2011). In recent years, the consideration of the conservation of the environment has also affected the selection of the muds and their additives in hydrocarbon drilling and production activities (Caenn et al. 2011). The advancement in technology in the drilling industry has brought a chance of the replacement of conventional mud additives with environmental friendly constituents. The preparation of the eco-friendly drilling mud is influenced by several factors such as the requirement of the complex mud treatment facility, high initial cost, and low availability of the raw material (Abdou and El-Sayed Ahmed 2011; Kok and Alikaya 2003; Lan et al. 2009; Li et al. 2002; Zhao et al. 2009). During drilling operations, hazardous vapors emit from drilling mud depending upon the type of additives in the mud may cause very detrimental effects on human health if they exceed safe exposure limits. Different materials such as brines, cleaning agents, solvents, and other fluids associated with drilling mud may damage the skin upon contact. The deleterious effect of the additives such as lubricants, viscosities, thinners, descalers, defoamers, stabilizers, corrosion inhibitors, and surfactants on living beings have been documented by many authors (Ameille et al. 1995; Candler et al. 1992; Greaves et al. 1997). The hazardous effects of the additives in the drilling mud give rise to a need for the development of environmental friendly replacements which may have the same efficiency as the conventional makeup of the drilling mud (Apaleke et al. 2012). The toxicity of the drilling mud depends upon the nature of the additive, which opens the chance of water-based drilling fluid as environmentally hazardous as well. The level of damage to the environment depends upon the proportion of the type of drilling fluid with a concentration of the additive and the rate at which consumed drilling mud is discharged open in the

environment. Recent research has been focused on the preparation of a substitute for oil-based drilling fluid which may act as more environmental friendly. The increase in environmental restrictions has boosted the idea of oil-based drilling fluid replacement (Wajheuddin 2014). Recent studies have shown that additives such as diesel-based/mineral-based fluids have high toxicity ranges (Dosunmu 2010; Duchemin et al. 2008; Rana 2008). There are five (5) drilling properties such as viscosity, the density of drilling mud, filter cake, solids content, and quality of water makeup which as a part of a drilling program are closely checked during drilling activity (Hossain and Wajheuddin 2016). Support Vector Machine can be used for the study of metabolization in the feed of the animals (Ahmadi and Rodehutsord 2017). Zhang et al. (2018) used the computer-aided algorithm for the study of uncertainty analysis in managed pressure drilling (MPD). Bhandari et al. (2015) and Mohaghegh (2015) studied the application of machine learning in the prediction of blow-out and drilling anomalies in operations.

The concentration of the oil and gas industry has been diverted towards natural environmentally friendly additives to get rid of the hazardous effects of conventional additives in the drilling mud for rheological improvement. Several researchers have worked on the application of natural materials such as rice husk, sugar cane ash, coconut shells, cocoa beans shells, date seed powder, fibers, grass, and a wide range of other locally available materials for the improvement of drilling mud properties. One of the main applications of natural materials in drilling operations is to control loss circulation. As shown in Table 1, many researchers have successfully used the environmental friendly materials as a source for loss circulation prevention and improvement of

the flowing properties of the drilling mud. The natural vegetable is used for the improvement of resistance of drilling mud against temperature variations. The temperature resistance of the drilling mud is enhanced by 40% with the use of vegetable gum as a drilling mud additive (Li et al. 2002).

Natural polymers such as carboxymethyl cellulose, guar gum, and starch are found in large amounts in the environment and are very cheap which can be used for reduction in loss circulation (Kok and Alikaya 2003). Cellulose from groundnut husk can be a very good material for fluid loss control if used in high concentrations (Dagde and Nmegbu 2014). On the other hand, the grass is the main feed for all animals and is present in large amounts on the earth which are considered as a major producer of cellulose (Nmegbu and Bekee 2014). The average composition of grass includes cellulose, hemicellulose, and lignin which are useful as an additive in the drilling mud while major elements are Calcium, Potassium, and Chlorine as per results from XRD analysis (Table 2). Wajheuddin and Hossain (2018) studied

**Table 2** XRF analysis of grass (Hossain and Wajheuddin 2016)

Element	(wt %)
Calcium	53.80
Potassium	19.83
Chlorine	15.54
Sulfur	3.89
Silicon	3.13
Iron	2.46
Phosphorus	1.24
Sulfur	0.11
Manganese	0.12

**Table 1** Use of environmental friendly additives in drilling mud

References	Natural material	Function
(Nestle 1952)	Tree bark (douglas fir)	Filtration controlling
(Morris 1962)	Ground peach seeds	Filtration controlling
(Lummus and Ryals 1971)	Ground nutshells and nur flour	Filtration controlling
(Burts Jr 2001)	Corn cob outers	Filtration controlling
(Green 1984)	Ground cocoa bean shells	Lost circulation material
(Burts Jr 1992)	Rice fractions (rice hulls, rice tips, rice straw, and rice bran)	Lost circulation material
(Cremeans and Cremeans 2003)	Cottonseed hull	Lost circulation material
(Sharma and Mahto 2006)	Tamarind gum	Viscosifier
(Macquoid and Skodack 2004)	Coconut coir	Lost circulation material
(Sampey 2006)	Sugar cane ash	Filtration Controlling
(Ghassemzadeh 2011)	Fibers	Lost Circulation material
(Okon et al. 2014)	Rice husk particles	Filtration controlling
(Dagde and Nmegbu 2014)	Groundnut Husk	Filtration controlling
(Amanullah et al. 2016)	Date seed powder	Filtration controlling
(Hossain and Wajheuddin 2016)	Grass	Rheological improvement
(Wajheuddin 2014)	Grass	Rheological improvement

the effects of grass as a natural additive in the drilling mud of 8.6 ppg and found that the flowing properties of the drilling mud are improved using grass as a natural additive. It has been reported by Hossain and Wajheuddin (2016) that grass can be used to handle fluid loss problems keeping the mud environmental friendly. The addition of grass in the drilling mud improved the viscosity and gel strength of the drilling mud efficiently when a particle size of 300  $\mu\text{m}$  was used. Grass not only modifies the rheological properties but also acts as material for fluid loss. Starch is relatively less effective as compared to grass as a naturally available environmental friendly additive. The amount of salt concentration increases with the addition of grass in the drilling mud that modifies the rheological properties of the drilling mud (Al-Hameedi et al. 2019). Al-Saba et al (2018) used many naturally available drilling additives to evaluate the effects of the rheology of the drilling mud and concluded that soybean skin and coconut shell powder are most effective against the fluid loss.

In this study, the performance of the machine learning techniques has been assessed in terms of the prediction of the rheological properties of the drilling mud. The grass is used as an environment-friendly additive to improve the properties of the drilling mud. The characterization of the grass is done by physio-chemical experimentation. The influence of the grass as an environmental friendly additive has been studied on water-based drilling mud of 8.5, 8.6, and 8.7 ppg. Multivariate Linear Regression Analysis, Artificial Neural Network, Support Vector Machine Regression, k-Nearest Neighbor, Decision Stump, Random Forest, and Random Tree approaches are used to predict the rheological properties of the drilling mud by application on the experimental data.

## Statistical and machine learning models

Machine learning is divided into three classes as supervised, unsupervised, and reinforced learning; supervised learning includes the learning from target data, in unsupervised learning the model finds a pattern in the data, and the unsupervised learning model is planned to make predictions without contribution from target data (Schmidt and Lipson 2009). Supervised learning estimations linked with the trained model which is controlled by target data. A lot of the algorithms are present for the estimations based on supervised learning such as decision algorithms, regression analysis, Support Vector Machine, and kernel machines (Jordan and Mitchell 2015). In unsupervised learning, the target data is missing in the model formulation and prediction as compared to supervised learning. One example of unsupervised learning is cluster analysis in which the algorithm clusters the data by identifying the distinct patterns in the input data.

In reinforcement learning, the model takes actions grounded on the information acquired from the environment provided to maximize the accuracy (Kaelbling et al. 1996).

*Multivariate linear regression analysis* (MVLA) is one of the first approaches for the formulation of the predictive model. The advantage of this technique as compared to simple regression is more accuracy and the capability to summarize more information. MVLA can reveal the correlation between dependent and independent variables (Eq. 1). MVLA is dependent upon the average trend between independent and dependent variables (Ismail et al. 2017a, b; Kumar et al. 2019). The regression coefficient was improved by 27% when MVLA was employed for the prediction of shear wave velocity (Du et al. 2019). MVLA has been used successfully in petrophysics and well logging for the estimation of shear velocity

$$Y = a_0 + a_1 V1 + a_2 V2 + a_3 V3 \dots a_4 V4 \quad (1)$$

whereas  $Y$  is a dependent variable  $a_0, a_1, a_2, a_3, a_4$  are coefficients determined using Multivariate Linear Regression modeling, and  $V1, V2, V3, V4$  are independent variables.

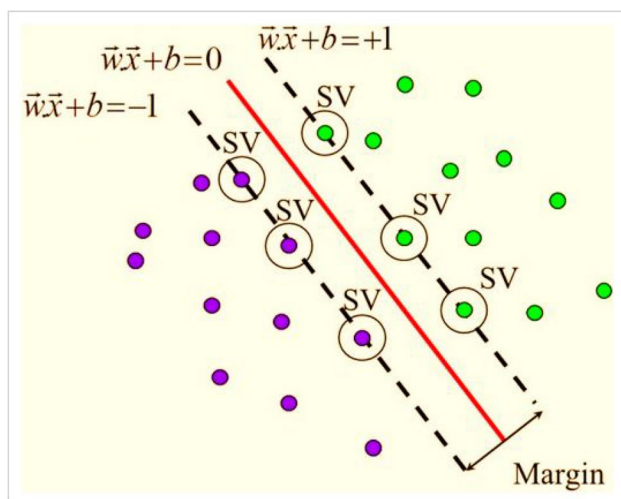
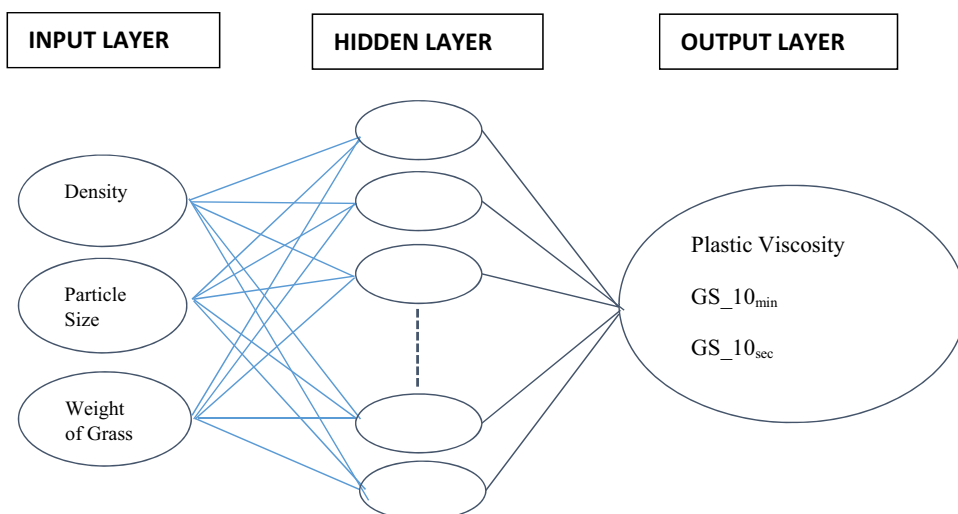
*Artificial neural network* (ANN) is a process that is based on the human brain simulation in which there is the ability to identify the relationship between input and target data. ANNs can make alternations, connections, and predictions from the input data. The neural network is considered an important source to model the nonlinear trend and complexity of the data when conventional mathematical modeling fails. Neural networks are extensively used in the petroleum engineering domain to model fluid and rock properties (Noshi and Schubert 2018). The neural network belongs to supervised learning that is structured into multilayers which compute the functions between input data (Bengio 2009; Schmidhuber 2015).

The implicit relationship developed by ANN is a nonlinear function based on backpropagation modeling in which it is not possible to express as an explicit expression (Shi et al. 2004) (Eq. 2).

$$\text{Dependent variable} = \text{ANN}(\text{variable 1, variable 2} \dots \text{variable } n) \quad (2)$$

The architecture of the ANN consists of a large number of layered neurons with adjusted weights (Helle et al. 2001). The working of the neural network includes the training and prediction phase. In training, the model is trained to adjust the weights between layers of neurons depending upon the output target data as shown in Fig. 1. The trained model is further applied to the input data to make predictions for the dependent variable. The backpropagation learning algorithm is used for the learning phase in the prediction from the Artificial Neural Network (Haykin 1994; Lim 2003).



**Fig. 1** Architecture of the neural network**Fig. 2** Diagrammatic representation of the Support Vector Machine (Cortes and Vapnik 1995)

*K-nearest neighbor* (KNN) is built by a comparison between training and test data set using an unsupervised learning approach. In KNN, the data is classified using recognized patterns into clusters which are further used for computations purposes (Araghinejad 2013; Cho 2018; Imandoust and Bolandraftar 2013). KNN can be used for regression modeling by a selection of property for each nearest neighbor cluster. The weight is adjusted in such a way that the contribution of the nearest cluster is considered greater as compared to far ones. The distance is measured from the target data to the nearest neighbor by using an appropriate approach. Usually, the Euclidean approach is used to measure the above-mentioned distance (Imandoust and Bolandraftar 2013).

*Support Vector Machine* is a simple nonlinear machine learning algorithm. The approach defines a hyper-plane

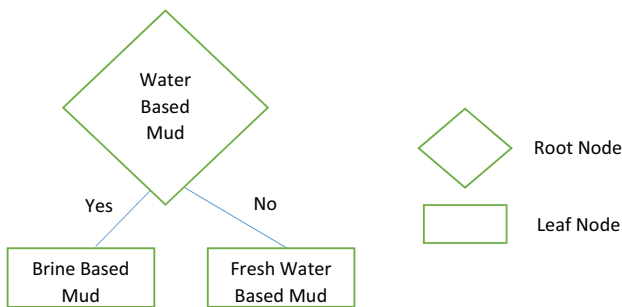
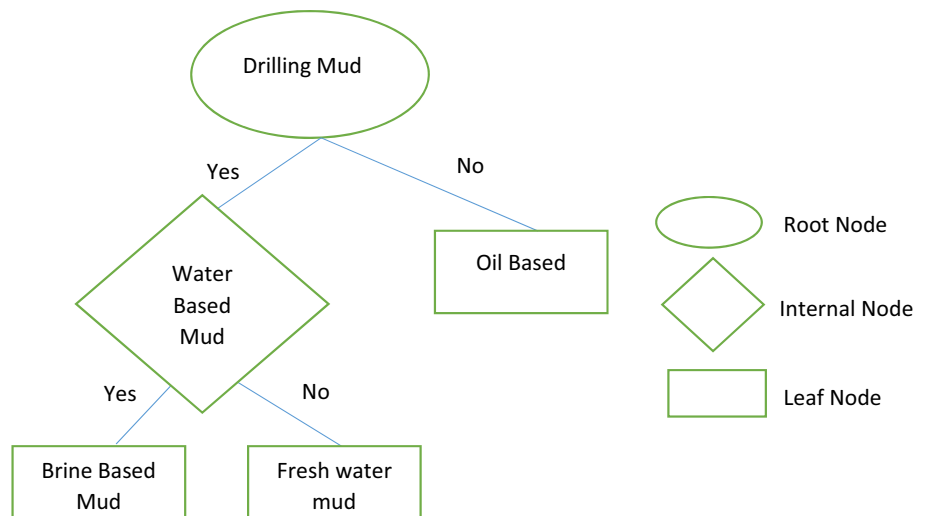
that splits the input data into distinct classes. Hyperplane increases the distance between the two classes and is defined by the distance between input data as shown in Fig. 2. The separations boundaries are called a support vector.

where SV is a support vector, and  $w$  and  $b$  are the parameters of the linear function. The parameter  $x$  represents the relation between the training set and decision function margin maximization. The application of a proper kernel approach boosts the hyperplane margin between classification instances (Bishop 2006; Zhang et al. 2018). Categorically, the Support Vector Machine is divided into support vector classification and support vector regression. The generalized performance of the model is achieved by reducing the generalization error in the support vector regression approach. The application of SVM to solve problems has evolved in many fields of study (Awad and Khanna 2015). SVM is considered efficient in the classification approaches, but the rules obtained as a product of SVM are very difficult to understand (Zhang et al. 2018).

A *decision tree* is an approach to show an order of the conditions that constitute a product (Peng et al. 2009). Decision trees are structured on three parts known as the root node, internal node, and leaf node as shown in Fig. 3. The startup information is known as the root node while the leaf node is the terminal node. The nodes located in between the above-mentioned nodes are called internal nodes. There is a lot of decision tree algorithms that are used for the classification of the data. Those algorithms include ID3, AD Tree, REP, J48, FT Tree, LAD Tree, decision stamp, LMT, random forest (Sewaiwar and Verma 2015).

*Decision Stump* is an algorithm in machine learning in which there is only one root node. The root node is divided into several leaf nodes depending upon the attributes of the data. The complex relationship between variables can be modeled using Decision Stump. This algorithm also has the ability to distinguish the nature of variables in terms of importance and takes

**Fig. 3** Structure of Decision Tree (For brine-based mud as a target)



**Fig. 4** Structure of Decision Stump (For brine-based mud as a target)

less time in the training phase as compared to neural networks (Ali et al. 2012). The Decision Stump tree is also called 1-rule due to one-level decision structure as shown in Fig. 4.

*Random force algorithm* is the ensemble process of all Random Trees generated from the application of attribute selection on samples of the training phase. The prediction from the Random Forest is done by cumulating the dominated output from decision trees. Interpretation problems arise in the model due to the average of the decision trees but the overall performance is enhanced in the Random Forest algorithm (Breiman 2001).

Considering  $M$  trees, ensembling of these trees in the Random Forest will be as  $\{T_1(X) \dots T_M(X)\}$  Where  $X = \{x_1, \dots, x_p\}$ . Due to the ensemble, we have  $M$  products  $\{\hat{Y} = T_1(X), \dots, \hat{Y}_M = T_M(X) \text{ where } \hat{Y}_k, k = \hat{1}, \dots, M\}$ . The output of all trees is combined to get the final prediction ( $\hat{Y}$ ). The prediction in classification problems is the decision by the majority of the trees while in regression the prediction of each tree is averaged (Svetnik et al. 2003).

*The Random Tree* belongs to supervised classification in which distinct learners are generated using the ensemble learning approach. In this technique, bootstrap aggregation of the decision trees is done. Random tress takes input,

then classification is done with all the available decision trees, and output is decided by analyzing the majority trend. A Random Tree algorithm can be used for classification as well as regression approaches (Mishra and Ratha 2016). The Random Tree is the mixture of a single model tree and the Random Forest approach. Decision trees have a linear model which is modified according to information held by each leaf. The performance of the model is modified on the basis of decision tree performance (Kalmegh 2015; Shajahaan et al. 2013). The Random Tree approach has been widely applied efficiently in many studies as a part of machine learning. The Random Tree can be used for classification as well as the regression of the data (Basak 2010).

## Material and methods

The detail of the methodology followed in this research work is as follows;

### Additive preparation

Domestic grass was selected for its potential use as an environmental friendly additive in drilling mud. The grass was allowed to dry in sunlight for three days and subsequently ground in a food processor to make moisture-free grass powder. The main purpose of using grass as an additive was to introduce a widely available cheap material for improvement in the drilling mud efficiency. The powdered grass material was sealed in an airtight bottle to avoid any possible contact with atmospheric moisture. Sieve analysis was performed on the prepared grass powder for particle size distribution analysis. Different grades of the sizes of the grass powder were separated using the mesh size of sieves used in particle size distribution analysis. On the basis of sieve analysis, 38, 75,

**Table 3** Particle size distribution of grass sample

Sieve size (μm)	Weight retained (%)
1190	3.29
600	8.10
425	5.05
300	6.03
250	47.14
150	22.75
75	3.50
50	1.00
38	3.14

**Table 4** Composition of Bentonite in prepared samples

Density of mud (ppg)	(W) Weight of bentonite/350 ml of Water (g)
8.5	16.73
8.6	26.79
8.7	37.00

and 300 μm particle sizes were used in this research as given in Table 3.

### Mud sample preparation

A mud mixer was used to mix the bentonite with water to get water-based drilling mud of 8.5, 8.6, and 8.7 ppg (Table 4).

$$W = \frac{42D_A(D_F - D_o)}{D_A - D_F} \quad (3)$$

The values of the added materials used in the preparation of the samples (Eq. 3) are as under;

$D_A = 21.7$  ppg (Density of bentonite).

$D_F = 8.5, 8.6, 8.7$  ppg (Final mud density).

$D_o = 8.33$  ppg (Density of water).

### Physio-chemical and rheological testing

In this work, Paqualab photometer and titration analysis were used for the estimation of chlorine, calcium, sulfate,

ammonia, aluminum, magnesium, and fluoride as shown in Table 5. The electrical conductivity of the grass extract was 9.13mS measured with Lovibond electrical conductivity meter while 6.4 pH was measured using Hanna pH meter. Mud was prepared with densities of 8.5, 8.6, and 8.7 ppg. The designed density of the mud was confirmed using mud balance and for further adjustments. The mud was mixed with specific amounts of grass additive with the grain size of 38, 75, and 300 μm. Fann VG meter was used to measure the viscosity and gel strength of the prepared mud samples.

### Comparison of machine learning approaches

In this study, the performance of the machine learning techniques has been assessed in terms of the prediction of the rheological properties of the drilling mud. Multivariate Linear Regression Analysis, Artificial Neural Network, Support Vector Machine Regression, k-Nearest Neighbor, Decision Stump, Random Forest, and Random Tree approaches are used to predict the rheological properties of the drilling mud by application on experimental data.

Computer-aided approaches have been used for the modeling of the rheological behavior of the drilling mud. Supervised and unsupervised machine learning is used to study the effects of grain size and density of drilling mud on major rheological properties of the drilling mud. Interactive Petrophysics and Weka suite were used to conduct this study. In this study, 30 rheological tests were performed on the grass sample with varying particle size and drilling mud density. A tenfold cross-validation method was used to test the prediction performance of each developed machine learning model.

### Results and discussions

The diverse concentrations of the grass were added in the drilling mud of different densities to study its effects as an environmental friendly additive on rheological properties of the drilling mud. Machine learning approaches are used to study the application of machine learning in the prediction of rheological properties.

### Effect of grass as additive on rheological properties of the drilling mud

The chemical composition of the grass determines the effects on the drilling fluid as a natural additive. The grass is considered as one of the natural environmental friendly additives

**Table 5** Elemental composition of the grass extract

Calcium (mg/L)	Chlorine (mg/L)	Sulphate (mg/L)	Ammonia (mg/L)	Aluminium (mg/L)	Magnesium (mg/L)	Flouride (mg/L)
44	444	648	0.96	3.28	23	0.4

which improve the rheological properties of the drilling mud without creating toxic chemicals as a consequence (Table 1). The rheology of the drilling mud can be modified using grass as a natural additive (Hossain and Wajheuddin 2016). In this study, the grass was added into mud samples of 8.6, 8.7, and 8.8 ppg mud to study the effects of that additive on viscosity and gel strength of the water-based drilling mud. For the assessment of the behavior of grass on the viscosity and gel strength, different concentration of grass was mixed in mud density of 8.5, 8.6, and 8.7 ppg.

The viscosity and gel strength of the drilling mud was measured with addition of different amount and particle size of grass powder in mud density of 8.5, 8.6, and 8.7 ppg as shown in Table 6.

Three types of particle size (75, 150, and 300  $\mu\text{m}$ ) of grass for each mud density were used as a natural additive. The contribution of the grass additive in the improvement of the viscosity was significant in water-based drilling muds with

higher density ranges. The effect of the grass concentration was highest in the drilling mud with 8.7 ppg drilling mud. It was observed that the viscosity of the drilling mud does not increase exponentially with the increase of the particle size of the grass. In the absence of any additive, the density of the 8.5 ppg drilling mud was 2 cp which was increased to the highest value of 3 cp when 300  $\mu\text{m}$  of 0.5 g grass were used as an additive in the drilling mud. The average of the values of viscosity at 150  $\mu\text{m}$  particle size of grass shows that viscosity is at the highest value for any density of the drilling mud when the grass of 150  $\mu\text{m}$  particle size is used as a rheological modifier. The viscosity of the drilling mud was reduced in the presence of 75 and 300  $\mu\text{m}$  in 8.7 ppg drilling mud. The grass modified the gel strength of the drilling mud mainly at a density greater than 8.7 ppg. The viscosity and gel strength of the drilling mud was affected by both the physical and chemical properties of the grass powder. In all samples of drilling mud, the highest achieved viscosity

**Table 6** Effects of the weight of grass on rheological properties of drilling mud

Density	Particle size	Weight of grass (g)	R <sub>300</sub> (cp)	R <sub>600</sub> (cp)	PV = R <sub>600</sub> - R <sub>300</sub> (cp)	GS <sub>10 sec</sub> (lbs/100ft <sup>2</sup> )	GS <sub>10 min</sub> (lbs/100ft <sup>2</sup> )	
8.5 ppg	–	0	2	4	2	1	1	
		0.25	2	4	2	1	1	
		0.5	2	4	2	1	1	
	150 $\mu\text{m}$	1	2	5	3	1	1	
		0.25	2.5	5	2.5	1	1	
		0.5	3	5	2	1	1	
		1	2	5	3	1	1	
		75 $\mu\text{m}$	0.25	3	5	2	1	1
		0.5	2	4	2	1	1	
	8.6 ppg	–	0	2.5	4.5	2	1	1
			0.25	3	5	2	1	1
			0.5	3	6	3	1	1
150 $\mu\text{m}$		1	3.5	6	2.5	1	1	
		0.25	3.5	6	2.5	1	1	
		0.5	3	5.5	2.5	1	1	
		1	3.5	6	2.5	1	1	
		75 $\mu\text{m}$	0.25	3	5	2	1	1
		0.5	3	5	2	1	1.5	
8.7 ppg		–	0	4.5	7.5	3	1	2.5
			0.25	5	8	3	1	2.5
			0.5	5	8.5	3.5	1.5	3.5
	150 $\mu\text{m}$	1	5	8	3	1.5	2.5	
		0.25	5	9	4	1.5	3	
		0.5	4	8	4	1.5	2	
		1	4	8	4	1.5	2.5	
		75 $\mu\text{m}$	0.25	4	6	2	1	1.5
		0.5	4	7	3	1.5	2	
	75 $\mu\text{m}$	1	4	7	3	1	2.5	



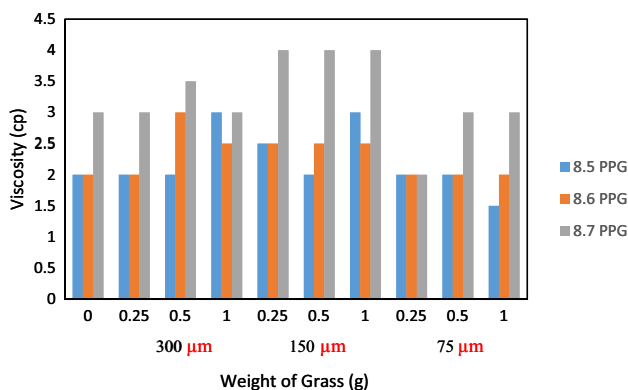


Fig. 5 Effect of grass additive on the viscosity of drilling mud

was 4 cp for drilling mud of 8.7 ppg density (Fig. 5). The results of the viscosity behavior of the drilling mud showed that the influence of the grass powder on the viscosity of the drilling mud depends upon the type of grass as well because the viscosity of the drilling mud with 8.5 ppg density was nearly double in the presence of grass additive as reported by Wajheeuddin and Hossain (2018). The non-uniform chemical composition of the grass around the globe has a wide

Fig. 6 Effect of grass additive on the gel strength of drilling mud of 8.6 ppg density

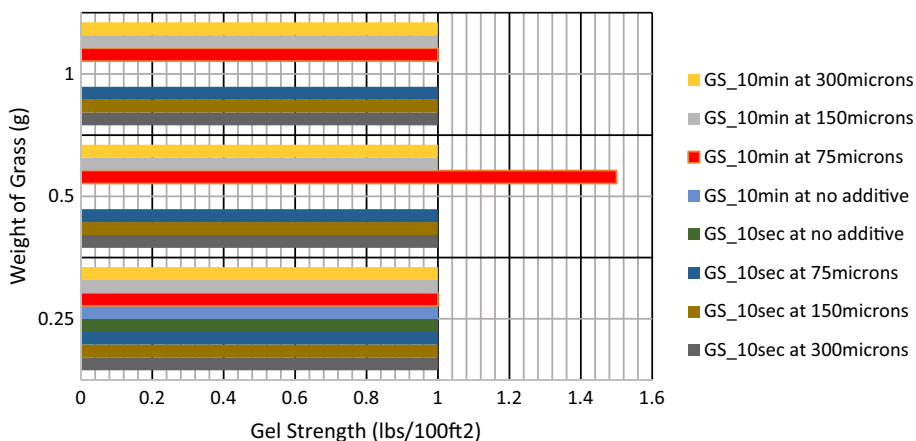
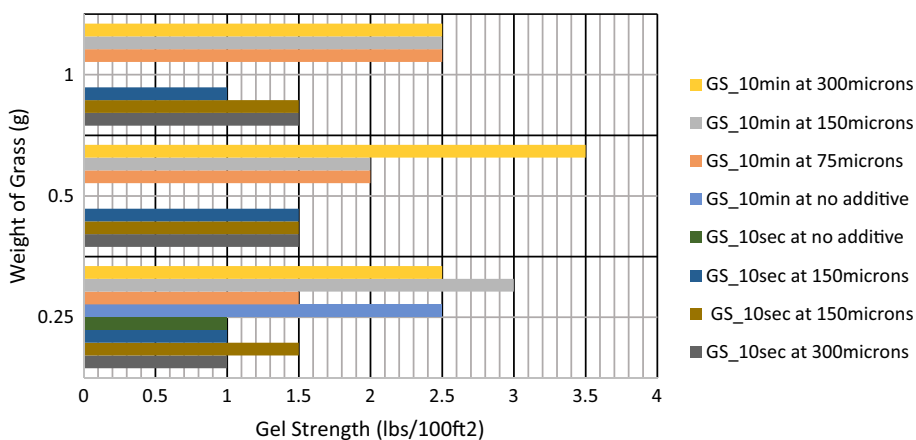


Fig. 7 Effect of grass additive on the gel strength of drilling mud of 8.7 ppg density



range of possibilities in terms of the improvement of the rheological properties of the drilling mud.

Gel strength is an important property of the drilling mud which plays a significant role during the tripping and makes connection operations for suspension of cuttings in the borehole. The used grass did not affect the gel strength of the drilling mud of 8.5 ppg. The 10-s dial gauge reading value was 1 lbs/100ft<sup>2</sup> which remained the same for 10 min reading for the mud density of 8.5 ppg. Due to the low value of GS<sub>10sec</sub>, the mud falls into the flat gel mud but this will create the problem of the settling of solids. This same trend remained consistent for 8.6 ppg mud until the use of 75 μm particle size of grass in 0.5 g weight.

The gel strength of the 8.6 ppg mud was 1 lbs/100ft<sup>2</sup> in the absence of grass additive. The addition of the grass of different particle sizes had no effect on the gel strength of the drilling mud but for the grass of particle size of 75 μm in 0.5 g weight (Fig. 6).

The major improvement of the gel strength was seen in the drilling mud of 8.7 ppg density. The GS<sub>10sec</sub> and GS<sub>10min</sub> were 1 lbs/100ft<sup>2</sup> and 2.5 lbs/100ft<sup>2</sup>, respectively in the absence of any additive addition in the drilling mud. No significant effect on the GS<sub>10sec</sub> was seen with the

addition of grass as additive with the particle size of 75, 150, and 300  $\mu\text{m}$ . The highest improvement in the  $\text{GS}_{10_{\min}}$  was up to the value of 3.5 lbs/100ft<sup>2</sup> when 0.5 g of grass with 300  $\mu\text{m}$  particle size was used as an additive in the drilling mud (Fig. 7). The low 10 s values and  $\text{GS}_{10_{\min}}$  with no significant increase in the values show that the drilling mud is low flat gel mud, which is most desirable. The improvement in the gel strength of the mud was increasing with the increase in the density of drilling mud with the use of grass as an environmental friendly additive.

### Application of machine learning in rheological properties of drilling mud

Machine learning and regression modeling are widely used for the forecasting and modeling of complex engineering problems in the industry. Neural networks are considered an important tool for the selection of bit and drilling fluid problems. In this study, Multivariate Linear Regression Analysis, Artificial Neural Network, Support Vector Machine Regression, k-Nearest Neighbor, Decision Stump, Random Forest, and Random Tree were used to model the effect of grass as an additive to improve the rheological properties of the drilling mud.

### Multivariate linear regression analysis

In Multivariate Linear Regression Analysis, the density of mud, the particle size of grass, and the weight of the grass are related to plastic viscosity,  $\text{GS}_{10_{\text{sec}}}$ , and  $\text{GS}_{10_{\min}}$ . The confidence interval of 95% was selected for the statistical model formulation. It was seen that adjusted  $R^2$  for the statistical model of the plastic viscosity was 42% confirming low reliability of the estimation of PV in terms of density, particle size, and weight of the grass. The regression constant was also low (0.48). It was found that density was the major contributing variable in the relationship, having more than 95% contribution in the estimation of the plastic viscosity (Eq. 4).

$$\begin{aligned} \text{Plastic Viscosity} = & -42.9 + 5.25 \text{ Density} \\ & + 0.0013 \text{ Particle Size} \\ & + 0.264 \text{ Weight of Grass}; R^2 = 0.48 \end{aligned} \quad (4)$$

The statistical analysis of the  $\text{GS}_{10_{\text{sec}}}$  with the density of drilling mud, particle size, and weight of the grass showed that independent variables have very low capacity to estimate the gel strength. The adjusted  $R^2$  for the relationship was 34% which is a very low for a moderate to a good statistical model. The major reasons for this behavior are due to the lack of data points, absence of any average trend in the independent variables corresponding to dependent variable

gel strength. These statistical models show the limitation of the regression analysis in such cases (Eq. 5).

$$\begin{aligned} \text{GS}_{10_{\text{sec}}} = & -11.87 + 1.5 \text{ Density} + 0.00022 \text{ Particle Size} \\ & + 0.06 \text{ Weight of Grass}; R^2 = 0.34 \end{aligned} \quad (5)$$

The regression model of the  $\text{GS}_{10_{\min}}$  had a relatively good-adjusted  $R^2$  of 60% as compared to  $\text{GS}_{10_{\text{sec}}}$  and plastic viscosity. The statistical model is shown in Eq. 5 reveals that the independent variables were predicting the dependent variable with 64% accuracy. The multiple  $R^2$  was high as compared to adjusted  $R^2$  and  $R^2$  because the adjusted  $R^2$  penalized the variables which do not contribute positively to the model. The  $R^2$  assumes that every variable is contributing to the statistical model but adjusted  $R^2$  only improves if the added variables in the model are improving the prediction strength of the model. The standard error of this relationship is 47.3%.

$$\begin{aligned} \text{GS}_{10_{\min}} = & -60.95 + 7.25 \text{ Density} + 0.00072 \text{ Particle Size} \\ & - 0.02 \text{ Weight of Grass}; R^2 = 0.64 \end{aligned} \quad (6)$$

The regression coefficient of all the regression models was low (Eqs. 5–6). Multivariate Linear Regression modeling is not a suitable methodology for the prediction of rheological properties of the drilling mud when the number of data points is low. This happens because the model is unable to mimic the true average relationship between independent and dependent variables for generalization. Linear regression fails if there exists a nonlinear relationship between independent and dependent variables. The performance of linear regression analysis is more affected by outliers as compared to machine learning approaches. In this study, due to the lack of a proper linear relationship between regression variables, the overall regression coefficients were low. Hence, suitable machine learning approaches need to be applied for the prediction of the rheological behavior of drilling mud.

### Artificial neural network

Virtual approaches such as artificial neural network (ANN) are possible replication of the working of the brain. The set of layers of information is called artificial neurons which are roughly the neurons of the human brain. The information is propagated from one layer to another. In ANN, only the information of input and output layers are assessable, but the hidden layer is not clear in the model. The training pass defines how many times the neural network will be trained each time the training phase of the model is done. In each training run, the neural network can sometimes get 'stuck' with what it (erroneously) thinks are the best results. In the used software (Interactive Petrophysics suite), the value of 3 was selected to avoid this problem. 100 epochs per pass were selected to present the training data to neural networks

during each training phase. The magnitude of the closeness of fit (Cfit\_nn) values shows the difference between target data and predicted data. One hidden layer was used in neural network modeling for the prediction of the rheological properties of the drilling mud. The consideration of the high number of hidden layers for a low non-linearity behavior may cause the problem of overfitting. Keeping the simple trend of the input layers into consideration, a shallow neural network (hidden layer=1) was used in this study. Hit and trial methodology was used in the training phase to match the training neural network with target data.

In this approach, firstly the model is forced to learn the behavior of the target data against independent variables in the training phase, and weight is assigned to each neuron. The training model is based upon the discrete attributes of each data point corresponding to target data. In this study, the density of drilling mud, the particle size of grass, and the weight of the

grass are related to the plastic viscosity of the drilling mud using a backpropagation neural network. The training intervals are shown in the second track of Fig. 8. From Cfit\_nn, it can be assessed that the difference between the predicted plastic viscosity and target plastic viscosity is varying between 0–1 cp.

In this study, 70% of the data was used for the model formulation and 30% of the data was used for validation of the ANN models. It was seen that there was good agreement between independent parameters and target parameters in the training phase of the model which can be assessed by close values of target data (PV) and trained data (PV\_nnt). The Artificial Neural Network has successfully predicted the plastic viscosity (PV\_nn) of the drilling mud in the prediction zone. The accuracy of the ANN model was assessed by the bivariate regression model between measured plastic viscosity and predicted plastic viscosity of the test data set.

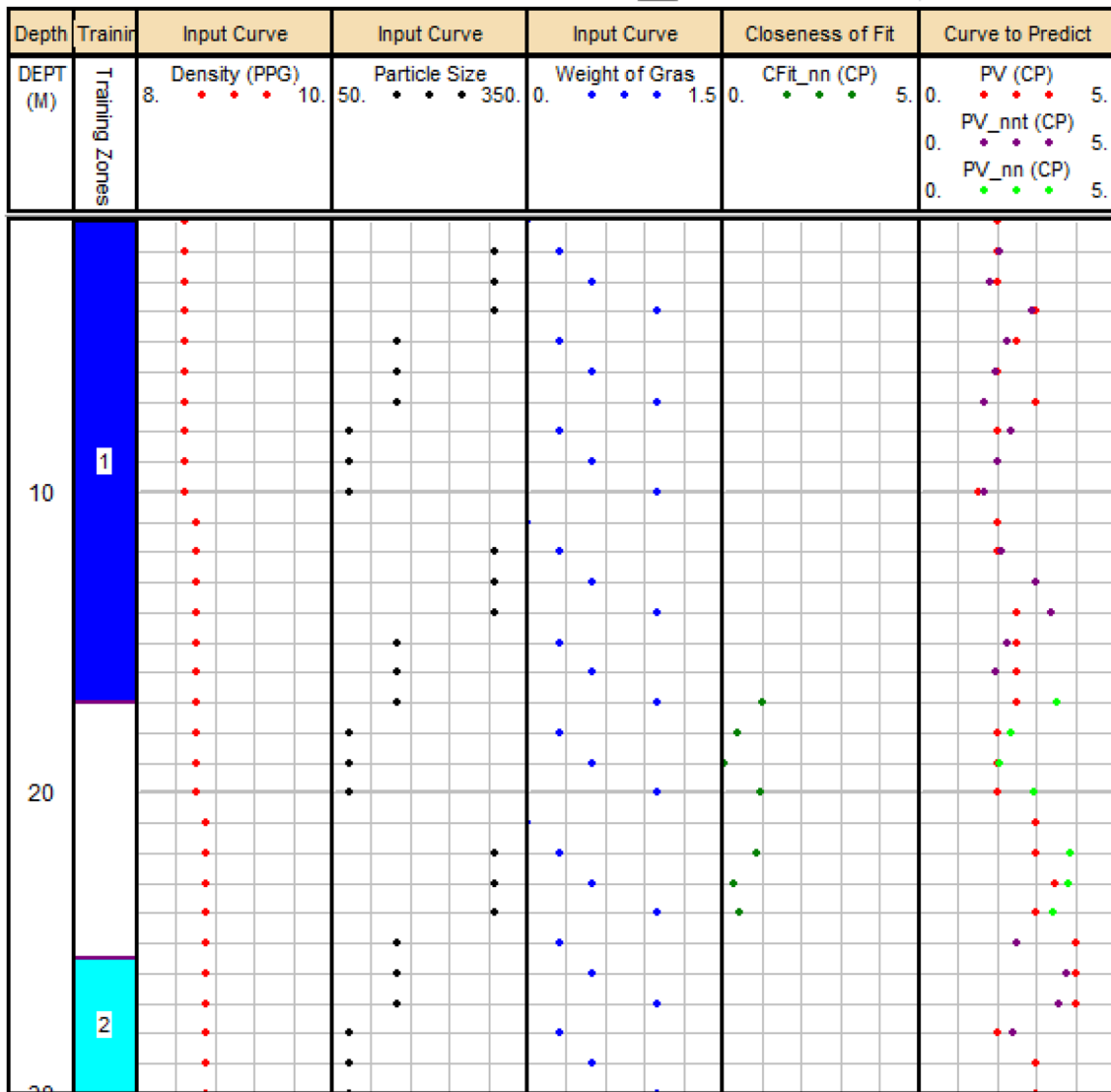


Fig. 8 Application of Artificial Neural Network for Prediction of Plastic Viscosity

The bivariate correlation of the predicted plastic viscosity using an Artificial Neural Network showed that the model was predicting the plastic viscosity with 72% accuracy considering the density of mud, the particle size of grass additive, and weight of the added grass in the drilling mud to improve the rheological properties (Fig. 9).

For GS<sub>10sec</sub>, the ANN model was trained accurately to learn the behavior of input variables and target data. The model was predicting the values of GS<sub>10sec</sub> with the highest accuracy as compared to other parameters. The overlapping of the values of GS<sub>10sec\_nn</sub> and measured GS<sub>10sec</sub> were showing very good agreement with each other as shown in the last track of Fig. 10. The application of ANN can be very effective in experimental studies related to drilling fluids if the model is well trained.

The ANN model successfully predicted the values of GS<sub>10sec</sub> which was validated with the accuracy plot. The accuracy plot showed that the model was predicting the values with 98% accuracy (Fig. 11). The higher accuracy of the model was driven by the accurate training of the predictive model.

Keeping the same parameters, backpropagation Artificial Neural Network was used to formulate a nonlinear neural network model by training on the data set shown by training zone 1 and training zone 2 in the second track (Fig. 12). The model successfully predicts the values of GS<sub>10min</sub> as shown in the last track. At 8.7 ppg density of drilling mud, the model was predicting the values of GS<sub>10min</sub> with higher accuracy as compared to 8.8 ppg (Fig. 12). The target data was representing the major trends in the data set; no effect on the gel strength and improvement in the gel strength of the drilling mud by the addition of grass of specific concentration.

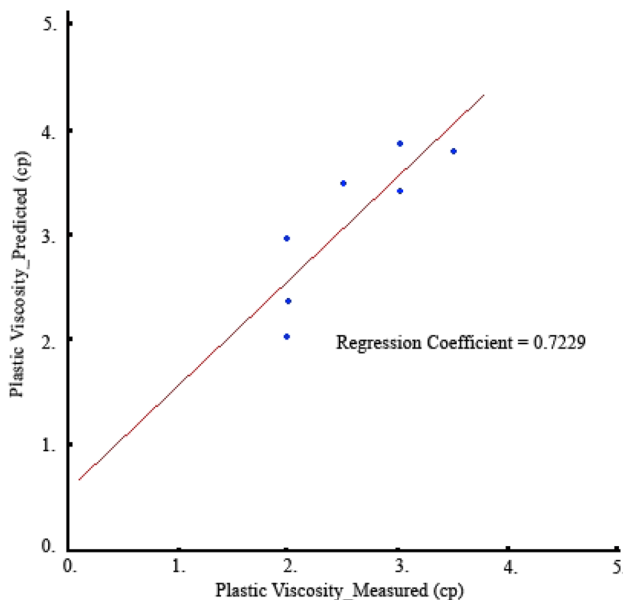


Fig. 9 Accuracy plot of plastic viscosity ANN model

It was seen that the ANN model of GS<sub>10Min</sub> was predicting the values of the prediction interval with 76% accuracy (Fig. 13). The overall efficiency of the neural network is highly dependent upon the limit of the accuracy of the training zone, which is a very difficult job.

### K-nearest neighbor

K-nearest neighbor (kNN) is an example of supervised machine learning which can be easily used for classification as well as regression models. In this study, the selection of the number of neighbors (k) in each model of PV, GS<sub>10sec</sub>, and GS<sub>10min</sub> was based upon the highest efficiency in the prediction. The Euclidian distance was used as a distance metric in the kNN approach methodology. The tenfold cross-validation was used to assess the accuracy of the kNN models. The models were assessed for different k-values. Different models of kNN were formulated for PV for different values of k as shown in Table 7. It was seen that the PV model was predicting the plastic viscosity with 58% accuracy during the tenfold cross-validation process when the value of 2 was selected for the nearest neighbor. To avoid the possible problem of overfitting, a higher nearest neighbor value can be selected. There was no significant effect on the increment of k-value on the coefficient of regression in the k-fold cross-validation process. The optimum value of k=3 was selected in the kNN model of GS<sub>10sec</sub>. The model was estimating the improvement of gel strength as a function of grass additive with a 64% coefficient of regression as evident from the cross-validation approach. The accuracy of the model was decreasing with an increase in the value of k greater than 3. ANN and kNN models of gel strength were predicting the gel strength of the drilling mud with the highest accuracy. The kNN model of GS<sub>10min</sub> was giving the values of GS<sub>10min</sub> with 76% accuracy. The optimum value of 3 was selected in the model formulation of the gel strength of drilling mud as given in Table 7.

### Support vector machine regression (SVMR)

The Support Vector Machine is a very popular machine learning tool for classification and regression models. It is considered a non-parametric approach due to the application of kernel in the development of models. In this study, a Support Vector Machine is used for the regression analysis of PV, GS<sub>10Sec</sub>, and GS<sub>10Min</sub> with the density of drilling mud, the particle size of the grass additive, and the weight of the grass additive. In this study, the polynomial kernel is used in the SVMR. The models were developed for normalized values of the parameters used in SVMR. The hyperplane in the SVMR is affected by the scale of the input feature recommending the possible normalization of the trained data. The regression model was made by utilizing the Support Vector Machine results by relating attributes of the grass additive with rheological properties of the drilling mud. The tenfold cross-validation methodology was used to check the accuracy of the developed models.

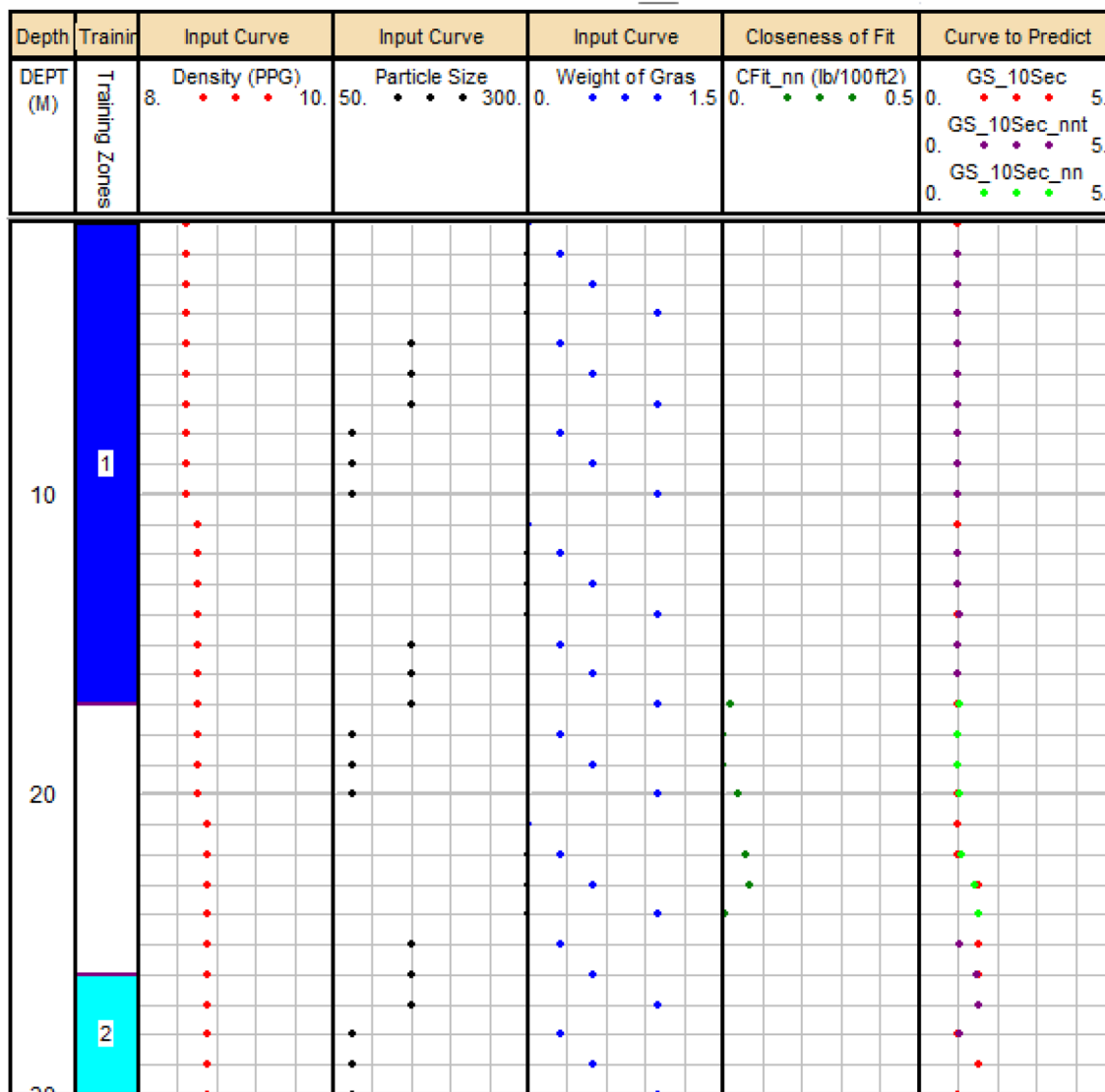


Fig. 10 Application of Artificial Neural Network for the Prediction of Gel Strength at 10 s

The application of the SVMR on the test data set showed that the model was predicting the plastic viscosity of the drilling mud with 63% accuracy. The accuracy of the SVMR for  $GS_{10_{sec}}$  was very low as compared to the plastic viscosity

model. The coefficient of regression for the accuracy plot of  $GS_{10sec}$  was 29%. The prediction ability of the  $GS_{10Min}$  SVMR model was highest as compared to PV and  $GS_{10Sec}$  in the Support Vector Machine approach. The regression model was predicting the test data with 75% accuracy.

$$PV_{NORM} = 0.1999 + 0.3996 (\rho_{mud})_{NORM} + 0.0019 (PS)_{NORM} - 0.0014 (W_{grass})_{NORM} \tag{7}$$

$$GS_{10_{sec} NORM} = -0.0011 + 0.0028 (\rho_{mud})_{NORM} + 0.0007(PS)_{NORM} + 0.0005(W_{grass})_{NORM} \tag{8}$$

$$GS_{10_{min} NORM} = -0.0014 + 0.4029 (\rho_{mud})_{NORM} + 0.0017(PS)_{NORM} - 0.0010(W_{grass})_{NORM} \tag{9}$$



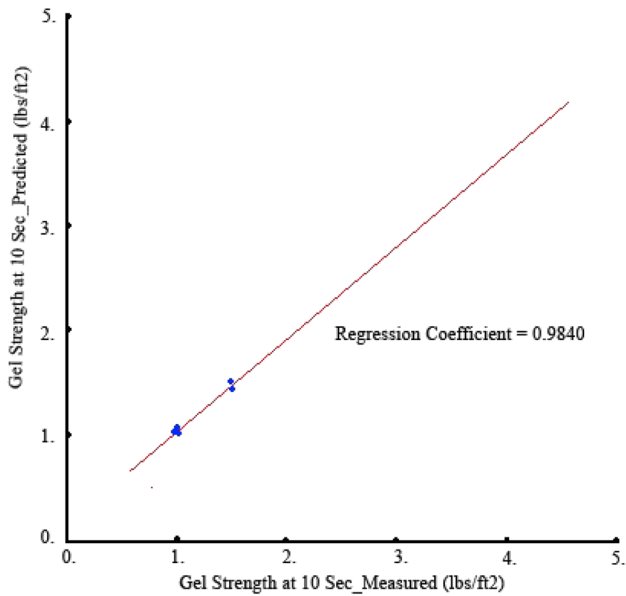


Fig. 11 Accuracy plot of GS\_10Sec ANN model

The normalized regression model based on the Support Vector Machine was predicting the rheological properties of the drilling mud with higher accuracy as compared to Multivariate Linear Regression Analysis (Eqs. 7–9). The regression models of GS\_10Min were predicting the dependent variables with higher accuracy as compared to other developed models.

### Decision stump

Decision Stump is a one-level approach used for the prediction of data by splitting the parent data into different subsets. The process is repeated on all subsets until no new information is generated or the nodes have the same attributes of parent data. In this study, Decision Stump models were developed by modeling the density of drilling mud and grass additive with rheological properties of the drilling fluid. The tenfold cross-validation was used to assess the accuracy of the Decision Stump models. The details of the cutoff selected in the Decision Stump modeling of the rheological properties of the drilling mud are shown in (Table 8). The tenfold cross-validation showed that the model for that was predicting the plastic viscosity with a 61% coefficient of regression. A decision tree is an algorithm that predicts the data on the basis of a single input feature. In this study, the Decision Stump models were based on the density of the drilling mud.

The tenfold cross-validation approach was used to check the efficiency of the Decision Stump model in the prediction of the properties of the drilling fluid. The decision model of GS\_10<sub>min</sub> was predicting the rheological properties with

an 89% coefficient of regression. The prediction scope of the gel strength-based decision models had almost the same ability to predict the properties of drilling fluid (Table 8).

### Random tree

In a Random Tree approach, a tree is based on randomly chosen information at each node according to the feed dataset. In this process, on pruning is done. The value of 1 was considered for the selection of attributes. The Random Tree approach was predicting the plastic viscosity with 56% of the coefficient of regression in the tenfold cross-validation. There were 35 nodes in the formulated Random Tree model of plastic viscosity. The accuracy of the Random Tree model for GS\_10<sub>sec</sub> was low as compared to plastic viscosity. The regression coefficient of the tenfold cross-validation was 37%. GS\_10<sub>min</sub> Random Tree model has the highest  $R^2$  of 61% for the estimation of GS\_10<sub>min</sub> in the application of Random Tree models for the prediction of rheological properties of the mud.

### Random forest

Random Forest is a group of Random Trees. Bagging with 100 iterations was used in Random Tree modeling. The Random Forest model of PV was predicting the plastic viscosity with 72% accuracy. The model for GS\_10Sec was predicting the gel strength with a 61% regression of coefficient. It was seen that the accuracy of the model for GS\_10Min was highest among all other Random Forest tree algorithms. The regression coefficient of the GS\_10Min model was 82% in the tenfold cross-validation approach.

### Comparison of the machine learning approaches

In this study, it was seen that the Artificial Neural Network has the highest ability to predict the rheological properties of the drilling mud under the influence of grass additives. Random Forest was the second most accurate algorithm for the prediction of rheological properties of the drilling mud. The training phase is a very important part of the application of Artificial Neural Networks because the prediction of the data is highly dependent on the quality of the training phase. The ANN approach was predicting the GS\_10<sub>sec</sub> with excellent accuracy of  $R^2 = 0.98$  while Random Forest was predicting the same property with moderate accuracy of 0.61 regression coefficient. In application, Random Forests is an easy approach as compared to a neural network because there is no training required in its application. Random Forest was estimating the GS\_10<sub>min</sub> with greater accuracy as compared to ANN. Most of the machine learning approaches were failed to model the behavior of GS\_10<sub>sec</sub> except Artificial Neural Network. Decision Stump falls in the same class

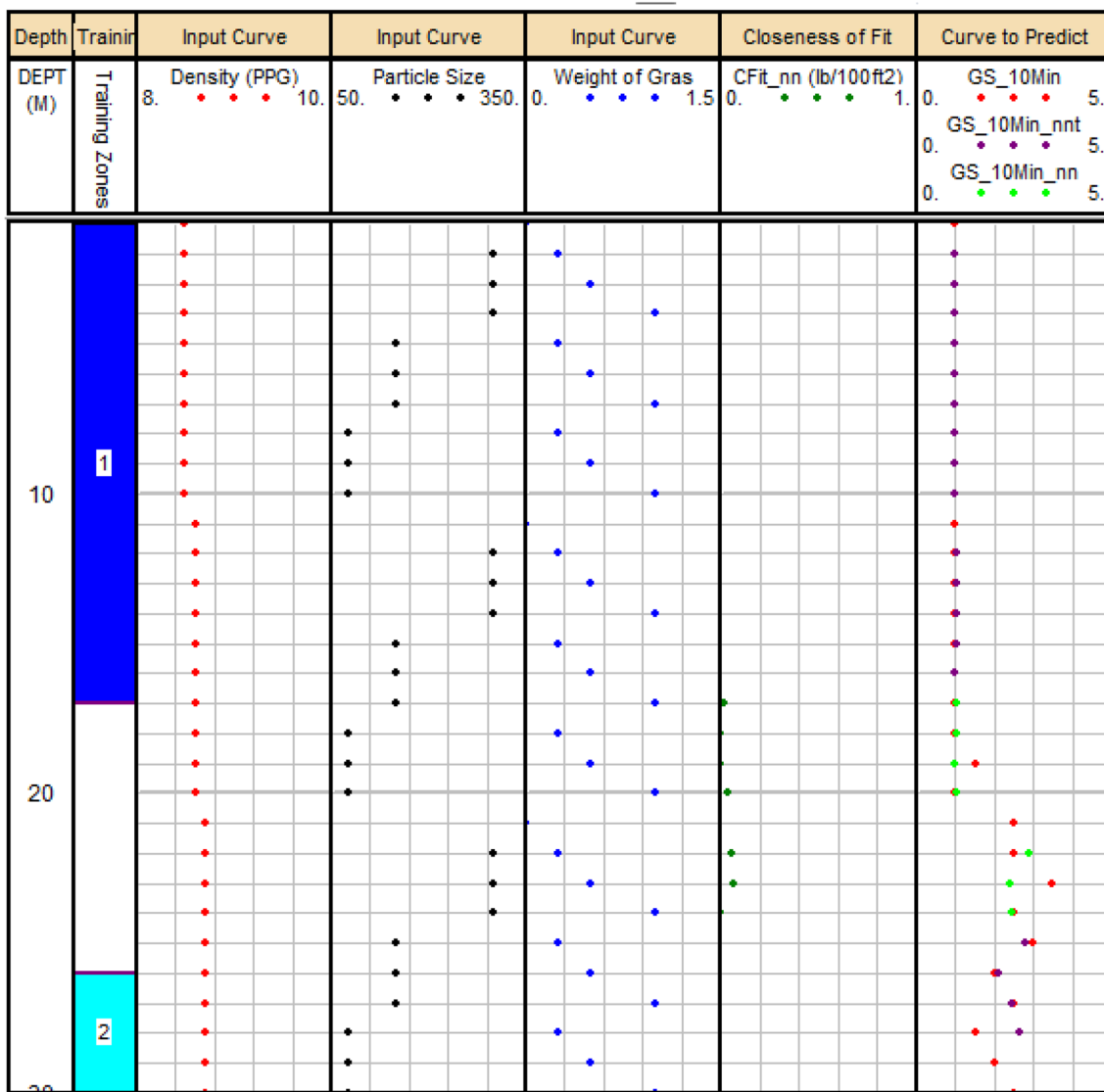


Fig. 12 Application of Artificial Neural Network for Prediction of Gel Strength at 10 min

as Random Forest but it’s a one-level decision tree. The overall accuracy of the Decision Stump was in third place in the estimation of the properties of drilling mud. After the Artificial Neural Network, Decision Stump was the most accurate in the prediction of  $GS_{10_{min}}$ . Most of the models were predicting the values of  $GS_{10_{min}}$ . The models based on Support Vector Machine and Random Tree were not able to predict the  $GS_{10_{sec}}$  (Table 9).

### Conclusions

In this study, machine learning was used to predict the rheological properties of the drilling mud under the influence of grass as an environmental friendly additive. Firstly, the grass

was characterized by the performance of physio-chemical laboratory testing of the grass and rheological properties of the drilling mud. The generated data were used to generate statistical and virtual models and their validation.

1. The improvement of the rheological properties of the drilling mud depends on the type of grass used for this purpose. The highest improvement in the plastic viscosity was seen against the weight of grass greater than 0.25 g of 150  $\mu m$  particle size in drilling mud of 8.7 ppg. There was no effect of the gel strength of the drilling mud for densities 8.6 and 8.7 ppg but 8.7 ppg. At tested conditions of the weight of grass, the particle size of grass, and density of drilling mud, it was seen that the

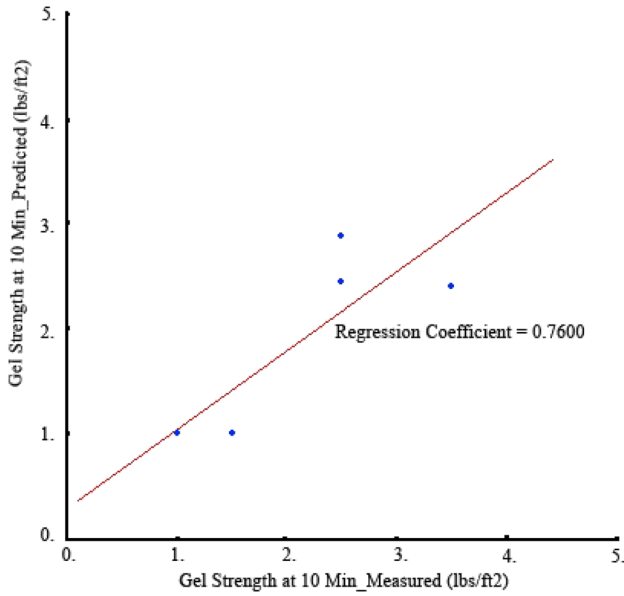


Fig. 13 Accuracy plot of GS\_10Min ANN model

added grass had a significant effect on  $GS_{10min}$  in 8.7 ppg drilling mud.

- Multivariate Linear Regression Analysis was failed to model the relation of plastic viscosity and gel strength with a density of mud, weight, and particle size of grass additive. All the MLRA models of rheological properties had a low regression coefficient because there was a lack of linear relationships between the dependent and independent variables.

Table 7 Response of  $R^2$  against selected k in kNN models

Plastic Viscosity					
K	1	2	3	4	5
$R^2$	0.38	0.58	0.56	0.56	0.58
$GS_{10Sec}$					
K	1	2	3	4	5
$R^2$	0.34	0.58	0.64	0.55	0.56
$GS_{10Min}$					
K	1	2	3	4	5
$R^2$	0.79	0.68	0.76	0.74	0.74

- Artificial Neural Network and Random Forest were predicting the plastic viscosity with the highest accuracy with  $R^2$  of 0.72. The best approach to estimate the  $GS_{10sec}$  is Artificial Neural Network which was predicting the gel strength with  $R^2 = 0.98$  on the application on test data. Random Forest was the approach from machine learning which was predicting the  $GS_{10min}$  of the grass additive with the highest accuracy evident from the regression coefficient of tenfold cross-validation.
- The Artificial Neural Network, Decision Stump, and Random Forest machine learning techniques were predicting the rheological properties of the drilling mud with acceptable accuracy but the most applicable model in the current study was backpropagation Artificial Neural Network.

Table 8 Attribute selection in Decision Stump

Rheological Property	Selected Rules			Coefficient of Regression
PV	Density $\leq$ 8.649: 2.25	Density $>$ 8.649: 3.25	Density is missing: 2.583	$R^2 = 0.61$
$GS_{10Sec}$	Density $\leq$ 8.649: 1.0	Density $>$ 8.649: 1.3	Density is missing: 1.5	$R^2 = 0.62$
$GS_{10Min}$	Density $\leq$ 8.649: 1.025	Density $>$ 8.649: 2.45	Density is missing: 1.5	$R^2 = 0.89$

Table 9 Comparison of the machine learning approaches used in this study

Approach	Regression coefficients of validation testing		
	Plastic viscosity	Gel strength at 10Sec	Gel strength at 10Min
Artificial neural network	0.72	0.98	0.76
K-nearest neighbor regression	0.58	0.64	0.76
Support vector machine regression	0.63	0.29	0.75
Decision stump	0.61	0.62	0.89
Random tree	0.56	0.37	0.61
Random forest	0.72	0.61	0.82

**Acknowledgements** The authors are thankful to the Department of Geological Engineering and the Department of Petroleum & Gas Engineering, University of Engineering and Technology Lahore for permission to use their laboratory resources in this work.

**Author contribution** Regarding the journal policy of “Measurements”, the contribution of the authors in the submitted research article “Application of machine learning for the prediction of the rheological properties of water-based drilling mud: An experimental study on grass as an environmental friendly additive” is as following: Atif Ismail; Conceptualization, Methodology, Investigation, Software, Validation, Formal analysis, Writing—Original Draft. Hafiz Muhammad Awais Rashid; Investigation, Writing—Review & Editing. Raof Gholami; Writing—Review & Editing, Supervision, Arshad Raza; Writing—Conceptualization, Review & Editing, Supervision, Visualization.

**Funding** The author(s) received no financial support for the research, authorship, and/or publication of this article.

#### Declaration

**Conflict of interests** On behalf of all authors, it has been declared that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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

## References

- Abdou M, El-Sayed Ahmed H (2011) Effect of particle size of bentonite on rheological behavior of the drilling mud. *Pet Sci Technol* 29(21):2220–2233
- Ahmadi H, Rodehutsord M (2017) Application of artificial neural network and support vector machines in predicting metabolizable energy in compound feeds for pigs. *Front Nutr* 4:27
- Al-Bulushi N, King PR, Blunt MJ, Kraaijveld M (2009) Development of artificial neural network models for predicting water saturation and fluid distribution. *J Petrol Sci Eng* 68(3–4):197–208
- Al-Hameedi ATT, Alkinani HH, Dunn-Norman S, Al-Alwani MA, Alshammari AF, Albazzaz HW, Alkhamis MM, Alashwak NF, Mutar RA (2019) Insights into the application of new eco-friendly drilling fluid additive to improve the fluid properties in water-based drilling fluid systems. *J Petrol Sci Eng* 183:106424
- Ali J, Khan R, Ahmad N, Maqsood I (2012) Random forests and decision trees. *Int J Comput Sci Issues (IJCSI)* 9(5):272
- Al-Marhoun M, Nizamuddin S, Raheem AA, Ali SS, Muhammadain A (2012) Prediction of crude oil viscosity curve using artificial intelligence techniques. *J Petrol Sci Eng* 86:111–117
- Al-Saba MT, Amadi KW, Al-Hadramy KO, Dushaishi MF, Al-Hameedi A, Alkinani H (2018) Experimental investigation of bio-degradable environmental friendly drilling fluid additives generated from waste. In: *SPE International Conference and Exhibition on Health, Safety, Security, Environment, and Social Responsibility*. OnePetro
- Amanullah M, Ramasamy J, Al-Arfaj MK, Aramco S (2016) Application of an indigenous eco-friendly raw material as fluid loss additive. *J Petrol Sci Eng* 139:191–197
- Ameille J, Wild P, Choudat D, Ohl G, Vaucouleur JF, Chanut JC, Brochard P (1995) Respiratory symptoms, ventilatory impairment, and bronchial reactivity in oil mist-exposed automobile workers. *Am J Ind Med* 27(2):247–256
- Anifowose F, Adeniyi S, Abdulraheem A (2014) Recent advances in the application of computational intelligence techniques in oil and gas reservoir characterisation: a comparative study. *J Exp Theor Artif Intell* 26(4):551–570
- Apaleke AS, Al-Majed AA, Hossain ME (2012) State of the art and future trend of drilling fluid: an experimental study. Paper presented at the SPE Latin America and Caribbean petroleum engineering conference
- Araghinejad S (2013) Data-driven modeling: using MATLAB® in water resources and environmental engineering. Springer
- Asadisaghani J, Tahmasebi P (2011) Comparative evaluation of back-propagation neural network learning algorithms and empirical correlations for prediction of oil PVT properties in Iran oilfields. *J Petrol Sci Eng* 78(2):464–475
- Awad M, Khanna R (2015) Support vector regression. *Efficient Learning Machines* (pp. 67–80): Springer
- Barbosa LFF, Nascimento A, Mathias MH, de Carvalho Jr JA (2019) Machine learning methods applied to drilling rate of penetration prediction and optimization-A review. *J Pet Sci Eng* 183:106332
- Barros C, Andrade A (2013) Determination of water saturation by angular competitive neural network. *J Petrol Sci Eng* 102:47–56
- Basak D (2010) Support vector regression. *Neural Inform Process* 11(10):203–224
- Bengio Y (2009) Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, 2(1), 1–127
- Bhandari J, Abbassi R, Garaniya V, Khan F (2015) Risk analysis of deepwater drilling operations using Bayesian network. *J Loss Prev Process Ind* 38:11–23
- Bishop CM (2006) *Pattern recognition and machine learning*. Springer
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Burts Jr BD (1992) Lost circulation material with rice fraction: Google Patents
- Burts Jr BD (2001) Well fluid additive, well fluid made therefrom, method of treating a well fluid, method of circulating a well fluid: Google Patents
- Caenn R, Darley HC, Gray GR (2011) *Composition and properties of drilling and completion fluids*. Gulf professional publishing
- Candler J, Leuterma A, Wong S, Stephens M (1992) Sources of mercury and cadmium in offshore drilling discharges. *SPE Drill Eng* 7(04):279–283
- Cho HJ (2018) Efficient Shared Execution Processing of k-Nearest Neighbor Joins in Road Networks. *Mobile Inform Syst*
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
- Cremeans KS, Cremeans JG (2003) Lost circulation material and method of use: Google Patents
- Dagde KK, Nmegbu CGJ (2014) Drilling fluid formulation using cellulose generated from groundnut husk. *Int J Adv Res Technol* 3:65
- Davenport T, Kalakota R (2019) The potential for artificial intelligence in healthcare. *Future Healthc J* 6(2):94

- Dimiduk DM, Holm EA, Niezgodna SR (2018) Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. *Integr Mater Manuf Innov* 7(3):157–172
- Dosunmu A (2010) Development of environmentally friendly oil based mud using palm-oil and groundnut-oil. Paper presented at the Nigeria annual international conference and exhibition
- Du Q, Yasin Q, Ismail A, Sohail GM (2019) Combining classification and regression for improving shear wave velocity estimation from well logs data. *J Pet Sci Eng* 182:106260
- Duchemin G, Mojtahid M, Bicchi E, Gaultier M, Jorissen F, Durrieu J. et al (2008) A New Monitoring Tool for Assessing Environmental Impact of Offshore Drilling Activities: Benthic Foraminifera. Paper presented at the 9th International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production
- Dutta S, Gupta J (2010) PVT correlations for Indian crude using artificial neural networks. *J Petrol Sci Eng* 72(1–2):93–109
- Ghassemzadeh J (2011) Lost circulation material for oilfield use: Google Patents
- Greaves IA, Eisen EA, Smith TJ, Pothier LJ, Kriebel D, Woskie SR et al (1997) Respiratory health of automobile workers exposed to metal-working fluid aerosols: respiratory symptoms. *Am J Ind Med* 32(5):450–459
- Green PC (1984) Use of ground, sized cocoa bean shells as a lost circulation material in drilling mud: Google Patents
- Gul S, van Oort E (2020) A machine learning approach to filtrate loss determination and test automation for drilling and completion fluids. *J Pet Sci Eng* 186:106727
- Gurina E, Klyuchnikov N, Zaytsev A, Romanenkova E, Antipova K, Simon I, Makarov V, Koroteev D (2020) Application of machine learning to accidents detection at directional drilling. *J Pet Sci Eng* 184:106519
- Haykin S (1994) *Neural networks: a comprehensive foundation*. Prentice Hall PTR
- Hegde C, Gray K (2018) Evaluation of coupled machine learning models for drilling optimization. *J Nat Gas Sci Eng* 56:397–407
- Hegde C, Millwater H, Gray K (2019) Classification of drilling stick slip severity using machine learning. *J Petrol Sci Eng* 179:1023–1036
- Helle HB, Bhatt A, Ursin B (2001) Porosity and permeability prediction from wireline logs using artificial neural networks: a North Sea case study. *Geophys Prospect* 49(4):431–444
- Hossain ME, Wajheuddin M (2016) The use of grass as an environmentally friendly additive in water-based drilling fluids. *Pet Sci* 13(2):292–303
- Imandoust SB, Bolandraft M (2013) Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *Int J Eng Res Appl* 3(5):605–610
- Ismail A, Yasin Q, Du Q, Bhatti AA (2017b) A comparative study of empirical, statistical and virtual analysis for the estimation of pore network permeability. *J Nat Gas Sci Eng* 45:825–839
- Ismail A, Yasin Q, Du Q (2017) *3D Reservoir Characterization in Heterogeneous Sand*. Paper presented at the 公益社団法人物探学会学術講演会講演論文集 = Proceedings of the SEGJ Conference
- Jordan MI, Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. *Science* 349(6245):255–260
- Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: a survey. *J Artif Intell Res* 4:237–285
- Kalmegh S (2015) Analysis of weka data mining algorithm reptime, simple cart and randomtree for classification of indian news. *Int J Innov Sci Eng Technol* 2(2):438–446
- Kersting K (2018) Machine learning and artificial intelligence: two fellow travelers on the quest for intelligent behavior in machines. *Frontiers in Big Data* 1:6
- Kok VM, Alikaya T (2003) Rheological evaluation of polymers as drilling fluids. *Pet Sci Technol* 21(1–2):113–123
- Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: A review of classification techniques. *Emerg Artif Intell Appl Comput Eng* 160:3–24
- Kumar CV, Vardhan H, Murthy CS (2019) Multiple regression model for prediction of rock properties using acoustic frequency during core drilling operations. *Geomech Geoeng* 15(4):1–16
- Lan Y, Soh YC, Huang GB (2009) Ensemble of online sequential extreme learning machine. *Neurocomputing* 72(13–15):3391–3395
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
- Li CX, Huang JJ, Xu Y (2002) A new high temperature stable wettability reversal agent in oil base mud. *J-Southwest Pet Inst* 24(5):22–24
- Lim JS (2003) Reservoir permeability determination using artificial neural network. *J Korean Soc Geosyst Eng* 40:232–238
- Lumms JL, Ryals JN (1971) Preventing loss of drilling fluid to drilled formations: Google Patents
- Macquoid M, Skodack D (2004) Method for using coconut coir as a lost circulation material for well drilling: Google Patents
- Mahmoud AA, Elkhatatny S, Al-AbdulJabbar A (2021) Application of machine learning models for real-time prediction of the formation lithology and tops from the drilling parameters. *J Pet Sci Eng* 203:108574
- Maniar H, Ryali S, Kulkarni MS, Abubakar A (2018) Machine-learning methods in geoscience. Paper presented at the 2018 SEG International Exposition and Annual Meeting
- Mishra AK, Ratha BK (2016) Study of random tree and random forest data mining algorithms for microarray data analysis. *Int J Adv Electr Comput Eng* 3(4):5–7
- Mohaghegh SD (2015) System and method providing real-time assistance to drilling operation: Google Patents
- Morris GR (1962) Drilling fluid: Google Patents
- Nestle AC (1952) Drilling fluid: Google Patents
- Nmegbu CGJ, Bekee B (2014) Evaluation of corn cob cellulose and its suitability for drilling mud formulation. *Int J Eng Res Appl* 4(5):112–117
- Noshi CI, Schubert JJ (2018) The role of machine learning in drilling operations; a review. Paper presented at the SPE/AAPG Eastern Regional Meeting
- Okon AN, Udoh FD, Bassey PG (2014) Evaluation of rice husk as fluid loss control additive in water-based drilling mud. Paper presented at the SPE Nigeria Annual International Conference and Exhibition
- Peng W, Chen J, Zhou H (2009) An implementation of ID3-decision tree learning algorithm. From web. arch. usyd. edu. au/wpeng/DecisionTree2. pdf Retrieved date: May, 13
- Provost F, Fawcett T (2013) Data science and its relationship to big data and data-driven decision making. *Big Data* 1(1):51–59
- Rana S (2008) Facts and data on environmental risks-oil and gas drilling operations. Paper presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition
- Sampey JA (2006) Sugar cane additive for filtration control in well working compositions: Google Patents
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
- Schmidt M, Lipson H (2009) Distilling free-form natural laws from experimental data. *Science* 324(5923):81–85
- Sewaiwar P, Verma KK (2015) Comparative study of various decision tree classification algorithm using WEKA. *Int J Emerg Res Manag Technol* 4(2015):2278–9359
- Shajahaan SS, Shanthi S, ManoChitra V (2013) Application of data mining techniques to model breast cancer data. *Int J Emerg Technol Adv Eng* 3(11):362–369



- Sharma VP, Mahto V (2006) Studies on less expansive environmentally safe polymers for development of water based drilling fluids. Paper presented at the SPE Asia Pacific Oil & Gas Conference and Exhibition
- Shi G, Zhou X, Zhang G, Shi X, Li H (2004) The use of artificial neural network analysis and multiple regression for trap quality evaluation: a case study of the Northern Kuqa Depression of Tarim Basin in western China. *Mar Pet Geol* 21(3):411–420
- Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP (2003) Random forest: a classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci* 43(6):1947–1958
- Wajheuddin M, Hossain ME (2018) Development of an environmentally-friendly water-based mud system using natural materials. *Arab J Sci Eng* 43(5):2501–2513
- Wajheuddin M (2014) Development of an environmentally-friendly drilling fluid using date seeds and grass
- Waqas U, Ahmed MF, Arshad M (2020) Classification of the intact carbonate and silicate rocks based on their degree of thermal cracking using discriminant analysis. *Bull Eng Geol Environ* 79(5):1–13
- Zhang L, Wu S, Zheng W, Fan J (2018) A dynamic and quantitative risk assessment method with uncertainties for offshore managed pressure drilling phases. *Saf Sci* 104:39–54
- Zhao S, Yan J, Wang J, Ding T, Yang H, Gao D (2009) Water-based drilling fluid technology for extended reach wells in Liuhua Oilfield, South China Sea. *Pet Sci Technol* 27(16):1854–1865

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