Petroleum and Coal

Article **Open Access**

Optimizing Dogleg Severity for Improved Directional Drilling using Machine Learning

Anthony Kerunwa[1], David Chimezirim Duru[1], Stephen G. Udeagbara[2], Nnaemeka Uwaezuoke[1] and Angela Nkechinyere Nwachukwu[1]*

[1] Department of Petroleum Engineering, Federal University of Technology, Owerri, Imo State, Nigeria
[2] Department of Chemical and Petroleum Engineering, Afe Babalola University, Ado-Ekiti, Ekiti State, Nigeria

Received July 23, 2023; Accepted October 17, 2023

Abstract

One of the greatest challenges of directional drilling (DD) is the accurate dogleg severity (DLS) estimation for selection of drive systems for DD. Several conventional techniques has been employed which either over estimates DLS leading to hole problems and inability to follow trajectory which greatly contribute to non-productive time thereby resulting to financial losses or are expensive to maintain and time consuming. In this work, three different machine learning models (MLMs) were developed to optimize DLS, to wit: Decision tree (DT), Random Forest (RF) and Support Vector Machine (SVM). Two scenarios were evaluated, viz: a case without optimization and the optimized case. In the unoptimized case, R^2 was used to rank the accuracy of the MLMs with RF having R^2 of 91.4% with error margin of 14%-22.6%, DT had R^2 of 85.7% with error margin of 18%-29.2% while SVM had R^2 of 83.3% with error of 16%-31.6% for all the tested error metrics. For the optimized case, grid Search optimization algorithm was implemented on DT and RF while genetic optimization algorithm was implemented on SVM. From the result, the optimized SVM-MLM yields the same DLS as the actual DLS from drilling survey data except for a single value that was less than the actual DLS with a difference of 0.001deg/100ft in comparison to those obtained using DLS conventional estimation techniques. The optimization improved DT's R^2 from 85% to 90% with error margin of 16%-24%, RF from 91.4% to 98.1% with error margin of 5.8% -10.7% and SVM from 83.3% to 98.4% with error margin of 0.94%-2.2%. Thus, depending on the best suited priorities the SVM-MLM well path should be selected. The result indicates that DD operation should be more economical, easier and safer with the well trajectory planned by utilizing the approach proposed in this study.

Keywords: *Dogleg severity; Machine learning; Decision tree; Random forest; Genetic algorithm; Support vector machine.*

1. Introduction

The demand for cost-effective drilling operations in directional and horizontal wells in the oil and gas sector is ever growing [1-2]. One of the important aspects to achieving the aforementioned challenge is determining the optimal well trajectory or directional path to reach the desired/target point [2]. The most important optimization objective is to achieve an optimal well trajectory in safe and stable drilling conditions [3-4]. However, Determination of optimal well trajectory is quite a difficult task because there are many parameters to consider such as dogleg severity (DLS), azimuth and inclination angle [5]. In the early stages of directional drilling, trajectories were designed manually, which makes it impossible to find an optimal solution quickly and efficiently. Mistakes in design may lead to expensive procedures that would be required to be carried out to fix emergencies at the drilling stage [6]. Determination of optimal well trajectories constitutes one of the problems of field development planning. This problem must be solved when a well pad pattern is being selected; all existing engineering constraints must be taken into consideration; at the same time, the cumulative length of all trajectories must be minimal to reduce capital costs for drilling [7]. Well trajectories are classified according to the number of wellbore intervals. A number of parameters define every well section: length, DLS, inclination angle change. The values of parameters can be changed within the limits of the engineering constraints [6-7]. Hydrocarbon field development planning is a complex process with many decision variables being considered [7-8]. This process requires long-term predictions and capital investments strongly linked with decisions about different components of an integrated process: facilities, production strategies, number of rigs, health, safety and environmental constraints, improved and enhanced oil recovery (EOR) strategies [8].

In the context of a hard economical constrained situation and cost efficiency, one of the most important parts in field development is planning and drilling the wells. More specifically, well trajectories, number of wells to be drilled along with choices of locations plays a central role in the decision making process of optimizing a field development plan [8-9]. Planning a new well involves creating a trajectory that passes through several target points in the reservoir and yet remains deliverable. The deliverability is assessed based on a large number of criteria including well length and DLS [10-11]. In the world of model based reservoir optimization, most attention has been given to life cycle production optimization, aiming to find optimal operating strategies for fixed well configurations. Relatively few studies focus on the optimization of directional well path or trajectory and much fewer studies focuses on implementing machine learning (ML) techniques for the optimization of the important parameters that determine the trajectory of the well which is mostly DLS [8].

DLS is one of the most important parameters which affect the determination of optimal well trajectories in directional drilling [12-13]. DLS is determined at the dogleg. It is the measurement of azimuth or inclination change usually expressed in 30 m of CL (course length) or degrees/100 ft. DLS describes the wellbore's curvature and smoothness and as such responsible for the side forces like bending forces that act on DS. Majority of directional wells do not follow smooth paths as planned, consequently containing crookedness that cause drilling problems like stuck pipe [14]. Improvement of this indicator actually means choosing the best conditions for the DD in order to reach the target point [12]. Selection of high levels of the DLS actually means minimizing well trajectory, but on the other hand, increases fatigue in drill string (DS), increases torque and drag, particularly in the rotation mode [12,15]. Therefore the aim is to define the index in an optimal range which meets both requirements [15]. As always, drilling can be optimized by using existing well data, but because the DLS optimization has so many governing factors, it is difficult to create a model that can simulate DLS utilizing previous experiences that can be applied to future wells [16]. Any of the optimization algorithms can be employed for the optimization of the DLS. The purpose of this algorithm is to minimize the trajectory length by optimizing the DLS [1,17]. The minimum parameters will reduce the DLS, which in turn reduce the chances for operational problems like high torque and drag [1].

2. Literature review

Wilson, [18] conducted a study on improved method for computing directional surveys. He presented Radius of Curvature method (RCM)) (equation 1) as an improved method used for computing directional surveys which he showed, performed better than the tangential method which was the previous method used for computing directional surveys. He discovered that the interpretation of calculations made with the tangential method created problems in accurately depicting Louisiana Gulf Coast reservoirs. He pointed out computer calculations indicated the tangential method to be in error by about 12ft in vertical depth and about 40ft in departure in horizons encountered above the point at which return to vertical was started. He cited example where, computations with single-shot data indicated a spread of 40ft in the vertical depth of an oil-water contact encountered by several S-type wells now known to have penetrated the same reservoir. He applied the RCM in the computation of directional surveys and discovered that the RCM eliminated inherent errors in vertical depth, horizontal departure, direction coordinates and DLS that occurred when tangential method was applied. He concluded that computations with the RCM reduced the spread of contact depths to only 5ft. The utilized radius of curvature equation is stated below:

$$D = 100\sqrt{C^2 \sin^4 \varphi + b^2} \quad (1)$$

where: D - DLS at the point at which ϕ is determined.

The limitation to radius of curvature is that the result yields a trajectory with high DLS, and, consequently, high drags and torque on the DS compared to constant curvature method (CCM).Guo *et al.*, [6] conducted a study on CCM for planning 3-D Directional Well. He proposed a new method for planning 3-D directional well path, which had several advantages over the conventional Radius-of-Curvature Method and Constant-Turn-Rate Method (CTRM). The new method yields constant curvature well path sections which are compatible with the directional performance of deflection tools, it also yields less DLS of well trajectories, and, consequently, less drags and torque on the DS. Mathematically, the formulation of the method involves integrals that do not have closed form solutions and need to be estimated numerically. To avoid the numerical integrations, they presented two alternative approximations to the exact solution, namely Piecewise- RCM and Piecewise-CTRM. The proposed CCM assumes constant well path curvature is compatible with the directional performance of deflection tools and is also consistent with the geometrical relationship between the tool face orientation angle, rate of build and DLS. He utilized this formula to calculate DLS

$$DLS = \frac{I_2 - I_1}{L_2 - L_1}\sqrt{1 + \left[\frac{A_2 - A}{ln\left[\frac{tan\left(^{I_2}/_2\right)}{tan\left(^{I_1}/_2\right)}\right]}\right]^2} \quad (2)$$

From their result, the CCM yielded a better result in comparison to other methods utilized. The limitation to CCM is that, it is time consuming because of the numerous computations for each section of the well.

Hosseini *et al.*, [2] utilized Particle Swarm Algorithm to study DLS in directional oil wells. They performed optimization of well path and directional drilling parameter with measured depth as the objective function, horizontal depth and target TVD as constraints related to well path, maximum tensile strength as constraint related to DS mechanical properties and maximum well angle as constraint related to operational condition. Particle Swarm Optimization (PSO) algorithm was implemented, with measured depth as the objective function.

$$f(x) = MD = D_1 + \frac{100\alpha}{DLS} + \frac{D_3 - D_1 - R\sin\alpha}{\cos\alpha} \quad (3)$$

The result showed that the final measured depth was less than the proposed program and the fatigue criteria were also satisfied. The limitations or disadvantages of PSO algorithm are that it is easy to fall into local optimum in high dimensional space and has a low convergence rate in the iterative process.

Shokir *et al.*, [1] investigated the use of GA for optimal 3-D horizontal and directional wells planning. Two constraints were considered in the work: operational constraints and non-negative constraints. The operational constraints were imposed on the well design due to the different types of formation to be drilled, casing setting depths, limitations of variable equipment or technology, and other operation related limitations. The non-negativity constraints were imposed on the well design model to ensure that certain components of the model are always positive. They discovered from their result that GA reduced the measured depth of the well by about 70ft less than the conventional method in the first application. For the second application, the GA reduced the measured depth by about 110ft less than the conventional technique. The limitation to this method is that it is time consuming and also expensive.

Brechan *et al.*, [19] carried out an in-depth study on planning horizontal and extended reach wells. They developed a model to determine DLS from course length and dogleg angle as shown in equation 4. Nkengele [14] posited that the technique for DLS determination developed by Brechan, *et al.*, [19] known as the Brechan *et al*. method is the most reliable, functional, easy to implement and highly accurate in comparison with other models for DLS calculation.

$$DLS = \frac{\varphi\ x30}{CL} \quad (4)$$

where: φ - dogleg angle; CL - course length or dogleg's length (m).

$\varphi = cos^{-1}[cos\,\alpha_n\,cos\,\alpha_{n+1} + sin\,\alpha_n\,sin\,\alpha_{n+1}\,cos(\beta_{n+1} - \beta_n)]$ and $CL = DMEA_{n+1} - DMEA_n$

where: DMEA –Hole depth (MD), n = 1, 2, 3, …. (survey measurement points).

3. Materials and methods

3.1. Materials

The materials used for this work are python and dataset: (i) Python is an object-oriented and high-level programming language in scientific computing, because of its data-oriented feature packages that can speed up and simplify data processing, thus saving time. Python supports modules and packages, which encourages program modularity and code reuse. (ii) Dataset used in this work shown in Table 1, was obtained from drilling survey report available at the volve data repository. The drilling survey report contains the following features: Measured depth (MD), Inclination, Azimuth, True Vertical depth (TVD), N/S and E/W coordinates, Vertical Section and DLS.

3.2. Method

The method used in this study include: data preprocessing, outliers, exploratory data analysis and thereafter model development or simply modeling

3.2.1. Data preprocessing

Missing values and outliers impact the accuracy of the MLM. Data preprocessing was performed to handle the missing values and outliers in the dataset. This is because; the outliers could be taking place as a result of measurement errors.

3.2.2. Outliers

The outliers were identified by making boxplots of all the parameters in the dataset as depicted in Figures 1 and 2, and thereafter, they were removed them from the data so that they will not affect the model's decision. Figures 1 and 2, reveals the outliers that were present in DLS and inclination respectively.

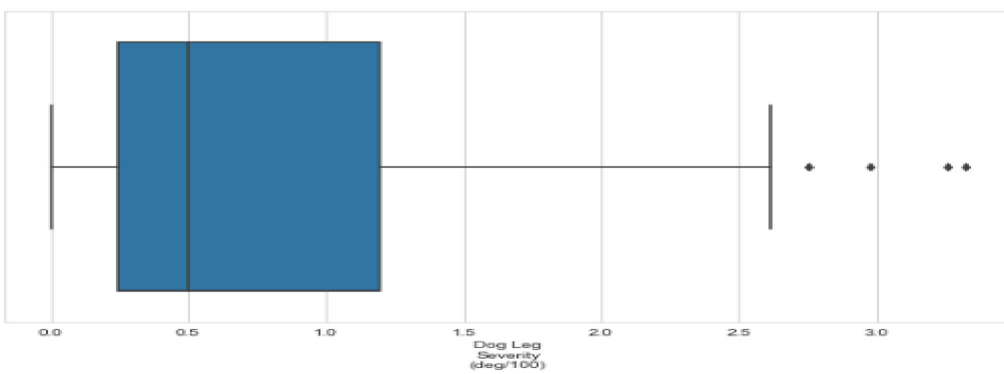

Fig. 1. Boxplot showing the outliers of DLS present in the data.

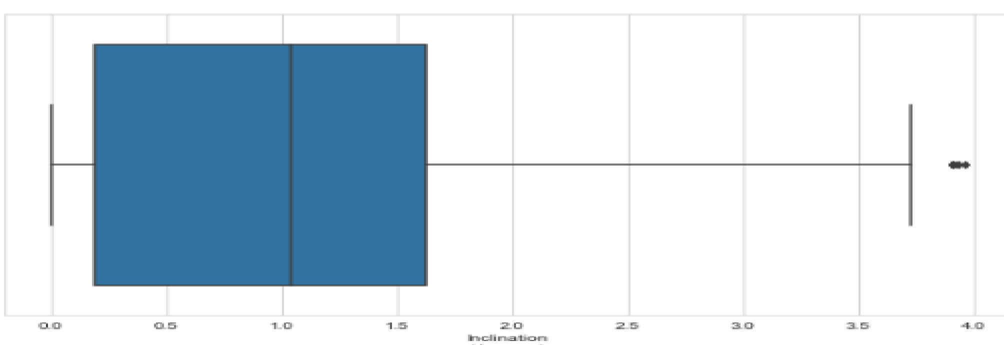

Fig. 2. Boxplot showing outliers of inclination present in the data.

3.2.3. Exploratory data analysis (EDA)

The accuracy of the MLM depends on the correlations between the various features used in training the model with the target feature as depicted in Figure 3. Determination of the relative important features for DLS optimization from the various input features in the data is therefore crucial. EDA was then implemented on the data to discover relationships/correlations that exist between the various input features and the target feature. Depicted in Figure 4 is the Scatter plot of DLS against true TVD. The pairplot in Figure 5 shows the relationship that exists between all the features in the dataset.

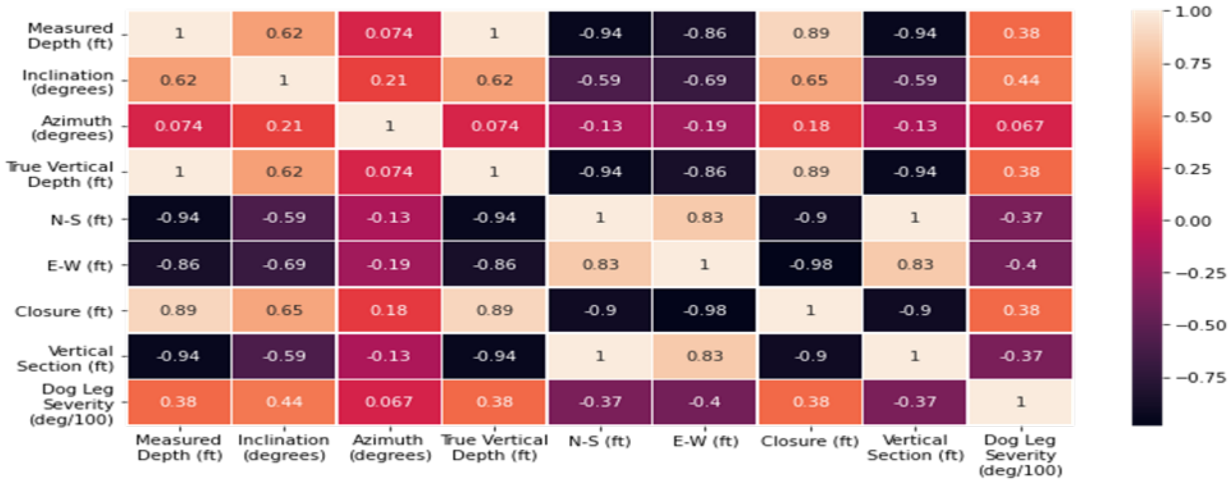

Fig. 3. Heatmap showing correlations between the dataset features.

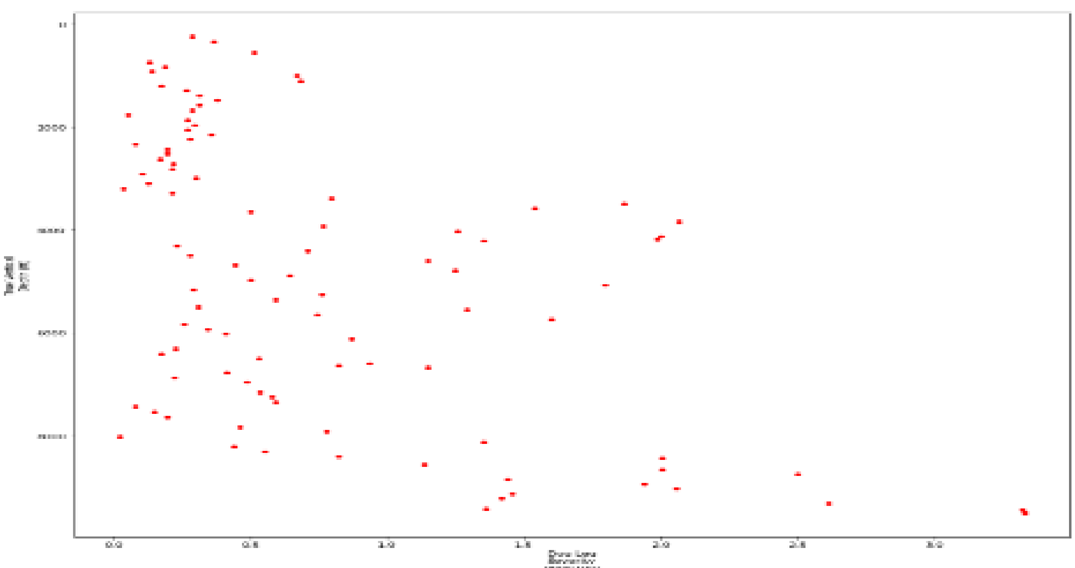

Fig. 4. Relationship between TVD and DLS.

3.2.4. Modeling

Three different machine learning models (MLMs) were developed for this study. The models are: Support Vector Machine (SVM), Decision tree (DT) and Random Forest algorithm (RFA). Grid search optimization method was applied to DT and Random Forest (RF) for tuning the hyper parameters of these models in order to obtain the optimal input parameters of these models. Genetic Algorithm (GA) was applied on SVM for tuning the SVM's hyper parameters. Theses optimization techniques were used to update the input features of the model.

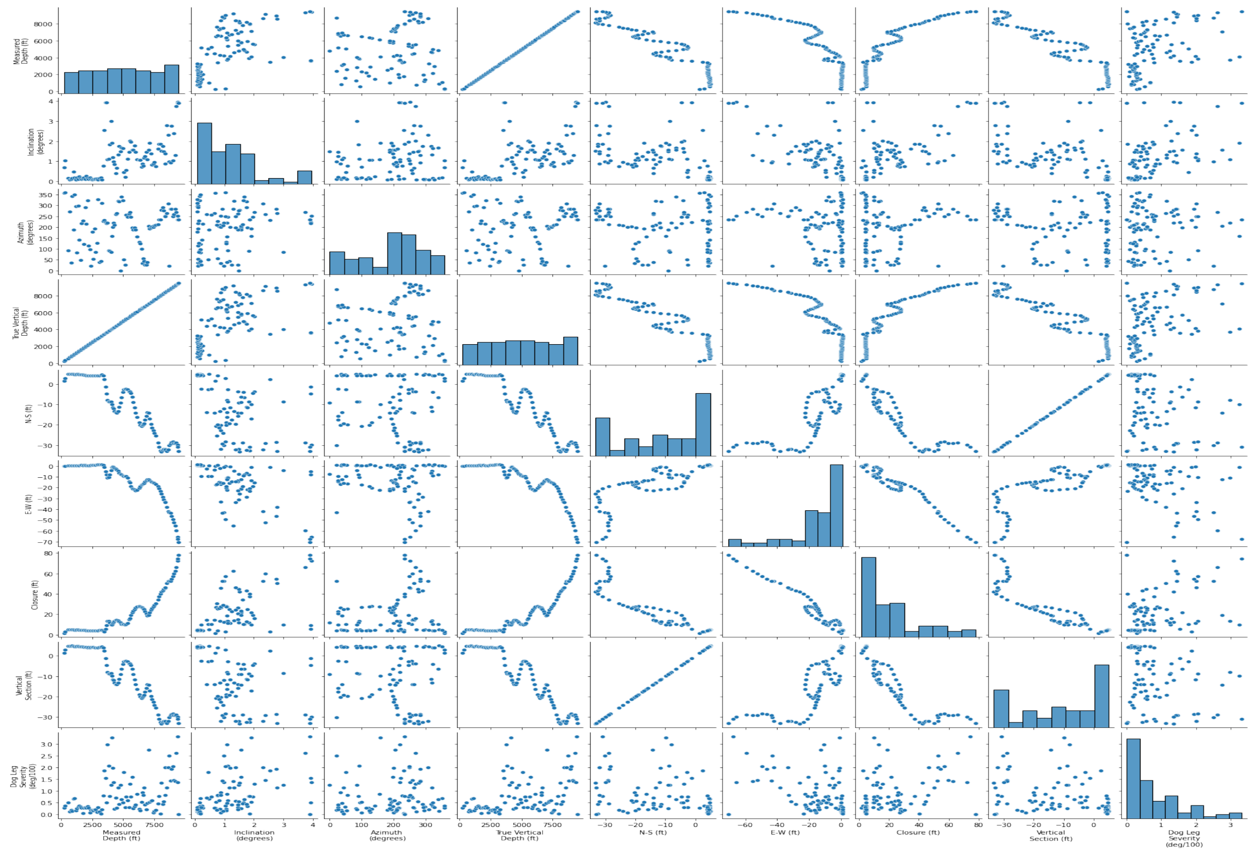

Fig. 5. Pairplot showing relationship between features of the dataset.

3.2.5. SVM and genetic algorithm

Different predictions for DLS were first obtained using different input features of the SVM. These input features were encoded in a binary form called chromosomes. After obtaining the predictions, the GA was implemented for the optimization of DLS using the binary encoded chromosomes. The GA performed a heuristic search on the input parameters of the models to obtain the optimal input parameters for all the models. The optimized input parameters of the different models were then applied to train the different model using the same dataset.

GA optimized the SVM using these steps: (i) Initial Population**:** The GA first generated an initial population of possible solutions called chromosomes. These solutions were different predictions of DLS at any depth using the SVM. Value chromosome encoding was selected using real numbers to encode the problem variables on GA chromosomes. Then, each variable (gene) had a value to build an initial population. Each chromosome represented a solution to the problem. The population size was 20 chromosomes. (ii) Evaluation Function: After building the initial population, each chromosome (solution) in the population was evaluated according to fitness values of the objective function. Assessment of the accuracy of the possible solutions was carried out using an objective function which depicted the accuracy of the MLM. (iii) Selection: The best performing chromosomes were selected by the GA which formed the parents The selection method utilized in the GA was Tournament selection, to select two parents (chromosomes) from the population to produce two children from them by reproduction operators. Tournament selection ensured that only the best performing chromosomes were selected as parents. (iv) Crossover: The crossover type used in the model was one-point crossover. In this type, one point was selected randomly, and cutting the two parent chromosomes at this point then exchange alternate pairs of sections between the first part of one parent and the last part of the other parent. After this crossover, two children chromosomes (new solutions) were produced. The selected probability of crossover was 100%. The population of offsprings generated had better solutions than the parents. (v) Mutation: After crossover and producing two children chromosomes, one gene was selected from each child chromosome to mutate its value. This mutation technique occurred by changing the value of the variable (gene) by adding a random value to its old value. The Mutation induced diversity within the offspring to further improve its solution. (vi) New Population: After producing the offspring, two chromosomes of the two parents and the two children chromosomes were inserted into the population. This was done by determining the best two chromosomes (the two chromosomes that had the best fitness value) from the two parents and the two children chromosomes, and were inserted into the population to improve the population (group of solutions).The process was repeated for several generations until an optimal solution was achieved. Depicted in Figure 6 is the flowchart of the GA.

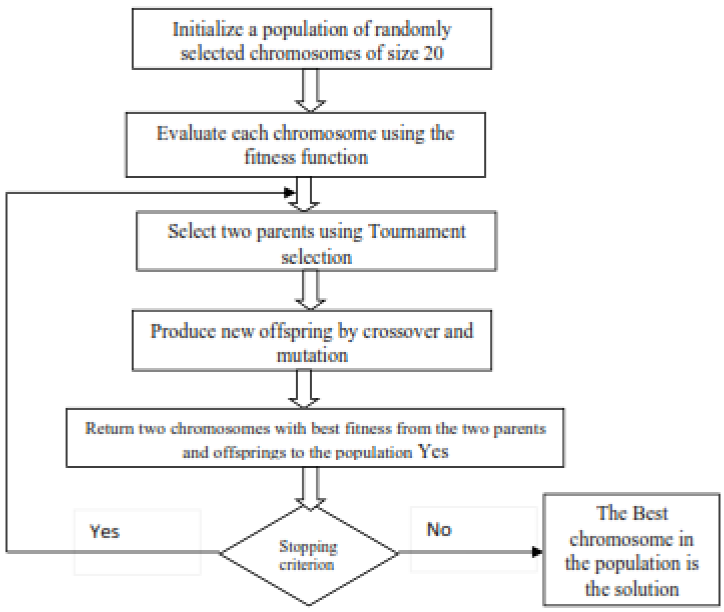

Fig. 6. GA flowchart

3.2.6. Decision tree (DT) and grid search

DTs are general class of MLMs that are used for both classification and regression. The trained models resemble a tree, complete with branches and nodes. At the top of the tree is the root node. This node is split to form two branches. Observations that satisfy the criterion printed at the top of the box is moved to one branch while the rest to the other. For regression, the partitions are picked to reduce the variance of sample labels. For the tree displayed below, node splits were chosen to lead to an overall reduction of the Gini metric. The nodes that do not branch off are called terminal nodes or leaves. The process of constructing a decision tree for regression is that splits are chosen to produce nodes with an overall reduction in variance of the training labels. Depicted in Figure 7 is the structure of the DT model. The max depth in a decision tree is one of the most important hyper parameter of DT algorithm which determines the performance of the model. Grid search was implemented to search through maximum depth range of 1 – 20, in order to determine the best maximum depth parameter of the DT that gives the best result for the model as shown in Figure 7. There was a sharp drop in error when initially increasing the maximum depth feature of the DT. In general, as the maximum tree depth increases, performance increases. After maximum tree depth of 13, there was no significant reduction in the error metric. Increasing the max depth further, could lead to overfitting of the model. So a maximum depth of 13 was chosen for the DT.

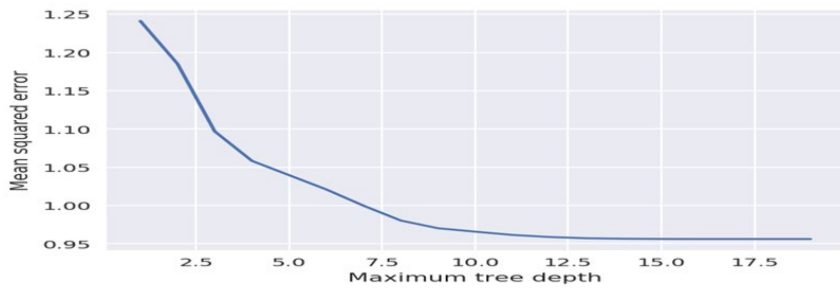

Fig. 7. Grid search for best maximum tree depth parameter for DT.

3.2.7. Random forest (RF) and grid search

The performance of a single DT is limited. Thus, instead of relying on one tree, a better approach used was to aggregate the predictions of multiple trees. On average, aggregation performs better than a single predictor. RF was implemented for this purpose. In order for RF to be effective, the model utilized a diverse collection of trees. There were variations in the chosen thresholds for splitting the number of nodes and branches. The number of trees in RF is one of the most important hyper parameter of RF which determines the performance of the model. To optimize the performance of the RF model, grid search was implemented to search over 50 trees and to determine the number of trees that gives the best result for the RF as shown in Figure 8. The metric employed to keep record of the RF performance when a specific number of trees were utilized was MSE. There was a sharp drop in error when initially growing the forest. In general, as the number of trees increases, performance also increases. The initial drop in error was due to large increase of diverse trees when the forest is small. In other words, the additional trees were very different from the previous trees simply because the forest was small. The increase in tree diversity drives predictive power. As the forest grew beyond 10 trees, newer trees were not significantly different from the previous pool of trees. As a result of this, there was no much reduction in the error metric as the forest grew beyond 10 trees. Depicted in Figure 9 is the snapshot of Python code used to develop the RF model. Python codes were also utilized for DT and SVM models development

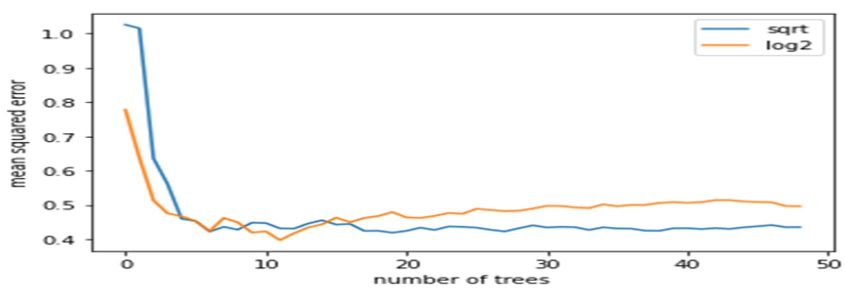

Fig. 8. Grid search for best number of trees parameter for RF.

RANDOM FOREST Optimized with Grid Search

```
In [72]:
from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor(n_estimators=10)

model.fit(x, y)
pred = model.predict(x)
# pred

In [73]:
from sklearn import metrics
import numpy as np
print('mean absolute error:', metrics.mean_absolute_error(y, pred))
print('mean squared error:', metrics.mean_squared_error(y, pred))
print('root mean squared error:', np.sqrt(metrics.mean_squared_error(y, pred)))
print('R^2: %f' % metrics.r2_score(y, pred))

mean absolute error: 0.05798078756260574б
mean squared error: 0.011354163301008757
root mean squared error: 0.10655591631161902
R^2: 0.981057

In [74]:
accuracy = model.score(x,y)
error = 1-(accuracy)

print('accuracy:', accuracy)
print('error:', error)

accuracy: 0.9810565663999914
error: 0.018943433600008608
```

Fig. 9. Snapshot of Python code utilized for RF model development.

Table 1. Dataset used for this study.

Measured depth (MD) (ft)	Inclination (degrees)	Azimuth (degrees)	TVD (ft)	N-S (ft)	E-W (ft)	Closure (ft)	Vertical section (ft)	DLS (deg/100)
0	0	359.98	0	0	0	0	0	-
241	0.69	358.29	241	1.5	0	1.5	1.5	0.286
337	1.04	359.49	337	2.9	0.1	2.9	2.9	0.365
545	0.17	93.92	545	4.8	0.2	4.8	4.8	0.513
731	0.07	271.76	731	4.8	0.4	4.8	4.8	0.129
822	0.12	37.38	822	4.8	0.4	4.8	4.8	0.187
914	0.15	342.6	914	5	0.4	5	5	0.138
1005	0.47	187.27	1005	4.8	0.3	4.8	4.8	0.67
1101	0.19	350.91	1101	4.5	0.3	4.5	4.5	0.682
1195	0.07	294.81	1195	4.7	0.2	4.7	4.7	0.172
1291	0.19	82.64	1291	4.8	0.3	4.8	4.8	0.263
1386	0.12	297.71	1386	4.8	0.3	4.8	4.8	0.312
1480	0.24	141.52	1480	4.7	0.4	4.7	4.7	0.376
1576	0.17	48.65	1576	4.6	0.6	4.7	4.6	0.314
1670	0.1	216.87	1670	4.7	0.7	4.7	4.7	0.286
1765	0.14	231.05	1765	4.5	0.5	4.6	4.5	0.052
1861	0.18	123.07	1861	4.4	0.6	4.4	4.4	0.271
1955	0.1	292.13	1955	4.3	0.6	4.4	4.3	0.297
2050	0.26	216.12	2050	4.2	0.4	4.2	4.2	0.268
2145	0.15	323.76	2145	4.1	0.2	4.1	4.1	0.355
2240	0.12	115.81	2240	4.2	0.2	4.2	4.2	0.276
2335	0.18	98.85	2335	4.1	0.5	4.1	4.1	0.078
2431	0.11	21.4	2431	4.2	0.6	4.2	4.2	0.197
2526	0.08	179.69	2526	4.2	0.7	4.2	4.2	0.197
2620	0.1	52.94	2620	4.2	0.7	4.2	4.2	0.171
2715	0.12	192.1	2715	4.1	0.8	4.2	4.1	0.217
2810	0.13	85.79	2810	4	0.9	4.1	4	0.211
2905	0.14	127.62	2905	4	1.1	4.1	4	0.102
2981	0.09	327.33	2981	4	1.1	4.1	4	0.298
3101	0.15	39.36	3101	4.2	1.2	4.3	4.2	0.124
3196	0.18	42.83	3196	4.4	1.3	4.6	4.4	0.033
3289	0.08	311.16	3289	4.5	1.4	4.7	4.5	0.214
3384	0.76	228.36	3384	4.1	0.9	4.2	4.1	0.794
3480	2.55	231.96	3479.9	2.4	-1.3	2.7	2.4	1.867
3575	3.93	223.19	3574.8	-1.3	-5.2	5.3	-1.3	1.539
3642	3.93	218.32	3641.6	-4.7	-8.2	9.4	-4.7	0.498
3737	1.1	218.9	3736.5	-8	-10.8	13.4	-8	2.979
3832	1.11	93.69	3831.5	-8.8	-10.4	13.6	-8.8	2.065
3927	1.83	89.39	3926.5	-8.8	-8	11.9	-8.8	0.766
4022	3.01	85.27	4021.4	-8.6	-4	9.5	-8.6	1.255
4117	1.92	159.41	4116.3	-9.9	-0.9	9.9	-9.9	3.259
4181	1.19	199.7	4180.3	-11.5	-0.8	11.5	-11.5	1.987
4213	0.99	220.09	4212.3	-12	-1.1	12.1	-12	1.354
4308	0.86	209.26	4307.3	-13.3	-1.9	13.4	-13.3	0.229
4403	0.4	258.72	4402.3	-14	-2.6	14.2	-14	0.708
4498	0.66	264.54	4497.3	-14.1	-3.5	14.5	-14.1	0.279
4593	1.33	319.25	4592.2	-13.3	-4.8	14.1	-13.3	1.148
4688	1.75	318.56	4687.2	-11.4	-6.4	13.1	-11.4	0.443
4783	1.48	0.52	4782.2	-9.1	-7.4	11.7	-9.1	1.246
4879	1.68	339.26	4878.1	-6.5	-7.9	10.2	-6.5	0.641
4974	1.65	322.94	4973.1	-4.1	-9.2	10.1	-4.1	0.498

Measured depth (MD) (ft)	Inclination (degrees)	Azimuth (de-grees)	TVD (ft)	N-S (ft)	E-W (ft)	Closure (ft)	Vertical section (ft)	DLS (deg/100)
5069	0.45	52.1	5068.1	-2.8	-9.7	10.1	-2.8	1.794
5164	0.21	25.93	5163.1	-2.4	-9.3	9.7	-2.4	0.292
5259	0.57	255.65	5258.1	-2.4	-9.7	10	-2.4	0.762
5354	1.12	247.76	5353.1	-2.8	-11	11.4	-2.8	0.59
5499	1.46	260.8	5498	-3.7	-14.2	14.7	-3.7	0.308
5545	2.01	268.39	5544	-3.8	-15.6	16	-3.8	1.293
5640	1.97	247.92	5639	-4.5	-18.7	19.3	-4.5	0.745
5736	1.56	198.04	5734.9	-6.3	-20.7	21.6	-6.3	1.598
5832	1.57	189.05	5830.9	-8.9	-21.3	23.1	-8.9	0.256
5927	1.77	197.93	5925.8	-11.5	-21.9	24.8	-11.5	0.344
6022	1.47	189.12	6020.8	-14.1	-22.6	26.7	-14.1	0.41
6117	1.1	155.48	6115.8	-16.2	-22.4	27.6	-16.2	0.867
6307	1.06	132.51	6305.7	-19	-20.4	27.9	-19	0.227
6402	1.07	123.66	6400.7	-20.1	-19	27.6	-20.1	0.173
6497	0.71	100.21	6495.7	-20.7	-17.6	27.2	-20.7	0.532
6592	0.88	33.24	6590.7	-20.2	-16.7	26.2	-20.2	0.935
6635	0.77	56.73	6633.7	-19.8	-16.2	25.6	-19.8	0.82
6675	1.18	44.22	6673.7	-19.3	-15.7	24.9	-19.3	1.149
6770	1.25	25.97	6768.7	-17.7	-14.6	22.9	-17.7	0.412
6865	1.46	26.18	6863.7	-15.7	-13.6	20.8	-15.7	0.221
6960	1.09	38.77	6958.6	-13.9	-12.5	18.7	-13.9	0.486
7056	1.57	205.96	7054.6	-14.3	-12.5	19	-14.3	2.754
7151	1.92	193.85	7149.6	-17.1	-13.5	21.7	-17.1	0.533
7246	1.38	197.6	7244.5	-19.7	-14.2	24.3	-19.7	0.579
7341	0.83	203.94	7339.5	-21.4	-14.8	26	-21.4	0.592
7436	0.89	200.92	7434.5	-22.7	-15.4	27.4	-22.7	0.079
7531	0.95	208.89	7529.5	-24.1	-16	28.9	-24.1	0.149
7639	0.74	209.63	7637.5	-25.5	-16.8	30.5	-25.5	0.195
7830	1.62	212.6	7828.4	-28.9	-18.8	34.5	-28.9	0.462
7925	1.87	235.65	7923.4	-30.9	-20.9	37.2	-30.9	0.778
8020	1.85	235.65	8018.4	-32.6	-23.4	40.1	-32.6	0.021
8116	1.52	279.59	8114.3	-33.3	-25.9	42.2	-33.3	1.351
8211	1.76	291.62	8209.3	-32.5	-28.5	43.3	-32.5	0.44
8306	1.49	275.63	8304.2	-31.9	-31.1	44.5	-31.9	0.553
8402	0.87	250.94	8400.2	-32	-33	46	-32	0.821
8428	1.39	249.84	8426.2	-32.1	-33.5	46.4	-32.1	2.002
8558	2.78	264.46	8556.1	-33	-38.1	50.4	-33	1.136
8653	2.31	307.21	8651	-32.1	-42	52.8	-32.1	2.006
8743	0.99	21.37	8741	-30.2	-43.1	52.7	-30.2	2.501
8843	0.92	283.71	8841	-29.2	-43.6	52.5	-29.2	1.438
8938	2.76	282.65	8935.9	-28.6	-46.6	54.6	-28.6	1.937
9033	1.05	250.44	9030.9	-28.3	-49.6	57.1	-28.3	2.056
9127	2.4	258.85	9124.8	-29	-52.4	59.9	-29	1.457
9222	1.31	284.52	9219.8	-29.1	-55.4	62.6	-29.1	1.416
9317	3.73	270.04	9314.7	-28.9	-59.5	66.1	-28.9	2.614
9412	3.96	250.99	9409.5	-29.9	-65.7	72.2	-29.9	1.36
9444	3.91	235.47	9441.4	-30.9	-67.6	74.4	-30.9	3.322
9497.5	3.91	235.47	9494.7	-33	-70.6	78	-33	0
9500	3.91	235.47	9497.3	-33.1	-70.8	78.1	-33.1	0

3.2.8. Models validation

The MLM was validated using data set extracted from Table 1. The same extracted data set was also used to validate Equations 2, 4 and 5 which are CCM, Brechan *et al.* [19] method and CTRM. CCM, Brechan method and CTRM are conventionally utilized models for DLS estimation. The comparison of these models is presented in the result section.
CTRM:

$$DLS = \sqrt{\left(\frac{I_2-I_1}{L_2-L_1}\right)^2 + \left(\frac{A_2-A_1}{L_2-L_1}\right)^2 sin^2(\bar{I})} \quad (5)$$

4. Results and discussions

4.1. Modeling without optimization

Some performance metrics were applied to track the performance of the models. The models without the optimal input structures were first tested. Cross-validation was applied to avoid overfitting the model. Figure 10 depicts: accuracy of the different models before optimization while Table 2 depicts error metrics of the different models before optimization. From Figure 10, DT had R^2 of 85%, RF had R^2 of 91.4% while SVM had R^2 of 83.3%. From Table 2, DT had MAE of 0.1912, MSE of 0.1852 and RMSE of 0.2919, RF had MAE of 0.1408, MSE of 0.1512 and RMSE of 0.2263 while SVM had MAE of 0.1686, MSE of 0.1996 and RMSE of 0.3156.

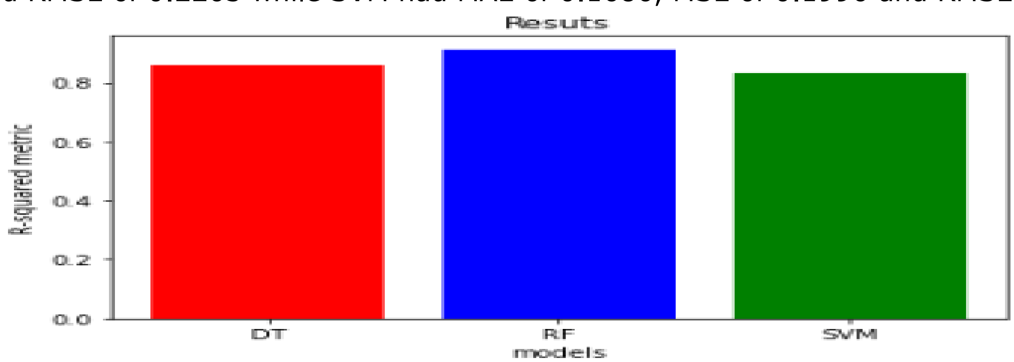

Fig. 10. Accuracy of the different models before optimization.

From Table 2, it is evident that RF had the least error of all the tested error metrics. From Figure 10 and Table 2, it is also evident that RF had the best performance among the three un-optimized MLMs and therefore gave a better DLS estimation. Depicted in Figure 11 is the un-optimized RF-MLM trajectory (model trajectory) and data trajectory of the horizontal section of the directional well. Since RF performed better than DT and SVM, the estimations from RF were used to plot the trajectory of the well's horizontal section. The data trajectory (plot of Table 1) is shown in red colour while the model trajectory is shown in blue colour. From Figure 11, the data trajectory increased with decreasing measured depth until -12.5, it decreased sharply with decreasing depth until -20.7, and it started increasing again with decreasing depth until -0.9 and decreased again from -0.9 to -10.8, thereafter increased from -10.8 to 1.1 and sharply reduced to 0. The model trajectory had a sharp increase with decreasing measured depth from -56 to 0, and thereafter decreased with depth to about -4. From Figure 11, it is obvious that there is much difference between the two trajectories. From Figure 11, the data path/trajectory to achieve maximum angle occurred at 9412 ft depth while that of model path/trajectory occurred at 9322 ft depth. The difference of 90 ft between the two cases indicates the average drilling speed achieved when rotary drilling assemblies are utilized. From Figure 11, the model trajectory satisfies the fatigue condition of the DS. Considering the two cases, depending on priorities considered the RF-MLM well path should be selected. Thus the model trajectory reasonably satisfied DS's fatigue condition.

Table 2. Error metrics of the different models before optimization.

	DT	RF	SVM
Mean Absolute Error (MAE)	0.1912	0.1408	0.1686
Mean Square Error (MSE)	0.1852	0.1512	0.1996
Root Mean Square Error (RMSE)	0.2919	0.2263	0.3156

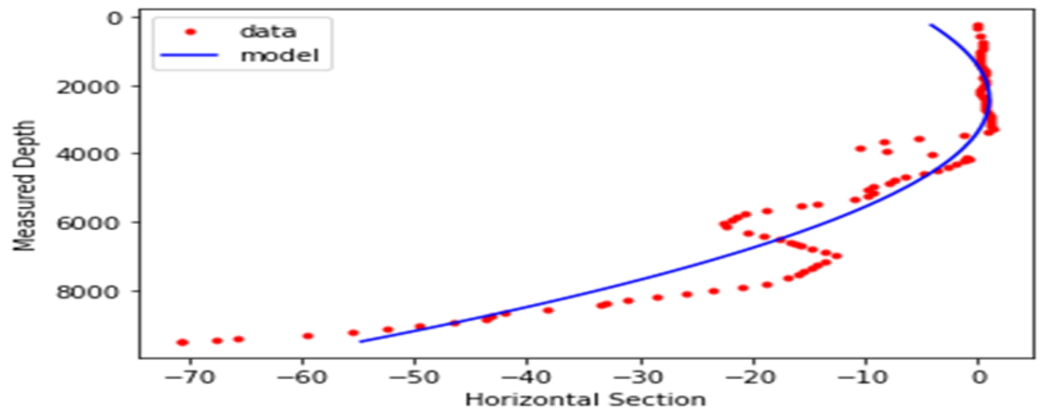

Fig. 11. Trajectory for the unoptimized model.

4.2. Optimized modeling

After testing the performance of these MLMs, optimization techniques were applied on these models to improve their performances. Grid Search was implemented on DT and RF, while GA was used to determine the optimal input structures of the SVM. The same performance metrics were applied to track the performances of these models with determined optimal input parameter. The results of the optimization obtained from these models are as depicted in Figure 12, Table 3 and Figure 13. Figure 12 depicts the accuracy of the different models after optimization while Table 3 depicts error metrics of the different models after optimization. From Figure 12, with the implementation of grid search algorithm on DT, R^2 of 90% was achieved, also with the implementation of grid search algorithm on RF, R^2 of 98.1% was achieved while implementation of GA on SVM improved the R^2 to 98.4%. In comparison with the unoptimized model result of Figure 10, R^2 of DT increased from 85% to 90%, that of RF increased from 91.4% to 98.1% while that of SVM increased from 83.3% to 98.4%. From table 3, DT had MAE of 0.1640, MSE of 0.0574 and RMSE of 0.2395, RF had MAE of 0.0579, MSE of 0.0113 and RMSE of 0.1065 while SVM had MAE of 0.0216, MSE of 0.0094 and RMSE of 0.0097. From Table 3, it is evident that SVM had the least error of all the tested error metrics. From Figure 12 and Table 3, it is also evident that implementation of optimization techniques like grid search and GA, improved the performances of the MLMs with SVM having the best performance of the three MLMs and therefore optimized DLS better than the other two. Figure 13 depicts the optimized SVM-MLM trajectory (model trajectory) and data trajectory of the horizontal section of the directional well. Since SVM performed better than DT and RF, the estimations from the SVM were used to plot the trajectory of the well's horizontal section. From Figure 13, it is obvious that there is again much difference between the two trajectories. From Figure 13, the data path/trajectory to achieve maximum angle occurred at 9412 ft depth while that of model /trajectory occurred at 9276 ft depth. The difference of 136 ft between the two cases indicates the average drilling speed achieved when rotary drilling assemblies are utilized. From Figure 13, the model trajectory satisfies the fatigue condition of the DS much better than that of Figure 11. Thus, implementation of GA on SVM greatly improved the previously unoptimized modeled trajectory. Considering the two cases, depending on the best suited priorities the SVM-MLM well path should be selected. This is an indication that the model trajectory satisfied the DS's fatigue condition. From Figure 13, consideration of the optimization process indicates that irrespective of the hole size, ML optimization algorithm can be implemented in DLS optimization in any hole size.

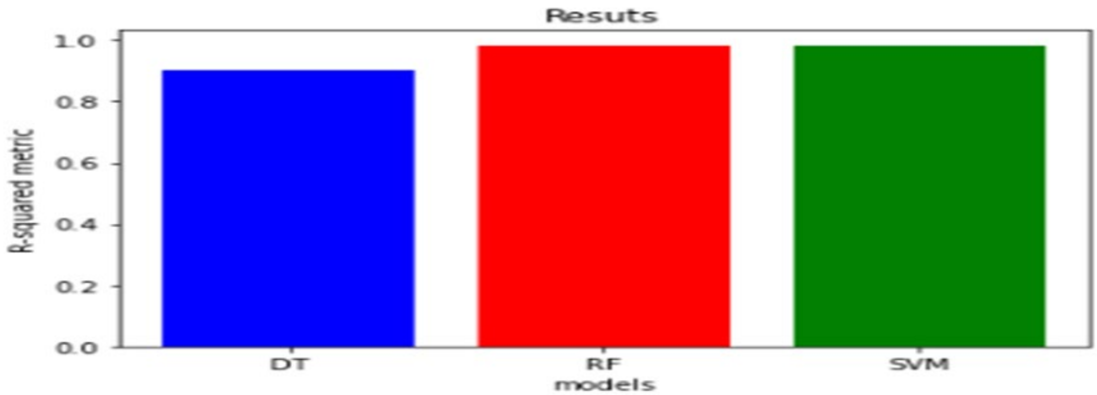

Fig. 12. Accuracy of the different models after optimization.

Table 3. Error metrics of the different models after optimization.

	DT	RF	SVM
MAE	0.1640	0.0579	0.0216
MSE	0.0574	0.0113	0.0094
RMSE	0.2395	0.1065	0.0097

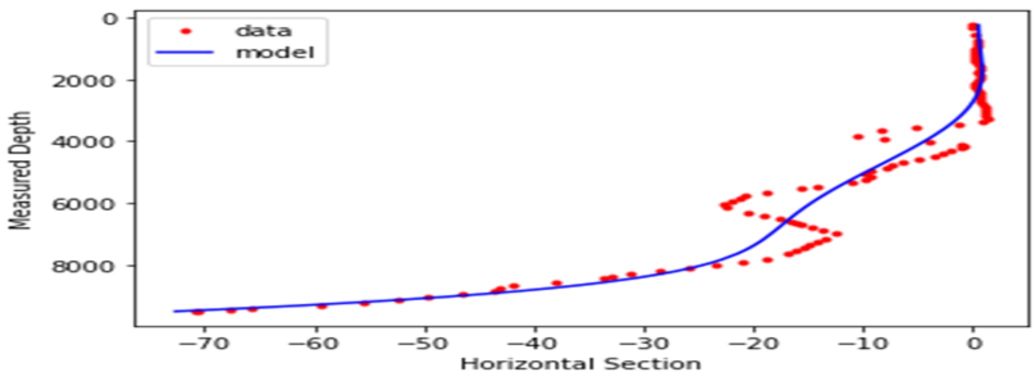

Fig. 13. Optimized model trajectory.

4.3. Models validation

After optimization of the three MLMs, SVM optimized by GA had the best performance as shown in Table 4. The optimized SVM-MLM was used to compare against conventional techniques for calculating DLS. The three mathematical correlations used for this comparison are: CCM, Brechan method [19] and CTRM. A separate validation dataset was utilized for the validation of the models and comparison thereafter made. Depicted in Table 4 are the calculated DLS of CCM, Brechan method [19], CTRM and SVM-MLM. From Table 4, CCM, Brechan *et al*. [19] method, CTRM and the optimized SVM gave different estimates of DLS with all the DLS calculated using optimized SVM-MLM being almost the same as the actual DLS, unlike the DLS calculated from CCM and CTR which are less in some cases and in other cases higher than the actual DLS. Brechan method also performed reasonably well and better than CCM and CTR [19]. It was expected that the SVM-MLM should yield DLS completely the same as the actual DLS with zero difference, except for a single value that was less than the actual DLS. This occurrence could be attributed to improper data recording by LWD/MWD tools at a particular depth or the defect could be due to issues related to sorting and processing of survey and RTDD. The estimated DLS from SVM-MLM is in agreement with error metrics obtained for SVM as depicted in Table 3. From Table 4, it is evident that the optimized SVM-MLM performed better than the three other mathematical correlations, followed by Brechan method [19]. The implementation of GA on SVM yielded DLS almost the same as the actual DLS, which in turn will reduce the chances of operational problems like high torque and drag. This is in agreement with work of [1]. Shokir *et al.* [1] pointed out that reduced DLS means reduced chances of operational problems like high torque and drag.

Table 4. Calculated DLS from validation

DLS calculated using CCM (deg/100ft)	DLS calculated using Brechan method [19] (deg/100ft)	DLS calculated using CTR (deg/100ft)	DLS calculated using SVM-MLM (deg/100ft)	Actual DLS from drilling survey data (deg/100ft)
0.011	0.365	0.365	0.365	0.365
0.108	0.130	0.207	0.129	0.129
0.172	0.172	0.186	0.172	0.172
0.266	0.263	0.517	0.263	0.263
0.288	0.376	0.537	0.376	0.376
0.372	0.220	0.221	0.221	0.221
0.027	0.080	0.079	0.078	0.079
0.077	0.149	0.149	0.149	0.149

5. Conclusion

In this study ML was utilized to compute DLS. The work is therefore is quite significant because the estimated DLS is in agreement with the actual DLS from survey data. From this study, the following conclusions are drawn:

- The proposed approach is highly compatible with deflection tools' directional performance;
- The implementation of GA on SVM yielded excellent DLS value that is comparable with actual DLS;
- The SVM-MLM R^2 value of 98.4% is also in agreement with computed DLS;
- Processing of survey and RTDD data could cause unrealistic outputs production.

Nomenclature

A1	*Azimuth 1 algorithm*	*CL*	*Course length*	*RFA*	*Random Forest*
A2	*Azimuth 2 data*	*DLS*	*Dogleg severity*	*RTDD*	*Real Time Drilling*
I1	*Inclination 1*	*DMEA*	*Hole depth (MD)*	*DS*	*Drill String*
I2	*Inclination 2 machine*	*DS*	*Drill String*	*SVM*	*Support Vector Machine*
L1	*Length 1*	*MD*	*Measured Depth*	*DT*	*Decision tree*
L2	*Length 2*	*GA*	*Genetic Algorithm*		
β	*Azimuth*	*PSO*	*Particle swam optimization*		
α	*inclination*	*MLM*	*Machine learning model*		
φ	*Dogleg angle*	*DMEA*	*Hole depth (MD)*		

References

[1] Shokir EM, Eid MS, and Waly AA. Optimal 3-D directional and horizontal wells planning using genetics algorithm. 2002; SPE 79164, 2002

[2] Hosseini S, Ghanbarzadeh A, and Hashemi A. Optimization of Dogleg Severity in Directional Drilling Oil Wells Using Particle Swarm Algorithm. Journal of Chemical and Petroleum Engineering, 2014; 48 (2): 139–151

[3] Sampaio JH. Designing Three-Dimensional Directional Well Trajectories Using Bézier Curves. J. Energy Resour. Technol., 2016; 139 (3): 032901-032901-8.

[4] Sampaio JHB. Planning 3-D well trajectories using spline-intension functions. J. Energy Resour. Technol., 2007; 129 (4):289–299.

[5] Gong H, Teo KL, Liu C, and Feng E. Horizontal well's path planning: An optimal switching control approach. Applied Mathematical Modelling, 2015; 39 (14):4022–4032.

[6] Guo B, Miska S, and Lee R. Constant curvature method for planning a 3-D directional well. 1992; SPE-24381-MS.

[7] IIyasov RR, Svechnikov LA, and Karimov MR. Automation of optimal well trajectory calculations. 2014; SPE- 171326-MS.

[8] Ajmi A, Adel M, and Zimmerman W. A new well path optimization model for increased mechanical borehole stability. Journal of Petroleum Science and Engineering, 2009; 69(1-2): 53–62.

[9] Farah A, and Omar F. Directional well design, trajectory and survey calculations, with a case study in Fiale, Asal Rift, Djibouti. United Nations University, 2013.

[10] McCann RC, and Suryanarayana PVR. Horizontal Well Path Planning and Correction Using Optimization Techniques. J. Energy Resour. Technol., 2001; 123(3): 187-193.

[11] Noynaert SF. AIMR (Azimuth and Inclination Modeling in Realtime): A Method for Prediction of Dog-Leg Severity based on Mechanical Specific Energy. PhD. Thesis, Texas A&M University, 2013; USA.

[12] Shengzong J, Xilu W, Limin C, and Kunfang L. A new method for designing 3D trajectory in sidetracking horizontal wells under multi-constraints. 1999; SPE-57282-MS.

[13] Sinem K. Prediction of dog-leg severity using artificial neural network. M.Sc. Thesis, Middle East Technical University, 2019; Turkey.

[14] Nkengele J. Determination of Dogleg Severity and Side Force for Stuck Pipe Prevention. Master's thesis, Norwagian University of Science and Technology, 2019; Norway.

[15] Li A, Feng E, and Gong Z. An optimal control model and algorithm for the deviated well's trajectory planning. Applied Mathematical Modelling, 2009; 33 (7): 3068–3075.

[16] Li C, and Cheng C. Prediction and Optimization of Rate of Penetration using a Hybrid Artificial Intelligence Method based on an Improved Genetic Algorithm and Artificial Neural Network. 2020; SPE-203229-MS

[17] Samuel GR. Ultra-Extended-Reach Drilling (u-ERD: Tunnel in the Earth)—A New Well-Path Design. 2009; SPE-119459-MS

[18] Wilson GJ. An Improved Method for Computing Directional Surveys. J. Pet. Tech., 1968; (20): 871-876.

[19] Brechan BA, Corina AN, Gjersvik TB, Sangesland S, and Skalle P. Compendium; TPG4215 Drilling Engineering, Drilling, Completion, Intervention and P&A –design and operations; Revision 1, 2017; pp1-23, Trondheim

To whom correspondence should be addressed: Dr. Anthony Kerunwa, Department of Petroleum Engineering, Federal University of Technology, Owerri, Imo State, Nigeria; e-mail: anthonykerunwa@rocketmail.com