Artificial Intelligence in Oil & Gas Drilling Operations: Managing Key Performance Indicators through Machine Learning Predictive Model

Assad Sheraz
M.Sc. in Business Analytics
Abu Dhabi School of Management
Abu Dhabi, UAE
0009-0003-5272-4158

Shahjahan Khan
Drilling and Completion Department
MOL Pakistan Oil and Gas B.V
Islamabad, Pakistan
0009-0006-6260-6560

Adnan Manzoor
Drilling and Completion Department
MOL Pakistan Oil and Gas B.V
Islamabad, Pakistan

https://doi.org/10.63962/XSQK7927

***Abstract*—This paper presents the applications of Artificial Intelligence (AI) and Machine Learning (ML) in optimizing oil and gas drilling operations and then presents how macro-level drilling Key Performance Indicators (KPIs) are managed through machine learning predictive model. Drilling operations are challenging in nature whereas drilling in Pakistan's northern region is more challenging due to complex geological and tectonic conditions. CRISP-DM methodology is adopted to develop ML models for the prediction of macro-level drilling KPIs such as Dry Hole Drilling Days (DHDD), Dry Hole Drilling Cost (DHDC), and Clean Time (CT). Historical data of the Wells located in Pakistan's northern region are used to train, test, and evaluate ten different ML algorithms. Two top performing models for each KPI are then used to develop ML based predictive calculator in Google Colab. Deployment results from unseen data of six Wells show that the predictions are either better or complimenting the traditional methods. These deployment results show the effectiveness of ML methods and potential of AI & ML in enhancing drilling efficiency, reducing costs, and overcoming challenges of drilling oil & gas Wells.**

Keywords—Artificial Intelligence, Machine Learning, Drilling, Oil & Gas Wells, Key Performance Indicators, Dry Hole Drilling Days, Dry Hole Drilling Cost, Clean Time

I. Introduction

This paper intends to summarize the ongoing applications of Artificial Intelligence (AI) and Machine Learning (ML) in oil & gas drilling operations, followed by a real case example of the deployment of machine learning predictive models for managing the macro-level drilling Key Performance Indicators (KPIs). Previous applications of AI and ML in the field of oil & gas drilling mainly focused on micro-level KPIs like drilling parameters, downhole vibration and Rate of Penetration (ROP), whereas this study focuses on the prediction of macro-level drilling KPIs.

Oil & gas industry play an important role in global economy and serve as a backbone for the world's development and growth. In oil & gas industry, drilling is the Capital Expenditure (CAPEX) intensive and one of the most important domain as it enables the extraction of oil & gas present deep below the subsurface. Oil and gas drilling operations are complex in nature due to their reliance on subsurface data. Drilling in Pakistan's northern region is even more challenging since the geology in this region is complex and tectonically active [1]. Even the Wells located close to each other are entirely different in operational difficulties and hence simple prediction methods produce inaccurate results [2]. In order to tackle these drilling challenges, proactive approach is needed and hence accurate prediction and optimization of Key Performance Indicators (KPIs) is very important. There are two main categories of drilling KPIs, which are micro-level KPIs and macro-level KPIs [3].

A. *Related Work*

Artificial intelligence (AI) was applied to predict gamma-ray (GR) logs in real time during drilling operations, addressing delays caused by logging-while-drilling (LWD) tools measuring already-penetrated formations. Using surface drilling parameters and 4609 data entries from three different Wells, Support Vector Machine (SVM) and Random Forests (RF) models were developed and validated. SVM was better than RF, having correlation coefficient (R) of 0.98 and an average absolute percentage error of 1.42%. These models helped to predict real-time GR and identify formation lithology more efficiently [4]. AI-based workflow was created to enhance drilling performance by controlling and reducing the drill-string vibrations. Support Vector Regression (SVR), Multi-Layer Perceptron (MLP), and Decision Tree Regression (DTR) were utilized to predict & optimize key drilling parameters like Bit RPM, Rate of Penetration (ROP), and torque. SVR was the best model and helped to achieve 43% increase in ROP and reduced the drilling time from 66 to 31 hours [5].

Mechanical Specific Energy (MSE) optimization was carried out through machine learning for improving the drilling performance in one of the offshore field. Random Forest (RF) proved to be the best model for ROP. It also performed better than traditional methods and helped to make real-time adjustments that saved time and cost [6]. In order to detect

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

downhole vibrations (axial, torsional, and lateral) using real-time Rig surface data, machine learning models were developed. 5750 drilling data points were used to develop Radial Basis Function (RBF), Support Vector Machine (SVM) and Adaptive Neuro-Fuzzy Inference System (ANFIS). This resulted in achieving high accuracy with R values ranging from 0.91 to 0.98 and average absolute percentage error from 1.1% to 7.3%. In addition to this, validation was also done through the unseen data. This system provided real-time vibration monitoring and helped to overcome the failures and reduce costs [7].

For extracting data from Rig reports, a conversational AI chatbot using generative AI was designed. It was designed through large language models (LLMs) and machine learning models (ML) to answer the queries related to drilling. In addition to this, it also performed diagnostic analysis and provided operational recommendations. This work showed that how generative AI can change energy sector operations by improving decision-making [8]. In order to perform directional drilling tool selection like Rotary Steerable Systems (RSS) or Positive Displacement Motors (PDM) a machine learning model using offset Wells data was developed. Multiple input variables like Rig hydraulics, inclination-ROP relationships, and tripping quality were used. XGBoost proved to be the best model for this task [9]. Machine learning was applied on data from four Wells to predict & optimize drilling parameters like Weight on Bit (WOB), Flowrate (FR), and Rotation per Minute (RPM). This helped to improve drilling efficiency with the increase in Rate of Penetration (ROP) [10].

Later part of this paper presents the deployment results of the machine learning predictive model developed for macro-level drilling KPIs.

II. Methodology

A. *Machine Learning Predictive Model for Macro-Level Drilling KPIs*

A machine learning model is developed in order to predict macro-level drilling KPIs like Dry Hole Drilling Days (DHDD), Dry Hole Drilling Cost (DHDC), and Clean Time (CT). CRISP-DM (Cross Industry Standard Process for Data Mining) methodology is applied, utilizing historical data from Wells drilled in Pakistan's complex geological region. CRISP-DM (Cross Industry Standard Process for Data Mining) is shown in Figure 1. Multiple input variables are used in order to develop these machine learning models. Ten machine learning algorithms are trained, tested, and evaluated where Support Vector Machine (SVM) & Random Forest (RF) remained the best models for DHDD, Random Forest (RF) & Grid Search CV XGBoost remained best for DHDC and Stacking (SVM Base & LR Meta) & Grid Search CV XGBoost remained best for CT. Single algorithm could not best predict all the KPIs and hence ML based predictive calculator is developed based on two top performing algorithm for each KPI. Best and second best algorithm can vary based on the data used for training, testing and evaluating the machine learning model [11].

B. *Deployment*

In order to manage the Key Performance Indicators (KPIs) through the developed machine learning predictive model, a machine learning model based calculator is developed in Google colab. In order to make the calculator user friendly, unseen input data for carrying out the prediction is also asked to be inserted in the form of csv prediction file. As shown in Figure 2, DataF.csv is the file used to input the data for training, testing and evaluating the machine learning algorithms, whereas Prediction.csv is the file used to input unseen data for prediction. Snapshot of the prediction file showing the required input variables is shown in Figure 3. Once the model is run, it train, test and evaluate the model based on DataF.csv file uploaded in it and presents the two best machine learning models for each of the KPI. Thereafter, using the Prediction.csv file it produces the prediction results for each KPI separately from the two best algorithms.

III. Results & Discussion

A. *Best Performing ML Models*

Data used in deployment phase for training, testing and evaluating the machine learning model, resulted in Support Vector Machine (SVM) & Random Forest (RF) as the best performing models for DHDD with the error percentages of 14.2% & 14.9% respectively, Random Forest (RF) and Grid Search CV XGBoost as the best performing models for DHDC with the error percentages of 11.2% & 14.5% respectively and Stacking (SVM Base & LR Meta) & Grid Search CV XGBoost as the best performing models for CT with the error percentages of 10.56% & 10.64% respectively. Summary of these results is shown in Figure 4.

B. *Deployment Results*

During the deployment phase, six Wells are selected, which are not part of the data that is used to train, test and evaluate the machine learning models. Well 1 to 4 are already drilled, so their planned as well as the actual results are available for comparison, whereas Well 5 & 6 are not drilled completely, so only planned values; from traditional method; are available. ML based calculator made prediction for each of the three KPIs using two best performing models for that specific KPI as mentioned in the previous section. Prediction results for all three KPIs are shown in Table 1. Whereas, Table 2 is showing the planned and actual values for these three KPIs for the same six Wells. Planned values mentioned in this table are from the traditional methods currently in practice, whereas actual values for these KPIs are from the actual drilling of these Wells.

Figure 5 presents the comparison of traditional method & ML model results for DHDD. It can be observed that actual values are exceeding the planned values for all the four Wells which are already drilled, which shows the need of improvement in traditional method adopted for calculating the planned values of DHDD. Predicted values from best performing machine learning model SVM are matching with planned values in Well 1, 2, 5 and 6, whereas they are matching with the actual values in Well 3 and 4. Figure 6 presents the comparison of traditional method & ML model results for DHDC. It can be observed that actual DHDC on two instances is less than the planned DHDC, despite of the fact that actual DHDD are more than the planned DHDD for all the Wells here. Where Actual vs Planned DHDC trend is matching with the Actual vs Planned DHDD trend, predicted values from best performing model RF is matching more closely with the actual DHDC values. Figure 7 presents the comparison of traditional method & ML model results for CT. Since the error percentages for both the models are almost same, the predicted values of both the models Stacking (SVM Base & LR Meta) & Grid Search CV XGBoost are matching closely with the actual CT values.

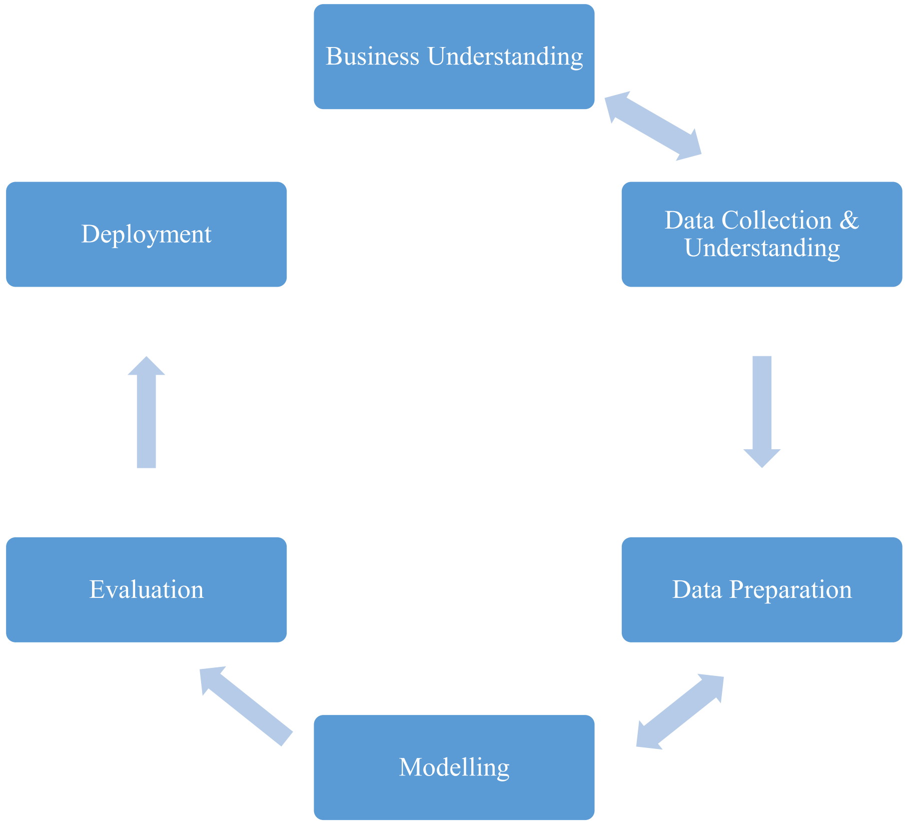


Fig. 1. CRISP-DM Framework [11]

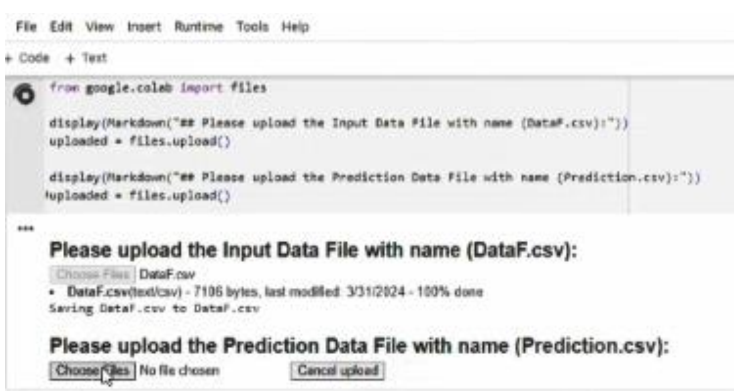
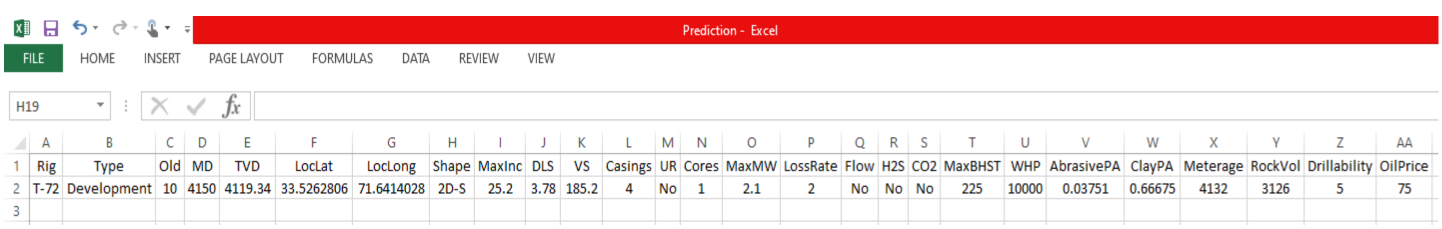


Fig. 2. Snapshot from ML Based KPI Prediction Calculator



Rig	Type	Old	MD	TVD	LocLat	LocLong	Shape	MaxInc	DLS	VS	Casings	UR	Cores	MaxMW	LossRate	Flow	H2S	CO2	MaxBHST	WHP	AbrasivePA	ClayPA	Meterage	RockVol	Drillability	OilPrice
T-72	Development	10	4150	4119.34	33.5262806	71.6414028	2D-S	25.2	3.78	185.2	4	No	1	2.1	2	No	No	No	225	10000	0.03751	0.66675	4132	3126	5	75

Fig. 3. Snapshot of the Prediction.csv File

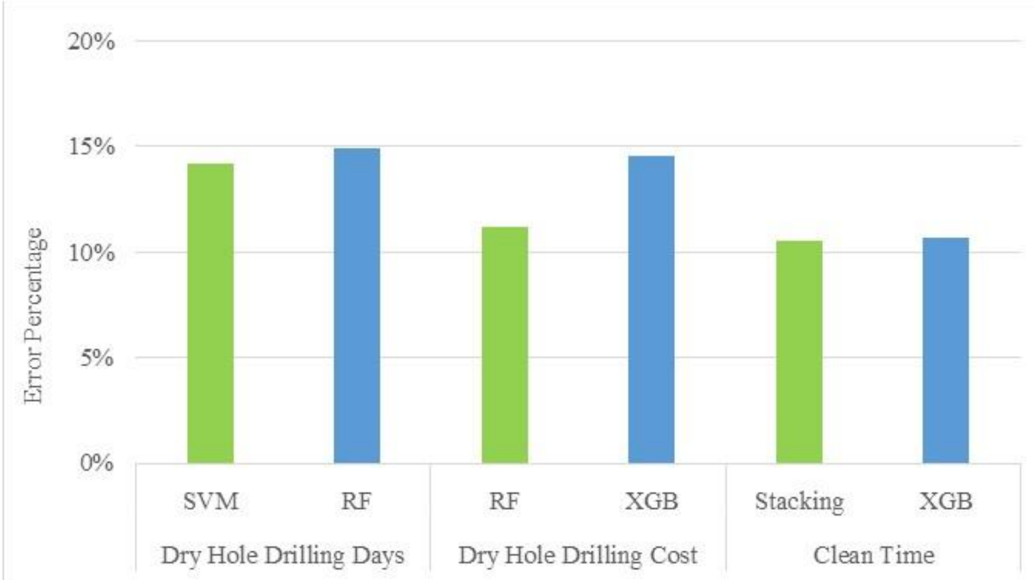


Fig. 4. Two Best Performing Machine Learning Models for each KPI

TABLE I. PREDICTION RESULTS OF TWO BEST PERFORMING ML MODELS

Well Name	**DHDD (days)**		**DHDC (m$)**		**CT (days)**	
	SVM	**RF**	**RF**	**XGB**	**Stacking**	**XGB**
Well 1	135	194	15.4	7.2	115	101
Well 2	218	266	21.7	18.2	162	189
Well 3	199	280	22.7	16.1	214	174
Well 4	233	272	21.7	19.9	191	205
Well 5	108	203	16.0	13.1	121	112
Well 6	116	113	10.1	11.1	88	100

TABLE II. PLANNED & ACTUAL VALUES OF THE KPIS

<table>
<tr><th rowspan="2">Well Name</th><th colspan="2">DHDD (days)</th><th colspan="2">DHDC (m$)</th><th colspan="2">CT (days)</th></tr>
<tr><th>Planned</th><th>Actual</th><th>Planned</th><th>Actual</th><th>Planned</th><th>Actual</th></tr>
<tr><td>Well 1</td><td>133</td><td>162</td><td>12.5</td><td>13.6</td><td rowspan="6">Not Applicable</td><td>101</td></tr>
<tr><td>Well 2</td><td>225</td><td>235</td><td>17.4</td><td>15.5</td><td>179</td></tr>
<tr><td>Well 3</td><td>160</td><td>183</td><td>15.6</td><td>12.7</td><td>175</td></tr>
<tr><td>Well 4</td><td>208</td><td>249</td><td>17.6</td><td>17.8</td><td>181</td></tr>
<tr><td>Well 5</td><td>129</td><td>Not Available</td><td>13.8</td><td>Not Available</td><td>Not Available</td></tr>
<tr><td>Well 6</td><td>110</td><td>Not Available</td><td>10.5</td><td>Not Available</td><td>Not Available</td></tr>
</table>

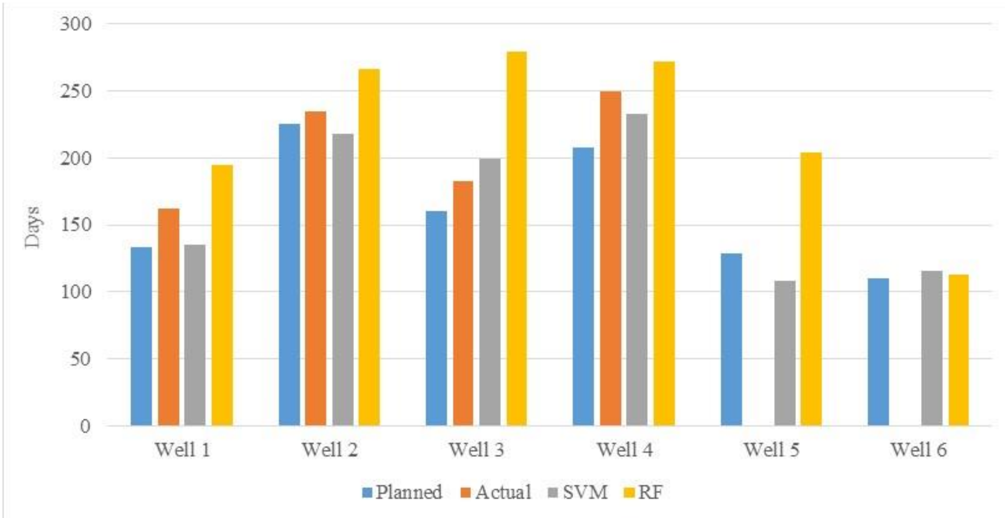


Fig. 5. Comparison of Traditional Method & ML Model Results for DHDD

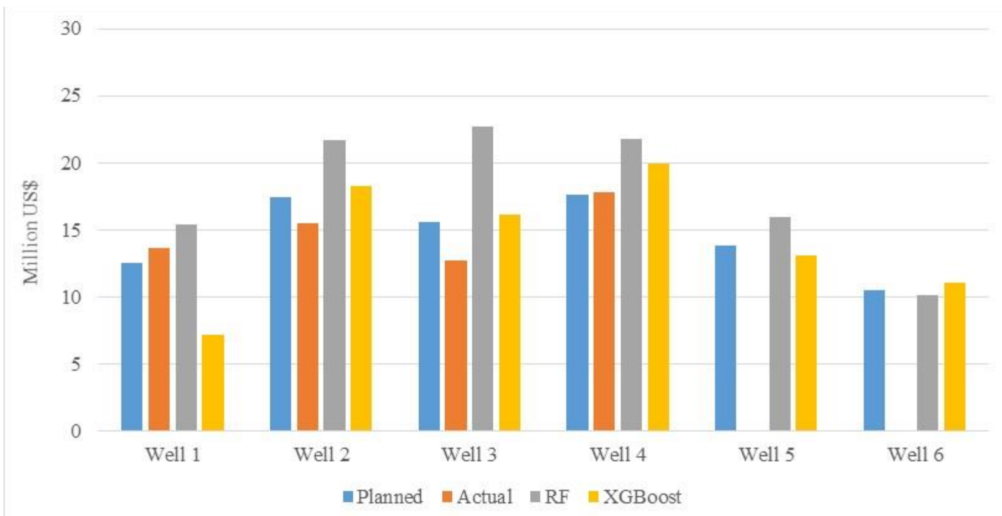


Fig. 6. Comparison of Traditional Method & ML Model Results for DHDC

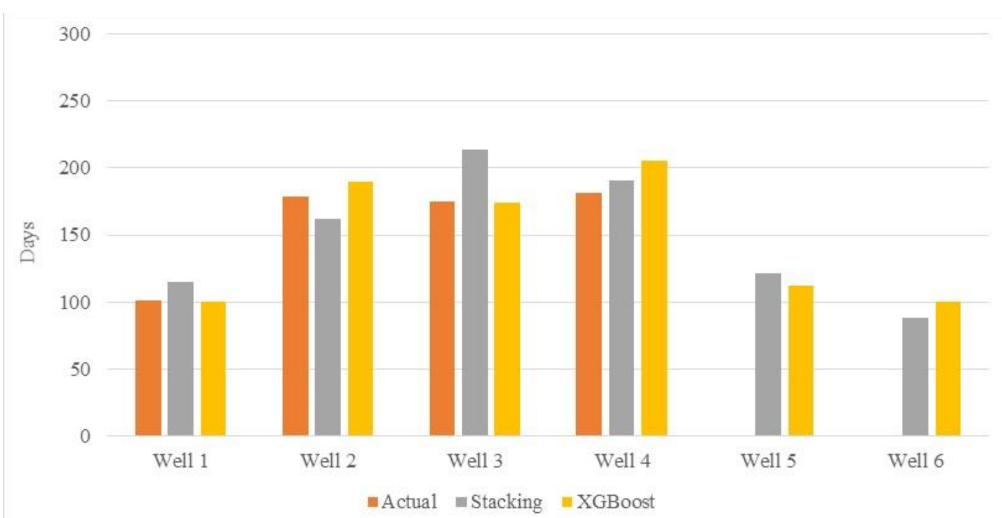


Fig. 7. Comparison of Traditional Method & ML Model Results for CT

IV. Conclusion

This paper shows the application of AI & ML in optimizing oil and gas drilling operations, particularly in complex geological regions like Pakistan's northern region. Machine learning predictive models are developed to predict macro-level drilling KPIs such as DHDD, DHDC, and CT and accordingly then deployed by predicting the KPIs of the unseen data. Development of a user friendly ML based predictive calculator in Google Colab assists in the deployment phase. The results from six Wells show that the predictions from these models either complement or perform better than the traditional methods, highlighting the potential of AI & ML in enhancing drilling efficiency and reducing the cost. This paper focuses on macro-level drilling KPIs and hence contributes to AI & ML growing applications in the field of oil & gas drilling which has previously focused on micro-level drilling KPIs.

Since this study is based on the data availability from one of the fields from Pakistan's northern region. Future study can focus on training, testing, evaluating and deploying the same ML based predictive calculator on other fields and geological settings. This study will help to expand this application in other drilling regions. Moreover, future study can also be conducted to reduce variance between the planned / actual DHDC and ML predicted DHDC value by including some additional cost related input variables. This will help to refine DHDC prediction.

Acknowledgment

Authors express their heartfelt gratitude to MOL Pakistan for their support in publishing this research paper and for sharing the actual Wells data, which played a pivotal role in this research study and develop machine learning models. Special thanks to Abu Dhabi School of Management (ADSM) for providing the platform to conduct this research study.

References

[1] S. Khan, A. N. Awan, Z. Sarfraz, and I. Muhammad, "Geomechanics Application in the North of Pakistan: Unlocking Drilling Excellence-A Comprehensive Analysis of 10 Years," in *SPE/PAPG Pakistan Section Annual Technical Conference*, pp. SPE-219509, 2023.

[2] J. Haneef and A. Sheraz, "Development of well complexity calculator and its integration into standard well engineering management system/well delivery system," *Journal of Petroleum Exploration and Production Technology*, vol. 12, no. 6, pp. 1727–1757, 2021.

[3] J. Haneef and A. Sheraz, "A Comparative Analysis of Well Key Performance Indicators (KPIs) with Well Complexities Using Well Complexity Calculator," *Arabian Journal for Science and Engineering*, pp. 1–18, 2022.

[4] F. Ibrahim and S. Elkatatny, "Real-time GR logs estimation while drilling using surface drilling data; AI application," *Arabian Journal for Science and Engineering*, vol. 47, no. 9, pp. 11187–11196, 2022.

[5] F. S. Boukredera, M. R. Youcefi, A. Hadjadj, C. P. Ezenkwu, V. Vaziri, and S. S. Aphale, "Enhancing the drilling efficiency through the application of machine learning and optimization algorithm," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 107035, 2023.

[6] Nautiyal and A. K. Mishra, "Machine learning application in enhancing drilling performance," *Procedia Computer Science*, vol. 218, pp. 877–886, 2023.

[7] R. Saadeldin, H. Gamal, and S. Elkatatny, "Detecting downhole vibrations through drilling horizontal sections: Machine learning study," *Scientific Reports*, vol. 13, no. 1, p. 6204, 2023.

[8] Singh, T. Jia, and V. Nalagatla, "Generative AI enabled conversational Chatbot for drilling and production analytics," in *Abu Dhabi International Petroleum Exhibition and Conference*, p. D021S065R002, 2023.

[9] M. Nour, S. K. Elsayed, and O. Mahmoud, "A supervised machine learning model to select a cost-effective directional drilling tool," *Scientific Reports*, vol. 14, no. 1, p. 26624, 2024.

[10] Elahifar, "Real-Time Artificial Intelligence-Enhanced Machine Learning Technique for Accurate Drilling Parameter Prediction and Optimization," in *SPE Annual Technical Conference and Exhibition*, p. D021S017R001, 2024.

[11] Sheraz, S. Khan, N. Abdelhamid, and A. Manzoor, "Application of Machine Learning for Comprehensive Predictive Modelling of Drilling Key Performance Indicators Using Historical Drilling Data - Collaborative Case Study Between Academia and Industry," in *Oman Petroleum & Energy Show*. SPE, 2025.