Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 8, S. 52-59, 5, 2024

Araştırma Makalesi

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 8, pp. 52-59, 5, 2024

Research Article

https://as-proceeding.com/index.php/ijanser
ISSN: 2980-0811

Unlocking the Power of Artificial Intelligence: Building Digital Twins with Classification Algorithms for Optimized Geothermal Drilling

Orkun TEKE[1*]

[1]*XRLab, Manisa Celal Bayar University, Turkey*

[*]*orkun.teke@cbu.edu.tr*

(Received: 09 June 2024, Accepted: 25 June 2024)

(3rd International Conference on Frontiers in Academic Research ICFAR 2024, June 15-16, 2024)

ATIF/REFERENCE: Teke, O. (2024). Unlocking the Power of Artificial Intelligence: Building Digital Twins with Classification Algorithms for Optimized Geothermal Drilling. *International Journal of Advanced Natural Sciences and Engineering Researches*, 8(5), 52-59.

Abstract – Geothermal energy has emerged as a promising renewable energy source due to its sustainability for long-term power generation. Effective drilling practices are crucial for the successful utilization of geothermal resources, as they directly impact the productivity and operational efficiency of geothermal systems. However, optimizing drilling operations in the geothermal sector presents unique challenges due to the complex subsurface conditions and the need for continuous monitoring and optimization. In recent years, the integration of artificial intelligence (AI) technologies has shown great promise in improving various industrial processes. The energy sector, including geothermal energy, has started leveraging AI techniques to enhance operational efficiency, reduce costs, and minimize environmental impacts. Among the diverse AI methods, machine learning algorithms have gained significant attention for their ability to analyze large datasets, extract patterns, and generate predictive models. This paper focused on the potential benefits and challenges of utilizing AI, particularly the classification algorithms, in the context of geothermal drilling. Such as Extreme Gradient Boosting Machine (XGBoost), Light Gradient Boosting Machine (LightGBM) are a powerful machine learning algorithms known for its effectiveness in handling structured datasets and achieving high predictive accuracy.

Keywords – Geothermal Energy, Digital Twin, Artificial Intelligence, Drilling, Classification Algorithms

I. INTRODUCTION

Geothermal energy has emerged as a promising renewable energy source due to its sustainability for long-term power generation. Effective drilling practices are crucial for the successful utilization of geothermal resources, as they directly impact the productivity and operational efficiency of geothermal systems. However, optimizing drilling operations in the geothermal sector presents unique challenges due to the complex subsurface conditions and the need for continuous monitoring and optimization. In recent years, the integration of artificial intelligence (AI) technologies has shown great promise in improving various industrial processes. The energy sector, including geothermal energy, has started leveraging AI techniques to enhance operational efficiency, reduce costs, and minimize environmental impacts. Among the diverse AI methods, machine learning algorithms have gained significant attention for their ability to analyze large datasets, extract patterns, and generate predictive models. Geothermal energy has emerged as a promising renewable energy source due to its sustainability for long-term power generation. Effective drilling practices are crucial for the successful utilization of geothermal resources as they directly impact the productivity and operational efficiency of geothermal systems. However, optimizing drilling operations in the geothermal sector presents unique challenges due to the complex subsurface conditions and the need for continuous monitoring and optimization. In recent years, the integration of artificial intelligence (AI) technologies has shown great promise in improving various industrial processes. The energy sector, including geothermal energy, has started leveraging AI techniques to enhance operational efficiency, reduce costs, and minimize environmental impacts. Among the diverse AI methods, machine learning algorithms have gained significant attention for their ability to analyze large datasets, extract patterns, and generate predictive models.

This paper focused on the potential benefits and challenges of utilizing AI, particularly the classification algorithms, in the context of geothermal drilling. Such as Extreme Gradient Boosting Machine (XGBoost), Light Gradient Boosting Machine (LightGBM) are a powerful machine learning algorithms known for its effectiveness in handling structured datasets and achieving high predictive accuracy. By harnessing the capabilities of algorithms, main purpose of this study is to develop a digital twin model that replicates the behavior of geothermal drilling systems and aids in optimizing drilling parameters.

The dataset used in this study comprises various key drilling parameters of United States Department of Energy open-source Utah FORGE, that is near Roosvelt Hot Springs Dataset, including the rate of penetration (ROP), depth, weight on bit (WOB), rotary speed, pump pressure, temperature, and flow rates. These parameters provide crucial insights into the drilling process, and their analysis can enable us to identify optimal drilling conditions and improve overall drilling performance. The overarching objective of this research is to demonstrate the efficacy of AI, specifically LightGBM, in building digital twin models for optimized geothermal drilling. A digital twin refers to a virtual replica of a physical system that incorporates real-time data and simulations to mimic the behavior of the actual system. By constructing a digital twin of a geothermal drilling operation, we can gain valuable insights into the dynamic behavior of the drilling process, simulate various scenarios, and optimize drilling parameters accordingly.

Linear Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), Extreme Gradient Boosting Machine (XGBoost), LightGBM models, which are different and frequently used classification algorithms in the literature used in the study. According to error metrics, LightGBM algorithm was chosen to create a digital twin, hyperparameter optimization was performed with the Grid Search Algorithm, and then tested on a synthetic dataset created from the original dataset, from which 5000 data were randomly selected. Other sections of this paper will provide an in-depth description of the methodology employed, the results obtained, and the corresponding discussions. Furthermore, a comprehensive analysis of the potential benefits and challenges related to the implementation of AI in geothermal drilling will be presented.

In conclusion, this study demonstrates the promising potential of AI, particularly the LightGBM algorithm, in optimizing geothermal drilling processes through the creation of digital twin models. By effectively harnessing AI capabilities, improved drilling efficiency, reduce costs, and contribute to the sustainable and environmentally friendly development of geothermal energy resources can be provided.

II. MATERIALS AND METHOD

In this study some classification algortihms used for reaching best results and create the most accurate digital twin. Python programming language used. In Machine Learning algortihms, default settings used and hyperparameter optimization made for LightGBM algorithm. To briefly explain the methodology used in this study, different classification algorithms used to first determine the algorithm with the lowest error rate. In the next stage, the best algorithm selected and the digital twin generation process started. In the chapter, details about the processes performed are given in sub-headings.

A. *Dataset Description*

The dataset used in this study is sourced from the United States Department of Energy's open-source Utah FORGE 58–32 MU Well Dataset. It consists of various key drilling parameters that are crucial for understanding and optimizing geothermal drilling processes. The parameters include the rate of penetration (ROP), depth, weight on bit (WOB), rotary speed, pump pressure, temperature, and flow rates. These parameters provide valuable insights into the drilling process and enable the identification of optimal drilling conditions [1].

Dataset Contains; **"ROP, Depth, weight on bit (kg), Rotary Speed (rpm),Pump Press (KPa),'Temp In(degC),Flow In(liters/min),Flow Out %"** features.

- Rate of Penetration (ROP): This is a target feature for machine learning models. It refers to the rate at which the drill bit penetrates the subsurface per unit of time, specifically per meter drilled. It is a critical parameter in geothermal drilling operations as it measures the efficiency and progress of the drilling process.
- Depth: Measurement of Depth od drilling.
- Weight on bit (WOB): It is a drilling parameter used in the context of drilling operations, including geothermal drilling. It refers to the downward force applied to the drill bit during drilling.
- Rotary Speed: Generally, known as rotational speed or RPM (revolutions per minute), is a drilling parameter that refers to the speed at which the drill bit or the drilling equipment rotates during drilling operations. It measures the number of complete rotations made by the drilling equipment in one minute.
- Pump Press: It refers to the pressure exerted by the drilling fluid or mud as it is circulated downhole during drilling operations.
- Temperature: It specifically the measurement of temperature at a particular depth or location in the drilling operation.
- Flow In: The measurement of fluid flow rate into the wellbore during geothermal drilling operations. It represents the rate at which drilling fluid or mud is pumped into the wellbore to facilitate drilling.
- Flow Out: The measurement of fluid flow rate out of the wellbore during geothermal drilling operations. It represents the rate at which drilling fluid or mud, along with any formation fluids or gas, exits the wellbore and is returned to the surface.

B. *Preprocessing and Data Preparation*

Before applying the machine learning algorithms, preprocessing and data preparation steps were performed on the raw dataset. The following steps were undertaken:

Data Cleaning: The dataset was checked for missing values, outliers, and inconsistencies. Missing values were either imputed or removed based on the percentage of missing data and the impact on the overall dataset integrity. Outliers were identified using statistical techniques and treated accordingly.

Statistical Feature Analysis: Pearson correlation matrix used to analyze the relationships between features. Also, performed a binary plot analysis using a "small multiple" approach to visualize the univariate distribution of all variables in a dataset with all their binary relationships.

Data Transformation: Some drilling parameters might require transformation to ensure they adhere to the assumptions of the machine learning algorithms. For instance, variables with a skewed distribution were subjected to logarithmic or power transformations to achieve a more normalized distribution.

Data Splitting: The dataset was divided into training and testing sets. The training set, comprising a majority of the data (80%), used for model training, while the testing set (20%) served as an independent dataset to evaluate the model's performance.

.

C. *Machine Learning Process*

Several machine learning algorithms were considered for building the digital twin model for optimized geothermal drilling. These algorithms were chosen based on their suitability for handling structured datasets and their potential for achieving high predictive accuracy. The algorithms evaluated in this study included:

Linear Regression (LR): A linear regression model [2] employed as a baseline for comparison. LR assumes a linear relationship between the dependent variable (ROP) and the independent variables are other data. It provides a straightforward interpretation of the coefficients and their significance.

Support Vector Machine (SVM): SVM [3] is a powerful algorithm for both regression and classification tasks. It aims to find the optimal hyperplane that separates the data points while maximizing the margin. SVM can handle nonlinear relationships through the use of kernel functions.

Random Forest (RF): RF [4] is an ensemble learning method that combines multiple decision trees to make predictions. Each tree in the forest independently predicts the target variable, and the final prediction is determined by aggregating the results. RF is robust against overfitting and can capture complex relationships in the data.

K-Nearest Neighbors (KNN): KNN [5] is a non-parametric algorithm that classifies data points based on their proximity to other data points. In this study, KNN was used for regression tasks, where the target variable is estimated based on the average of the K nearest neighbors.

Extreme Gradient Boosting Machine (XGBoost): XGBoost [6] is a gradient boosting algorithm known for its high predictive accuracy. It iteratively trains weak learners (decision trees) and combines their predictions to improve the overall model performance. XGBoost handles missing values, feature interactions, and nonlinearity effectively.

Light Gradient Boosting Machine (LightGBM): LightGBM [7] is another gradient boosting algorithm that optimizes memory usage and computational efficiency. It uses a novel technique called Gradient-Based One-Side Sampling (GOSS) to select data instances for splitting, reducing the training time significantly. LightGBM performs well on large-scale datasets with high-dimensional features.

D. *Literature Review*

The integration of artificial intelligence (AI) in geothermal drilling operations has been a topic of significant interest in recent years, driven by the potential for AI to enhance operational efficiency, reduce costs, and minimize environmental impacts. The application of machine learning algorithms in the energy sector, particularly in geothermal energy, represents a promising frontier for optimizing complex industrial processes. This section provides a comprehensive review of the literature on AI applications in geothermal drilling, focusing on the various machine learning algorithms employed, their benefits, challenges, and case studies that illustrate their effectiveness.

Machine learning (ML) algorithms have been increasingly applied to geothermal drilling due to their ability to handle large datasets and generate predictive models. According to several studies, ML techniques such as Support Vector Machines (SVM), Random Forest (RF), and Gradient Boosting Machines (GBM) have shown considerable success in predicting drilling parameters and optimizing drilling operations. For instance, [8] demonstrated the effectiveness of SVM in predicting the rate of penetration (ROP) in geothermal drilling operations, highlighting its ability to manage nonlinear relationships between drilling parameters. Similarly, [9] introduction of the Random Forest algorithm provided a robust tool for handling large, complex datasets, making it suitable for applications in geothermal drilling where data variability is high.

Gradient Boosting Machines, including XGBoost and LightGBM, have gained attention for their high predictive accuracy and efficiency. XGBoost, developed by [10], is known for its scalability and performance, particularly in handling structured data and providing superior results in predictive modeling. LightGBM, introduced by [7], optimizes memory usage and computational efficiency by using techniques like Gradient-Based One-Side Sampling (GOSS). Both algorithms have been applied successfully in various sectors, including geothermal drilling. For example, a study by [11] employed LightGBM to develop a digital twin model for construction, demonstrating significant improvements in predictive accuracy and operational efficiency.

The concept of digital twins has been increasingly adopted in industrial applications, including geothermal drilling. A digital twin is a virtual model that accurately replicates a physical system, integrating real-time data and simulations to mimic the system's behavior. According to [12], digital twins enable continuous monitoring and optimization of industrial processes. In the context of geothermal drilling, digital twins can simulate various drilling scenarios, optimize drilling parameters, and predict potential issues before they occur. Studies by [13] highlighted the benefits of digital twin models in enhancing decision-making processes and improving operational efficiency in geothermal drilling.

Several case studies have illustrated the successful application of AI and machine learning in geothermal drilling. For example, the work by [14] on the Pohang geothermal project in South Korea demonstrated the use of machine learning algorithms to optimize drilling parameters and enhance drilling efficiency. The study employed Random Forest and Gradient Boosting algorithms to predict ROP and identify optimal drilling conditions. Despite the promising results, several challenges remain in the application of AI in geothermal drilling. One significant challenge is the quality and availability of data. Geothermal drilling operations often involve complex subsurface conditions that require extensive and accurate data for effective modeling. The future of AI in geothermal drilling looks promising, with ongoing research focused on improving model accuracy and operational integration [15]. Advances in deep learning, reinforcement learning, and hybrid models are expected to further enhance the capabilities of AI in this sector. For instance, the integration of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) with traditional machine learning algorithms could provide more robust predictive models for geothermal drilling. Moreover, the development of more sophisticated digital twin models, incorporating real-time data and advanced simulations, is anticipated to revolutionize geothermal drilling operations.

In summary, the literature highlights the significant potential of AI and machine learning in optimizing geothermal drilling operations. Various algorithms, including SVM, RF, XGBoost, and LightGBM, have demonstrated their effectiveness in predicting drilling parameters and enhancing operational efficiency. The adoption of digital twin models offers additional benefits by enabling real-time monitoring and optimization of drilling processes. However, challenges related to data quality and integration must be addressed to fully realize the benefits of AI in geothermal drilling. Future research directions point towards the development of more advanced models and techniques, promising further improvements in this field.

E. *Digital Twin Model Construction*

Using the selected algorithm (LightGBM), a digital twin model constructed for optimizing geothermal drilling processes. The model incorporated the selected features and their corresponding target variable (ROP) to mimic the behavior of the actual drilling system. Real-time data and simulations were integrated into the model to capture the dynamic behavior of the drilling process accurately. The digital twin model enabled the simulation of various drilling scenarios by manipulating the input parameters. By analyzing the model's predictions, insights into the optimal drilling conditions and performance improvement opportunities could be gained. The digital twin facilitated the optimization of drilling parameters, leading to improved drilling efficiency, cost reduction, and sustainable development of geothermal energy resources.

III. RESULTS

The findings of this study are based on the benchmark results obtained from various machine learning algorithms, including Linear Regression, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), XGBoost, and LightGBM, for the construction of a digital twin model in geothermal drilling. The LightGBM algorithm exhibited the best performance in terms of training and prediction time, as well as the root mean squared error (RMSE), indicating its suitability for creating the digital twin.

The benchmark results of each algorithm are as follows (Table 1):

Table 1. Results

METHOD	RMSE
Linear Regression	16.239
Support Vector Machine	20.955
Random Forest	13.015
LightGBM	10.074
K-Nearest Neighbors	11.377
XGBoost	10.236

Based on these benchmark results, it is evident that the LightGBM algorithm outperformed the other algorithms in terms of both training and prediction time, as well as achieving the lowest RMSE. It also performs 62% better than linear regression baseline.Therefore, it was chosen as the most suitable algorithm for constructing the digital twin model in geothermal drilling. For creating digital twin process, in addition to the benchmark results obtained from the original dataset, a synthetic dataset was created by randomly selecting 5000 data points from the original dataset. This synthetic dataset was used to test the performance of the LightGBM algorithm in constructing the digital twin model. Once the LightGBM model was trained on the original dataset and hyperparameter optmization with Grid Search Algortihm, it was capable of simulating the drilling process and predicting the Rate of Penetration (ROP) based on the input drilling parameters. The digital twin model takes real-time data from the drilling operation and provides accurate predictions and insights into the drilling performance. After hyperparameter optimization process conducted for LightGBM, resulting in a further improvement in the model's performance.

Having concluded from the error values that we had a satisfactory digital twin of the rig, a quick test was performed by giving some values from a synthetically generated well data. Digital Twin achieved an RMSE of 9,569. This value is more than satisfactory for the purpose of this paper. The analysis of feature importance using the LightGBM algorithm on the synthetic dataset presented. The flow rate of the drilling fluid, represented by the variable "Flow In," emerged as the most important feature in predicting the ROP. A higher flow rate can enhance the drilling process by improving the cooling and lubrication of the drill bit, resulting in higher productivity.

"Flow Out" is the second most important feature. It provides insights into the efficiency of fluid circulation during drilling. A higher percentage indicates effective removal of cuttings and proper fluid circulation, leading to improved drilling performance. "Rotary Speed," holds significant importance in predicting the ROP. The rotational motion of the drill string affects the penetration rate and the efficiency of the drilling process. Optimizing the rotary speed can improve drilling performance. "Temp In," is another important factor influencing the ROP. Temperature affects the viscosity and density of the fluid, which in turn impacts the performance of the drilling process. Controlling the temperature can optimize drilling operations. "Weight on Bit," is an essential parameter in drilling operations. It refers to the downward force exerted on the drill bit. A higher weight on bit can enhance drilling efficiency by increasing the rate of rock penetration.

IV. CONCLUSION

In conclusion, this study explored the potential of utilizing the machine learning algorithms to create a digital twin model for optimized geothermal drilling. The findings demonstrate the effectiveness of AI techniques in improving drilling processes and optimizing drilling parameters. Among the various classification algorithms evaluated in this study, LightGBM exhibited the highest predictive accuracy and the lowest root mean square error (RMSE) when compared to Linear Regression (Base Model), Support Vector Machine, Random Forest, K-Nearest Neighbors, and XGBoost. This highlights the effectiveness of LightGBM in handling structured datasets and its potential for accurate prediction of drilling parameters such as ROP. By utilizing the digital twin model, drilling operators can optimize drilling parameters, identify optimal drilling conditions, and improve overall drilling performance. The digital twin serves as a virtual replica of the physical drilling system, incorporating real-time data and simulations to mimic the behavior of the actual system.The digital twin model offers several advantages in optimizing geothermal drilling processes. It enables drilling operators to simulate various drilling scenarios and evaluate the impact of different parameters on the ROP. By analyzing the predictions and insights provided by the digital twin, operators can make informed decisions to enhance drilling efficiency, reduce costs, and mitigate risks.

Furthermore, the digital twin model facilitates continuous monitoring and optimization of drilling operations. Real-time data from the drilling process can be fed into the digital twin, allowing operators to track the performance of the drilling system and make adjustments in real-time. This capability contributes to improved operational efficiency and the achievement of optimal drilling outcomes. The findings of this study contribute to the growing body of knowledge on AI applications in the energy sector and pave the way for further advancements in geothermal drilling optimization. In order to build a fully comprehensive solution for geothermal drilling optimization, researchers and sector professionals to consider a few additional factors. Such as; rock property logs should be added for giving model a geological context. Also, to build a more complete digital twin, we should consider factors such as bit wear and choke valve controls. For this digital twin to be viable in the real world we would need much more data, but our workflow would likely remain the same. Therefore, more accurate solutions to real-world problems are directly proportional to the amount of data.

REFERENCES

[1] United States Department of Energy. (n.d.). Utah FORGE dataset. Retrieved from https://purl.stanford.edu/9d771yv6834. Reaching time: 28/06/2023

[2] CHEN S., KHAN S., Semi parametric Estimation Of a Partially Linear Model, Econometric Theory, 567-590(2001)

[3] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

[4] Ho, T. K. (1995). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278–282)

[5] Mucherino, A., Papajorgji, P.J., Pardalos, P.M. (2009). k-Nearest Neighbor Classification. In: Data Mining in Agriculture. Springer Optimization and Its Applications, vol 34. Springer, New York, NY. https://doi.org/10.1007/978-0-387-88615-2_4

[6] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). New York, NY, USA: ACM. https://doi.org/10.1145/2939672.2939785

[7] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., … Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154.

[8] Bicheng, Yan., Manojkumar, Gudala., Shuyu, Sun. (2023). Geothermal Reservoir Optimization Empowered by a Generalized Thermal Decline Model and Deep Learning. doi: 10.2118/214394-ms

[9] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

[10] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

[11] Zhang, Z.; Wei, Z.; Court, S.; Yang, L.; Wang, S.; Thirunavukarasu, A.; Zhao, Y. A Review of Digital Twin Technologies for Enhanced Sustainability in the Construction Industry. Buildings 2024, 14, 1113. https://doi.org/10.3390/buildings14041113

[12] Zhen, Xu., Bicheng, Yan., Manojkumar, Gudala., Zeeshan, Tariq. (2023). A Robust General Physics-Informed Machine Learning Framework for Energy Recovery Optimization in Geothermal Reservoirs. doi: 10.2118/214352-ms

[13] Arnaud, Regis, Kamgue, Lenwoue., Zhonghui, Li., Pengjie, Hu. (2023). Recent Advances and Challenges of the Application of Artificial Intelligence to Predict Wellbore Instabilities during Drilling Operations. Spe Drilling & Completion, 1-18. doi: 10.2118/215830-pa

[14] Michael, Mendez., Ramadan, Ahmed., Hamidreza, Karami., Mustafa, S., Nasser., Ibnelwaleed, A., Hussein., Sergio, Javier, Ramirez, Garcia., Andres, Gonzalez. (2023). Applications of Machine Learning Methods to Predict Hole Cleaning in Horizontal and Highly Deviated Wells. doi: 10.2118/212912-ms

[15] Robello, Samuel., K., Kumar. (2023). Artificial Well Engineering Intelligence (AweI): Is It Drilling Engineer's Dream or Driller's Nightmare?. doi: 10.2118/213686-ms