PROCEEDINGS, 49th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 12-14, 2024
SGP-TR-227

Machine Learning-Based Rock Facies Prediction Using Geothermal Data: A Comparative Analysis of Algorithms

Zeming Hu, Cesar Vivas, Salehi Salehi, Orkhan Khankishiyev

Hot Lab, 1101 Lexington Ave, Norman, OK 73069, The university of Oklahoma

zeming.hu-1@ou.edu, cesar.vivas@ou.edu, salehi@ou.edu, orkhan@ou.edu

Keywords: *Rock Facies, Geothermal Exploration, Machine Learning, well logs*

ABSTRACT

An approach to predicting rock facies has been generated. Machine learning techniques have generated a novel approach to predicting rock facies. The objective was to develop a reliable facies predictor capable of categorizing rocks into 5 distinct facies: Sand, sandy shale, shady sand, volcanic sand, and shale. To achieve this, electric logs, including gamma-ray, resistivity, and density, were utilized to generate a synthetic facies log that could accurately represent the true geological facies. The study evaluated the performance of four machine learning algorithms: k-nearest Neighbors (KNNs), Random Forest (RF), Decision Tree, and Stochastic gradient descent (SGD). These algorithms were employed to build predictive models using the Utah FORGE geothermal project data as input. The goal was to determine which algorithm performed better in accurately predicting rock facies based on the electric logs. The study results indicate that the developed facies prediction model successfully generated a synthetic facies log that closely matched human-defined facies logs.

Moreover, the comparative analysis revealed insights into the strengths and weaknesses of the various machine learning algorithms in the context of rock facies prediction. This research has significant implications for geology and geothermal energy exploration, as it offers a data-driven approach to enhance our understanding of subsurface rock formations. By leveraging machine learning, we can improve the accuracy and efficiency of facies prediction, ultimately aiding geothermal resource assessment and development.

1. Introduction

Geothermal energy has the potential to play an essential role in the global transition to renewable and sustainable energy sources. By tapping into the Earth's natural heat via underground reservoirs, geothermal power generation emits no greenhouse gases and operates continuously regardless of weather conditions or time of day. Despite the known advantages, geothermal energy comprises only 0.24% of U.S. energy consumption, with petroleum and natural gas dominating at 68% (Khankishiyev et al 2023).

As we venture into the era of advanced exploration and production techniques, geothermal resources worldwide are being unlocked, revealing a renewable resource with untapped potential for reliable, baseload energy. The traditional approach of analyzing lithology logs from well drilling has served us well in characterizing subsurface formations. Nevertheless, the advent of machine learning techniques promises a revolution in classifying and interpreting these logs. These cutting-edge computational methods hold the potential to significantly enhance the accuracy and efficiency of geothermal reservoir characterization, thereby optimizing resource development. This innovative approach could be the key to overcoming the limitations imposed on conventional data collection by harsh downhole environments.

Machine learning has already proven its worth in several research studies that utilized surface drilling data and well logs to predict rock properties in geothermal wells. For instance, Ishitsuka et al. (2021) successfully employed neural networks to predict deep temperature distribution based on resistivity data. Similarly, Vivas and Salehi (2021) presented a machine learning approach to estimate thermal conductivity using surface drilling parameters. Hu et al. (2021) took it a step further by using machine learning and computer vision techniques to predict the thermal diffusivity of subsurface rocks. Shahdi et al. (2021) explored the use of machine learning methods to predict subsurface temperature and geothermal gradient in the Northeastern United States using oil and gas well logs data and Khankishiyev et al. (2024) proposed machine learning algorithms to predict undesirable events happening during geothermal production. These practical applications of machine learning techniques underscore their potential in revolutionizing geothermal reservoir exploration.

Machine learning algorithms such as k-Nearest Neighbors (KNN), Decision Trees (DT), Random Forests (RF), and Stochastic Gradient Descent (SGD) have emerged as powerful tools for lithology classification based on well logs. Each of these algorithms brings unique advantages to the table. KNN, for instance, leverages the inherent patterns and relationships present in well log data to autonomously categorize lithological formations, reducing the reliance on manual interpretation and subjective analysis. DT and RF, on the other hand, provide interpretable insights into lithology classifications and enhance accuracy by aggregating the predictions of multiple trees. SGD optimizes lithology classification through iterative optimization, adjusting model parameters to minimize classification errors. Collectively, these algorithms represent a significant advancement in geothermal reservoir exploration.

The KNN algorithm, a non-parametric and instance-based approach, identifies lithological patterns by comparing the similarities between data points in the feature space (Guo et al., 2003). DF utilizes a tree-like model to partition data based on key features, providing interpretable insights into lithology classifications (Song and Ying 2015). RF, an ensemble method of Decision Trees, further

enhances accuracy by aggregating the predictions of multiple trees (Neale and Kremer 2011). SGD optimizes lithology classification through iterative optimization, adjusting model parameters to minimize classification errors (Rajkumar and Agarwal 2012).

Applying machine learning methods, such as KNN, DT, RF, and SGD, in geothermal reservoir exploration is significant due to their ability to efficiently analyze vast amounts of well-log data. These algorithms can discern intricate lithological patterns, identify anomalies, and predict subsurface compositions with precision that traditional methods may need help to achieve. This streamlines the exploration process and enhances the understanding of geological complexities within geothermal reservoirs.

2. METHODLOGY

The workflow is shown below. Our method started at collecting well log, the preprocessing data, such as rock facies re-labeling. Then we processed the data for generating model. The optimal hyperparameters are essential for the model, so we will select the hyperparameters. Then we will split and train data with 7:3 ratio. Then we will compare the results from different models.

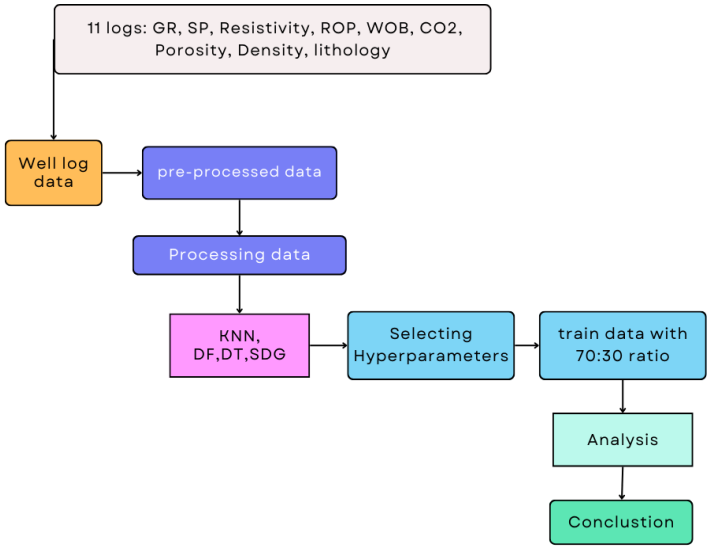

Figure 1: The Workflow

2.1 Data Collection

We utilized the well logs and lithology logs of well 21-31 from Utah forge geothermal project Location of the 21-31 to learn, generate, evaluate a lithology log in Figure 2. The depth of well logs is from 130 – 6180 ft.

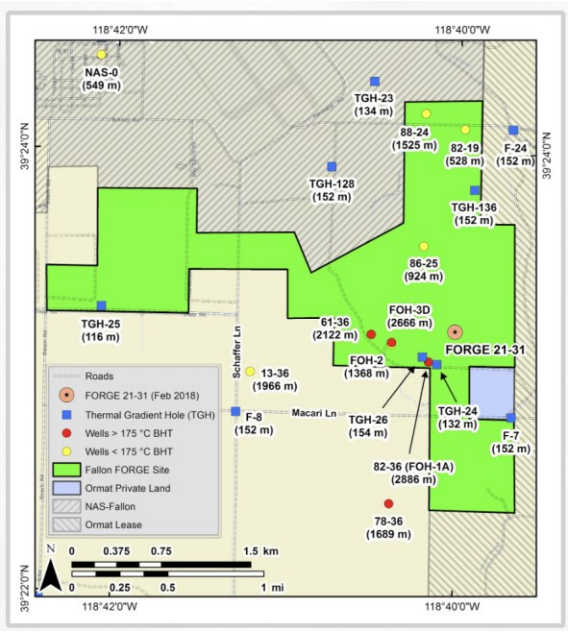

Figure 2: The Location of Well 21-31 in Red Box (Blankenship et al 2016)

All the information collected from the FORGE project is public and available on the Geothermal Data Repository website (https://gdr.openei.org). For this specific study, the data analyzed was collected by Sandia National Laboratories (2020).

2.2 Data Pre-Processing and Cleaning

The dataset contains 14 variables and 11 rock classifications. Table 1 shows the summary of the dataset. The total count for each varibles are 5668.

Table 1: the summary of the data with statistics analysis

	DEPTH	Lable	ReLable	SP	GR	NPHI	RHOB	DPHI	SPHI	HMIN	HMNO	ROP	WOB	CO2
count	5668.000000	5668.000000	5668.000000	5668.000000	5668.000000	5668.000000	5668.000000	5668.000000	5668.000000	5668.000000	5668.000000	5668.000000	5668.000000	5668.000000
mean	3135.750882	3.939308	3.275406	-654.605912	44.688051	0.370088	2.353913	0.179447	-0.534044	-1.697423	-2.000047	94.838215	10.415667	566.010762
std	1638.252982	2.666417	1.875190	45.820234	14.834046	0.178386	0.356702	0.216183	26.543237	92.929264	92.785401	109.860512	8.273990	232.757679
min	301.000000	1.000000	1.000000	-694.640600	18.031000	0.091200	1.435400	-1.199300	-999.250000	-999.250000	-999.250000	1.000000	-4.000000	181.000000
25%	1717.750000	1.000000	1.000000	-673.853400	35.514450	0.230300	2.066275	0.023775	0.062350	1.241675	1.115175	28.000000	2.000000	407.000000
50%	3134.500000	5.000000	5.000000	-667.259600	40.416500	0.324000	2.442500	0.125750	0.149850	3.134850	3.059350	49.000000	9.000000	575.000000
75%	4551.250000	6.000000	5.000000	-663.044600	47.977200	0.473500	2.610800	0.353725	0.259875	6.608475	6.670075	122.000000	18.000000	710.000000
max	5978.000000	11.000000	5.000000	-479.899000	158.114200	0.981600	4.628800	0.736100	0.431600	126.705800	114.293600	881.000000	30.000000	1519.000000

Rock types, showing in Fiugure 3 contains sand, clay, clayst, sandst, volcsedi, andesite, tuffcrys,basalt,rhyolite,dacite, and tuffash. In order to keep data balanced, the rock types have been shrunk into 5 classes from 12 classes, which are the clay, sand, clays, sands, and vouched. They are stored in ReLable varible. The count of rock types in the ReLable variable is shown in Figure 2. Clay (1) has 1239, sand (2) has 64, Clasyst (3) has 228, stanstone (4) is 63, and the volcanic sandstone (5) is 2883.

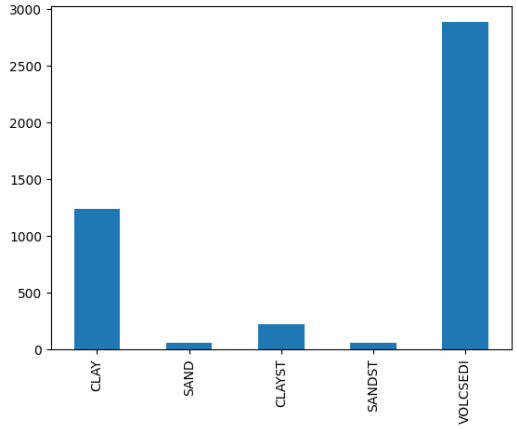

Figure 3: The Distribution of rock type in ReLable

Table 2 shows the description of well logs that applied for the prediction.

Well logs	Description	Unit
GR	Gramma Ray	API
NPHI	Netroun porosity	%
SP	Spontaneous potential logging	mV
RHOB	Density	gcc
DPHI	Density Porosity	%
HMIN	HIP Micro Inverse Resistivity	OHMM
HMNO	HIP Micro Normal Resistivity	OHMM
ROP	Rate of Penetration	ft/hr
CO2	CO2	N/A

WOB	Weight of bit	klb
Lable	Lables of rock types	N/A
Relable	Adjusted lables of rock types	N/A

The statistical technique employed in this study to assess the relationship between two variables was the Pearson correlation coefficient, as described by Benesty et al. (2008). This coefficient quantifies the strength of a linear correlation between the variables under investigation. Coefficients closer to 1 indicate a strong positive correlation, coefficients closer to -1 indicate a strong negative correlation, and coefficients closer to zero suggest no significant correlation between the variables. Then, by checking the impacts of variables on the rock types, SPHI does not have a high positive or negative impact. Also, the rest of the variables are below 30% in the impact factor. Therefore, only variables higher than 20 and standard well logs are selected to predict the rock lithology log, as shown in Figure 4. So, only Sonic porosity, HMIN, and HMNO will not be on the list to predict the lithology. Also, CO2 will not be on the list due to the lack of knowledge of CO2.

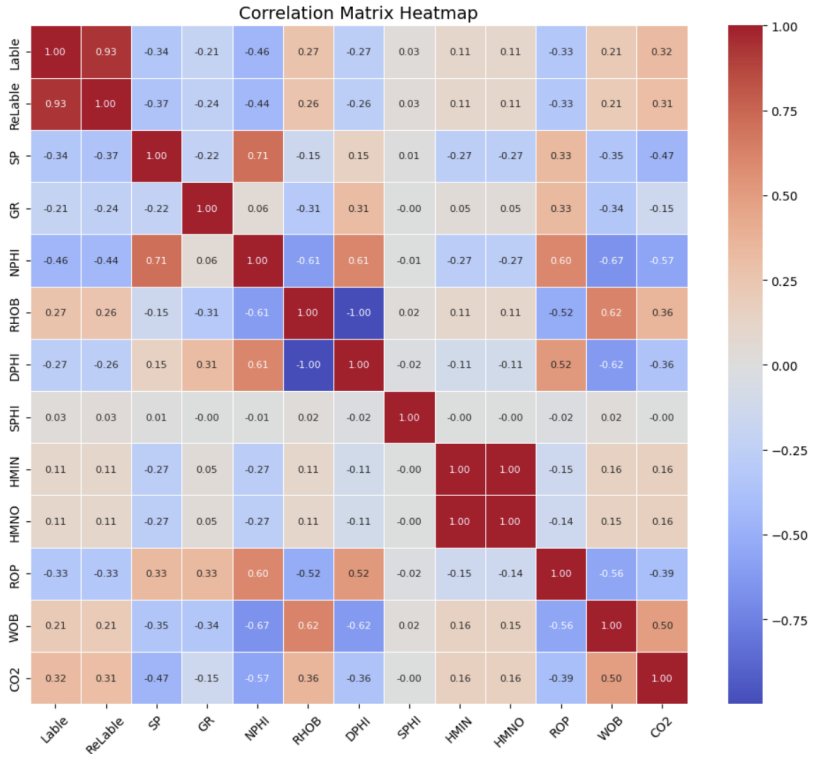

Figure 4: Heat map showing the correlation of the different variables.

Then, the outlier detection has been operated, and found that the values are in the appropriate range, showing in figure 5.

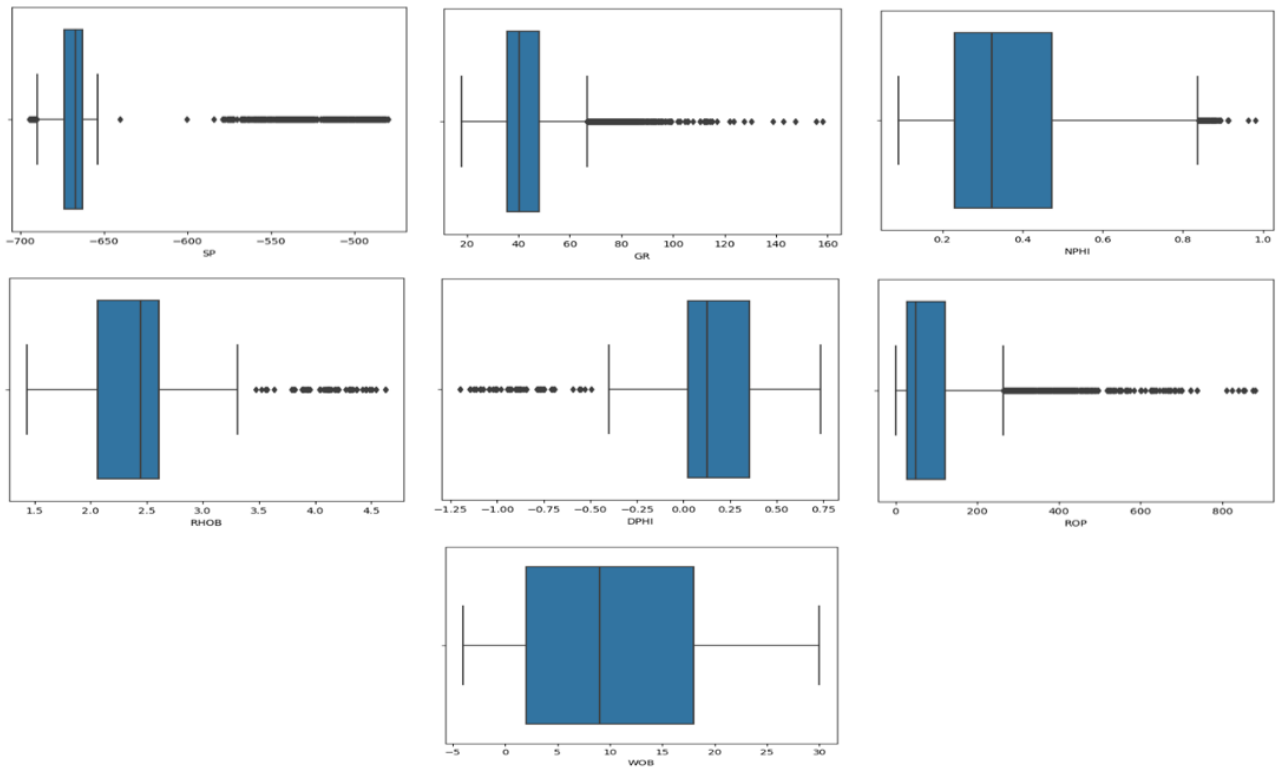

Figure 5: The boxplot for showing variables.

The better way to understand the relationship between two variables is to plot as 2D graph. In general view, most of them do not have a dominant trend, expect the relationship between DPHI and RHOB, the relationship between DPHI and NPHI. They are in an inverse proportional relationship.

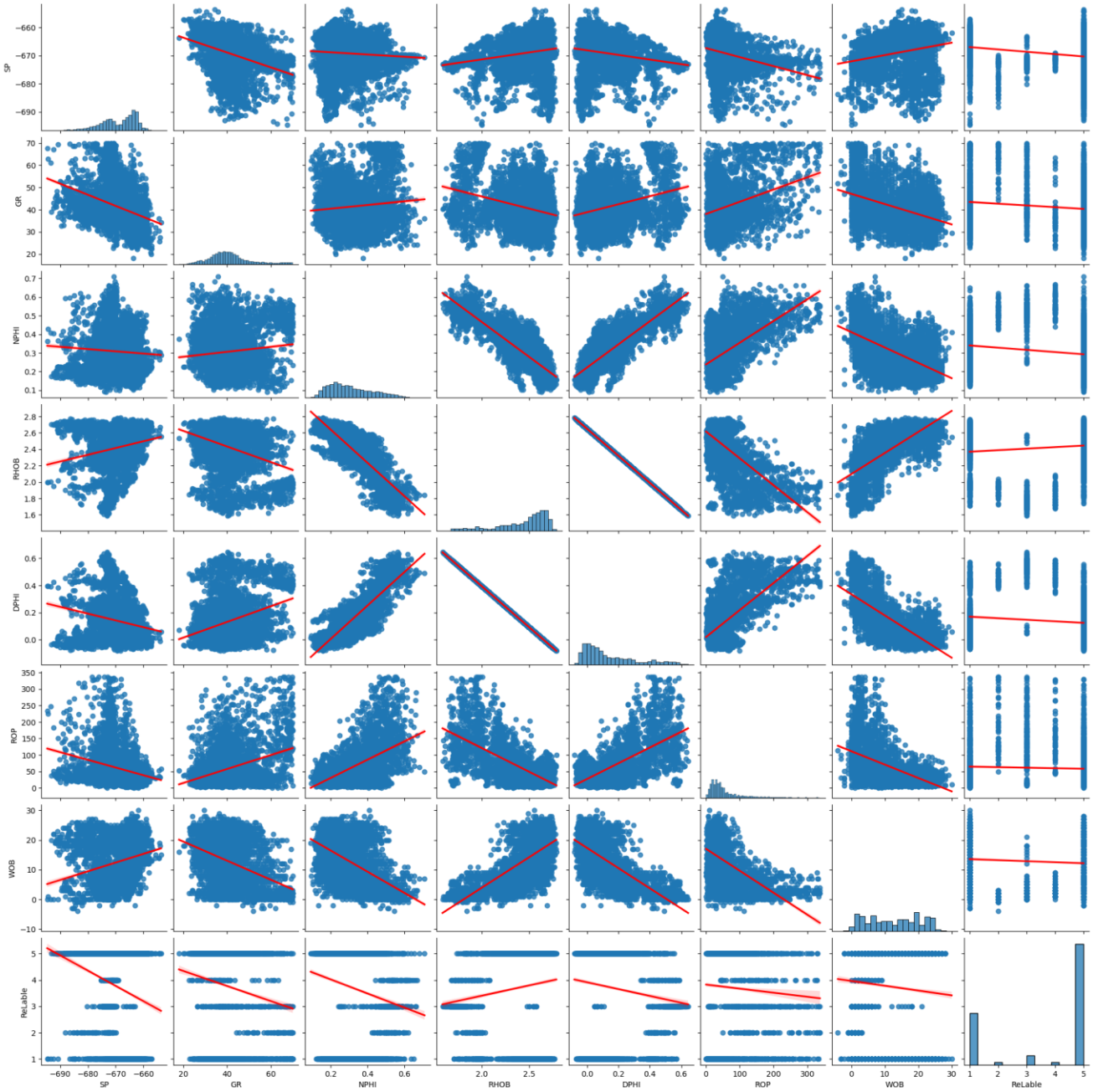

Figure 6. The relationship among the rock type, and well logs

Upon identifying influential factors, the subsequent step involves determining optimal hyperparameters for each model. Notably, the KNN model assumes prominence in this context, as selecting the optimal K value is crucial for achieving higher accuracy. We used accuracy and prediction with different K methods to find optimal K. The graphical representation of accuracy across different K values reveals a noteworthy trend – as K increases, there is a corresponding decrease in model accuracy. This observation underscores the importance of cautiously selecting K to maximize the model's predictive capabilities.

Further analysis involves plotting predictions across different K values for each rock type, facilitating the identification of the K value that yields superior accuracy. Upon examination of the plotted data, it becomes evident that at K = 3, each rock type exhibits a relatively high accuracy, as illustrated in Figure 7. This specific K value is a favorable choice for the KNN model in this lithology classification task. The visual representation of accuracy across various K values and rock types is a valuable tool in refining the hyperparameter selection process, ultimately enhancing the overall performance and reliability of the classification models.

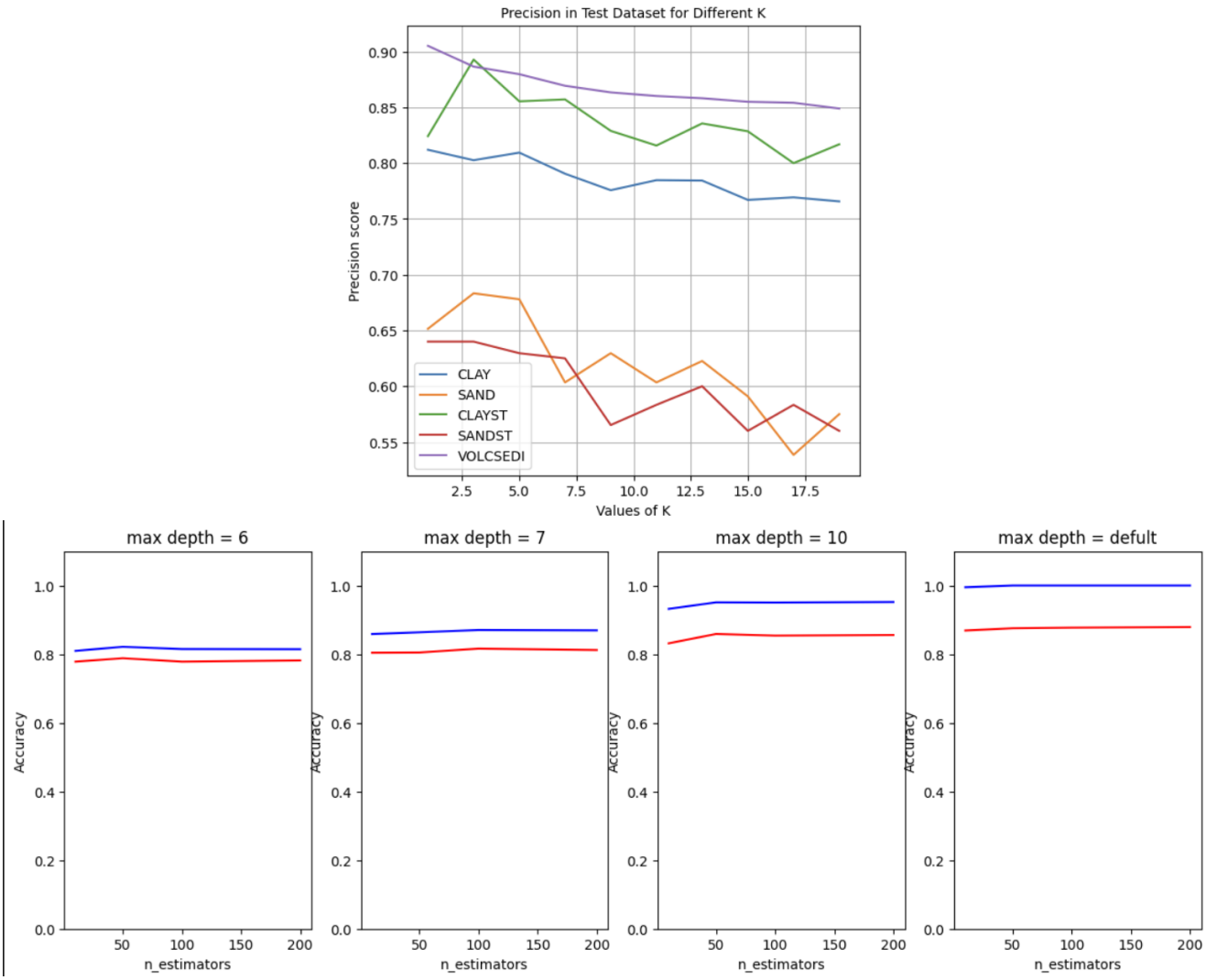

Figure 7. The selection of Optimal K and topmial depth for RF

In addition, determining the optimal depth for the Random Forest model is crucial. Through an accuracy plot varying the depth parameter, it was observed that the default maximum depth provided the highest accuracy. Moreover, configuring the model with 200 estimators contributed to optimal performance. The plot's trend indicates that default depth settings are most favorable for this lithology classification task. The use of 200 estimators enhances the model's robustness. Table 2 also includes the rest of hyperparameters in the each model.

Table 2. The hypermeters in each model

Model	Hypermeters
KNN	K = 3
SGD	Loss = hinge; penalty=l2; max_iter=5
Random forest	Maximum depth = default; n_estimator = 200
Decision tree	Min_sample_leaf = 1; Min_sample_split = 2;gini

Then 70% of data was trained using 4 models and 30% of data was tested based on these models. After that, the performance of the model was compared.

3 ANALYSIS/EVALUATION OF CLASSIFICATION METHODS

3.1 KNN Performance

In Figure 7 (left plot), the provided classification report presents a detailed analysis of the model's performance across multiple classes. Precision measures the accuracy of optimistic predictions; recall assesses the model's ability to capture all positive instances, and the F1-score balances precision and recall. The model's overall accuracy is 0.85, indicating a solid classification performance across all classes. The macro-average F1-score (0.75) and weighted average F1-score (0.85) suggest a generally balanced and accurate model. However, the discrepancies in precision and recall for specific classes, notably sand, and sandstone, could be higher due to the small dataset. The plot on Figure 7 (right), which is the confusion matrix plot, shows that the prediction for each rock type has a high accuracy of the value of each rock type.

	precision	recall	f1-score	support
1	0.81	0.82	0.81	614
2	0.65	0.62	0.64	69
3	0.82	0.79	0.81	95
4	0.64	0.57	0.60	28
5	0.91	0.91	0.91	879
accuracy			0.85	1685
macro avg	0.77	0.74	0.75	1685
weighted avg	0.85	0.85	0.85	1685

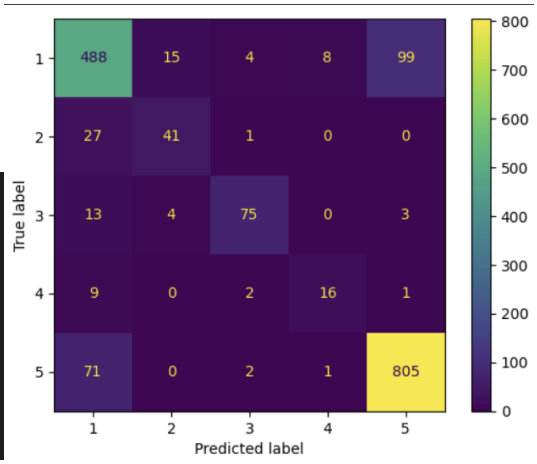

Figure 7. The performance of the KNN model

3.2 RF Performance

In Figure 8, the model demonstrates overall solid performance with an accuracy of 0.86 across five classes. Class 5 shows excellent precision, recall, and F1-score (0.89, 0.94, and 0.92), indicating high accuracy in identifying instances of this class. Class 1 also exhibits good precision, recall, and F1-score (0.83, 0.82, and 0.83). However, there are notable variations in performance across other classes. Class 2 has a high precision (0.88) but a lower recall (0.52), resulting in an F1-score of 0.65. Class 3 shows balanced precision and recall (0.79 and 0.77) with an F1 score of 0.78. Class 4 has a lower precision (0.68) and recall (0.46), contributing to an F1-score of 0.55. The macro-average F1-score (0.75) and weighted average F1-score (0.86) suggest a generally balanced model performance. However, the lower recall in class 2 and class 4 indicates potential challenges in correctly identifying instances of these classes, warranting further investigation and potential model refinement for enhanced performance in those specific categories. The confusion matrix also shows this fact as well.

	precision	recall	f1-score	support
1	0.83	0.82	0.83	614
2	0.88	0.52	0.65	69
3	0.79	0.77	0.78	95
4	0.68	0.46	0.55	28
5	0.89	0.94	0.92	879
accuracy			0.86	1685
macro avg	0.82	0.70	0.75	1685
weighted avg	0.86	0.86	0.86	1685

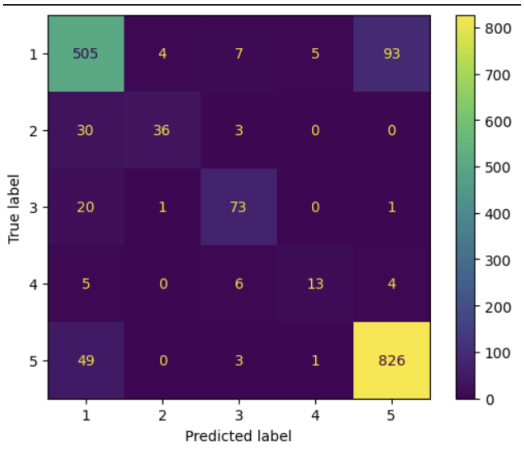

Figure 8. The performance of the RF model.

3.3 DT Performance

Figure 9 (left plot) outlines the model's performance across five classes, achieving an overall accuracy of 0.82. Class 5 has high precision, recall, and F1-score (0.86, 0.91, and 0.89), indicating its effective identification. Class 1 exhibits a slightly lower precision (0.80) and recall (0.72), yielding an F1-score of 0.76. Class 2 demonstrates moderate precision (0.61) and recall (0.59), resulting in an

F1 score of 0.60. Class 3 shows higher recall (0.82) but lower precision (0.67), contributing to an F1-score of 0.74. Class 4 presents balanced precision and recall (0.59 and 0.57) with an F1 score of 0.58. The macro-average F1-score (0.71) and weighted average F1-score (0.81) reflect a reasonably balanced model performance, with potential for improvement in specific classes through further analysis and refinement.

	precision	recall	f1-score	support
1	0.80	0.72	0.76	614
2	0.61	0.59	0.60	69
3	0.67	0.82	0.74	95
4	0.59	0.57	0.58	28
5	0.86	0.91	0.89	879
accuracy			0.82	1685
macro avg	0.71	0.72	0.71	1685
weighted avg	0.82	0.82	0.81	1685

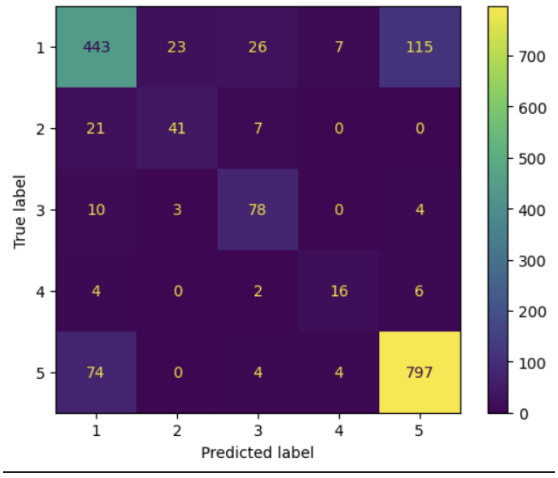

Figure 9. The performance of DT

3.4 SGD Performance

In Figure 10, this model achieved an overall accuracy of 0.71. Class 5 stands out with solid precision, recall, and F1-score (0.82, 0.79, and 0.80), indicating accurate identification within this class. However, challenges are apparent in other classes, particularly in class 2, where precision (0.20) and recall (0.01) are minimal, resulting in a low F1-score of 0.03. Class 4 shows negligible precision, recall, and F1-score, indicating limitations in identifying instances of this class. In the confusion matrix, it shows a similar result.

	precision	recall	f1-score	support
1	0.64	0.73	0.68	614
2	0.20	0.01	0.03	69
3	0.41	0.58	0.48	95
4	0.00	0.00	0.00	28
5	0.82	0.79	0.80	879
accuracy			0.71	1685
macro avg	0.41	0.42	0.40	1685
weighted avg	0.69	0.71	0.70	1685

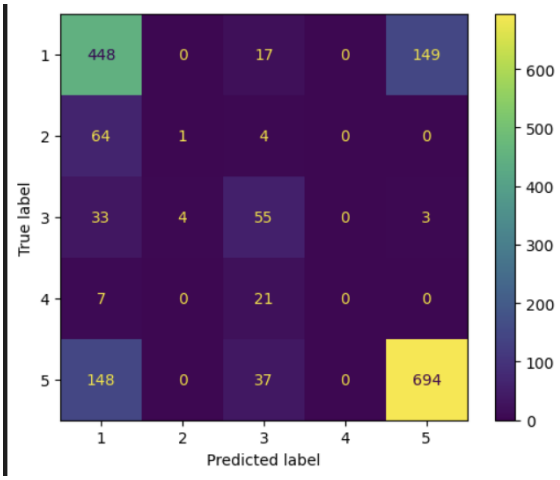

Figure 10, The performance of the SGD

Combined Figure 11 and Figure 12, the model, KNN, SGD, RF, and DT, reveals distinct strengths and weaknesses across various classes. KNN and RF emerge as robust performers, exhibiting high precision for dominant lithological classes such as VOLCSEDI and CLAYST. KNN achieves a balanced accuracy of 0.8 and a prediction accuracy of 0.8, showcasing its reliability. On the other hand, SGD encounters challenges, displaying low precision across all classes and resulting in the lowest accuracy of 0.7 and prediction accuracy of 0.4. With impressive precision for VOLCSEDI (0.94198) and overall accuracy of 0.4, RF stands out as a promising model. However, both SGD and DT models share similar accuracy metrics, suggesting potential areas for improvement, particularly in precision. This comprehensive analysis offers valuable insights into each model's performance, guiding recommendations for further refinement and optimization, such as tuning hyperparameters and evaluating the importance of features to enhance lithology classification accuracy.

	KNN	SGD	RF	DT
CLAY	0.781759	0.521173	0.820847	0.721498
SAND	0.594203	0.057971	0.521739	0.594203
CLAYST	0.778947	0.136842	0.778947	0.821053
SANDST	0.642857	0.071429	0.535714	0.571429
VOLCSEDI	0.922639	0.94198	0.94198	0.906712
Accuracy	0.8	0.7	0.4	0.4
Prediction	0.8	0.4	0.82	0.4

Figure 11. The Summary of the each model with different rock type.

RF emerges as the optimal model for lithology log prediction, primarily attributed to its elevated accuracy. Despite its efficacy, certain limitations warrant consideration. One significant challenge arises from the inherent imbalance in rock types, potentially leading the model to misinterpret certain rock types, akin to the issues encountered with SGD. Additionally, the model's reliance on a solitary well for training introduces limitations associated with dataset scarcity. An ideal scenario involves utilizing multiple wells for training and validation, ensuring a more representative dataset. Furthermore, incorporating a separate well for blind testing is crucial to assess the model's predictive prowess in real-world scenarios accurately. This comprehensive approach, encompassing balanced rock types and a diverse dataset, is pivotal for refining and validating lithology prediction models with enhanced reliability and generalization capabilities.

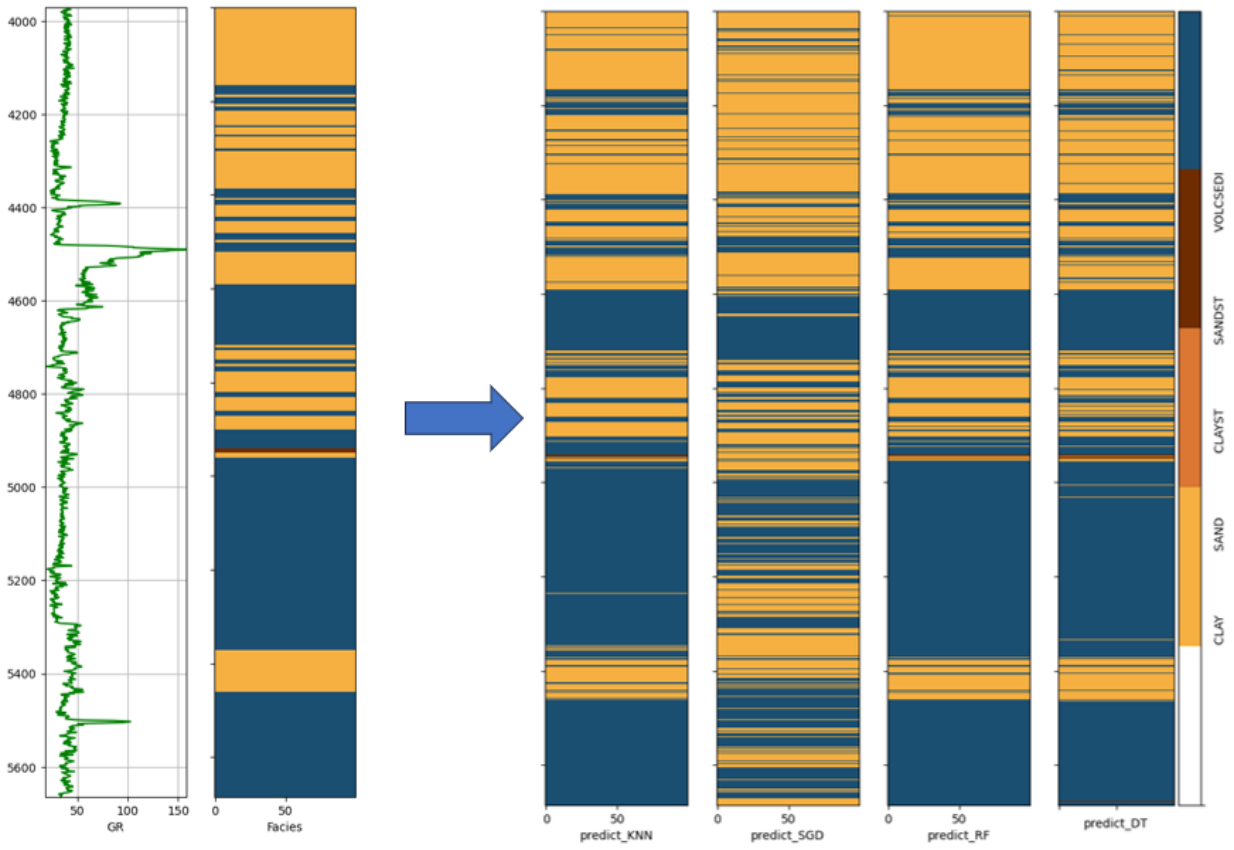

Figure 12. Comparison of prediced lithology log with different models.

4. CONCLUSION

Our study employed four algorithms to train lithology log prediction models, ultimately selecting KNN as the most effective. Despite this success, limitations like data imbalance and dataset size persist. Future studies will prioritize addressing these issues to enhance

model robustness. Balancing data representation, particularly for less prevalent lithological classes, is crucial to improving predictive accuracy. Additionally, recognizing the importance of dataset size, our strategy involves expanding with more drilling logs and diverse well-log types. This enrichment aims to provide a more nuanced understanding of geological variations, ensuring our models excel across broader scenarios. Our further study lies in refining and expanding the dataset continuously. By mitigating biases, incorporating additional information, and leveraging diverse features, we aim to elevate the accuracy and applicability of our lithology prediction models. This iterative approach ensures adaptability to the complexities of various geological formations and drilling conditions, advancing our capacity to deliver accurate predictions in geothermal well applications

REFERENCES

Benesty, J., Chen, J., Huang, Y., Cohen, I., 2008. Pearson's Correlation Coefficient, in: Noise Reduction in Speech Processing. Springer Netherlands, Dordrecht, pp. 1090–1091. https://doi.org/10.1007/978-1-4020-5614-7_2569.

Blankenship, D.A., Blake, K., Calvin, W., DeOreo, S., Faulds, J.E., Glen, J., Hickman, S., Hinz, N., Kaven, O., Lazaro, M. and McCulloch, J., 2016. Frontier Observatory for Research in Geothermal Energy: Phase 1 Topical Report West Flank of Coso, CA (No. SAND2016-8930). Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).

Guo, G., Wang, H., Bell, D., Bi, Y. and Greer, K., 2003. KNN model-based approach in classification. In On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings (pp. 986-996). Springer Berlin Heidelberg.

Hu, Z., Vivas, C., and Salehi, S. 2023. Computer Vision in Predicting Thermal Diffusivity of Subsurface Rocks. Geothermal Resources Council Transactions 47, 2299–2308.

Ishitsuka, K.; Kobayashi, Y.; Watanabe, N.; Yamaya, Y.; Bjarkason, E.; Suzuki, A.; Mogi, T.; Asanuma, H.; Kajiwara, T.; Sugimoto, T. 2021. Bayesian and neural network approaches to estimate deep temperature distribution for assessing a supercritical geothermal system: Evaluation using a numerical model. Nat. Resour. Res. 2021, 30, 3289–3314.

Khankishiyev, O., Salehi, S., Vivas, C., Nygaard, R., and Rehg, D. Techno-Economic Investigation of Geothermal Development in Sedimentary Basins ASME Open J. Engineering ASME. January 2023, DOI: https://doi.org/10.1115/1.4062412

Khankishiyev, O., Salehi, S., Mammadzada, V. Identification of Undesirable Events in Geothermal Fluid/Steam Production using Machine Learning. In Proceedings of the 49th Workshop on Geothermal Reservoir Engineering Standford University, Standford, CA, USA, February 2024

Neale, D.B. and Kremer, A., 2011. Forest tree genomics: growing resources and applications. Nature Reviews Genetics, 12(2), pp.111-122.

Rajkumar, A. and Agarwal, S., 2012, March. A differentially private stochastic gradient descent algorithm for multiparty classification. In Artificial Intelligence and Statistics (pp. 933-941). PMLR.

Sandia National Laboratories. 2020. Fallon, NV FORGE well 21-31 Lithology, Mineral, Image Log, and Injection Test data [data set]. Retrieved from https://dx.doi.org/10.15121/1632115.

Shahdi, A., Lee, S., and Karpatne, A. 2021. Exploratory analysis of machine learning methods in predicting subsurface temperature and geothermal gradient of Northeastern United States. Geotherm Energy 9, 18. https://doi.org/10.1186/s40517-021-00200-4

Song, Y.Y. and Ying, L.U., 2015. Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry, 27(2), p.130.

Vivas, C., Salehi, S., Tuttle, J.D., Rickard, B., 2020. Challenges and Opportunities of Geothermal Drilling for Renewable Energy Generation. Geothermal Resources Council Transactions 44, 904–918.