JOURNAL OF PETROLEUM GEOMECHANICS
(JPG)

Extended Abstract

Application of Advanced Machine Learning Models for Lithological Facies Prediction in a Southern Iranian Oilfield

Hamid Ghalibaf Mohammad abadi[1] , Naser Hafezi Moghaddas*[2] , Elham Mahdipour[3], Mojtaba Hydarizad[4] Hossein Talebi[5]

1-Department of AL and Robotics, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
2-Department of Geology, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran
3-Computer Engineering Department, Khavaran Institute of Higher Education, Mashhad, Iran
4-State Key Laboratory of Marine Geology, Tongji University, Shanghai, 200092, China
5-Department of Geological Survey & Development, National Iranian South Oil Company, Iran

Received: 13 October 2023; Accepted: 11 November 2023

DOI: 10.22107/JPG.2023.420597.1213

Keywords

lithological facies, Extra Trees Classifier, F1-Score, Gradient Boosting Classifier, Random Forest Classifier, Confusion Matrix, ROC Plots

Abstract

This study aims to employ supervised Advanced machine learning for the classification of lithological facies from geophysical log data in wells without drilling core samples. For this purpose, a dataset from seven wells in a training set from one of the oil fields in southern Iran has been utilized. This dataset includes natural gamma ray (SGR), corrected gamma ray (CGR), bulk density (RHOB), neutron porosity (NPHI), compressional wave slowness (DTSM), and shear wave slowness (DTCO), which directly influence the classification of geomechanical facies. These parameters are employed as independent variables, while lithological facies serve as the dependent variable for classification. This dataset pertains to depths ranging from 3000 to 4000 meters in the Ilam and Sarvak fractured limestone formations (Bangestan Limestone) of the subsurface. As the title suggests in this article, Initially, through artificial intelligence clustering methods and laboratory studies, these formations were categorized into five distinct lithological facies After this stage, eight supervised machine learning methods were employed, including Logistic Regression, K Neighbors Classifier, Decision Tree Classifier, Random Forest Classifier, Gaussian NB, Gradient Boosting Classifier, Extra Trees Classifier, and Support Vector Machine (SVM), to predict lithological facies in wells without existing classifications. The dataset of these wells underwent training and testing stages with each of these algorithms to construct an appropriate model. As a result, facies labels were predicted. The performance of the models was evaluated using multiple metrics including Accuracy, Precision, F1-Score, and Recall through confusion matrices and ROC curves. The Extra Trees Classifier, Gradient Boosting Classifier, and K Neighbors Classifier showed superior results among these methods. Finally, the model's performance in predicting lithological features of wells outside the training set or unseen wells is presented.

1. Introduction

In recent years, the use of machine learning methods for the classification of geomechanical units or lithofacies using well log data has seen significant research advancements [1-2]. Many researchers have proposed solutions using artificial intelligence algorithms to estimate unknown parameters such as lithological facies[3].

2. Methodology

n this study, a wide spectrum of advanced artificial intelligence methods, including Logistic Regression, K Nearest Neighbors, Decision Tree, Random Forest, Gaussian NB, Gradient Boosting, Extra Trees, and SVM, has been employed for the prediction of unclassified lithological facies in wells. For this purpose, geophysical data and well log data from 7 training wells in one of the oil fields in southern Iran, consisting of Natural

* **Corresponding Author:** nhafezi@um.ac.ir

Gamma Ray (SGR), Corrected Gamma Ray (CGR), Density (RHOB), Neutron Porosity (NPHI), Shear Sonic Slowness (DTSM), and Longitudinal Sonic Slowness (DTCO), directly influencing the determination of geomechanical facies, were used as independent variables. The lithological facies were classified as dependent variables, separated into five categories through artificial intelligence classification methods and laboratory studies, and then applied to model wells. Given that these lithological units are characterized based on parameters such as porosity, sonic wave slowness, and density, it is unreasonable to assign specific names, as each rock may exhibit these properties. Therefore, these lithofacies have been labeled with numerical suffixes. The dataset covers depths from 3000 to 4000 meters corresponding to the Ilam and Sarvak formations (Bangestan Limestone).
In the first step of model construction, data preparation was carried out, including data visualization, feature engineering, handling missing values, and extracting important features, which were crucial steps in data preparation. The second important stage of this research involves building and validating a model. A baseline model was selected, and hyperparameters were tuned for efficient model performance. In this study, a grid search was employed to find optimal parameters. Finally, model evaluation, the most critical task in ML model development was conducted, and lithological facies labels were predicted using test data. The performance of the models was evaluated using various metrics, including Accuracy, Precision, F1-SCORE, and Recall through confusion matrices and ROC curves. Among these methods, the Extra Trees Classifier, Gradient Boosting Classifier, and K Nearest Neighbors Classifier demonstrated better results (Figure 1). Ultimately, the model's performance in predicting lithological facies for wells outside the model or unseen wells was presented (Figure 2).

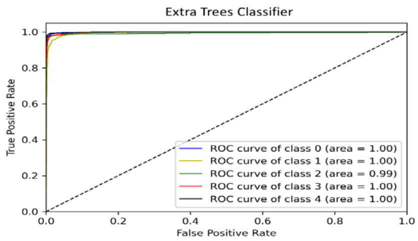

Fig. 1. The Average ROC Curve for the Extra Trees Classifier Algorithm generated during the model testing phase

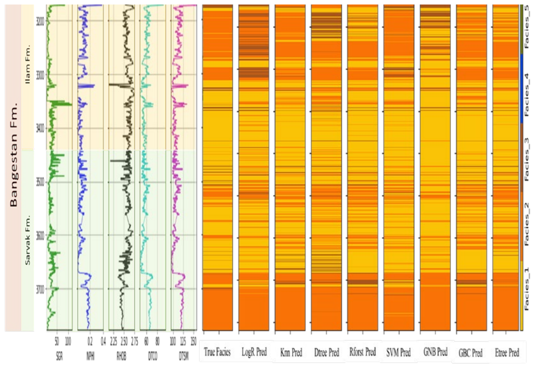

Fig. 2. Column chart of predicted lithofacies by machine learning models for the target well (True Facies). As evident, machine learning algorithms including Extra Trees, Random Forest, and k-Nearest Neighbors provide better results compared to other algorithms in identifying actual lithofacies.

3. Results and Conclusions

In this study, standardized and comprehensive steps were taken to select the best model and hyperparameters for predicting lithofacies in the dataset. Initially, the data was prepared for modeling, models were fitted, and through cross-validation, the model was validated. Predicted lithofacies labels were determined, and the model accuracy was evaluated using multiple metrics, including Accuracy, Precision, F1-Score, Recall, and ROC curves.

4. Acknowledgment

We express our gratitude to the senior researcher, Mr. Mordani, whose precision and expertise have significantly contributed to the improvement of the article, opening new avenues for our research.

5. References

[1] Ghalibaf, H., Hafezi Moghaddas, N., Lashkaripoor, G.R., Raoof G., Hossin T., (2022). Determination of geomechanical zones based on evaluation of Unsupervised Machine Learning algorithm methods “JOURNAL OF PETROLEUM GEOMECHANICS (JPG). (DOI): 10.22107/jpg.2022.329417.1158.

[2] Ghalibaf, H., Hafezi Moghaddas, N., Lashkaripoor, G.R., Raoof G., Hossin T., (2022). Estimation of Geomechanical Parameters, In Situ Stress Measurement Techniques, and Determination of Safe Mud Weight Windows Using Machine Learning Algorithm Methods “JOURNAL OF PETROLEUM GEOMECHANICS (JPG).

[3] Mardani, R., (2020). https://github.com/mardani72/Facies-ClassificationMachine-Learning