Well log and core integration to predict permeability in carbonates: A methodology using machine learning

Filipe de Moura Antonio Cordeiro[1,2], Cícero da Paixão Pereira[2] e Alexsandro G. Cerqueira[1,2] — [1]INCT-GP/UFBA [2]GAIA-UFBA

This paper was prepared for presentation during the 4th Joint SBGf/SEG Workshop on Machine Learning held in Rio de Janeiro, Brazil, 03-04 December 2024.

Abstract

Predicting permeability from geophysical logs is a crucial aspect of reservoir characterization in oil and gas exploration. As Guéguen and Palciauskas (1994) assert, permeability is a fundamental property that dictates the fluid transmission capacity of a reservoir. Traditional methods for estimating permeability, such as Nuclear Magnetic Resonance (NMR) logging and laboratory petrophysical analyses, are hampered by high costs and operational constraints. In response to these limitations, this study introduces a novel and cost-effective methodology that harnesses geophysical logs and machine learning (ML) techniques to predict permeability, with a specific focus on the Barra Velha Formation in the Santos Basin.

Both parametric and non-parametric ML algorithms, such as K-Nearest Neighbors (KNN), Random Forest, XGBoost, and Support Vector Regressor (SVR), were selected to evaluate how algorithms with different principles and complexities can handle non-linear relationships in permeability prediction. The dataset consisted of geophysical logs, including gamma rays, sonic slowness, density, photoelectric factor, porosity, resistivity, and permeability data from 12 wells in the Búzios Field. The preprocessing involved removing null values and outliers and correlating core sample depths with recorded measurements. Derived parameters, such as porosity calculated from density and sonic slowness, logarithms of resistivity, and their differences, were added to improve model performance.

A key step in the presented methodology was the feature selection process. The importance of each attribute was calculated by training a decision tree and using the DecisionTreeRegressor from the Scikit-learn library. This tool includes a module called feature_importances_, which allows for a numerical understanding of the most important variables for the entire decision tree. By calculating the contribution of each feature to the overall model performance, we prioritized the most relevant geophysical logs for permeability prediction, thereby enhancing the efficiency of the models. The adopted methodology followed several key steps: i) selection of the study area and conversion of files to the Standard Log ASCII (LAS) format, ii) Integration of core sample data with well depths using a one-meter interval and calculating their medians, iii) Parameter selection based on the feature importance analysis from the decision tree, iv) Training the models, v) Analysis of regression metrics to evaluate model performance, and vi) Evaluation of the blind well.

The model evaluation was conducted using the holdout cross-validation method, ensuring accurate estimation of generalization performance for the algorithms. For the KNN model, a neighbor value of 7 was used, while for Random Forest, XGBoost, and SVR, hyperparameter tuning was performed using the BayesSearchCV module from the Scikit-learn library. The Mean Absolute Error (MAE) achieved satisfactory results for the algorithms.

The results underscore the adaptability of the ML algorithms, particularly the non-parametric models like Random Forest and XGBoost, in generating reliable permeability estimates. These models effectively captured the main characteristics of the permeability curve with appropriate MAE values, even when using data from only a few wells. The robustness and efficiency of the methodology were clearly evident, with data from 12 distinct wells providing satisfactory permeability predictions. The potential for further improvement with additional well data further underscores the optimism about the model's performance.