Submission code: MR94M50XBW

See this and other abstracts on our website: https://home.sbgf.org.br/Pages/resumos.php

Machine Learning Applied to Petrophysics: Lithological and Porosity Classification in the Alagoas Basin Using Random Forest

José Nascimento dos Santos (UFPE-PRH 47.1(ANP)/Observatório Nacional), Mário Lima Filho (UFPE-PRH 47.1(ANP))

Machine Learning Applied to Petrophysics: Lithological and Porosity Classification in the Alagoas Basin Using Random Forest

The Alagoas Basin contains a substantial volume of digital petrophysical data derived from well logs in its onshore portion. Integrating this data with advanced computational methods, such as Machine Learning, enables the determination of lithologies and porosity zones in hydrocarbon fields. For this study, supervised machine learning prototypes were developed using the Random Forest algorithm for classification problems. These prototypes were configured with a forest of 100 decision trees, retaining default hyperparameter values while setting random_state = 42 to ensure reproducibility through fixed random number generation. The training data was sourced from the REATE 2020 repository (Program for the Revitalization of Oil and Natural Gas Exploration and Production in Onshore Areas) maintained by Brazil's National Agency of Petroleum, Natural Gas, and Biofuels (ANP). The initial dataset comprised over 1.5 million records from approximately 400 wells. After data analysis, cleaning, and balancing, the final dataset was reduced to 56,000 records representing six wells. Selected input features included GR (Gamma Ray), RT (Resistivity), NPHI (Neutron Porosity), and RHOB (Bulk Density) logs, with a 70-30 split for training and testing, following established literature methodologies. For binary lithological classification (sandstone vs. shale), the model achieved an accuracy exceeding 81%, with GR identified as the most influential feature in tree-based decisions. In the porosity zone determination model, NPHI and RHOB curves were normalized using MinMaxScaling to generate the porosity zone curve, yielding 79% accuracy. Both models were evaluated through confusion matrices, cross-validation (LOO-CV and K-fold), feature importance analysis, error assessment, and benchmarking against other classification algorithms (SVM, KNN, Decision Tree, Logistic Regression, and Random Forest itself). Finally, the models were tested on a previously unused well, achieving 84% accuracy, confirming their reliability and generalization capability. The results demonstrate Random Forest's effectiveness in lithology and porosity classification, advancing petrophysical analysis in the Alagoas Basin, northeastern Brazil.