Well logs Interpretation Using Machine Learning Workflow

Zainab Alali

University of Stanford

Abstract

Objectives/Scope: Well logs are used to measure acoustic, nuclear, or conductive properties of the subsurface. In turn, these properties are interpreted based on physical correlations which require prior knowledge of fluid properties. This paper presents a comprehensive data-driven approach utilizing best machine learning modeling practices and integrating supervised and unsupervised learning methods to predict three essential reservoir properties: water saturation, porosity, and volume of shale, without the prior knowledge of complex physical rock and fluid characteristics.

Methods, Procedures, Process: We utilized a real dataset consists of eight wells with raw measurements of neutron, density, gamma ray, spontaneous potential and resistivity. The interpretation process starts by identifying data anomalies using the isolation forest method. We then applied and compared the performance of various types of machine learning algorithm. The models were trained using the group KFold cross-validation technique. The best model then was benchmarked using mean squared error. To improve training efficiency and speed, hyperparameter tuning was performed using evolutionary grid search with genetic algorithm.

Results, Observations, Conclusions: The cross-validation performance of several machine learning models were evaluated including decision tree, support vector machine with radial-base kernel, adaboost, neural network, gradient boosting and random forest. The models were trained using the group KFold cross-validation technique by sequentially training on six wells and validating the prediction on the seventh well. The data from the eighth well was kept unseen for final evaluation of the modeling performance. The best prediction results were achieved by combining isolation forest to remove anomalies and gradient boosting to predict the three parameters independently: porosity, water saturation and volume of shale. To improve training efficiency and speed, hyperparameter tuning of the gradient boosting parameters was performed using evolutionary grid search with genetic algorithm. The evolutionary search was performed for 10 generations achieving an optimum Root Mean Square Error (RMSE) of 0.03 on the validation set with 800 trees, a maximum depth of 8 and a learning rate of 0.06 with Huber loss. This optimal model was then evaluated on the unseen eighth well producing a RMSE of 0.05 with R2 of 0.8.

Novel/Additive Information: This paper represents a state of art workflow for log prediction using machine learning algorithms eliminating human errors and biases. For ensuring best hyperparameters we added an evolutionary grid search to the workflow to tune the best machine learning model before final testing and evaluation.