Submission code: XDKBKLWN9K

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

Machine Learning-Based Prediction of Shear Wave Travel Time in North Sea Reservoirs: A Case Study from Viking Graben

Rodrigo Rodrigues (UFPA), João Rafael B. Da Silveira (UFPA), Jose Jadsom Sampaio de Figueiredo (UFPA)

Machine Learning-Based Prediction of Shear Wave Travel Time in North Sea Reservoirs: A Case Study from Viking Graben

Abstract Summary

This study explores machine learning techniques for predicting Shear Wave Travel Time (DTS) using data from the Viking Graben region. Six regression models were tested with robust preprocessing and hyperparameter tuning via methods like Genetic Algorithm and Bayesian Optimization. The Random Forest model achieved the best results ($R^2 = 0.9757$), with Bayesian Optimization offering the best trade-off between accuracy and efficiency. The findings highlight the potential of ML pipelines for DTS prediction and similar geophysical applications.

Keywords: Shear Wave Travel Time (DTS); Machine Learning; Regression Models; Hyperparameter Optimization; Random Forest; Bayesian Optimization.

Introduction

Shear Wave Travel Time (DTS) is a key parameter for reservoir characterization and wellbore stability, as it provides elastic properties derived from sonic logs and helps mitigate drilling risks (Olutoki et al., 2024; Rajabi et al., 2023). However, direct DTS acquisition is often limited by high costs and operational constraints.

To address this, Machine Learning (ML) techniques have emerged as effective alternatives for estimating DTS based on available well log data (Ahmed et al., 2022; Liu et al., 2021; Silveira et al., 2023). These models can capture complex, nonlinear relationships and improve predictive accuracy, even with noisy or incomplete datasets.

This study investigates the performance of ML regression algorithms for DTS prediction, emphasizing the role of hyperparameter optimization. It also compares different tuning strategies—such as Genetic Algorithm Optimization, Bayesian methods, and GridSearchCV—in terms of both accuracy and computational efficiency.

Methodology

This study followed a structured workflow comprising data preprocessing, feature selection, model training, hyperparameter optimization, and performance evaluation, as outlined in Figure 1.

Outlier detection was tested using four algorithms: Local Outlier Factor (LOF), Support Vector Machine (SVM), Isolation Forest (IF), and Mahalanobis Distance. Among them, Mahalanobis Distance produced the most consistent and geologically coherent results (Figure 2).

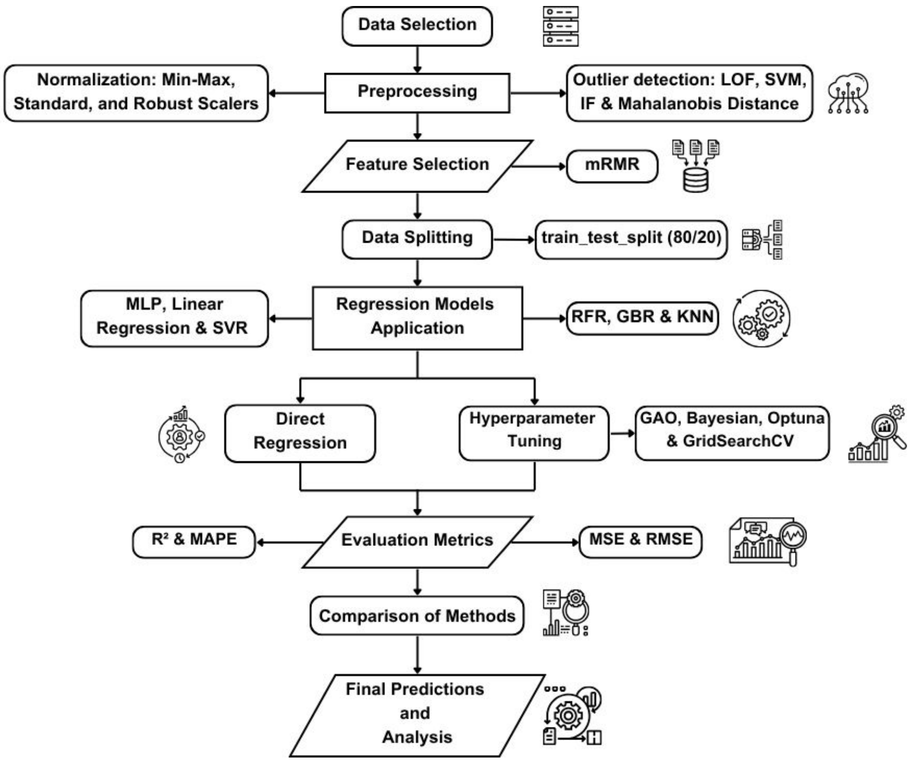

Figure 1: Flowchart of the methodological steps adopted in this study.

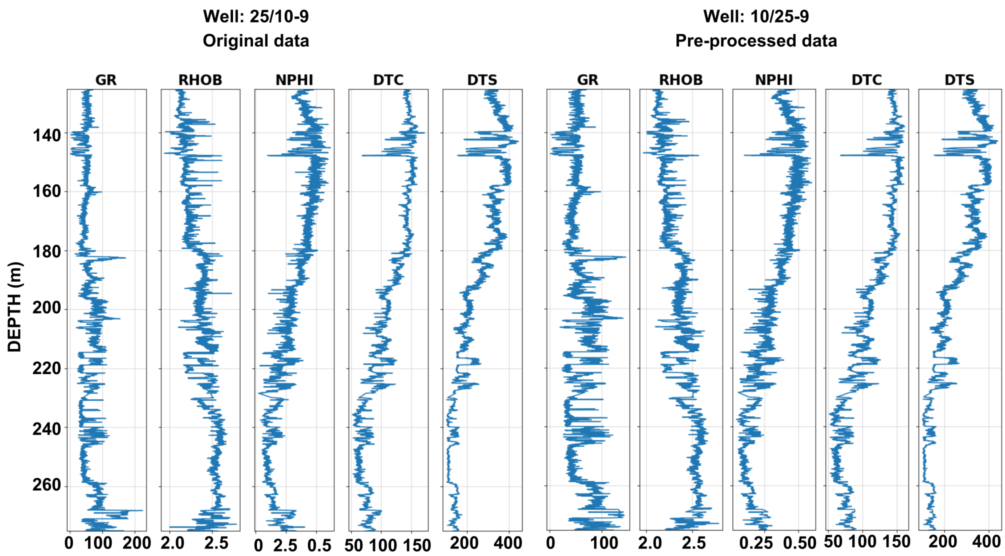

Figure 2: Comparison of the application of outlier detection before and after for well 25_10-9 using Mahalanobis Distance. The GR peaks indicate lithological variations such as shales, clays and radioactive minerals.

Six regression algorithms were evaluated: Linear Regression, K-Nearest Neighbors (KNN), Support Vector Regression (SVR), Multi-Layer Perceptron (MLP), Gradient Boosting Regressor (GBR), and Random Forest Regressor (RFR) (Breiman, 2001; Mrabet et al., 2022; Segal, 2003). Rather than detailing the theoretical basis of each model, the focus was placed on their application to DTS prediction using normalized and filtered well log data.

Model performance was assessed using metrics such as Mean Squared Error (MSE), Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The impact of different hyperparameter optimization techniques—Genetic Algorithm Optimization (GAO), Bayesian Optimization, Optuna, and GridSearchCV—was also systematically compared across the top-performing models.

Results

The performance evaluation focused on three key aspects: predictive accuracy of the models, the impact of hyperparameter tuning, and the effectiveness of different optimization strategies.

Among the models tested, RFR consistently outperformed others, achieving the lowest initial error (MSE = 0.0123; $R^2 = 0.9757$), followed by KNN and GBR. Linear models lagged significantly, reinforcing the superiority of nonlinear approaches for DTS modeling.

Hyperparameter tuning led to measurable improvements across all models. The GAO improved both RFR and GBR performance, while gains for KNN were smaller. Notably, Bayesian Optimization emerged as the most efficient technique overall—achieving high accuracy while reducing computational cost, especially relevant for models with multiple hyperparameters and large search spaces. For instance, RFR reached an MSE of 0.01182 and KNN obtained the lowest MAPE (2.92%) using this method.

Table 1 summarizes the predictive metrics before and after tuning for the top-performing models. The results confirm the effectiveness of combining nonlinear models with advanced optimization.

Table 1: Predictive performance (MSE, R^2, RMSE, MAPE) of the top three models before and after hyperparameter optimization.

Model	**Technique**	**MSE**	R^2	**RMSE**	**MAPE (%)**
RFR	Original	0.0123	0.9757	0.1108	3.36
	GAO	0.0118	0.9765	0.1088	3.13
	Bayesian	0.01182	0.97654	0.10873	3.18
	Optuna	0.01197	0.97625	0.10940	3.19
	GridSearchCV	0.01186	0.97647	0.10889	3.16
KNN	Original	0.0135	0.9732	0.1162	3.30
	GAO	0.0132	0.9738	0.1150	3.20
	Bayesian	0.01311	0.97399	0.11448	2.92
	Optuna	0.01365	0.97292	0.11683	3.06
	GridSearchCV	0.01311	0.97399	0.11448	2.92
GBR	Original	0.0138	0.9726	0.1174	3.30
	GAO	0.0122	0.9758	0.1105	3.44
	Bayesian	0.01311	0.97399	0.11448	2.92
	Optuna	0.01248	0.97523	0.11173	3.14
	GridSearchCV	0.01213	0.97593	0.11014	3.25

Conclusions

This study confirmed the superior performance of nonlinear regression models—particularly RFR, GBR, and KNN—over traditional linear approaches for predicting DTS. RFR achieved the highest accuracy, with an initial MSE of 0.0123 and $R^2 = 0.9757$, which improved further after hyperparameter optimization.

Among the tuning strategies tested, Bayesian Optimization offered the best trade-off between accuracy and computational efficiency, proving especially effective in scenarios involving large search spaces and multiple hyperparameters. The GAO also produced consistent improvements, particularly for ensemble-based models like RFR and GBR.

The proposed pipeline integrates robust preprocessing, model selection, and optimization stages, showing strong potential for generalization beyond DTS. Its modular structure allows for adaptation to other geophysical targets, such as velocity, porosity, or seismic-derived attributes, and to datasets with broader lithological variability or from different basins.

As future work, this approach could be extended to include more diverse geological settings, incorporate temporal sequences or well logs from larger fields, and explore deep learning architectures (e.g., 1D-CNNs) to enhance pattern recognition in complex subsurface data.

References

Ahmed, S. A., M. MonaLisa, M. Hussain, and Z. U. Khan, 2022, Supervised machine learning for predicting shear sonic log (dts) and volumes of petrophysical and elastic attributes, kadanwari gas field, pakistan: Frontiers in Earth Science, **10**.

Breiman, L., 2001, Random forests: Machine Learning, **45**, 5–32.

Liu, S., Y. Zhao, and Z. Wang, 2021, Artificial intelligence method for shear wave travel time prediction considering reservoir geological continuity: Mathematical Problems in Engineering, **2021**, 5520428.

Mrabet, Z. E., N. Sugunaraj, P. Ranganathan, and S. Abhyankar, 2022, Random forest regressor-based approach for detecting fault location and duration in power systems: Sensors, **22**, 458.

Olutoki, J. O., J. guo Zhao, N. A. Siddiqui, M. Elsaadany, A. E. Haque, O. D. Akinyemi, A. H. Said, and Z. Zhao, 2024, Shear wave velocity prediction: A review of recent progress and future opportunities: Energy Geoscience, **5**, 100338.

Rajabi, M., et al., 2023, Predicting shear wave velocity from conventional well logs with deep and hybrid machine learning algorithms: Journal of Petroleum Exploration and Production Technology, **13**, 19–42.

Segal, M. R., 2003, Machine learning benchmarks and random forest regression: Kluwer Academic Publishers. (Division of Biostatistics, University of California, San Francisco, CA).

Silveira, J. R. B. D., J. J. D. Figueiredo, C. R. Lima, J. F. V. Gonçalvez, and M. L. D. Amaral, 2023, S-wave log construction through semi-supervised regression clustering using machine learning: A case study of north sea fields: Journal of Applied Geophysics, **193**, 104353.