geoconvention
Calgary • Canada • May 11-13 2026

# Web App for Well Log Prediction Analysis: Random Forest Model

*Alex Falkovskiy*
*RM Seismic Software*

## Summary

New Python libraries like Streamlit enable faster software development for geophysical applications through easy coding of interface and simple implementation. Such apps make it possible to test numerous geophysical models in reasonable time to find optimal parameters. This work is based on a Python application with Streamlit developed for well log prediction that helped to investigate how prediction errors in a Random Forest algorithm are affected by input data and features.

## Theory and Workflow

This work uses machine learning methodology in application to geophysical data developed by Russel (2004, 2020). Some approaches to processing well logs like finding optimal Convolutional Neural Network architecture were discussed in Falkovskiy (2021). However, the Random Forrest (RF) method eliminates the necessity of finding optimum network architecture as it uses multiple decision trees. RF method came to machine learning in the mid 90th (Breiman, 2000), and later has been successfully applied to geophysical data by many researchers (Roure and Perez, 2025; Zhang et al., 2025).

For testing RF models on well log data and analysis of how the selection of features impacts prediction errors, a Python app with Streamlit was developed. Streamlit is a Python library developed by streamlit.io which enables Python developers to create web applications with less coding efforts and easily deploy them. The app is available at https://predictlog.streamlit.app

As a first step, this app loads a training dataset of LAS files for multiple wells, displays selected logs and histograms with log distribution, and correlation of features. After the user selects a target log to be predicted, it shows target vs. features crossplots. Figures 1-3 show screenshots of this app on the Blackfoot dataset provided by CREWES (Margrave et al., 1997)

After loading well logs, the next step in a workflow is training the RF model on ⅔ of the data, and leaving ⅓ of the data for validation. Figure 4 shows typical prediction results of this tool.

Calculation of Linear Regression (LR) for comparison with RF prediction was included in the workflow.

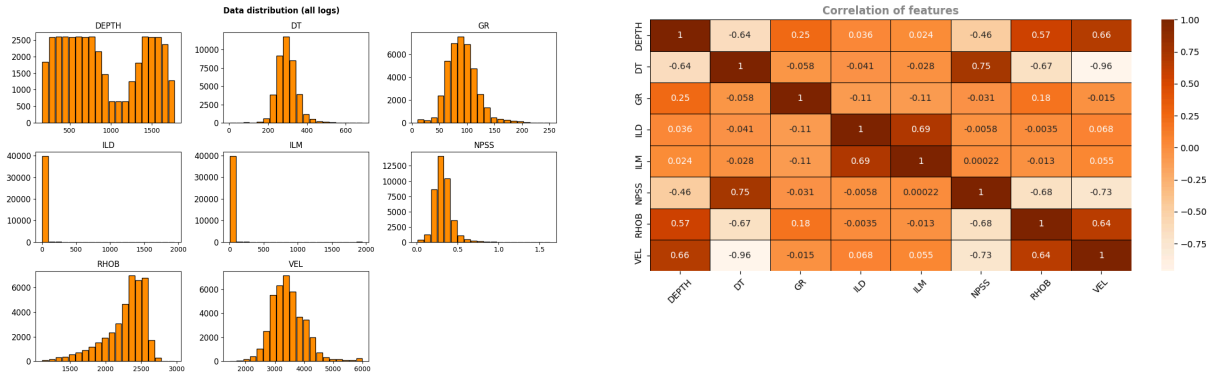
## Results

Since this RF app allows simple selection of several features and one target, the next intuitive step would be to analyse how selection of input features impacts the prediction errors.

For experiments with features and testing this app, Blackfoot dataset was used. Different combinations of input wells and well logs were used for testing which logs affect each target. Since any log may be used as a prediction target, various combinations were tested.
In a standard workflow, DT or seismic velocity would be typically used for prediction of density (Gardner et al. (1974), Quijada and Stewart (2007)). But what if we use density log to predict velocity? For RF prediction and different selections of features, tests showed that errors were larger when DT was predicted from RHOB, than when RHOB was predicted from DT. Flipping the target and feature showed different prediction errors.




Figure 1. RF app screenshot.




a. Data distribution

b. Correlation of features

Figure 2. Log data distribution and correlation of features.

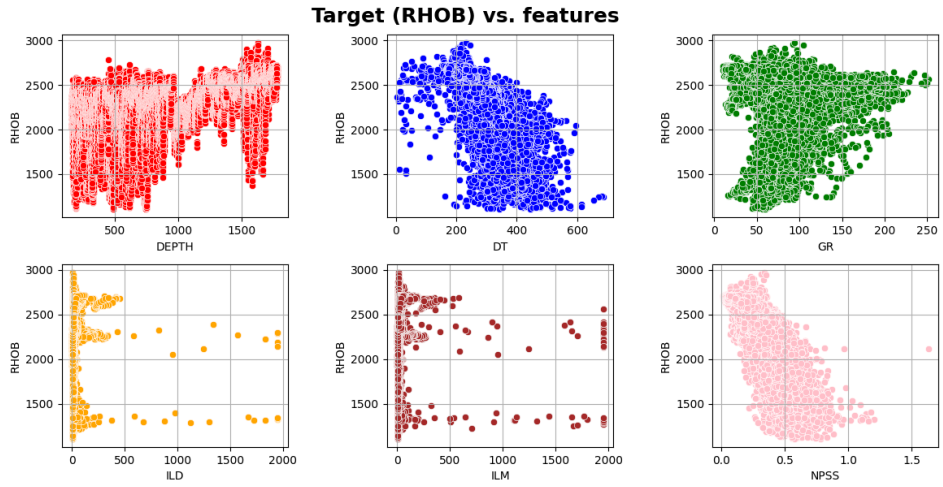
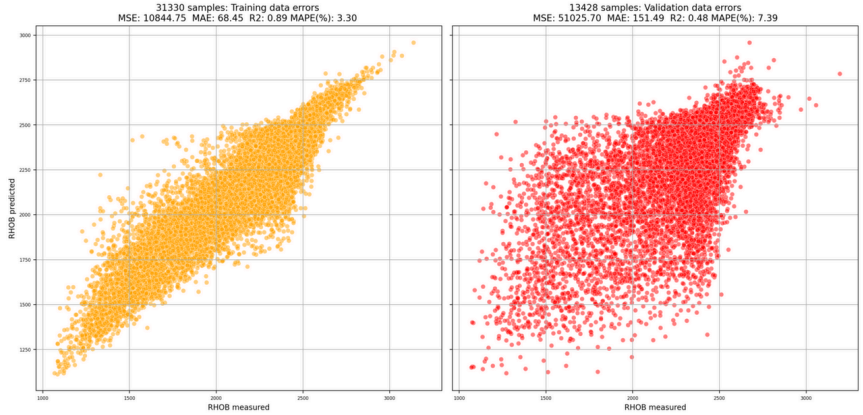
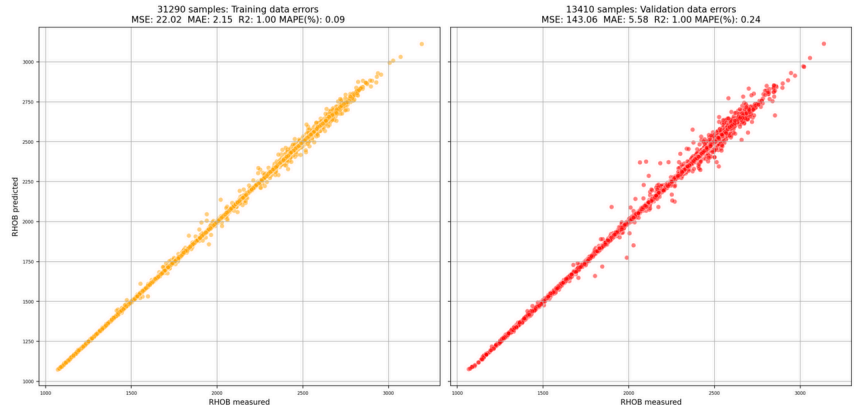



Figure 3. Target vs. features crossplots.

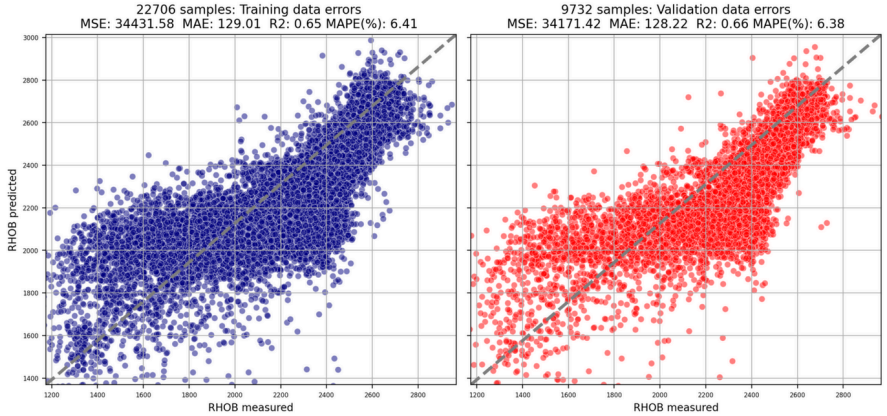



a.Target: RHOB, feature: DT MAPE: 7.39% on validation data




b. Target: RHOB, features: DT, DPSS, SP MAPE: 0.24% on validation data

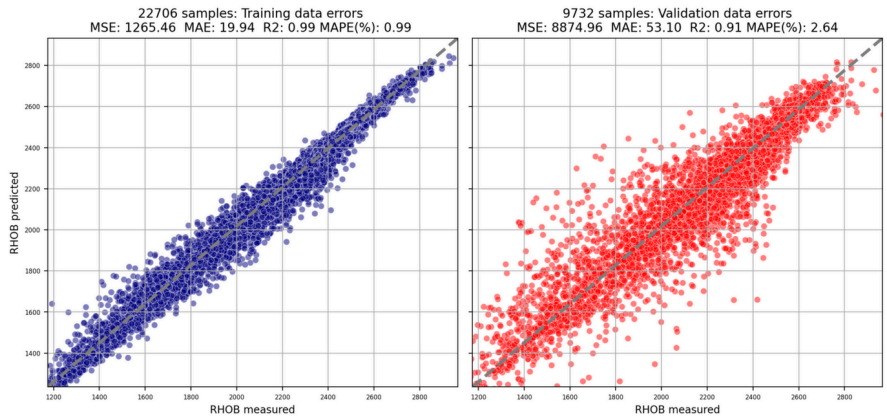
Figure 4. Predicted vs. measured RHOB (ρ) and prediction errors on different input features.




a.Predicted vs. measured RHOB on training and validation data

Linear Regression

Target: RHOB
features:
GR, DT, DEPTH, SP
MAPE: 6.41% on training data
MAPE: 6.38% on validation data




a.Predicted vs. measured RHOB on training and validation data

Random Forest

Target: RHOB
features:
GR, DT, DEPTH, SP
MAPE: 0.99% on training data
MAPE: 2.64% on validation data

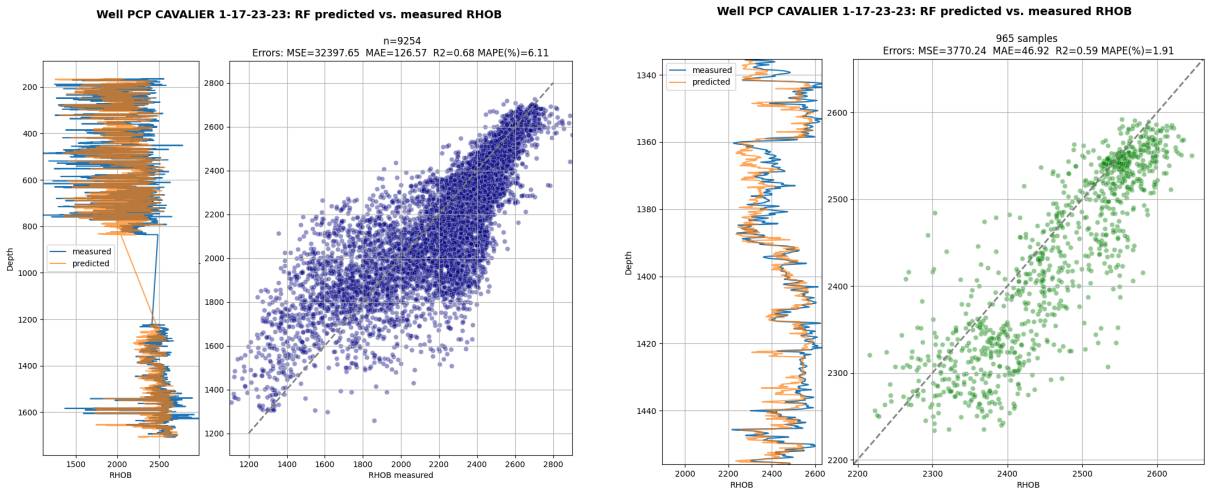
Figure 5. Predicted vs. measured RHOB (ρ) and prediction errors for Linear Regression and Random Forest methods.

LR method was added in this tool as a quality control of RF prediction. Comparison of RF and LR errors showed that on data with high correlation between feature and target the prediction errors were comparable for both methods. However, on data with low correlation between input features and target log, RF outperformed LR on validation data.

Figure 5 shows comparison of prediction results for RHOB on the same combinations of features for two methods: RF and LR. RF shows better prediction error 2.64% compared to 6.38% for LR on validation data.

Note that LR prediction errors are almost the same on training and validation data, but for RF prediction errors are much larger on validation data compared to training (2.64% vs. 0.99%), however it is still much better than LR errors.

Figure 5 shows typical predicted vs. actual density on one of Blackfoot wells.




a. All data

b. Zoomed lower interval

Figure 6. Predicted vs. measured RHOB.

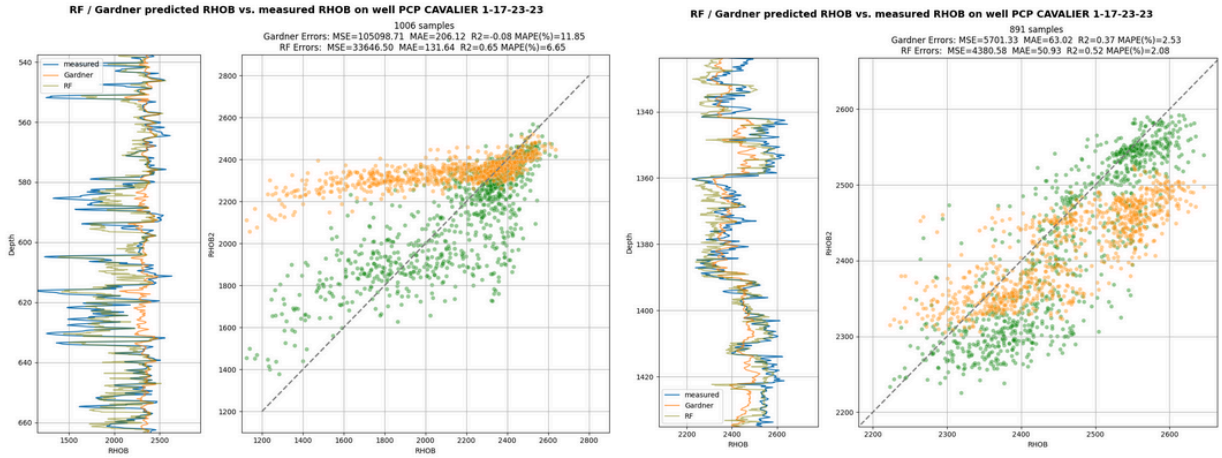



Figure 7. Gardner vs. RF prediction in different depth intervals.

Figures 6 and 7 show that for this dataset prediction errors depend on the depth. MAPE for RF prediction varies from 6.65% in a shallow zone of 550-660m, to 2.08% in a deeper zone of 1340-1420m. But it is also possible to conclude that errors depend on data dispersion. Crossplots in Figure 8 show the same trend - prediction errors decrease in deeper zones and decrease when interval standard deviation is small.

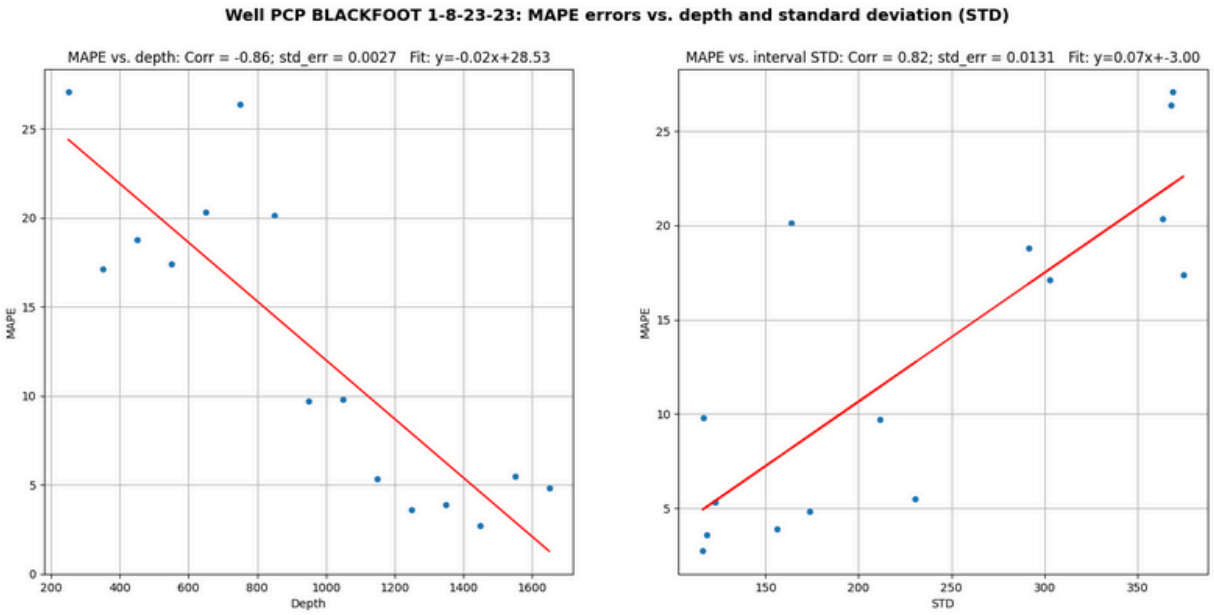



Figure 8. Well Blackfoot 1-8-23-23: MAPE errors vs. depth and interval standard deviation (STD).

## Conclusions

Prediction errors depend on the selection of input well logs (features) used for training and prediction. The general rule here is that the more input logs are used, the better the prediction. However, some logs almost did not affect prediction. The prediction errors (MAPE) may vary from 10% (or greater) to 0.25% depending on the selection of logs used as features.

On this dataset there were only 2-3 logs that strongly affected prediction error of a target, and the other may not be used for the RF model. LR prediction was as good as RF on well correlated data, but RF showed better results even for poor correlation.

Prediction errors depend on depth and dispersion of input data

## Acknowledgements

Well logs used in this research are from Blackfoot dataset courtesy of CREWES. The author thanks Petru Kiss Parciu for comprehensive testing of this software.

### References

Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32.

Falkovskiy, A. 2021. Convolutional Neural Networks for Density Log Prediction. 2021 CSEG GeoConvention.

Gardner, G., L. Gardner, and A. Gregory, 1974. Formation velocity and density—The diagnostic basics for stratigraphic traps: Geophysics, 39,770–780.

Pham, N., Wu, X. and Naeini, E.,Z. 2019. Missing well log prediction using convolutional long short-term memory network. GEOPHYSICS, VOL. 85, NO. 4 (JULY-AUGUST 2020); P.WA159–WA171.

Quijada, M.F. and Stewart, R.R. 2007. Density estimations using density-velocity relations and seismic inversion. CREWES Research Report — Volume 19 (2007).

Mardani, R.A. 2020. Estimation of Acoustic Impedance from Seismic Data in Well-log Resolution Using Machine Learning, Neural Network, and Comparison with Band-limited Seismic Inversion. 2020 CSEG GeoConvention.

Margrave, G.F., Lawton, D.C., Stewart, R.R., Miller, S., Yang, G., Simin, V., Potter, C., Zhang, Q., and Todorov, T. 1997. The Blackfoot 3C-3D seismic survey: A case study. CREWES Research Report — Volume 9 (1997).

Benjamin Roure and Marco Perez, 2025. Maximizing field development through rock physics informed hi-res ML seismic estimates of reservoir properties.

Russell, B.H., 2004. The application of multivariate statistics and neural networks to the prediction of reservoir parameters using seismic attributes. Ph.D. thesis, University of Calgary 2004.

Russell, B.H., 2020. Demystifying Machine Learning. 2020 CSEG GeoConvention.

Zhang, J.; Liu, G.; Wei, Z.; Li, S.; Zayier, Y.; Cheng, Y. Machine Learning-Based Prediction of Well Logs Guided by Rock Physics and Its Interpretation. Sensors 2025, 25, 836.