

Machine Learning-Augmented Well Log Data Processing to Predict and Generate Missing Sonic DT Log Values Using Available Well Log Data

(By Edy Irnandi Sudjana)


Background

In the upstream oil and gas industry, well log data (Lwd/mwd, open-hole, cased-hole and production logs) plays a crucial role in petrophysical analysis, reservoir characterization, well placement & drilling optimization and production optimization. However, it is not uncommon for certain well log measurements, such as sonic DT, to be missing or incomplete due to factors like measurement challenges, logging tool malfunctions, or environmental conditions.

Predicting missing sonic DT values can be a valuable tool for downstream analyses, including porosity estimation, lithology classification and time-depth conversion.

By applying the supervised learning technique of Random Forest Regression to a dataset from nearby or offset wells where both sonic and other well log measurements e.g. gamma ray, resistivity, and density are available, we can predict and generate the missing sonic DT values.

Please note that this ML-Augmented method can complement existing petrophysical analysis and specialized software-based modeling by offering a more automated, scalable solution for missing data imputation, which can be validated and integrated into broader geological models.

To see how this technique works in practice, feel free to explore my sample generate_SonicDT_log_x-val script on GitHub  <https://lnkd.in/d2XqjXGv>

The script freely available under the MIT license. Let's improve the reliability of well logging data together!

Implementation Guidelines

This Python script is designed to predict and generate missing Sonic DT log values in WELL_2, using a trained machine learning model Random Forest Regressor based on available data from WELL_1.

1. Prepare Your Input CSV Files

Ensure you have two well log datasets in CSV format:

- well_log_data_WELL_1.csv
- well_log_data_WELL_2.csv

Both should include the following columns:

- GammaRay, Resistivity, Density, NeutronPorosity, SonicDT

Important: Column names must match exactly as shown (case-sensitive). SonicDT may have missing values in WELL_2.

2. Install Python & Required Libraries

Make sure you have Python installed (preferably version 3.7+). Then install the required libraries:

```
pip install pandas numpy scikit-learn
```

Or with Conda:

```
conda install pandas numpy scikit-learn -c conda-forge
```

3. Run the Script

Get the script from GitHub: <https://lnkd.in/d2XqjXGv>

Place generate_SonicDT_log_x-val.py and both CSV input files in the same directory.

Open a terminal and run:

```
python generate_SonicDT_log_x-val.py
```

The script will:

- Load both WELL_1 and WELL_2 datasets.
- Train a Random Forest model on WELL_1.
- Predict missing SonicDT values in WELL_2.
- Output: well_log_data_WELL_2_SonicDT_added.csv

4. Optional - Review Model Performance

The script performs 5-fold cross-validation and prints:

- Individual Mean Squared Errors (MSE)
- Average Cross-Validation MSE

5. Notes & Tips

- You can modify the list of features in the script.
 - Data scaling (StandardScaler) is already included.
 - Ensure no missing values in predictor columns for WELL_1.
 - Validate predicted SonicDT against core or synthetic data if available.
-