Overview

This project focuses on the imputation of missing values in well log data and the classification of lithology types using machine learning models. The code processes well log data, applies regression models to impute missing values, and uses a classifier to predict lithology types. Finally, it visualizes the lithology distribution across depth.

Files

- `16_1-2.csv`: Sample well log data to be processed (replace with your own CSV file).
- `GR_regressor.pkl`, `RHOB_regressor.pkl`, `NPHI_regressor.pkl`, `DTC_regressor.pkl`, `RDEP_regressor.pkl`, `RXO_regressor.pkl`, `RSHA_regressor.pkl`: Trained regression models for imputing missing values.
- `lithology_classifier.pkl`: Trained classifier for predicting lithology types.

Dependencies

- `pandas`: For data manipulation and analysis.
- `numpy`: For numerical operations.
- `joblib`: For loading machine learning models.
- `matplotlib`: For data visualization.

Usage

1. **Load Data:** The script reads the well log data from a CSV file specified by `file_path`. Ensure that the CSV file is correctly formatted and located in the same directory or adjust the path accordingly.
2. **Preprocessing:**
    - Adjusts the data if the first row contains units (commented out).
    - Converts data to numeric and adds missing columns with NaN values if necessary.
    - Updates specific columns based on predefined conditions and sets outliers to NaN.
3. **Imputation:**
    - Loads pre-trained regression models for specified columns.
    - Imputes missing values in the dataset using these models.
4. **Lithology Classification:**
    - Uses a pre-trained classifier to predict missing lithology values based on other features.
5. **Save Results:**
    - Combines imputed data with other columns and saves the result to a new CSV file with the suffix `_Lithology.csv`.
6. **Visualization:**

- Creates a horizontal bar chart showing the distribution of lithology types across depth.
- Uses a color map to differentiate lithology types visually.