Well Log Feature Regression Model Training and Imputation

Overview

This project involves training regression models for well log features using data from multiple CSV files. The script processes various datasets to create a comprehensive model for each well log attribute: `GR`, `RHOB`, `NPHI`, `DTC`, `RDEP`, `RXO`, and `RSHA`. It then trains a RandomForestRegressor for each feature, evaluates the models, and saves them for future use. The script also includes functionality for imputing missing values using these trained models.

Files

- `31_2-7.csv`, `35_11-10.csv`, `34_10-19.csv`, `33_9-17.csv`, `25_11-24.csv`, `16_10-2.csv`, `15_9-14.csv`, `7_1-1.csv`: Input CSV files containing well log data.
- `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 each well log feature.

Dependencies

- `pandas`: For data manipulation and analysis.
- `numpy`: For numerical operations.
- `sklearn`: For machine learning models and metrics.
- `joblib`: For saving and loading machine learning models.

Usage

1. **Load Data:**
   - The script reads data from multiple CSV files and concatenates them into a single DataFrame. It also drops the first row of some datasets if it contains unit information.
2. **Data Preprocessing:**
   - Converts data to numeric, handles specific outlier values, and prepares the DataFrame for model training.
3. **Feature Scaling:**
   - Scales feature columns using `StandardScaler`.
4. **Model Training:**
   - Iterates over target columns (`GR`, `RHOB`, `NPHI`, `DTC`, `RDEP`, `RXO`, `RSHA`).
   - Splits data into training and testing sets.
   - Trains a `RandomForestRegressor` for each target column.
   - Evaluates the model using MSE and R² metrics.

o   Saves each trained model to a file.
5. **Model Loading and Imputation:**
   o   Loads the trained models.
   o   Imputes missing values in the DataFrame using these models.
6. **Save Results:**
   o   Saves the imputed DataFrame to a CSV file named
       `df_filtered_model_imputed.csv`.

Running the Script

Ensure all dependencies are installed and the required CSV files are available in the same directory as the script. Then, execute the script using Python:

```bash
Copy code
python your_script_name.py
```

Replace `your_script_name.py` with the actual name of your script file.

Example Output

The script generates and saves the following:

- **Trained Models:** `GR_regressor.pkl`, `RHOB_regressor.pkl`, `NPHI_regressor.pkl`, `DTC_regressor.pkl`, `RDEP_regressor.pkl`, `RXO_regressor.pkl`, `RSHA_regressor.pkl`
- **Imputed DataFrame:** `df_filtered_model_imputed.csv`

Customization

- Modify the list of CSV files in the script as needed to include your datasets.
- Adjust columns and model parameters as necessary for your specific use case.

Contributing

Feel free to fork the repository and submit pull requests with improvements or bug fixes.