XBioPred: **An Explainable Artificial Intelligence Framework to Accurately Predict Biogas Yields in Circular Bioeconomic Systems**

This project presents a step-by-step pipeline for predicting biogas yield in wastewater treatment plants using both artificial neural networks (ANN) and tree-based machine learning models with SHAP-based explainability. The full workflow includes preprocessing, model training, and interpretability analysis.

**Execution Sequence**

Run the following notebooks **in order** for a successful workflow:

**1. Preprocessing\_1.ipynb**

* Initial data loading
* Basic cleaning and formatting
* Splitting dataset for preprocessing

**2. Preprocessing\_2.ipynb**

* Missing value imputation
* Outlier reconstruction

**3. Preprocessing\_3.ipynb**

* Data denoising
* Feature engineering
* Data normalization

**4. TPEANNModel.ipynb**

* Training ANN using Tree-structured Parzen Estimator (TPE) for hyperparameter optimization
* Evaluating ANN performance on the test set
* Saving trained models and metrics

**5. TPETreeBasedModelsSHAPexplanation.ipynb**

* Training tree-based models (e.g., Decision Tree, Random Forest, and XGBoost) with TPE optimization
* Computing and visualizing SHAP values for feature importance
* Comparing model explainability and performance