

Tutorial 1. Predicting Bottom Dissolved Oxygen from Surface Variables

1.1 Introduction

1.2 Data Preparation

- 1.2.1 Overview
- 1.2.2 Raw Data Check
- 1.2.3 Layer Selection
- 1.2.4 Dataset Construction
- 1.2.5 Validate Layer Assignments
- 1.2.6 Reshape to Wide Format and Save

1.3 Random Forest (RF) Model

- 1.3.1 Load Cleaned Dataset
- 1.3.2 Train/Test Split
- 1.3.3 Train Random Forest Model
- 1.3.4 Visualize Predictions
- 1.3.5 Visualize a Tree
- 1.3.6 SHAP Interpretation
- 1.3.7 SHAP Bar Plot
- 1.3.8 SHAP Dependence Plots
- 1.3.9 *Bonus: Test RF on Another Station*

1.4 Multi-layer Perceptron (MLP) Model

- 1.4.1 Load and Scale Data
- 1.4.2 Build and Train the MLP
- 1.4.3 Plot Training vs Validation Loss
- 1.4.4 Predict and Evaluate MLP

1.4.5 SHAP Interpretation

1.4.6 SHAP Dependence Plots

1.4.7 Visualize MLP Architecture

1.4.8 *Bonus: Test MLP on Another Station*

1.5 Compare RF and MLP Predictions

Exercise:

Q1. Change the number of estimators in the Random Forest

Q2. Change the number of neurons in the MLP model

Q3. Add or remove layers in the MLP model