

Tutorial 2. Predicting Future DO

Tutorial 2.1. Preprocessing CBP Water Quality Data

2.1.1 Introduction

2.1.2 Data Loading & Inspection

2.1.3 Explore Each Dataset

2.1.4 Clean and Filter Data

2.1.5 Export Cleaned Data

Tutorial 2.2. Preprocessing USGS Discharge Data

2.2.1 Introduction

2.2.2 Set Up API Paths, Data Retrieval (*optional*)

2.2.3 Explore Raw Discharge Data

2.2.4 Calculate Accumulated Flow

2.2.5 Merge Accumulated Discharge Files

Tutorial 2.3. Preprocessing NOAA Meteorological Data

2.3.1 Introduction

2.3.2 Preview Raw Data Files

2.3.3 Load and Clean NOAA Meteorological Data

2.3.4 Convert Wind Direction into Cardinal Sectors

2.3.5 Aggregate Wind by Direction and Time Window

Tutorial 2.4. Merge Preprocessed Data

2.4.1 Introduction

2.4.2 Align and Merge Datasets

Tutorial 2.5. Model Training and Evaluation

2.5.1 Introduction

2.5.2 Define Station Groups and Mask Discharge

2.5.3 Add Prediction Target and Mask Irrelevant Discharge Columns

2.5.4 Train Gradient Boosting Model (with Current Features)

2.5.5 SHAP Interpretation - Simplified Wind Feature

2.5.6 Add Historical (Lagged) Features

2.5.7 Retrain Gradient Boosting with Lagged Features

2.5.8 SHAP Interpretation with Lagged Features

2.5.9 Long Short-Term Memory (LSTM) Modeling

Exercise

Q1. Gradient Boosting: Change Model Parameters

Q2. Gradient Boosting: Modify Lag Time Features

Q3. LSTM: Change Input Window Size

Q4. LSTM: Adjust Model Architecture