Environmental Data Science Term project proposal

Ting-Shao Chang 111322076

Yi-Cheng Lai 112322095

Quentin Lim 112502907

Topic: Comparing different methods to analyze the relationship between the Biochemical Oxygen Demand and other data.

Abstract:

This environmental data science study purposes on analyzing the water quality dataset, primarily considering the value of pH, Dissolved Oxygen (DO), Temperature, and Conductivity to forecast the Biochemical Oxygen Demand (BOD) index. Given that several parameters are considered as reference points in this study, we are particularly interested in identifying the most significant factor that could strongly influence the predictions. To achieve this, we plan to employ Linear Regression, Ridge Regression, and Lasso Regression models for training and assess the reliability of each model using the Root Mean Squared Error (RMSE) metric. In this way, we aim to find the most influential factor and select the model that best fits the dataset. Finally, we will engage in further discussions based on the results.

Data source: https://www.kaggle.com/datasets/shreyanshverma27/water-quality-testing

Keywords: Biochemical Oxygen Demand (BOD), Water Quality, Regression Analysis, Model Assessment.