

ChE 197/297: Intro to AI/ML for Chemical Engineers

Case Studies in ChemE

Instructions: Answer each problem then create a solution using Python code via Jupyter Notebook.

Problem: Predicting Chlorophyll-a content in Laguna Lake from Nitrogen and Phosphorus Concentrations using Neural Networks

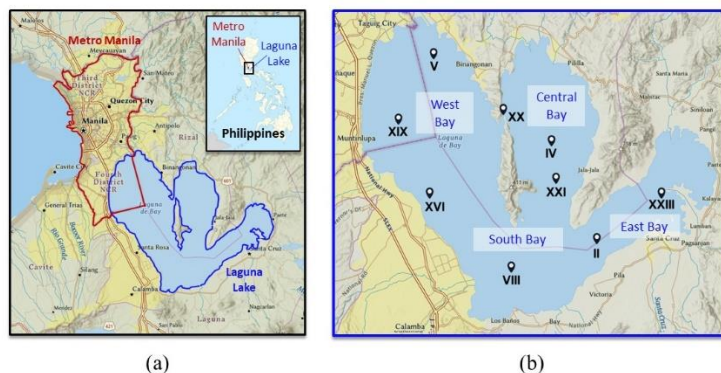


Figure 1. Map of stations in Laguna Lake where the data points were sampled.

You are given a data set of measurements of Chlorophyll-a (Chl-a) content in 9 sampling stations at Laguna Lake, along with the corresponding ion concentrations of $[\text{NO}_2^-]$, $[\text{NO}_3^-]$, $[\text{NH}_4^+]$, and $[\text{PO}_4^{3-}]$. It is well-known in literature that the total nitrogen and total phosphorus content in any waterbody has a strong relationship to Chl-a content. This is because Chl-a indicates the presence of algae, and algae needs these nutrients to survive. Chl-a prediction can help quantify the algal biomass in lakes, which is directly related to the extent of eutrophication.

Water treatment plants take water from Laguna Lake to turn into cleaner water, which is then distributed to nearby cities such as Las Pinas and Muntinlupa. If the lake has severe eutrophication, excessive algal biomass tend to clog the filters very quickly until the treatment operations are stopped. This can yield to occasional water shortages, while the filters are being serviced. If only we can predict algal blooms early (via Chl-a), water shortages can be prevented.

Our task is to use regression to predict Chl-a content. Do the following.

1. First, drop rows with missing values in the data set.
2. Split the data into training (60%) and testing (40%). Stratify this according to the sampling station.
3. Create a new data column called TN (Total nitrogen). Compute TN as the sum of $[\text{NO}_2^-]$ and $[\text{NO}_3^-]$. Meanwhile, we'll take the phosphate ion column as TP (Total phosphorus). From here on, we will use only 3 columns of data, Chl-a, TN, and TP.
4. Normalize the training data, then apply the same scaling to the validation and test data.
5. Train a 1-hidden layer ANN with 50 hidden neurons and "relu" activation to predict Chl-a from TN and TP. Evaluate the model on the test data, then report the accuracy as R^2 .

END OF EXERCISE