

Introduction to Bayesian Data Analysis

Term Project

PM2.5 Prediction

https://ci.taiwan.gov.tw/dsp/dataset_air.aspx

Please download the first dataset (i.e., 77 stations) under 環保署 (Environmental Protection Administration). Your job is to create a model that can predict 3 hours of PM2.5 readings based only on the historical PM2.5 readings and possibly other features from the same dataset.

The presentation of the term project will be held on 6/14, which accounts for 50% of the grade of your term project. By 6/19, each team will need to submit a term project report, which accounts for another 50% of the grade.

The time of the presentation will be 10 minutes per group. In the presentation, each group is expected to clearly state the data preprocessing, the model in Kruschke diagram, the PyMC3 code, and demo the performance of your model (i.e., MSE score of the station of our choice, which will be given 30 min prior to the start of the presentation).

For the report, each team needs to submit the following materials:

1. Your source code and the script(s) (if any) for data pre-processing, model building, sampling, and visualization.
2. A short report (up to 3 pages) – describing the key concepts, the performance of your model, and the modifications made to your model after receiving comments from the presentation. (Be concise in your report!)
3. A list of the tasks each teammate is responsible for.

Please zip all these items into a zip file and then submit via eclass before the midnight of 6/19/2023.

Important Note: Please make sure your model is a Bayesian model, as the aim of this project is to evaluate your learning on Bayesian Data Analysis.