More Details:

Problem:

The project has to provide deterministic forecasts for hourly load, ranging from 28 hours (2 days) in advance. 48 hourly predictions should be made, based on data available up to 8 AM of the previous day. The focus is to make 2-days ahead predictions of utility-scale load.

Data:

The data provided is from a metropolitan electric utility and represents the overall system load for the metropolitan area. You are given approximately four years of data, from March 18, 2017 to February 15, 2021. The data is presented in CSV format and includes hourly measurements and weather component forecasts for the following:

* Electricity demand in kilowatts
* Air pressure observations and forecasts in kilopascals
* Cloud cover observations and forecasts as a percentage
* Humidity observations and forecasts as a percentage
* Temperature observations and forecasts in degrees Celsius
* Wind direction observations and forecasts in degrees
* Wind speed observations and forecasts in kilometres per hour

The data is split into a training set and a test set. The training set contains data up to January 16, 2021. It is assumed that the weather forecasts were available 48 hours in advance and remained the same throughout the 48-hour period.

The test data is provided in daily batches over a period of 30 consecutive days, from March 15 to April 13, 2021. Each batch contained 24 hours of daily measurements up to 8 AM, as well as weather forecasts for the next 24 to 48 hours. The first batch of test data includes demand measurements up to 8 AM on January 17, 2021, while the last batch of test data includes demand measurements up to 8 AM on February 15, 2021.

Additional Notes:

**Must have:** forecast benchmarks are 1. Seasonal Naive 2. Naive 3. Random Walk with a drift

(these are available on R in forecast and fpp3 packages but other implementations of them also exist in other languages).

Train folder has two parts, you can either use them as train and evaluation respectively or just merge them together and split as desired.

The test files will be in individual files for each day ( jan 17 to feb 15, 2021)

The UI should be able to accept the forecast file from the test data zip files and the models’ prediction should be presented visually.

**Must have:** evaluation metrics would be MAE

The final report should be presented with a table showing the forecast errors (MAE along with 2 more metrics of your choice) for all the solutions and benchmarks over the test period (jan 17 to feb 15).

Regarding the retraining strategies, The students should come up with an automated process to monitor the model and retrain the model when needed.

An example of the monitoring strategy is data/concept drifts. When the monitoring strategy triggers an event saying the model needs to be trained, the retraining module should be executed (the students also have to review different methods of retraining such as fine tuning, cross validated training data selection, etc).