Bluwave-AI

Data Scientist Assignment

We would like you to develop an ML model to predict electricity load based on the provided dataset, which spans approximately 5 months. The dataset includes observed and forecasted weather data, as well as the corresponding observed load at each timestamp. Each column in the dataset is defined as follows:

Attribute Name	Description
event_timestamp	Timestamp
weatherkit_observed_temperature_C	Observed temperature in Celcius measured at event_timestamp
weatherkit_observed_humidity_pc	Observed humidity in Percentage measured at event_timestamp
weatherkit_observed_air_pressure_kPa	Observed air pressure in kPa measured at event_timestamp
weatherkit_observed_cloud_cover_pc	Observed cloud coverage in Percentage measured at event_timestamp
weatherkit_observed_wind_direction_deg	Observed wind direction in degrees measured at event_timestamp
weatherkit_observed_wind_speed_km_h	Observed wind speed in km_per_hour measured at event_timestamp
weatherkit_forecast_temp_C	Forecasted temperature in Celcius predicted for event_timestamp
weatherkit_forecast_humidity_pc	Forecasted humidity in Percentage predicted for event_timestamp
weatherkit_forecast_air_pressure_kPa	Forecasted air pressure in kPa predicted for event_timestamp
weatherkit_forecast_cloud_cover_pc	Forecasted cloud coverage in Percentage predicted for event_timestamp
weatherkit_forecast_wind_direction_deg	Forecasted wind direction in degrees predicted for event_timestamp
weatherkit_forecast_wind_speed_km_h	Forecasted wind speed in km_per_hour predicted for event_timestamp
load_MW	Measured electricity load in MW at event_timestamp

For the purpose of this exercise, please assume that during the prediction phase, only forecasted weather data will be available. However, for training the model, both observed and forecasted data can be utilized.
The predictions should be generated once per day at 8:00 am (UTC+05:30), specifically to predict the load for the following day in local time (UTC+05:30).
We understand that this exercise is time-limited and do not expect a perfect solution.

However, we would like to see how you approach the problem, including your model selection process and any code you create. In addition to the code you will create, we also request that you provide a presentation outlining the steps you took to approach the problem. This presentation should highlight the key stages of your solution. Please feel free to make simplifying assumptions if necessary to complete the task within a couple of days. If you have any questions or need further clarification, please don't hesitate to reach out to us. We look forward to seeing your code, which you can submit as Jupyter notebooks and/or Python scripts.