Assignment

Time series forecasting on solar irradiance data

- **Background**: Solar radiation is an important source for electricity generation. For effective utilization, it is important to precisely know the irradiance amount at different time horizons: minutes, hours, and days. Depending on the horizon, two main classes of methods can be used to forecast the solar radiation: statistical time series forecasting methods for short to midterm horizons and numerical weather prediction methods for medium- to long-term horizons.

- **Objective**: To forecast the next day solar irradiance (measured in W/m2) values using ClimaCell API data (6-hours per day) and real weather station data from a solar plant.

- **ClimaCell API Data** ('climacell_data_formodel.csv')
    - ClimaCell is forecasting tomorrow's irradiance value, ambient_temp and wind_speed from 6am to 12 am with 5 min frequency using their proprietary system.

    - Features in the file labelled as 'climacell_data_formodel.csv'
        - date_time -> IST date and time (format: 2020-08-20 06:05:00+05:30)
        - date -> dates are seperated from datetime stamp
        - time -> times are seperated from datetime stamp
        - irradiation -> irradiance value in 5 min duration/frequency
        - ambient_temp -> ambient temperature in 5 min duration/frequency in the used longitude and latitude
        - wind_speed -> wind speed in 5 min duration/frequency in the used longitude and latitude

- **Weather Station Data** ('weather_data_from_plant.csv')
    - Features in the file labelled as 'weather_data_from_plant.csv'
        - date_time -> IST date and time
        - irradiation -> irradiance value in 1 min duration/frequency
        - ambient_temp -> ambient temperature in 1 min duration/frequency in the used longitude and latitude
        - wind_speed -> wind speed in 1 min duration/frequency in the used longitude and latitude

- **Input**: input to the model are:
    - Time-based features to address seasonality and cyclic nature of irradiation data.
    - Raw data: irradiation, ambient_temp, wind_speed
    - Lagged features of solar irradiance
    - Resampling: chose a suitable duration/frequency – 5 min, 15 min or 30 min

- **Model and Approaches**:
    - Classical time-series model -> for example: ARIMA
    - Multilayer Perceptron (MLP)
    - Long Short-term Memory (LSTM)
    - Gated Recurrent Units (GRU)

- **Output**: output of the model will be solar irradiance values (W/m2) for the next day. You may choose to forecast it for hourly or 30-min, or 15-min, or 5-min frequency/duration.

- **Performance Measurements**:
    - R2
    - RMSE
    - Normalized RMSE (NRMSE)