Due Date: **See Webcampus**
How to submit: **Webcampus**

General Guidelines:

- Please prepare a **typed** report that describes what you did. The report should be as concise as possible while providing all necessary information required to replicate your plots.
- As there are many related papers on solar power generation forecast, you may want to review the literature and borrow ideas there. **Make sure you properly cite the references in the report. If you adopt methods from existing literature, you need to give credits to the authors by citing their papers.**
- Please provide, at the end of your report, a commented version of your python code files. **Python Notebook files are preferred. You may put the codes in a SINGLE ipynb file with necessary texts to explain your codes.**

Solar Power Generation Forecast

This project aims to develop solar power generation forecast models using the data to predict 24 h ahead solar power generation on a rolling basis for three solar power plants located in a certain region of Australia. The data is from Global Energy Forecasting Competition 2014 (see https://www.crowdanalytix.com/contests/global-energy-forecasting-competition-2014-probabilistic-solar-power-forecasting). A portion of the datasets will be used in our project.

1. Dataset Description.

You will find the data in the excel file "solar.csv" in the webcampus. The first 5 data entries are provided below:

ZONEID	TIMESTAMP	VAR78	VAR79	VAR134	VAR157	VAR164	VAR165	VAR166	VAR167	VAR169	VAR175	VAR178	VAR228	POWER
1	20120401 01:00	0.001967	0.003609	94843.63	60.22191	0.244601	1.039334	-2.50304	294.4485	2577830	1202532	2861797	0	0.754103
1	20120401 02:00	0.005524	0.033575	94757.94	54.6786	0.457138	2.482865	-2.99333	295.6514	5356093	2446757	5949378	0	0.555
1	20120401 03:00	0.030113	0.132009	94732.81	61.29489	0.771429	3.339867	-1.98254	294.4546	7921788	3681336	8939176	0.001341	0.438397
1	20120401 04:00	0.057167	0.110645	94704.06	67.77528	0.965866	3.106102	-1.44605	293.2615	9860520	4921504	11331679	0.002501	0.145449
1	20120401 05:00	0.051027	0.18956	94675	70.17299	0.944669	2.601146	-1.90449	292.7329	11143097	6254380	13105558	0.003331	0.111987

The data includes weather forecasts for 12 weather variables, as obtained from the European Centre for Medium-range Weather Forecasts (ECMWF). These variables are denoted as VAR attributes (from VAR78 to VAR228) in the dataset, which are summarized in Table 1.

The hourly measurements for each solar power plant from 20120401 01:00 to 20140701 00:00. are provided. ZONEID denotes the id of a solar power plant. As we have three solar power plants, ZONEID is from 1 to 3. The POWER attribute records the hourly solar power generation, which has been normalized. For example, the first data entry gives the hourly measurements for solar power plant 1 at 01:00 on April 1, 2012.

Table 1. 12 weather variables

Variable id.	Variable name	Units	Comments
078.128	Total column liquid water (tclw)	kg m^{-2}	Vertical integral of cloud liquid water content
079.128	Total column ice water (tciw)	kg m^{-2}	Vertical integral of cloud ice water content
134.128	Surface pressure (SP)	Pa	
157.128	Relative humidity at 1000 mbar (r)	%	Relative humidity is defined with respect to saturation of the mixed phase, i.e., with respect to saturation over ice below −23 °C and with respect to saturation over water above 0 °C. In the regime in between, a quadratic interpolation is applied.
164.128	Total cloud cover (TCC)	0–1	Total cloud cover derived from model levels using the model's overlap assumption
165.128	10-metre U wind component (10u)	m s^{-1}	
166.128	10-metre V wind component (10v)	m s^{-1}	
167.128	2-metre temperature (2T)	K	
169.128	Surface solar rad down (SSRD)	J m^{-2}	Accumulated field
175.128	Surface thermal rad down (STRD)	J m^{-2}	Accumulated field
178.128	Top net solar rad (TSR)	J m^{-2}	Net solar radiation at the top of the atmosphere. Accumulated field
228.128	Total precipitation (TP)	m	Convective precipitation + stratiform precipitation (CP + LSP). Accumulated field.

2. Tasks.

1. Split "solar.csv" into the training dataset "solar_training.csv" and the test dataset "solar_test.csv" as follows:

- The training dataset is from 20120401 01:00 to 20130701 00:00.
- The test dataset is from 20130701 01:00 to 20140701 00:00.

2. Build a 24 h ahead solar power generation forecast model using the training dataset.

To give you an idea, an example for one solar power plant is given as follows. Let $X_t = \{VAR78, VAR79, \ldots, VAR228, P\}$ denote a vector of all the attributes for one solar power plant. Let f denote the forecast model, which will predict 24 h ahead solar power generation by using the current and historical measurements. For example, $\hat{P}_{t+24} = f(X_t)$, which uses only the current measurements to predict 24 h ahead solar power generation, i.e., $\hat{P}_{t+24}$. You can explore the spatio-temporal correlation in the datasets to incorporate more relevant measurements as input for the prediction. You can build a forecast model for each plant separately.

3. Model Evaluation.

You need to evaluate your model using the test dataset. Let P_t and $\hat{P}_t$ denote the actual power generation and the predicted power generation at time t, respectively. You will use the following measures to evaluate your model:

- Mean absolute error (MAE)

$$MAE = \frac{1}{number\ of\ points} \sum_t \left|P_t - \hat{P}_t\right|$$

- Root mean squared error (RMSE)

$$RMSE = \sqrt{\frac{1}{number\ of\ points} \sum_t \left|P_t - \hat{P}_t\right|^2}$$

3. Report format.

When preparing your report, please use the following format:

- Cover page
  - Title, You name, CS458 or CS658
- Introduction
- Literature review
- Your methods
  - How you build your models (e.g., Data Preprocessing, Model selection, Parameter selection)
- Evaluation results
  - Need to provide a table of MAE and RMSE and figures to compare the curves of your predicted power with the actual power for days in different seasons.
  - Discussions of your results (e.g., pros and cons for your prediction models)

Table Example

ZONEID	1	2	3	Overall
MAE				
RMSE				

- Conclusions