IE 360 Statistical Forecasting and Time Series
Spring 2022 - Project

Solar Power Forecasting

This project is about providing hourly solar power prediction of KIVANC 2 GES (Güneş Enerjisi Santrali) for the next day. KIVANC 2 GES is located in Mersin (between 36-37° north latitude and 33-35° east longitude.) and further information related to the capacity and location is available on

- https://www.kivancenerji.com.tr/kivanc2ges.html
- https://www.enerjiatlasi.com/gunes/kivanc-2-ges.html

Google Earth link to the power plant is:
https://earth.google.com/web/search/G%c3%bclnar,+Mersin/@36.53293553,33.25506549,1133.49566978a,3417.68721644d,35y,-10.21136707h,0.67542784t,0r/data=CigiJgokCbDmFAthQkNAEbdEevHBMENAGY72bhshJkBAIVzzZJJPBEBA

The project is organized as a competition (like in platforms such as www.kaggle.com) in which you are expected to make submission everyday. For this purpose, we have built a system so that you will be able to make submissions via Google Forms and monitor your progress through Google Sheets. We will be informing you about the submission system once we finish our tests.

The prediction is needed in a very similar setting to real life needs of the energy traders performing operations in energy markets. In other words, the assumption is that on day *d*, the predictions are needed for day *d+1* and you know the production values until the end of day *d-1*. In real life, you have to submit your predictions before 12 pm of the day *d* for the next day (i.e. day *d+1*).

Provided data file includes the weather measurements for 9 grid points (coordinates) nearby the power plant. The grid points are provided in Table 1.

latitude	longitude
36.25	33
36.25	33.25
36.25	33.5
36.5	33
36.5	33.25
36.5	33.5
36.75	33
36.75	33.25
36.75	33.5

Table 1. The latitude and longitude pairs for weather variables

The weather variables that can help for solar power plant production forecasting are described below:

- **TEMP**: Temperature at the provided location. Temperature can represent the seasonality. Moreover it is known that high temperatures affect the solar panels and decrease their efficiency.
- **REL_HUMIDITY:** Relative humidity at the provided location. This variable can provide information about the rainy or cloudy times which potentially decreases the production. Further information about relative humidity can be reached through the following link: https://en.wikipedia.org/wiki/Humidity.
- **DSWRF**: This is the short version of downward shortwave radiation flux which is known to be highly related to the production level. This is further described on the following link: https://www.goes-r.gov/products/baseline-DSR.html#:~:text=The%20downward%20shortwave%20radiation%20(DSR,that%20reaches%20the%20Earth's%20surface.&text=DSR%20data%20are%20also%20employed,to%20drive%20ocean%20circulation%20models. On the other hand, I advise you not to spend too much time to understand the background information about this variable. Instead follow a data-driven approach to understand its relation to production levels.
- **CLOUD_LOW_LAYER**: This is total cloud cover data (in terms of percentage) for low-level type of clouds. Further information about cloud types are available on https://en.wikipedia.org/wiki/List_of_cloud_types however as advised for DSWRF data, please follow a data-driven approach to understand if the variable provides information on hourly production.

You are given this weather information that can be helpful in the forecasting task for 9 coordinates. This information is provided in so called “long format”. A sample view of the data is shown in Table 2. One can create a data set where we have the information of the weather variables at each location using reshaping operations. For R, you can follow the tutorial on https://cran.r-project.org/web/packages/data.table/vignettes/datatable-reshape.html if you use data.table package and https://www.statology.org/pandas-long-to-wide/ can help you to reshape your data in Python (and pandas). This operation can result in a data in the following format: “VARIABLENAME_LATITUDE_LONGITUDE” as illustrated in Table 3. This makes a tabular data set which is generally needed for regression approaches. This will make 9 coordinates multiplied with 4 variables = 36 weather variables in the wide format.

Performance measure

Error rate: Your model will be scored based on your weighted mean absolute percentage error over the whole test period during the competition phase.

Note that you will be compared to a baseline method as a competitor. Baseline uses the most recent day’s same hour’s production (i.e. 2 days ago, same hour). This is sometime called “persistence approach”. In the time series context, this will be lag 48 value.

Project Timeline

Test Phase: This is the time period between May 20-23, 2022 (both included) dedicated to testing the submission system. The aim is to make you familiar to online submissions.

Competition Phase: This is the period in which you will be evaluated based on your submissions. The period is between May 25-June 3, 2022 (both included).

Note that the dates indicate the forecasted dates. For example, competition phase starts on May 25th, 2022. This means you need to provide your submission before 12:00 pm of May 24th, 2022 for the next day (i.e. May 25th, 2022) in our prediction setting.

Prediction and Report

Note that 30% of your project grade will be determined by your final rank in this competition. First place will get full points (30 points) and this will decrease to a minimum of 15 points proportional to your deviation from the top performer (assuming that you did not miss any day during the competition phase). Moreover, you will lose 1 points (out of 30 points if you miss to submit a prediction during the competition phase.

You are required to submit a written report with a brief description of your final method, how you evaluated your methods, and you choose the parameter settings. Note that you can try different models during the competition phase of the project however you are expected to evaluate your approaches from March 1st to May 24th of 2022. You are allowed to work as a group of at most 3 members.

Your report should have the following format:

- Introduction: Problem description, summary of the proposed approach, descriptive analysis of the given data.
- Related literature: Summarize relevant literature if there is any
- Approach: Explain your approach to this problem.
- Results: Provide your results and discussion.
- Conclusions and Future Work: Summarize your findings and comments regarding your approach. What are possible extensions to have a better approach?
- Code: Provide the Github link for your codes at the end of your report.

	date	hour	lat	lon	variable	value
1	2021-02-01	0	36.25	33.00	DSWRF	0.0000
2	2021-02-01	1	36.25	33.00	DSWRF	0.0000
3	2021-02-01	2	36.25	33.00	DSWRF	0.0000
4	2021-02-01	3	36.25	33.00	DSWRF	0.0000
5	2021-02-01	4	36.25	33.00	DSWRF	0.0000
6	2021-02-01	5	36.25	33.00	DSWRF	0.0000
7	2021-02-01	6	36.25	33.00	DSWRF	0.0000
8	2021-02-01	7	36.25	33.00	DSWRF	0.0000
9	2021-02-01	8	36.25	33.00	DSWRF	0.0000
10	2021-02-01	9	36.25	33.00	DSWRF	3.0000
11	2021-02-01	10	36.25	33.00	DSWRF	50.0000
12	2021-02-01	11	36.25	33.00	DSWRF	80.0000
13	2021-02-01	12	36.25	33.00	DSWRF	140.0000
14	2021-02-01	13	36.25	33.00	DSWRF	180.0000
15	2021-02-01	14	36.25	33.00	DSWRF	190.0000
16	2021-02-01	15	36.25	33.00	DSWRF	230.0000
17	2021-02-01	16	36.25	33.00	DSWRF	290.0000
18	2021-02-01	17	36.25	33.00	DSWRF	270.0000
19	2021-02-01	18	36.25	33.00	DSWRF	200.0000
20	2021-02-01	19	36.25	33.00	DSWRF	150.0000
21	2021-02-01	20	36.25	33.00	DSWRF	120.0000
22	2021-02-01	21	36.25	33.00	DSWRF	100.0000
23	2021-02-01	22	36.25	33.00	DSWRF	0.0000
24	2021-02-01	23	36.25	33.00	DSWRF	0.0000
25	2021-02-02	0	36.25	33.00	DSWRF	0.0000

Table 2. Weather information in “long format”

	date	hour	CLOUD_LOW_LAYER_36.25_33	CLOUD_LOW_LAYER_36.25_33.25	CLOUD_LOW_LAYER_36.25_33.5	CLOUD_LOW_LAYER_36.5_33
1	2021-02-01	0	45.00	31.00	3.00	53.0
2	2021-02-01	1	44.00	26.00	19.00	64.0
3	2021-02-01	2	37.00	35.00	35.00	60.0
4	2021-02-01	3	43.00	45.00	46.00	51.0
5	2021-02-01	4	72.00	95.00	46.00	16.0
6	2021-02-01	5	69.00	95.00	31.00	24.0
7	2021-02-01	6	75.00	96.00	53.00	21.0
8	2021-02-01	7	79.00	97.00	65.00	16.0
9	2021-02-01	8	83.00	97.00	60.00	13.0
10	2021-02-01	9	86.00	97.00	67.00	20.0
11	2021-02-01	10	100.00	100.00	60.00	79.0
12	2021-02-01	11	100.00	100.00	59.00	71.0
13	2021-02-01	12	98.00	100.00	71.00	65.0
14	2021-02-01	13	97.00	100.00	69.00	73.0
15	2021-02-01	14	97.00	100.00	67.00	73.0
16	2021-02-01	15	93.00	98.00	71.00	77.0
17	2021-02-01	16	82.00	46.00	6.00	90.0
18	2021-02-01	17	58.00	33.00	3.00	94.0
19	2021-02-01	18	42.00	23.00	2.00	73.0
20	2021-02-01	19	32.00	17.00	1.00	60.0

Table 3. Weather information in "wide format" (due to large number of variables, partial table is shown)