Flexa

Flexa Data Science Challenge

This challenge consists of 2 tasks. The first task focuses on forecasting and data analytics, the second task focuses on optimization. The tasks give you an idea of the type of data and problems you will have to work with at Flexa. Therefore, they also test you on the skills we require from data scientists at Flexa.

Please read through the entire challenge before attempting any of the tasks. Feel free to answer the tasks in any order you like and to put more emphasis on the task that you feel addresses some of your strengths, while moving through the other task more quickly.

For both tasks, please provide us with the code that implements your solution. The code should be submitted in a way that it can run on a local machine, and should contain instructions on how to do so. Please also add documentation that provides sufficient details to understand the underlying logic of your code. Where no documentation is provided, please ensure that the code is self-explanatory.

Please find all the files needed for completing the challenge here:
https://drive.google.com/file/d/1PU6bs7whHwuNe5KmetY-ClLHRy0Joft4/view?usp=sharing

flexa

Task 1 - Forecasting of energy production and consumption

The goal of this task is to forecast, for every hour, the net energy (production minus consumption) for a set of sixteen pools. A "pool" is defined as a group of households located in the same geographical region, which both consume energy and produce energy via PV panels.

Please use as input for this task the file "1_measurements.csv", which contains data for the sixteen pools, from September 2021 to May 2023. The file has the following columns:

- datetime_utc: timestamp of energy measurement for interval (t, t+1), hourly sampled
- pool: the pool to which a measurement belongs to
- target: the energy measurement value (production or consumption)
- is_consumption: flag to indicate if the target represents consumption or production

For building the forecasting model, use the "target" column as your gold standard. As additional information for the prediction, you can also use the additional weather forecast data contained in the file "1_weather.csv" (a description of the data is included in the Appendix of this document).

For the analysis split the data in two sets:

- training set: data from September 2021 to December 2022
- test set: data from January 2023 to May 2023

To train your model you must use only the training set. However, at prediction time, for a target data point, your model can use any data point preceding the target data point.
For example, in order to predict the net energy for April 15, 2024 between 14:00 and 15:00, you can use all the training data, and also the test data from January 2023 to April 15, 2024 at 13:59.
Please report your results on the test set and use MAE (mean absolute error) as the main evaluation metric (if you like, you can also choose additional metrics).

Please make sure to perform (and report on) proper data analysis before getting into the problem solution, for example if you see any pattern in the data, correlation among features, or if you see and need to address data quality issues.

flexa

Task 2 - Optimization

The goal of this task is to optimize the profits of selling and buying energy on the market, given a large battery. Specifically, assume that we have a 1MW, 1MWh battery. This means the battery can charge and discharge with a maximum power of 1 MW. The battery can store at most 1 MWh, i.e. after charging with full power for 1h, the battery will be full.
To earn money with the battery we buy electricity on the wholesale market when it is cheap to do so and sell when it is expensive, storing the electricity in the battery in the meantime.

Please use as input for this exercise the file "2_electricity_prices.csv", which contains hourly forecasts of wholesale electricity prices from September 2021 to May 2023. Your task is to find the profit-maximizing times to charge and discharge the battery based on the available prices. Perform the optimization for each day separately and include the side constraint that the battery is discharged at the end of the day.

Illustrate your findings, showing for example how much revenue you would have earned, at what times of the day it is usually best to charge and when it is best to discharge, etc. Please list any assumptions you make, if any.

Optional subtask

Suppose we now have a 1 MW, 2 MWh battery. This means the battery can charge for 2 hours with full power when completely empty and discharge for 2 hours with full power when completely full. It can also charge and discharge to half full by charging/discharging with full power for 1 hour. How does your answer to task 2 change?

flexa

Further reads available to you

To immerse yourself a bit further with what we do, feel free to have a look at the following (not required for the task):

Our webpage
www.flexa.energy

Our linkedin
https://www.linkedin.com/company/flexaenergy/?viewAsMember=true

Newspaper article of Enpal announcing flexa (German)
https://efahrer.chip.de/news/so-viel-leistung-wie-ein-akw-enpal-baut-das-groesste-virtuelle-kraftwerk-europas_1020412

Podcast of Mario Kohle, founder of Enpal, mentions also flexa (German)
https://www.youtube.com/watch?v=iQDPbi02V50

General intro to the potential of battery storage:
https://www.energy-storage.news/software-and-batteries-can-enable-energy-industrys-amazon-prime-plus-moment/

Notes on the data of this challenge

The data used in this challenge is derived from Kaggle (1600 Amphitheatre Parkway, Mountain View, California, USA) and the case challenge is derived and adapted from the following Enefit Kaggle Challenge: https://www.kaggle.com/competitions/predict-energy-behavior-of-prosumers/ (last access: 22.08.2024)

Flexa

Appendix

Weather forecast data description

- `forecast_datetime` - The timestamp of the predicted weather. Generated from. This represents the start of the 1-hour period for which weather data are forecasted.
- `temperature` - in degrees Celsius
- `dewpoint` - The dew point temperature
- `cloudcover_[low/mid/high/total]` - The percentage of the sky covered by clouds in the following altitude bands: 0-2 km, 2-6, 6+, and total
- `10_metre_[u/v]_wind_component` - The [eastward/northward] component of wind speed measured
- `direct_solar_radiation` - The direct solar radiation reaching the surface on a plane perpendicular to the direction of the Sun accumulated during the hour, in watt-hours per square meter.
- `surface_solar_radiation_downwards` - The solar radiation, both direct and diffuse, that reaches a horizontal plane at the surface of the Earth, accumulated during the hour, in watt-hours per square meter.
- `snowfall` - Snowfall over hour in units of meters of water equivalent.
- `total_precipitation` - The accumulated liquid, comprising rain and snow that falls on Earth's surface over the described hour, in units of meters.