46040 - Introduction to energy analytics

Assignment 2: Forecasting Electricity Prices

Deadline: Tuesday, April 25, 2023 at 23.59 hrs.

The goal of Assignment 2 is to become familiar with statistical and machine learning tools for extracting value out of data, and apply them for forecasting.

Our specific goal is to design a forecasting tool for electricity prices. If you are able to forecast prices, then you can combine this with your optimization tool from Assignment 1 to determine the appropriate day-ahead bidding strategy for your battery.

- Who are you? A battery owner.
- Where are you? Somewhere in Denmark (zones DK1 or DK2).

You will use the electricity spot prices from 2019 until 2021 and a set of provided exogenous variables. These can be found in the following files:

- elspotprices_19to21.csv
- production_19to21.csv

Electricity spot prices are provided in DKK/MWh. Work only with UTC time, to avoid any problems with gaps or extra hours due to the change of summer and winter time. You can use the following code to load your data:

```python
import os
import pandas as pd

# Load the data
price_path = os.path.join(os.getcwd(),'elspotprices_19to21.csv')
df_prices = pd.read_csv(price_path)
df_prices['HourUTC'] = pd.to_datetime(df_prices["HourUTC"])

# Remove timezone information
df_prices["HourUTC"] = df_prices["HourUTC"].dt.tz_localize(None)

# Sort values
df_prices = df_prices.sort_values('HourUTC')
df_prices = df_prices.reset_index(drop=True)

# Load the data
Exog_path = os.path.join(os.getcwd(),'production_19to21.csv')
df_exog = pd.read_csv(Exog_path)
df_exog['HourUTC'] = pd.to_datetime(df_exog["HourUTC"])

# Sort values
df_exog = df_exog.sort_values('HourUTC')
df_exog = df_exog.reset_index(drop=True)
```

Logistics for the submission of your report:

1. Maximum length of the report: **6 pages, excluding the frontpage**
2. Upload one report per group as a .pdf in Assignment 2 at DTU Learn.
3. Please **mention how you shared the workload** either on the frontpage or at the end. This is necessary by the DTU guidelines.
4. You must also **provide running-code**, i.e., the user should be able to press run in your **Main.py** file and obtain answers to all the following questions. Please use short comments in the code to indicate what each part of the code does.
5. Upload all your code files in a single zip file.
6. You need to upload your report and code on DTU Learn the latest by **Tuesday, April 25, 2023, at 23:59 hrs**.
7. Any unjustified (i.e., without a serious and properly documented reason) late submission will result in a penalty of 10 % for each day.

Problem description

General Notes:

- Training and Testing Datasets:
  - **Training Dataset: 1/1/19 until 30/11/21**. You can use the whole or only part of the training dataset. Explore different options and see what gives you a good performance.
  - **Test Dataset: 1 December 2021 until 31 December 2021.**
    - **Hour-ahead prediction:** 1st Dec 2021 00.00-01.00 hrs; then get the real value of hour 1, include it in your input, and predict hour 2, i.e., 01.00-02.00 hrs. Do that for 24 hours, until 23.00-00.00 hrs, and for every day until the end of the month, i.e. until December 31st, 23.00-00.00 hrs.
    - **Day-ahead prediction:** 1st Dec 2021 00.00-23.00 hrs (all 24 hours at once; no update of the inputs as the day progresses); then get the real values for Day 1, include them in your input, and predict Day 2, i.e., 2nd of December. Do that for every day until the end of the month, i.e., until December 31st, 23.00-00.00 hrs.
- You do not necessarily need to train different ARIMA models to predict the hour-ahead or the day-ahead prices. You can use the same model for both options or different ones, it is up to you.
- **Exogenous Variables:** The exogenous variables we have available are for DK1 and DK2. Regardless of the area you are looking at, you can use these exogenous variables as additional features for your forecast. Besides those, you can also use any calendar features or transformations you think they help you achieve a better performance.
- Evaluation of the Assignment: The grade does not depend so much on the forecasting quality/performance you achieve (i.e., achieving the lowest RMSE). It depends more on the overall thinking and process you followed.

Task 1: Develop an ARIMA model to predict electricity prices. Your goal in all following tasks is to achieve the best possible performance. **In all sub-tasks, report the RMSE values you achieve with your models.** Pick a zone (DK1 or DK2) of your liking.

1.1 Use <u>NO</u> exogenous variables in your model and make **hour-ahead and day-ahead prediction** for your testing dataset. You can use a seasonal ARIMA or FourierFeaturizer and any data transformation you want in your model, but no features from the production_19to21 dataset. Establish a suitable persistence forecast, and report the RMSE values in all 4 cases (your model and persistence, and for both hour-ahead and day-ahead). Briefly discuss your results and the difference between the two forecasting cases.

1.2 Add any exogenous variables you want (maximum 3) and repeat the process (choose/optimize your model and evaluate it for the **hour-ahead and the day-ahead prediction**). What exogenous variables helped you improve the prediction and how did you choose the specific ones? Compare your results with those from task 1.1 and briefly discuss them.

Task 2: Develop a temporal Convolutional Neural Network (tCNN) to predict the electricity prices. Your goal in all following tasks is to achieve an as good performance as possible.

2.1 Pick a maximum of 3 exogenous variables and **perform the hour-ahead prediction** for the training dataset. Report the RMSE value you achieve with your model and compare with the persistence RMSE you established in Task 1.

2.2 Pick a maximum of 3 exogenous variables and **perform the day-ahead prediction** for the training dataset. Report the RMSE value you achieve with your model and compare with the persistence RMSE you established in Task 1.

Task 3: Compare the performance of your ARIMA and your tCNN models. What do you observe?

3.1 In your comparison, please consider the following points:

- How do the two models compare in accuracy when you perform the **hour-ahead prediction** and how when you perform the **day-ahead prediction**? Do you observe a difference? What are the reasons for the differences in performance you observe in each case?
- How much time did it take you to train the models? Is computation time comparable? If not, what are the differences that led to a different computation time?
- How many parameters did you have to tune in each model? Which model did you find easier to train?
- Any other point you consider relevant for the comparison.

3.2 Recommendations: Assume you own an engineering company that develops software solutions for energy forecasting. A small energy trading company comes to you and asks you to develop a tool for them that can predict electricity prices. Considering your assessment, what kind of model would you advise your client to buy/develop for day-ahead predictions and why?

Task 4: Combine the forecasting tool that predicts **for the next 24 hours** you developed with the optimization tool for the optimal bidding strategy for batteries you developed in Assignment 1, and calculate the total profits for the whole month of December 2021.

Consider the following for your battery system.

- The battery's state of charge can vary between 10% and 100%.
- The battery has a power capacity of 1 MW and an energy capacity of 2 MWh.
- Your charging and discharging efficiency is equal to 95%, i.e., $\eta_c = \eta_d = 0.95$.
- Each day you optimize the schedule of the battery between 00:00 and for the next 24 hours.
- You buy and sell energy at the spot price (do not consider any taxes or tariffs).

- Your state of charge at the start and end of each daily optimization you perform is equal to 50%.

4.1 Calculate the profits you achieve during the training dataset by using the bidding model from Assignment 1 and perfect knowledge of prices (you performed this task in Assignment 1). Next, use the forecasting model you developed and has the highest accuracy (ARIMA or tCNN) for day-ahead prediction, and use the forecasted prices instead of the actual ones to bid. Compare the profits you achieve with your forecasting model with the ones you achieve with perfect price knowledge. What do you observe?

    Note: To calculate profits when you use forecasted prices, you need to do the following. First, establish a schedule by optimizing your battery based on your **forecasted** prices. Once you have a schedule, use the **actual** prices to calculate your profit from that day.

4.2 Use the persistence model for day-ahead predictions you established in task 1.1 to forecast electricity prices, and use those to calculate your profits from optimizing the battery. Compare the profits you obtain from persistence forecasting with the profits you calculated in task 4.1 (perfect knowledge and forecasts via ARIMA or tCNN). Discuss the results and what is the added value of your forecaster, compared to simply using persistence.

4.3 Can you identify the key features you need to predict well in order to deliver a good profit-maximizing bidding strategy for a given day?

Lessons Learnt: During this assignment, there were definitely several issues that came up while trying to answer the questions or until you got your code running correctly. In no more than half a page, please list 2-3 main points that you think you should remember for the next time you design and code a bidding strategy. Please list at least one issue that had to do with coding, i.e., what should you remember to do in some specific way, or avoid, next time you code for a similar task? And please list at least one main takeaway from the design of the optimal bidding with forecasting and how it affected the results you obtained.