31762 - Introduction to energy analytics

Assignment 2: Forecasting Electricity Prices

Deadline: Tuesday, April 26, 2022 at 20.00 hrs.

The goal of Assignment 2 is to become familiar with statistical and machine learning tools for extracting value out of lots 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 or its neighboring areas: Sweden, Norway, or Germany (we avoid Great Britain for this Assignment, as it has still no direct connection to Denmark).

You will need to use elspot prices, which you can find here: `https://www.energidataservice.dk/tso-electricity/elspotprices`. Prices from 2019 until 2021 and relevant exogenous variables can also be found on DTU Learn in Assignment 2 as:

- elspotprices _19to21.csv
- exogenousvariables _19to21.csv

Once you open the file you will notice that it contains the electricity prices (in DKK/MWh) for every hour of the years 2019 to 2021 **(in UTC)** for various price zones. Work only with UTC time. Using time in UTC means that there are no gaps or extra hours, and you can disregard the issue of summer/winter time.

Logistics for the submission of your report:

1. Maximum length of the report: **4 pages + 1/2 page** for "Lessons Learnt" (see last task) + 1 page for bonus questions (optional; see Task X)
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 and obtain answers to all the following questions. The code must be a .py file for Python. Please use short comments in the code to indicate what each part of the code does.
5. You need to upload your report and code on DTU Learn the latest by **Friday, April 8, 2022, at 14:00 hrs**.

Problem description

General Notes:

- **You should use the same price area as the one you used in Assigment 1.**
- For Task 1, we will use the ARIMA, SARIMA, and the SARIMAX models. ARIMA stands for Auto-Regressive Integrated Moving Average, and SARIMA stands for Seasonal Auto-Regressive Integrated Moving Average. The package pdmarima in Python includes instead the SARIMAX model, which is a "superset" of ARIMA and SARIMA. SARIMAX stands for Seasonal Auto-Regressive Integrated Moving Average with eXogenous variables.

- Training and Testing Datasets:
  - **Training Dataset: 1 January 2019 until 30 November 2021.** 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 need to train different ARIMA/SARIMAX models to predict the hour-ahead or the day-ahead prices. You can use the same model for both options and compare how the same model performs for the prediction horizons you set. In a following optional step (see optional Task X), you can train different models (and use different parts of the dataset) to "optimize" the performance for the different prediction horizons.
- For the tCNN, you will probably need to train different models, one for the hour-ahead prediction and one for the day-ahead prediction.
- **Exogenous Variables:** The exogenous variables we have available are for DK1 and DK2. Independent 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. In some sub-tasks you need to limit the number of exogenous variables to 3. If you would like to use information contained in more than 3 exogenous variables, then you can consider adding them together, e.g. you could add OffshoreWindLt100MW_MWh for DK1 and DK2 together and use it as a single exogenous variable.
- 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.
- As mentioned above, you may choose to NOT use all the historical data as the training dataset.
- You should NEVER peak into the future and see the testing data or train upon that! That is why it is called testing dataset and not training dataset!

Based on historical values, you want to forecast the electricity prices for the next 24 hours. The better your forecast is, the better the bidding strategy of your battery can be to maximize your profits.

Task 1: Develop a SARIMAX model to predict the electricity prices. SARIMAX stands for Seasonal Auto-Regressive Integrated Moving Average with eXogenous variables. Your goal in all following tasks is to achieve an as good performance as possible. **In all sub-tasks, report the RMSE values you achieve with your models.**

1.1 Use <u>no</u> exogenous variables and <u>no</u> seasonality (this is equivalent to an ARIMA model), pick the appropriate values for the p, q, d parameters and do the **hour-ahead and day-ahead prediction**.

1.2 Add seasonality (this is equivalent to SARIMA), choose/optimize your model, and perform the **hour-ahead and the day-ahead prediction** again. Do you observe a change in performance?

1.3 Add maximum 3 exogenous variables and repeat the process (choose/optimize your model and evaluate your model for the **hour-ahead and the day-ahead prediction**). What exogenous variables helped you improve the prediction?

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**. Report the RMSE value you achieve with your model.

2.2 Pick a maximum of 3 exogenous variables and **perform the day-ahead prediction**. Report the RMSE value you achieve with your model.

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

3.1 In your comparison, please consider the following points:

- How do SARIMAX and tCNN compare in accuracy when you perform the **hour-ahead prediction** and how when you predict for 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 SARIMAX and tCNN in each case? Is the 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. The company is participating in both the intra-day market, e.g. need to forecast prices for the next 1 hour, and in the day-ahead market. Considering your comparison in the previous task, what tool would you advise your client to buy/develop?

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

4.1 Compare the profits you achieved with your model from Assignment 1 (i.e. a perfect forecast) with the profits you achieve if you combine it with the SARIMAX or tCNN forecasts. What do you observe?

4.2 How do the profits change based on the quality of your forecasts (e.g. SARIMAX vs tCNN, or SARIMAX with hour-ahead vs day-ahead prediction)?

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

(Optional) Task X:

X.1 Train different SARIMAX or tCNN models for the hour-ahead and the day-ahead prediction. When doing that, please consider (and address when you write your report):

- Did you need to use the same training dataset for both SARIMAX (or tCNN) models (hour-ahead and day-ahead)? Did you achieve a better performance if you used different training datasets? What kind/type of data are the most appropriate to achieve a good performance in each of the test datasets?
- Did you use the same hyperparameters parameters for both SARIMAX (or tCNN) models (hour-ahead and day-ahead)? Did you achieve a better performance if you used different hyperparameters? What are your insights? How do the hyperparameters affect your model performance?

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 code a bidding strategy or you have to design such a 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 a bidding strategy? And please list at least one main takeaway from the design of the optimal bidding and how it affected the results you obtained.