ETC3555 – Project 2022

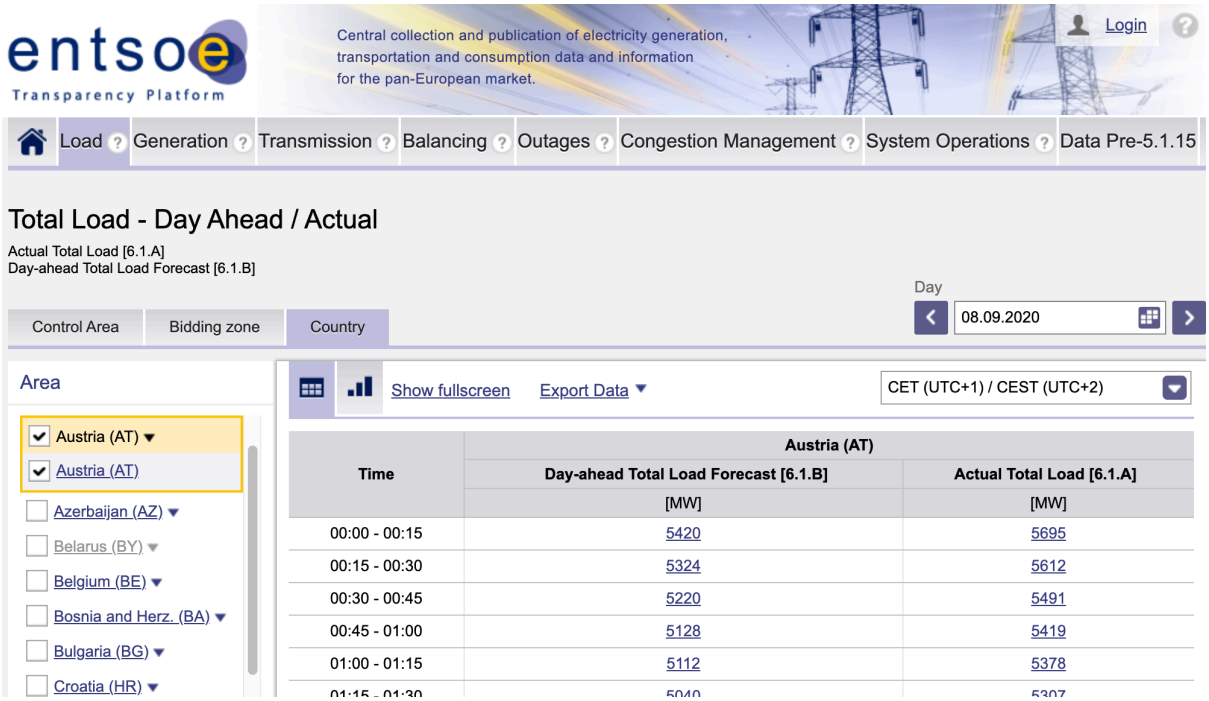
The goal of this group project is that you apply the concept learned in the lecture and in the lab on a real data-set. There are 2 Data-Sets to choose from.

1. Wikipedia click prediction: This data-set has daily visits per Wikipedia page and the goal is to forecast the future visits. **Note: for education purposes only**

2. Electricity demand prediction: This data-set has aggregated electricity demand data by hour for European countries. The data is provided as "raw" requiring significant more data-wrangling and matching skills compared to 1. We will take this into account when assessing your project. **Note: for education purposes only**

ETC5555 – Project 2022

1. Electricity demand prediction: This data-set has aggregated electricity demand data by hour for European countries up to May 2021. The data is provided as "raw". In addition you are required to benchmark your results against the provided benchmark data from https://transparency.entsoe.eu/ The data is not provided to you, part of the task is to figure out how to download it and incorporate it into your workflow. Hint:



2. Same as number 1 for ETC5555 with the additional challenge that you download the most recent data and try to predict electricity load during an energy crisis. (Note: this is more difficult, select it if you feel up for the challenge)

The Task:
You have to build one neural network model with Keras and one benchmark model of your choosing and compare the prediction performance. It is your choice to select a loss function, a train test split, and how to evaluate the performance of your final model. The goal is not to build the best model possible, more the process how you got to your result and document this. You need to submit a short report (max 5 pages), a presentation (max 10 slides) and give a presentation (7 minutes) in week 12 during the lecture. Every team member needs to present. If according to allocate+ you are assigned for in-person you have to present in person (every team member needs to cover a few slides), if you are allocated for online tutorials you need to present via Zoom. (If you get sick or any other special consideration reason on the presentation day, different arrangement will be made)
Remember: Cite academic papers if you us the same architecture, loss functions etc.

Project report and presentation
The data analysis report can be a maximum of 5 pages and must abide by the section structure described below.

Section 1: Model

This section presents the learning problem and the models you have considered. Some of the questions you could answer include “what are the parameters/hyperparameters?”, “how do you optimize these parameters?”, “explain how these parameters control the complexity/flexibility of the algorithm”, “does your method scale well with large data sets?”, etc.

Section 2: Experiment

This section describes the data set you have chosen or got assigned and the experiment you have performed, including a justification of your choices (evaluation metric, optimization procedure, etc). How did you select the benchmark method?

Section 3: Results and Discussion

This includes for example graphs and tables, as well as a discussion of the results. You should also present your model fitting, diagnostics, etc.

Section 4: Conclusion

This includes a summary of the findings.

Grading

Total points: 100
- Report: 80

- Presentation: 20

For your report, you should clearly explain what you have done, using figures to supplement your explanation. Your figures must be of proper size with labelled, readable axes. In general, you should take pride in making your report readable and clear. You will be graded both on statically/machine learning content and quality of presentation.

Deadlines

Do not wait until the last minute.

- October 14th: 4:30 pm: Upload to Moodle (i) your project report, one per group, (ii) the slides for your presentation and (ii) a zip file of the source code R files you used.
- October 19th: Present your slides in the lecture period.