**Deep Learning 2020 - Final Assignment**

**Main Task:**

The main task of this project is to predict the open price of an electricity contract for the next hour. Note that this means that the input of the model should by no means contain any data that is closer than one hour to the opening price of the next hour that you are predicting. Please also bear in mind, that the data consists of multiple contracts, that is, multiple time series, which can be parallel!

**Groups:**

For the project, you should form groups of 3-5 people. Please write all the team members with their full name and student ID on top of the assignment.

As in a real data science project, people can be fired from a team, if they do not perform. A team member who is removed from his/her group will have to perform the exercise by themselves. You must officially announce if you remove a group member via email and this has to take place no later than the 15th of October.

You should measure your predictive accuracy in terms of Mean Squared Error **and** Mean Average Error.

**Details**

You should proceed in terms of the data science loop:

1. Start with descriptive statistics to explore the data-set
2. Build a simple model using traditional statistical techniques to get a baseline
3. Build a more sophisticated model (if needed) using basic data-science techniques
4. Build a “cutting edge” model given the question at hand and what you have learned at the previous steps

(1) to (4), and particularly (4) may involve various sub-steps.

At each step you should write down into the notebook (a) the take-aways, i.e. what you have learned and (b) motivate the particular techniques you are using given the question posed and what you have learned from your previous analyses.

It is crucial that you motivate your modeling by the problem and its structure. Remember that bias is essential to build good models but the bias has to represent the key structural features of the problem.

**Data**

*All data has been anonymized. You must nevertheless not share this data-set with any outsiders*

You can find a notebook showing you how to pull the data and giving you some descriptions of how to preprocess the data here:

<https://colab.research.google.com/drive/1Cf4GwNN1FBSSLS6mm5xX6fXXdeFmznKa#scrollTo=qRNum_21P_Ne>

As the opening price you should take the first price of a contract in a given hour. See the notebook on how to fill prices in case there is no trade.

**Format of submission**

You need to submit the file of the Jupyter notebooks(or Google Colab) that you have worked on as well as a pdf of it and a requirements.txt. It may be that you have multiple notebooks as you are proceeding in a step by step manner and this is ok (submit all of them). Submissions will be made through Canvas. You can write a short report on your findings (no longer than 5 pages). Other than that what you are doing at each step (and why) and your findings should be extensively commented in the notebook.

**Due date:8th of November end of day to be uploaded in canvas**

**Late submission will be awarded a grade of 0. It is thus better to submit unfinished work than to submit late.**

**Grade**

Marks will be awarded based on the following three broad criteria:

(1) The more accurate the predictions are, on the test and validation data-sets, the better. However, you must make sure that you are not making any mistakes here e.g. by including information from the next hour in your independent variables. Also, analyze when the model is good and when and where it does not predict well. Thus, comment when the model can best be employed. Note that you need to validate the predictive accuracy of your model on a hold-out of unseen data that the model has not been trained with. For more details on this, please see the notebook.

(2) How well you proceeded, i.e. did you motivate each step of your analysis well based on the theory of the class and what you have learned in the previous steps of your analyses

(3) The consistency of your take-aways, i.e. what you have learned from the data. For a detailed overview of the evaluation criteria, please see [here](https://docs.google.com/document/d/1gL5r-tN3A1-f6H3kHrsAU8HfRxyy6ot7_y7KWrM1bTU/edit?usp=sharing).

Please make sure that you comment with # on the separate steps of the code you have produced, as well as utilize textual cells for longer discussion

**Plagiarism: The Frankfurt School does not accept any plagiarism. Data science is a collaborative exercise and you can discuss the research question with your classmates from other groups, if you like. You must not copy any code or text though. Plagiarism will be prosecuted and will result in a mark of 0 and you failing this class.**

**After carefully reading this document and having had a look at the data you may still have questions. Please submit those question to the public Q&A board in canvas and we will answer each question, so that anyone in this class can see the answer.**

***Confidentiality: The data obtained for this assignment must not be shared with anyone outside of the class and must be removed from your computer after you have finished the assignment.***