Meeting with Jeroen and Baldiri

Main question: is the way we approach the timestamp predictions the right approach?

* Try getting an indication by simply predicting the gaspower by just looking at the past days. This will give you a baseline as a reference for the errors we get when applying the neural networks. --> this will help actually rate the model better, since the prediction mean more to us
* The average resampling question: Jeroen wasn’t amazed (not sure if he understood the comment by Brian)
* Jeroen: try an arima model, this is another type of regression (looked into it already, don’t know if this will work but ok). According to Jeroen there probably is a library that can simply implement ARIMA
* How do you validate timestamp data?
  + Run several experiments: use something like 200 data points, try to predict 201st. Then take first 201 and then predict 202... and so forth
  + Use the parameters that you already used as a starting point, this will save you a lot of training time. When doing this you will still need to update it for the entire set, meaning that you still need to iterate over the model multiple times
* Thoughts about timeseriessplit: Jeroen didn’t really talk about it
* Multivariate regression: adding the hour of the day. Think this is done, but Daan doesnt seem too sure

Adjusting the nerual network:

* When training on the first half of the dataset, the NN will probably not consider the time trends since it is mostly temperature based. Might wanna look into this  
  However, his will probably not work since the weekdays are very different from the weekenddays.
  + Dummy variables might be the solutions: Boleans: Is it night or not? Add a certain weight. This should be done to each hour of the day. Therefore 24 different weights will be added to the network. This can also be applied to the regression we are doing.
  + However, in our case we will need 24x7 weights since the weekends are so different than the weekdays. This might also work for the seasons or something.
* Fourier: Might be more usefull to seek an explanation for the different gaspower levels. It might not be as good to predict the actual gaspower.
  + Baldiri: yeh exactly, that is the question. We were talking about Fourier series, not just the Fourier transformation
* Trying out Fourier and feeding the pattern recognition to the neural network might be a good thing to try.
* Other problem: Jeroen: Having one neural network, which doesnt adapt itself over time, is the best option. This will be fast and cheap.
  + Daan: yeh but all the houses are different...?
  + Jeroen: Well there are neural networks about grammarcorrection that are a simple algorithm, which can identify “how” someone is talking (e.g. spanish). This should not be a problem if done correctly
* Sequence to Sequence: When you want to apply the S2S, then you need to train it for the specifically thing you want to apply to the test-data.
  + Brian is already doing this. Jeroen didn’t fully understand what Brian was doing, it took some time to explain
  + Jeroen on this topic: you probably are overlooking something
  + Baldiri: The dummy variables will probably make a huge different with (all the) neural networks.
* Cofusional(?) network+LSTM cell: This might be another solution. When creating a triple layer neural networks, it might work.
  + Why?: When creating three layers, you take 3 hours (first layer) to get a pattern at that time (for example, T going down). Second layer (takes three outputs of first layer, meaning 9 hours) will probably predict parts of the day (morning/evening/night) and last layer will take these 9 hours to make a daily prediction.
  + The confusional networks might be a good candidate since it doesnt remember everything. Confusionals have no memory. This will be important since otherwise it might remember the vast majority of gaspower and it will apply this to the future houses.
* **Evaluation point**: The most realistic approach is training on the first 50 houses and predict the 60th.
  + Important point: MAKE SURE YOU ALREADY SPLIT THE HOUSES IN 50 TRAINING HOUSES AND 10 TEST HOUSES. IF DOING THIS TOO LATE, THE NEURAL NETWORK WOULD HAVE ALREADY SEEN THE PREDICTION HOUSES. THIS WILL AFFLICT THE PREDICTIONS.
  + Baldiri: That is the second question. First of all we need to answer the first one

Baldiri about the research question/delivirables:

* + 1 reasearch paper about predicting gspower with hourly frequenty
  + Baldiri wants to make a little paper about getting smartmeter data ready for research and data analysis

Next meeting: make a planning about what and when we want to do/achieve with the data analysis od multi regression and neural networks