

Beat the TSO: Forecasting of Energy Prices

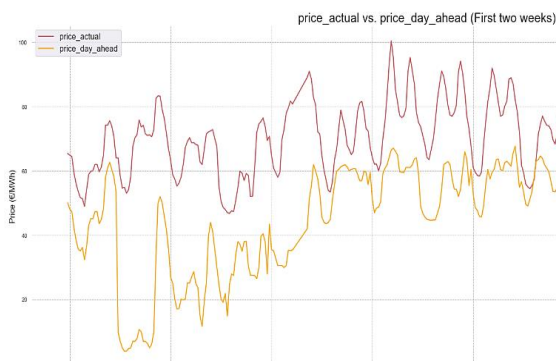
Willem Laumen – Machine Learning and Data Analytics – RWTH Aachen University

Motivation

- Large amount of studies focus on Tackling Climate Change with Machine Learning
- High potential of Machine & Deep Learning in the transition to renewable energy supply
- Forecasting in energy markets is one of the highest leverage contribution areas

The Energy Dataset

- The dataset contains hourly data for energy generation, load and the respective forecasts by the Spanish TSO (float numbers)
- Hourly forecasts on generation, load, and price
- Over 35.000 observations, 29 feature columns



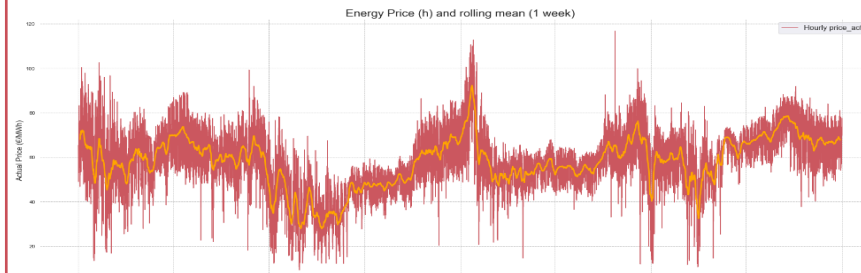
Features:

- Observation time (hourly)
 - 23 generation values
 - 5 TSO forecasts
 - Actual load
 - Actual price
- Time-dependencies in the dataset: Analysis showed seasonal & weekly trends

Methodology

Preprocessing and Feature Selection

- Removal of empty and unimportant columns/cells
- Addition of features that express time-dependency



- Feature Selection: Correlation analysis and gradient boosting to assess feature importance

Forecasting Approach 1: Multivariate Regression

- Neglect time dependency: Fit a linear regression to predict energy prices just based on energy features

Forecasting Approach 2: Time Series Forecasting with Univariate and Multivariate LSTM

- Fit two Long-Short-Term-Memory neural networks
- Multivariate considering selected features: *hour, month, weekday, fossil generation (gas + hard coal), biomass, total load*
- Batch size: 32, 50 epochs, 5 layers for both models.
- Baseline: Regression Forecast, TSO forecast and a naive day-ahead and hour-ahead price forecast

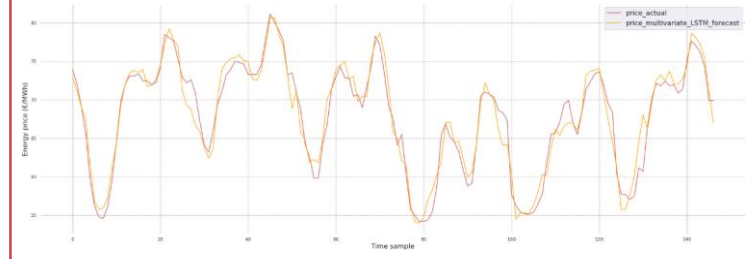
Analysis Outcomes

- TSO baseline: RMSE = 12.311
- Naive day ahead: RMSE = 5.798
- Naive hour ahead: RMSE = 3.242

Regression: Performed worst. Output used as another baseline for LSTM. RMSE = 15.562

Univariate LSTM: Outperformed all baselines: MSE = 4.55 , RMSE = 2.153

Multivariate LSTM: Outperformed all other predictions. Slightly better than univariate: MSE = 4.474 , RMSE = 2.115 ($\Delta_{\text{RMSE}} = 0.038$)



Conclusion & Outlook

- Regression with only electrical features not suitable for forecast of energy price
- Both univ. and multiv. LSTM are very much suitable and performed well
- Research Outlook: Using different ANN (CNN, MLP, Hybrid) & Hyperparam. tuning