

ExecutiveHour's Executive Summary

Members: I. Altiner, N. Fortier, S. Purohit, R. Modarresi Yazdi, Mentor: Julian Rosen

Electricity and energy markets are an integral part of the electrical grid, allowing for the matching of the supply and demand of electricity via market mechanisms. Accurate forecasting of electricity demand and price is crucial given that generated power needs to be consumed immediately and it cannot be stored. In this work, we focus on the day-ahead markets for electricity, where participants purchase or supply electricity to be delivered the following day. This mechanism accounts for 95% of overall transactions (NRG Energy, Inc.) and avoids the volatility of the real-time price.

The chosen market was determined to be the New York City wholesale electricity market, allowing for focus on a relatively small geographical area. We considered natural factors such as electricity demand and local weather conditions. The initial data set was constructed from the hourly day-ahead prices, grid-load data and local weather information (temperature, dew-point etc). This was augmented, based on a preliminary analysis of the time series, by price data at a certain number of previous time steps (in hours). The feature set was further expanded to include categorical temporal information (day of the week, weekday vs weekend etc), as well as monthly average natural gas prices in the country. The final data set contains these features for 27th of October 2020 to the 1st of October 2023.

Motivated by the preprocessing stage, the baseline was chosen to be a simple model where the predicted values for day-ahead price in the next 24 hours is the day-ahead price of the previous 24 hours. From here the team pursued different model classes: dense neural networks, convolutional networks, LSTMs, boosted decision trees and an autoregressive integrated moving average (ARIMA) model. The overall performance of the different models are quite close to each other, with convolutional and dense models doing marginally better than the others. The associated RMSE for these models, approximately (20, 20) \$/MWh for validation and test sets respectively, is slightly lower than the baseline results of (21, 20.4) \$/MWh.

This exploratory analysis has shown that models using simple time-forecasting features can beat a robust baseline model. The Neural Networks and Boosted Decision tree models could benefit from more precise exogenous data, such as actual industrial natural gas prices in the region, renewable energy generation history (and capacity), as well as load predictions provided by energy providers. These models could then be expanded to provide industrial actors insight into the advantage (or disadvantage) of buying energy in advance as opposed to buying it in the real-time energy market. Without exhaustive exogenous data, however, it is clear that most models, as different as they may be, only outperform our baseline predictor by ~5%.