

David Ribberholt Ipsen - Reel - Consumption forecast case

## Code

---

The code and plots are found by running `main.py`. Note that all plots are interactive for zooming, hovering etc. For visually inspecting the predictions, I expect the user to zoom in and browse him/herself. I've also converted the file to a rendered notebook in `main.ipynb` in case the `.py`-file doesn't work for you.

## On modelling

---

(See explanation of implementation below)

If you ask me, the real art of the modelling job is *not* being able to do the best within this narrowly defined problem, but to think outside of the box.

My general approach to such a problem (before doing any ML) is to try to understand the problem in depth:

1. What *causes* consumption? What *correlates* with consumption?
2. What of this data do we have: Consumption measurements (3d delayed)
3. What of this data don't we have, but could possible get:
  1. Weather data
  2. Holiday schedules - from client or national.

Questions to the scope:

1. Why do we have to bid into the day-ahead market a week ahead? If we could get around it, it would open up for a new variety of classical timeseries models, which we could add-on to our ML model.
2. What error metric do we want to minimize? MSE, MAE, etc. the choice depends on (assumptions on) imbalance price function.

## Short explanation of implementation

---

I implemented and evaluated 3 basic prediction algorithms:

1. A mean model (simplest possible baseline)
2. Linear regression (simple baseline)
3. XGBoost (Good all-round ML model)

XGBoost performed slightly better than the Linear Model on all metrics, but also had high variance in its predictions.

Next steps:

1. Hyperparameter tuning
2. Adding more features
3. Testing more sophisticated models

## Why I chose these models

We always want to start simple and build on top. XGBoost (from the CART family) is well-recognized in the time series domain where Deep Learning is still trying to find its way in.

This case is a classical timeseries problem. We have just one timeseries from which we want to predict future values. But due to the restrictions of the problem, we can't really use classical methods like ARIMA. By restrictions I refer to

1. We only forecast once a week
2. And thus we forecast a week ahead
3. (In this case) there are no other explaining variables

Ideally, we could add-on a timeseries model for on-line corrections of the forecast, if the horizon was shorter and frequency higher. Secondly, we would investigate the relation between consumption and other variables such as weather and holidays.