Final project: Demand Forecasting

Demand planners in the industry use commonly tools such as Microsoft Excel to forecast. Although Excel has many strengths and is a flexible tool, it is not well-suited whenever dealing with large amounts of time series. Demand planners struggle reviewing and manually adjusting those forecasts. As a result, they rely on aggregate level adjustments. In many cases, the aggregation is prone to mistakes and poorly reflects the underlying statistical process. Data Science can mitigate this process and make the prediction more precise.

Usually, the demand forecasting setting is to estimate $\boldsymbol{q_{t,i}}$ – the needed amount of a particular item $\boldsymbol{i}$ by the end of month $\boldsymbol{t}$. For many supply chain problems, the definition is extended to estimate $\boldsymbol{q^{f}_{t,i}}$ – the total quantity to be delivered by the end of month $\boldsymbol{t}$ through orders made $\boldsymbol{f}$ months in advance. Moreover, item $\boldsymbol{i}$ exists in a hierarchical relation and can have sibling items $\boldsymbol{s}$, children items $\boldsymbol{c}$ and/or parent item $\boldsymbol{p}$. Thus, adding an additional layer of complexity in forecasting its demand, as often it will depend on its hierarchical relation.

The provided data looks as follows:

- CSV file with a relationship of the items (parent–children)
- CSV file with the planned quantity of particular item $\boldsymbol{i}$ by the end of particular month $\boldsymbol{t}$ made $\boldsymbol{f}$ months in advance
- CSV file with a description of encoding and types of items

The main deliverables are:

1. Generating relevant features for the data set provided, specially exploring the hierarchical relation and the delivery date.
2. Building state of the art machine learning models for regression using neural networks, gradient boosting, and ensemble random forests and identify the best performing models.
3. Assess models using the out-of-sample validation data and report the accuracies.
4. Discuss the performance of the models