Demand forecasting

1. Problem description

The following task was part of the Data Mining Cup 2020 supported by prudsys.
An established retailer wants to optimize its inventory planning to not only significantly reduce storage space, but also its costs and need for logistical operations. It plans to restock its inventory every other week and only keep in stock the items that it has actually sold during that period. It is important to point out that some products will be promoted for limited periods of time. Products that are promoted during the simulation period will be earmarked. However, the transaction data needs to indicate whether a product is being promoted during the training period. Finally, the model does not need to be able to respond to price changes during the simulation period. To simplify matters, prices will not be changed during the period. In order to create this model, the teams obtain information about the exact time of every transaction during a period of six months and about other features that describe the products.

2. Data

Real anonymized data in the form of structured text files (csv) is provided for this task. There are three individual files containing master data ("items.csv"), transaction data ("orders.csv") and an info file ("infos.csv") for the simulation period. Here are some points to note about the files:

- Each data set is on a single line ending with "CR" ("carriage return", 0xD), "LF" ("line feed", 0xA) or "CR" and "LF" ("carriage return" and "line feed", 0xD and 0xA).
- The first line (top line) has the same structure as the data sets, but contains the names of the respective columns (data fields).
- A list of all the column names, which occur in the appropriate order, can be found in the "features.pdf" file as well as brief descriptions and value ranges of the associated fields.
- The top row and each data set contain several fields, which are separated from each other by the "|" symbol.
- Floating point numbers are not rounded. "." is used as the decimal separator.

The "items.csv" file is a master data set that contains descriptive features. The features may be categorical or numerical. The list of features is explained in the "features.pdf" file. Each data line contains the description for one single item. In addition to other information, the "orders.csv" file contains every order with its dedicated timestamp for the 6-month period. Each line displays

one transaction for one single item. All the attributes are described in the "features.pdf" file. The "infos.csv" file contains a list of all the items, their current sales price as well as the dates of scheduled promotions during the simulation period. Every line contains the price and promotion dates for one single item.

3. Task

Analyze and visualize the data in an appropriate way. Assess the variability of product demands and predict the demand for every product over the two-week period by selecting an appropriate forecasting method.