# Rossman & LalaProject charter

### Background

Currently Rossman (the iconic German drugstore chain) is thinking about **improving** their **sales forecasting tool**.

Quick facts: • owner-managed, internationally active family business

• 56,200 employees in Europe and 4,088 branches (2,196 in Germany).

• Turnover:10 billion € (2019).

The marketing departments is already using a BI tool (Microsoft Dynamics) and is doing the forecasting “manually” using some very basic MS Excel spreadsheets. Now the Marketing responsible wants to:

* verify the accuracy of the previous forecasts through a back-testing
* see if it is possible to improve the current system
* run a pilot project customized for the German stores and, if successful, extend it to the other locations (Poland, Spain, Hungary, Czech Republic, Turkey, Albania and Kosovo).

The company has two options:

- buy the integrated Azure Machine Learning services with the forecasting model

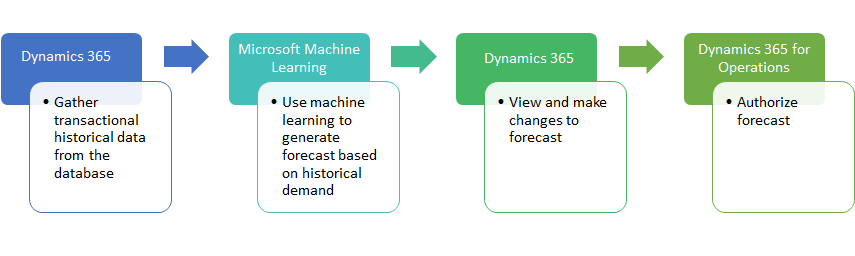
- or to buy a customizable tool from a third vendor.

The Azure Machine learning services would have the following advantages:

- they are easily integrated in the suite of Dynamics 365 with little costs of integration.

- they use the following statistical models ARIMA, ETS, STL, ETS+ARIMA, ETS+STL, ALL.

- they can remove outliers from historical transaction data when calculating a demand forecast



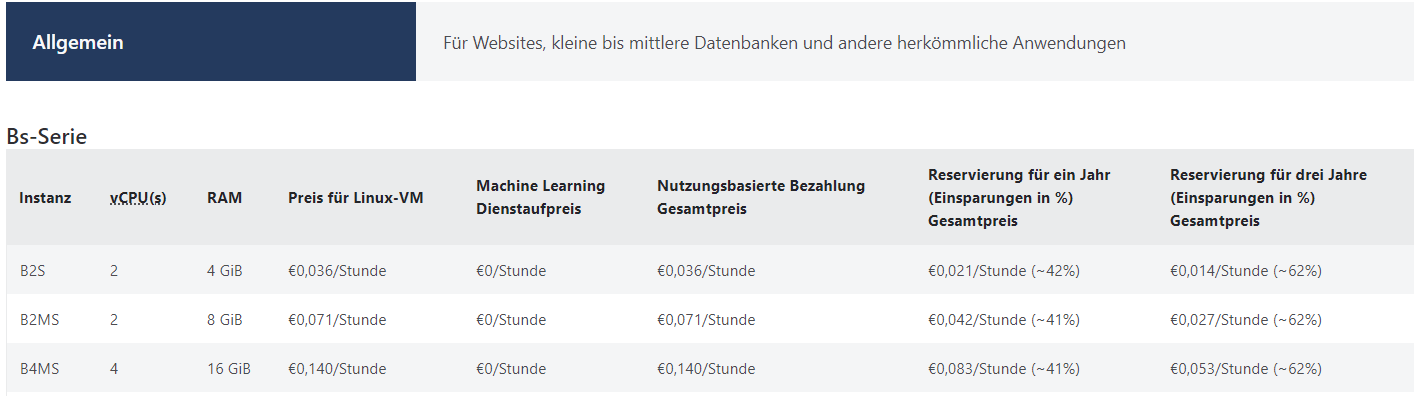
However:

- there are some running cost for using Azure Machine Learning server and Services: currently, the rates are low but there is no guarantee for similar cost in the future.

- the models are not customizable, are not inspectable (black box) and are not state-of-the-art

- the marketing personnel won’t be receiving any training and won’t know how these models are working

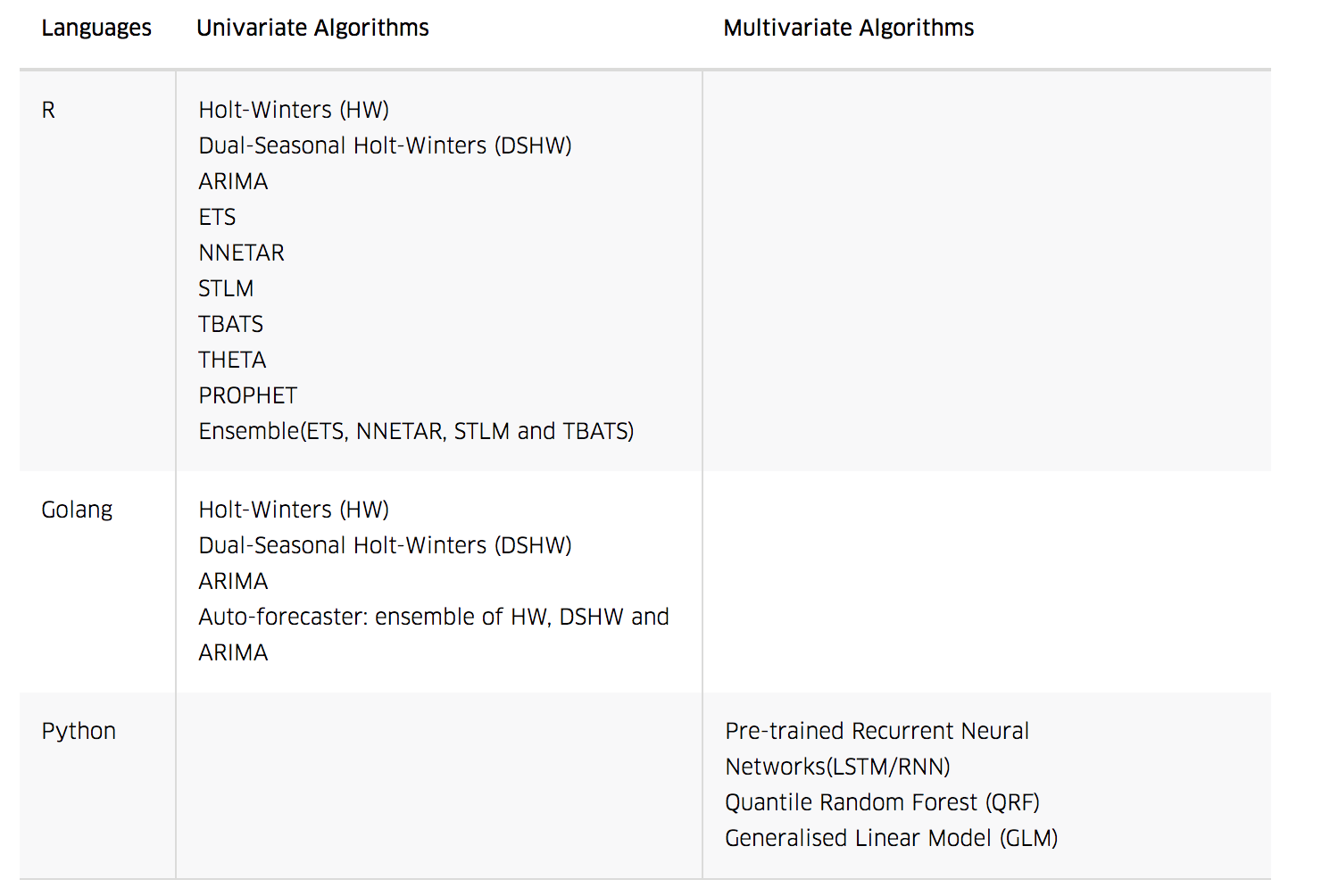


[[1]](#footnote-1)

For this reason, the company is considering the idea to let develop a tool by a third-party vendor and invest some money for acquiring the knowledge necessary for the maintenance and fine-tuning of it.

On the market there are available a lot of free-source algorithms, like the following ones:

|  |  |
| --- | --- |
| Classical & Statistical | Machine Learning |
| Autoregressive integrated moving average (ARIMA)  Exponential smoothing methods (e.g. Holt-Winters)  Theta | Recurrent neural networks (RNN)  Quantile regression forest (QRF)  Gradient boosting trees (GBM)  Support vector regression (SVR)  Gaussian Process regression (GP) |



**Data to be used:**

The tool should extract from the DWH the raw data (sales for each day in each department store, active promotions) and should enrich it with additional information external regarding such as:

* holidays (state holidays, school holidays)
* weather data
* macroeconomics indicators like:
  + GfK Consumer Confidence (measures the level of consumer confidence in economic activity)
  + Consumer confidence index (CCI)
* demographic indicators (new-borns, …)

The tool should:

* identify the characteristics of the time series (presence of trend, seasonality, serial correlation, heteroscedasticity, breaks points),
* take into consideration the external factors
* forecast the sales – possibly an ensemble of the results of different models

### Scope

Scope of this project is limited to:

* the statistical analysis of the daily sales of all 1115 Rossman stores for the period “2013-01-01” - “2015-07-31”
* the realization of a POC with use of Power BI and R & Python code that will consists of:
  + graphical visualizations (up to 6 Power BI slides)
  + the forecast matrix (daily forecast for each store for the following 6 weeks)
  + cross validation

The sales data to be used will be queried from the internal DWH with the collaboration of the customer.

The macroeconomic indicators and new-born statistics will be downloaded “one off” from the web by the software house.

Weather data won’t be use for the realization of the POC since web scraping is needed. However, they will be object of a possible extension of this project.

### Goals

The goals of the project are the following ones:

• dataset preparation / selection

• time series analysis

• model selection and cross validation (tests with different series and model calibration)

• POC tool selection & implementation

• test with the final users

• teaching how to use the tool, how to retrieve new data, how to fine-tune the model

### Future steps:

For the on-going use of the tool, it would be necessary to implement the following activities:

• data extraction/ integration for extracting the data from the DWH



• web scraping for the weather data

• web scraping for the external macroeconomic data

• extension of the project for the other European countries

### Metrics

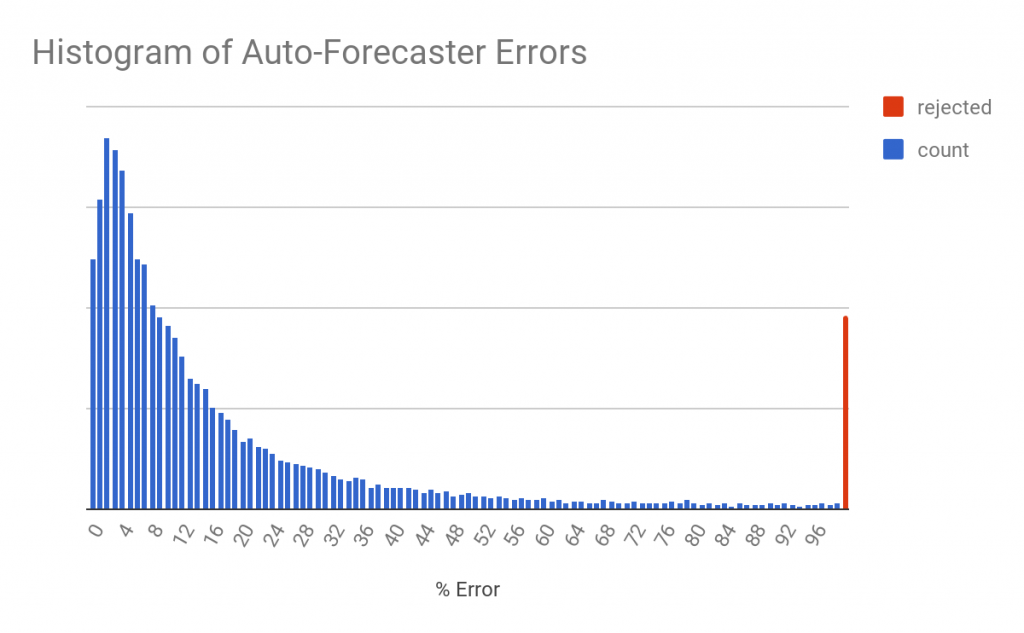
The actual system of forecasting is not performing a cross validation.

As cross validation metrics for the forecasts, the Mean Absolute Precision Error is proposed because of its simplicity and the widespread use.

 (A = Actual; F = forecast)

The proposed tool should reach the following target performance:

* **MAPE = < 50% in 95% of predictions**
* **running time** for the forecasts of all stores (6 weeks horizon) **<= 2h**



### Personnel involved

The software house will be engaging:

1 Project Manager

2 Data scientists (one specialized in statistics, the other in data management BI tools),

1 Account manager

1 System administrator

The Client will be engaging:

- Marketing department (analyst + Manager)

- Data Management analyst & CTO

### Key Stakeholders (other than personnel)

Client [Rossman]

Sponsor [Sales director]

Project manager [Claudio Contini]

Project team members [name], [name], [name], [name].

### Project Milestones

|  |  |  |  |
| --- | --- | --- | --- |
| Milestone | Start date | End date | Invoice date |
| M1 data extraction and time series analysis are performed | 1 month after pre-sale contract | start + 1 month | NA |
| M2 model selected and the cross-validation metrics is satisfied | M1 finish | M1 + 1 month | NA |
| M3 POC tool is selected | M1 start | M2 + 2 weeks | NA |
| M4 POC is realized | M3 finish | M3 + 1 month | NA |
| M5 test with the final users – UAT signed off | M4 finish | M4 + 1 month | end-of-month +1 month |

### Project Budget

The main project expenses are labour costs for the analysis and POC realization. The POC tool will be implemented using open software which doesn’t require any monthly costs. All the analyses and test of the tool will be done via virtual machines running on the customer’s server.

### Data Architecture

• the customer should provide with 2 virtual machines (one for each analyst) running Windows 10; MS Office Suite and MS Dynamics installed; the user should have privileges for downloading/installing R studio/Anaconda and the necessary packages

• the web access should grant access to github and other programming web sites and for downloading external regressor for the analysis

• The raw data regarding the sales will be queried via SQL on the DWH (Microsoft Dynamics)

• all the documentation and files for this project will be saved in a folder on the customer’s server

### Constraints, Assumptions, Risks and Dependencies

**Constraints** [data confidentiality agreement signed off by the whole software house team; the project manager should be in copy with all communications “customer ↔ software house”]

**Assumptions** [the project will start will with one month notice; the customer will provide within the start date 2 virtual machines with the access on the DWH and all the required privileges to access the data; the customer will be answering to all the questions of the team within 2 working days]

**Risks and Dependencies** [Data not available, performance target not reached, confidentiality breach lawsuit]

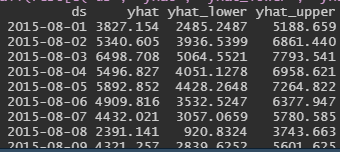
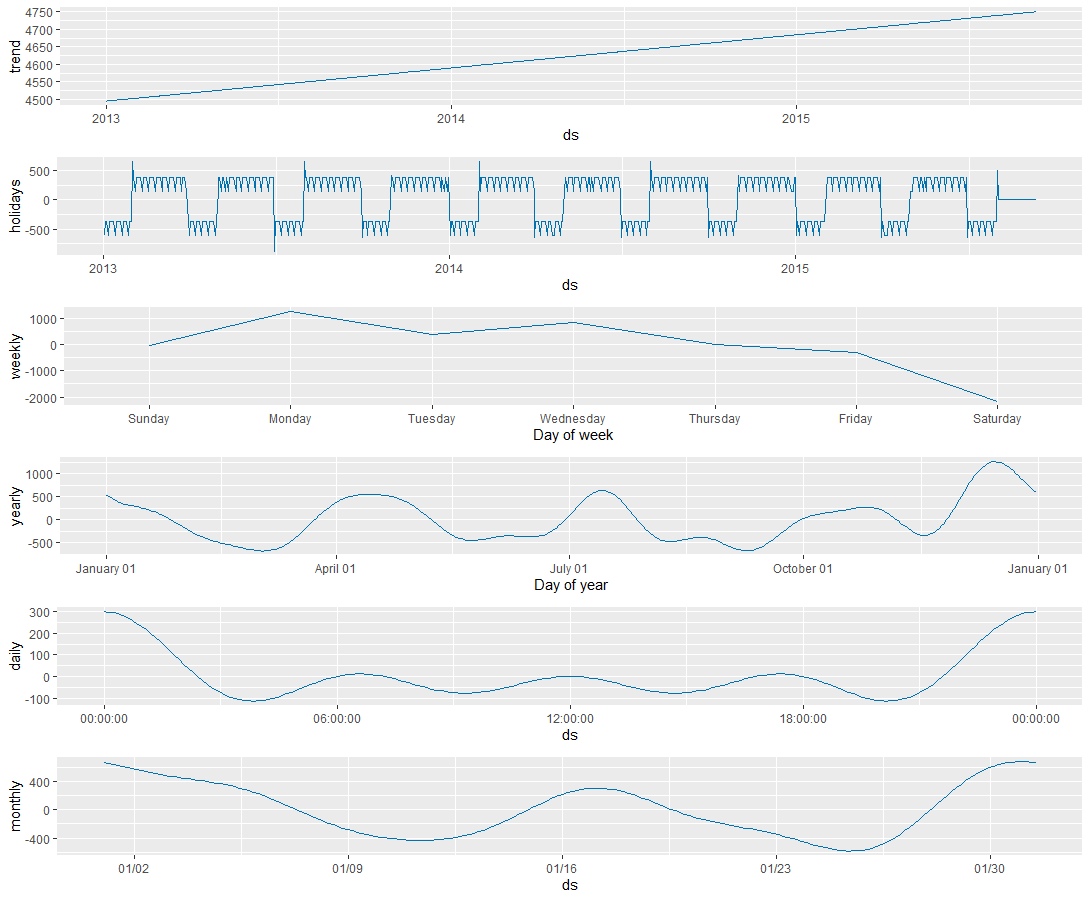
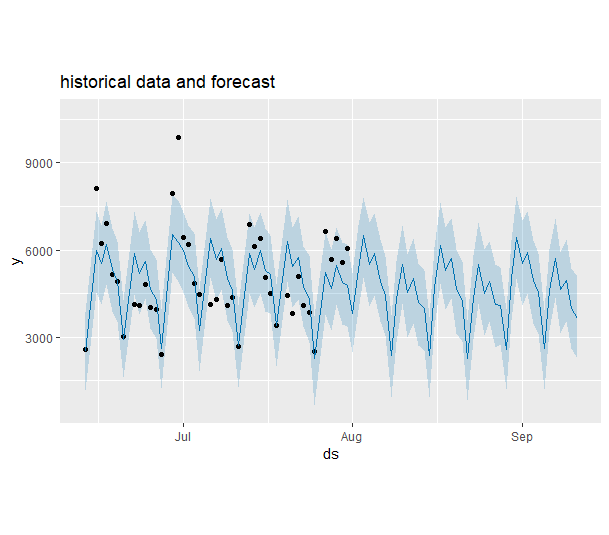
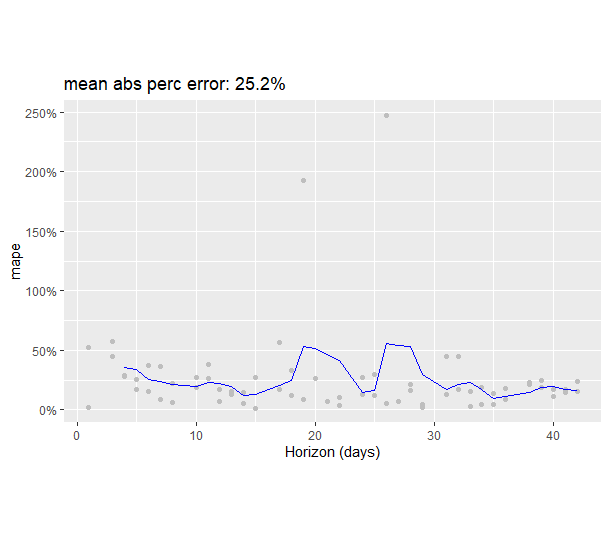
**Approval Signatures**

[Marketing Director Germany], Project Client [Global Sales Director], Project Sponsor

[Claudio Contini], Project Manager

# 2 Proof of Concept

Matrix with the forecast, plot of the different components and the cross validation

# 3 Full-fledged project Plan

1. https://azure.microsoft.com/de-de/pricing/details/machine-learning/ [↑](#footnote-ref-1)