PERSONAL

„Data Mining: My method of choice for practically all analysis situations"
A field report from Josef Schmid

Since Version 4, I have been using SPSS Modeler software almost daily and it never ceases to impress me. IBM SPSS Modeler, the leading data mining solution, is ideal for optimizing marketing campaigns, recruiting new customers, customer loyalty, cross and upselling, minimizing risks and detecting fraud.

Extensive data conversion possibilities

Data conversions and the collation of different data sources are increasingly important for Advanced Analytics: data sources have practically never been so easy to use. Likewise, creating new fields is often one of the most important factors of success for a good model. In this field, SPSS Modeler is extremely powerful. I can't remember a single time when a necessary conversion wasn't possible. With string, data and offset functions, Modeler seems to do whatever you want it to do. This has helped be out of a "spot" many times.

Time series – an extremely powerful algorithm that puts demands on data

The automatic creation of time series, i.e. the creation of forecasts, is a relatively unique feature of Modeler. I know many customers who successfully use this algorithm and get astonishingly accurate forecasts. Modeler can automatically and very quickly create individual forecast models for a large quantity of time series; 1,000 models in just a few seconds!
For time series, data must be processed in a particular way: the time must be in rows, the cases (products, customers, etc.) must be in columns. Modeler converts data into the right format without any problem – the restructure node combined with an aggregate node takes care of this in no time at all. But: Eating creates an appetite. In a real project, several million models have to be created. This is really not a problem, but in this particular case, the entire process must be performed within a database. Modeler supports this perfectly with SQL-Pushback. And normally, this is where the difficulties start: As described above, the cases should be in columns, which means we need several million columns. Certain databases only support 1,024 columns, which is only a fraction of the required amount!

Modeler Scripting – a step to full automation

Good advice is usually expensive, but with Modeler, a solution is quick to be found: as some databases only support 1,024 columns, and this was the case here, the modelling process must be divided into thousands of parts and repeated. This is where one of Modeler's functions comes into use, and in my experience only a few users are familiar with it: Modeler Scripting automatically divides the process into parts.
The implemented solution looks like this: a scheduler initiates a Modeler process on the Modeler Server, which automatically reads the data from the database, divides it into parts, automatically creates an individual forecasting model for each of the several million data sets and then uses this model to write concrete forecasts in the database.

The "Swiss Army Knife" of data analysis

Modeler creates even more surprises with the multitude of its abilities and offers individual solutions, which I didn't believe possible to begin with. Some people underestimate – due to the simplicity of the interface – the vast multitude of functions under the „bonnet". Modeler is really like a Swiss Army Knife of analyses – the impossible becomes possible and you can use it for practically anything! And, last but not least, the accuracy of the forecast models is astonishing. ●

Josef Schmid, Managing Partner Dynelytics AG is extremely impressed with the possibilities offered by this kind of data analysis with self-learning algorithms. His long-standing experience in customer projects in all fields and with a wide variety of issues allow him to find creative and efficient solutions for practically every data analysis problem. For Josef Schmid, it is clear that the demand and fields of application for data mining will continue to grow with new technological developments in the field of Big Data. **j.schmid@dynelytics.com, +41 (0) 44 266 90 30**