

## DATA QUALITY, YOUR KEY TOWARDS GREAT PREDICTIONS

As we all know, there is very classical saying on how a quality output cannot be achieved, if the ingredients to start with, are themselves not poor in quality: 'Garbage In, Garbage out'. This is true specifically for statistical analysis, machine learning and artificial intelligence, where the quality of the data which is used to build up predictive models, will to a very large extent determine the predictive power of the model which is being generated.

Also here, the saying holds true 'bad data in, bad prediction out'.

At Quantforce we make it our mission to bring the power of the artificial intelligence to your fingertips, without the need to be a data scientist or a wizard in mathematics or statistics. That in a certain way, is our contribution to helping you create predictive models which will drive your business forward.

One key element however which sits outside of our control, is the quality of the data file which you will be feeding into the Quantforce platform.

Given that the quality of the prediction, and therefore the impact on your business, depends largely on the quality of the input data which is loaded onto the platform, it is important that you ensure that the data file you will be loading, is of a high enough data quality standard.

Some questions you may ask yourself in preparing the data set are as follows:

- what is the best source from which to draw the data?
- are the data fields always nicely filled?
- has there been any problem in the past with the recording of the data in the system?
- in case 'raw' data is pulled from the system, it may need some additional cleaning before you can use it for predictive modeling. For example, in the case of invoices:
- you may want to exclude disputes
  - are the dates and values always in the right format? if not, you may need to reformat the dates.
  - are there any transactions which would fall outside of normal transactions [e.g. special intercompany invoices, special rebate transactions...]? If yes, you may want to exclude those from the data file

The stress we hereby put on the data quality does not need to be daunting - it is merely an advice to you as a user, to set aside some time to prepare the data files. You may need to seek internal or external help to achieve this, and often the preparation of the data relates to some simple questions and actions to ensure a quality file.

As our expertise has shown over the many years of building predictive models, it is more important to spend time on ensuring the quality of the data, than running hugely complex algorithms on faulty data. Just a top chef in a 3-star restaurant cannot prepare a high-quality menu with poor quality ingredients, so too the Quantforce platform will only be able to generate predictions can really drive your business, if the data is of good quality.

If you need any help, please let us know, we will take care.