Lessons from Automatic Modelling Processes

FICO

The demand for ever increasing numbers of predictive models, across ever expanding areas of business interest, in shorter time periods, with scarce expert resource has led to increased focus on the use of automated modelling processes. For a number of years FICO has been using a series of automatic modeling approaches with the aim of improving model development capabilities, speed, and efficiency within the big data and adaptive modeling space.

The latest iteration establishes a streamlined modeling process to enable the fast turnaround (in a few days) of large-scaled model developments (400+ models) using big data (millions of records and thousands of variables).

This covers all key components in a model development project including: data processing/sampling, model development, model performance evaluation, and out-of-time validation. With this new tool, the entire modeling process including model development, evaluation, tracking and validation can be conducted in dramatically less time and with considerably less manual work on the part of analysts.

Automated modelling design considerations have included the ability to deal with:
•	Multiple data sources - to consider/assess as potential predictive inputs, both in stand-alone form and in combination with each other
•	Multiple performance variables - to cover different credit products, over differing response windows
•	Multiple segmentation schemes - to be evaluated and incorporated in the model build

This session will look to share key learning from developing and using automated modelling processes, using a financial services marketing case study to illustrate areas such as: how to compile performance reports on hundreds of models into a meaningful summary, the need for error scanners to quantify the number of bugs in the process and identify their location within the code, the use of parallelized execution techniques to reduce run times, efficient and robust techniques for assessing thousands of variables, the use of subject matter expert driven model development parameters and out of time validation routines, to ensure appropriate models are produced first time.