



Contribution ID: 34

Type: **not specified**

Enabling Scientists and Engineers to deploy and exploit Data Science in the Process Industries

Tuesday, 18 May 2021 11:55 (20 minutes)

Statistical modeling is, perhaps, the apex of data science activities in the process industries. These models allow analysts to gain a critical understanding of the main drivers of those processes to make key decisions and predict future outcomes. Interest in this field has led to accelerating innovation in the field of model development itself. There is a plethora of different modeling techniques to choose from, from tree-based methods to neural networks, penalized regression (lasso, ridge) and Partial Least Squares, and many more. While theoretical knowledge or experience can sometimes direct analysts towards the technique most appropriate for their specific situation, very often it is not known in advance which method will produce the most accurate model. Even within these individual methods, decisions and assumptions must be made that can have a profound effect on the model's output, and by extension, the ability to control a process with any degree of certainty. This is further complicated by these two common industrial habits: relying on statistical teams that are removed from the subject matter to build the models, and shrouding the entire process in mystery from the perspective of the domain experts. The former risks creating bottlenecks as new data demands continuous refinement of existing models. The latter prevents subject matter experts, who are ultimately responsible for decision-making, from advising on the process. The Model Screening platform in JMP Pro 16 addresses all of this by simultaneously fitting and validating a plurality of techniques through an easy-to-use interface. This allows for model-building methods to be in the hands of the many, thus democratizing data science and integrating it with domain knowledge, and freeing statisticians to be the managers and enablers of these processes.

Primary author: MYERS, Hadley (JMP)

Presenter: MYERS, Hadley (JMP)

Session Classification: Stat Engineering

Track Classification: Data Science in Process Industries