



Contribution ID: 132

Type: **not specified**

## **Data Preparation and Model Evaluation: The Interface of Machine Learning and Statistics**

*Tuesday, June 28, 2022 9:30 AM (30 minutes)*

Although confirmatory modeling has dominated much of applied research in medical, business, and behavioral sciences, modeling large data sets with the goal of accurate prediction has become more widely accepted. The current practice for fitting and evaluating predictive models is guided by heuristic-based modeling frameworks that lead researchers to make a series of often isolated decisions regarding data preparation and model evaluation that may result in substandard predictive performance or poor model evaluation criteria. In this talk, I describe two studies that evaluate predictive model development and performance. The first study highlights an experimental design to evaluate the impact of six factors related to data preparation and model selection on predictive accuracy of models applied to a large, publicly available heart transplantation database. The second study uses a simulation study to illustrate the distribution of common performance metrics used to evaluate classification models such as sensitivity and specificity. This study shows that the metrics are sensitive to class imbalance as well as the number of classes and provides a simple R function for applied researchers to use in determining appropriate benchmarks for their model scenario. These studies are two illustrations of how statistical approaches can be used to inform the modeling process when fitting machine learning models.

### **Keywords**

**Primary author:** Dr MEGAHED, Fadel M. (Farmer School of Business, Miami University (OH))

**Co-authors:** Dr JONES-FARMER, Allison; RIGDON, Stephen E.; CHEN, Ying-Ju (Farmer School of Business, Miami University (OH))

**Presenter:** Dr MEGAHED, Fadel M. (Farmer School of Business, Miami University (OH))

**Session Classification:** INVITED North American

**Track Classification:** Other/special session/invited session