



Contribution ID: 20

Type: **not specified**

A permutation-based solution for Machine Learning model selection

Tuesday, 18 May 2021 11:15 (20 minutes)

In a regression task, the choice of the best Machine Learning model is a critical step, especially when the main purpose is to offer a reliable tool for predicting future data. A poor choice could result in really poor predictive performances.

Fast moving consumer goods companies often plan consumer tests to gather consumers' evaluations on new products and then are interested in analysing these data to predict how these products will perform on the market. Companies therefore need the final Machine Learning model to be as accurate as possible in predicting customers' reactions to new products.

In this paper, by taking advantage of a consumer survey and a brief simulation study, we propose an innovative method for choosing the final Machine Learning model according to multiple error metrics. We exploit nonparametric methods and in particular the NonParametric Combination technique (NPC)¹ and the ranking procedure proposed by Arboretti et al. (2014)², which are flexible permutation-based techniques. Using these tools, a ranking of the considered models based on multiple error metrics can be achieved, so that the solution significantly outperforming the others can be chosen.

1. Pesarin F, Salmaso L. Permutation tests for complex data: theory, applications and software. Wiley. 2010.
2. Arboretti R, Bonnini S, Corain L, Salmaso L. A permutation approach for ranking of multivariate populations. *Journal of Multivariate Analysis*. 2014; 132: 39 –57.

Primary author: Mr CECCATO, Riccardo (University of Padova)

Co-authors: Prof. ARBORETTI, Rosa (University of Padova); Mr PEGORARO, Luca (University of Padova); Prof. SALMASO, Luigi (University of Padova)

Presenter: Mr CECCATO, Riccardo (University of Padova)

Session Classification: DoE and ML for product and process innovation

Track Classification: Data Science in Process Industries