

# Predictive Modeling with *R* and the caret Package

Max Kuhn<sup>1</sup>

1. Nonclinical Statistics, Pfizer Global R&D

\*Contact author: [max.kuhn@Pfizer.com](mailto:max.kuhn@Pfizer.com)

**Keywords:** Machine Learning, Pattern Recognition, Classification, Regression

This course will provide an overview of using *R* for supervised learning. The session will step through the process of building, visualizing, testing and comparing models that are focused on prediction. The goal of the course is to provide a thorough workflow in *R* that can be used with many different modeling techniques. A case study is used to illustrate functionality.

## Outline:

- Introduction (philosophy, case study)
- General Strategies (data splitting, resampling, model tuning)
- Data PreProcessing (transformations, variable filtering)
- Conventions in *R* (OOP, function interfaces, consistency)
- Building and Tuning Models (performance metrics, trees, kernel methods, comparing models)
- Other Topics (as time allows) (parallel processing, variable importance)

## Required Background Knowledge

Basic understanding of *R* (matrices, data frames, functions, etc) is needed. Some basic understanding of regression techniques is helpful.