# Machine Learning with Tidymodels

Level: Intermediate

Duration: 6 hours

Machine learning is the process of applying statistical techniques to gain systematic information about a quantity of interest. We will be specifically focusing on how we can use the {tidymodels} suite of packages to implement these techniques. We cover key reasons for model fitting, such as prediction and inference, on quantitative and qualitative responses.
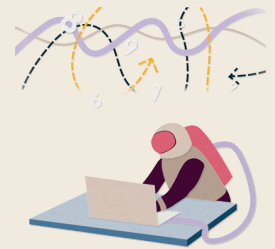
Course Outline

- **Introduction**: A general introduction into statistical modelling techniques and the {tidymodels} R package.
- **Quantitative model fitting**: Using {parsnip} to perform simple and multiple linear regression, allowing us to fit a model.
- **Qualitative model fitting**: Using {parsnip} to perform logistic regression and fit the model.
- **Classification**: Brief introduction to the K-nearest neighbours (KNN) technique for cases where a categoric response has more than two possible outcomes.
- **Model assessment**: Testing the reliability of model predictions and finding optimum models using resampling methods like the validation set approach.
- **Model pre-processing with {recipes}**: Including variable transformations, one-hot encoding and normalisation.

Learning Outcomes

Session 1:

By the end of session 1 participants will...

- understand Machine Learning with statistical models including prediction, inference, regression and classification.
- understand the purpose, aims and benefits of performing machine learning in the Tidymodels framework.
- be able to perform fundamental machine learning methods in the Tidymodels style using {parsnip}.
- know how to perform prediction and classification using linear and logistic regression models.
- be able to classify with K-nearest neighbours models.

Session 2:

By the end of session 2 participants will...

- be able to create training and testing datasets using {rsample}.
- be able to compare models and assess predictive performance using {yardstick}.
- be able to evaluate classification performance metrics such as accuracy, specificity and sensitivity.
- have enough knowledge to apply model pre-processing using {recipes}.
- comfortably create a workflow using the {workflows} package.

This course does not include

- Using {parsnip} to create Linear Discriminant Analysis models.
- Using V-fold cross validation data splits and bootstrap samples for model assessment and parameter tuning.
- Applying penalised regression techniques on models with a high number of predictor variables.
- Using tree-based models in regression & classification problems.

The above are covered in our advanced machine learning with Tidymodels course.

Prior Knowledge

It will be assumed that participants are familiar with R. For example inputting data, basic visualisation, basic data structures and use of functions. Attending the Introduction to R course will provide a sufficient background, but the programming with R will be helpful.

Attendee Feedback

- "Difficult concepts were well-explained and I never felt like I couldn't keep up"
- "Teaching staff was great and really patient in answering questions."
- "The course was brilliant. Theo covered every topic in just enough detail and he had a really engaging style. He created a great atmosphere where everyone felt comfortable enough to ask questions. I have never taken such a comprehensive course that was so well designed (even at Uni)."

Contact

hello@jumpingrivers.com