

# Machine Learning in R

## Course Overview - 1-day course

### Learn to use machine learning algorithms in R

This course is designed to equip you to start using cluster analysis and random forest classification and regression in R. You will build on your basic knowledge of R, gain an overview of machine learning and the machine-learning process, learn the specifics of performing cluster analysis and using random forests and how these can be incorporated into a Power BI workflow.

#### What do I need to know to attend?

- Attended our Introduction to R programming course or have basic familiarity with R. You will not be expected to code unassisted but will learn more and participate more easily if you have at least a fundamental understanding of R syntax.
- Basic familiarity with Power BI is desirable (data import and creating basic visuals).
- Basic understanding of statistics (mean, median, standard deviation, variance).

#### Learning Outcomes

- Generate insights from your data using cluster analysis
- Create predictive models from your data using random forests
- Assess the predictive accuracy of your classification and regression models
- Leverage models to make predictions to guide decision-making
- Incorporate R scripts in your Power BI workflow

---

### Cluster analysis

Learn the basics of cluster analysis and the k-means algorithm and how to implement this method in R. We will use cluster analysis to perform customer segmentation, a commonly used business application of this method.

### Random forests

Learn how to use random forests for prediction with both classification and regression problems. This popular technique is relatively easy to understand and use and is known for its ability to produce relatively accurate predictions for many datasets.

### R Scripts in Power BI

Power BI is an easy-to-use, interactive tool for data visualisation. It simplifies the processes of connecting to large datasets and creating and sharing beautiful reports and dashboards.

Learn how to use your R scripts in Power BI. Data refresh triggers execution of R code, keeping results up to date. Results from R analyses can be displayed in Power BI reports and dashboards and shared within your organisation.

# Machine Learning in R

## Detailed Content

### Introduction

- Introduction to machine learning
- Supervised vs unsupervised learning
- Preparing data for analysis

### Cluster analysis

- Purpose of cluster analysis
- Real-world applications
- K-means
- How the algorithm works
- Data preparation and feature engineering
- How many clusters?
- Performing k-means clustering in R
- Customer segmentation with cluster analysis

### Random forests

- Classification vs regression
- Basics of tree-based models
- The bias-variance trade-off
- From trees to (random) forests
- Ensemble learning: bagging to reduce overfitting and improve predictive accuracy
- The process of supervised machine learning
- Feature engineering
- Splitting data into training and test sets
- Training the model
- Improving the model
- Using the model for prediction
- Evaluating the final model
- Classification vs regression metrics
- The process of creating a random forest model
- Random forests in R
- Classification tree and random forest classification model
- Regression tree and random forest regression model
- Improving the model

### R scripts in Power BI

- Why bring a machine learning model into Power BI?
- Setting up
- Cluster analysis in Power BI
- Random forest models in Power BI
- R visuals in Power BI