

## R Advanced: Predictive Analytics Answering “What will happen?”



**Length:** 3 full days (6 half days)

**Overview:** This course continues to teach about the applications of statistics and data science in R by building machine learning and other advanced statistical models and understanding how they can be used in a business setting. It is designed for students who have a foundational understanding of R and want to explore advanced statistical techniques and machine learning. The course will cover essential topics such as regression methods, time series analysis, and various machine learning algorithms. Students will also learn about model evaluation and the importance of feature selection. By the end of the course, participants will be equipped to develop and implement predictive models, enabling them to make data-driven decisions in a business context.

**Objectives:**

- Apply various regression techniques, including linear, logistic, Ridge, and LASSO, to make predictions.
- Conduct time series analysis using ARIMA and SARIMA models for forecasting.
- Implement machine learning algorithms, including decision trees, random forests, and support vector machines.
- Evaluate model performance using cross-validation and other assessment techniques.
- Utilize advanced simulations and Bayesian models to enhance predictive accuracy.
- Understand and apply ensemble methods to improve model predictions.
- Pre-process data effectively for machine learning applications.
- Explore unsupervised learning techniques, including clustering and neural networks.

### Prediction/Cross Validation introduction

- Linear/Logistic regression
- Ridge/LASSO
- Splines
- Bayesian models

### Advanced Simulations

- More in depth about MCMC
- More difficult/applicable examples

### Time Series

- Setting up data for time series
- AR (Autoregressive)
- MA (Moving Average)
- ARIMA
- SARIMA
- Introduction to forecasting packages (prophet/forecast)
- Prediction for time series

### Machine Learning

- Supervised Machine Learning
  - Pre-processing of data
  - Decision Trees
    - Intuition
    - Feature Importance
    - Advantages/Disadvantages
    - Predictions
  - Random Forest/Baggings
    - Intuition
    - Feature Importance
    - Advantages/Disadvantages
    - Predictions
  - K-nearest neighbors
    - Intuition
    - Feature Importance
    - Advantages/Disadvantages
    - Predictions

- Support Vector Machines
  - Intuition
  - Feature Importance
  - Advantages/Disadvantages
  - Predictions
- Gradient boosting
  - Intuition
  - Feature Importance
  - Advantages/Disadvantages
  - Predictions
- Neural nets
  - Intuition
  - Feature Importance
  - Advantages/Disadvantages
  - Predictions
- Ensemble models
- Stacking models
- Unsupervised Machine Learning
  - Clustering
    - K-means
    - Hierarchical
    - Model based clustering
  - Neural nets
    - Autoencoders
    - Hierarchical
    - Self-organizing maps

### **Prerequisites**

R Beginner: Data Prep & Descriptive Analytics; R Intermediate: Diagnostic Analytics

### **Materials**

- All students will receive slides with lecture material and data and labs.
- Software needed on each student PC:
  - Microsoft Excel 2010 or later (2013 or later recommended)
  - Internet access
  - R (installed and fully operational)
  - RStudio (installed and fully operational) • Related data and lab files will be provided

**Other R Courses Available:** R Beginner: Data Prep & Descriptive Analytics;  
R Intermediate: Diagnostic Analytics