

Data Science -Foundations & Advanced: Data Mining and Predictive Analytics

Course Objectives

This two days course prepare analyst to take the knowledge gained and apply it to their own respective data mining problems, solving them quickly and easily. The lessons learnt will be applicable to areas such as customer analytics, targeted marketing, social media analytics, fraud detection, predictive maintenance, resource management, etc. This course is suggested for analysts and data scientists.

What Will You Learn?

- Perform all common data preparations
- Build sophisticated predictive models
- Evaluate model quality with respect to different criteria
- Deploy analytical predictive models

Target Audience

This course is suggested for analysts and data scientists.

Training Methodology

Hands-on exercise, lecture, group discussion, and case study.

Training Outline

Foundations

[Overview](#)

- Business scenario
- Analytics Taxonomy & Hierarchy
- CRISP-DM & Data mining in the enterprise

[Basic Usage](#)

- User interface
- Creating and Managing RapidMiner repositories

- Operators and processes
- Storing data, processes, and results sets

[EDA: Exploratory Data Analysis](#)

- Loading Data
- Quick Summary Statistics
- Visualizing Data & Basic Charting

[Data Preparation](#)

- Basic Data ETL (Extract, Transform, and Load)
- Data Types & Transformations of value types
- Handling missing values
- Handling attribute roles
- Normalization and standardization
- Filtering examples and attributes

[Building Better Processes](#)

- Organizing
- Renaming
- Relative Path
- Sub-processes
- Building Blocks
- Breakpoints

[Predictive Model's Algorithms](#)

- K-Nearest Neighbour
- Correlations
- Naive Bayes
- Linear Regression
- Rules
- Decision Trees

[Model Construction and Evaluation](#)

- Machine Learning Theory: Bias, Variance, Overfitting & Underfitting
- Split and Cross Validation
- Applying models
- Optimization and Parameter Tuning
- Splitting data
- Evaluation methods & Performance criteria

[Additional Workshops](#)

- Outlier Detection
- Random Forests
- Ensemble Modeling

Prerequisite

Basic knowledge of computer programs and mathematics.