

Quandatics Academy

Enterprise Training for Big Data Analytics

RapidMiner & DataScience: Foundations & Advanced Data Mining & Predictive Analytics with RapidMiner

Course objective

This four 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
- ✓ Utilize more complex functionality of RapidMiner Studio
- ✓ Apply more sophisticated analytical approaches

Pre-requisite

Basic knowledge of computer programs and mathematics.

Training methodology

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

Course Outline: Foundation

1. Overview

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

2. Basic usage

- User interface
- Creating and Managing RapidMiner repositories
- Operators and processes
- Storing data, processes, and results sets

3. EDA: Exploratory Data Analysis

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

4. 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

5. Building better processes

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

6. Predictive models algorithms

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

7. 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

8. Additional Workshops

- Outlier Detection
- Random Forests
- Ensemble Modeling

Course Outline: Advanced

1. Overview

- Business case
- Intro course review
- Loading new data

2. EDA: Exploratory Data Analysis

- Multiple sources
- Joins & Set Theory
- Understanding new attributes

3. Data preparation

- Advanced Data ETL (Extract, Transform, and Load)
- Aggregation & Multi-level aggregation
- Pivot & De-Pivot
- Calculated values
- Regular Expressions
- Changing value types
- Feature Generation and Feature Engineering
- Loops
- Macros

4. Predictive models algorithms

- Support Vector Machines
- K-Means Clustering
- Neural Networks
- Logistic Regression

5. Model Construction and evaluation

- Advanced performance criteria
- ROC plots
- Comparison between models
- Sampling
- Weighting
- Feature Selection: Forward Selection
- Feature Selection: Backward Elimination
- Validation of preprocessing and preprocessing models
- Optimization & Logging results

6. Additional Workshops

- Principal Components Analysis
- Logistic Regression
- Performance (Cost) Model Optimization