

## RapidMiner Training



## Data Mining and Predictive Analytics with RapidMiner

### What is this training about?

This training is a highly practical, application focused four-days course (RapidMiner Basics Pt. 1 & Pt. 2), exploring the possibilities of performing data mining and predictive analytics with RapidMiner. Where Part 1 of the course takes a clean, simplified business example to build a strong foundation, Part 2 explores a similar business case with some of the messiness of the real world added in.

This class resembles a mentor-mentee relationship with the entire group performing as members of a data science team. After successfully completing this course, participants will have a decent understanding of how RapidMiner advanced analytics platform is used, and be able to prepare data and create predictive models in data environments typically found within most analyst positions. Participants will also be ready to extend their knowledge with advanced topics such as text/web mining, operationalization, and big data.

Practical exercises during class prepare the participants to transfer the knowledge gained and apply it to their own data mining problems, solving them quickly and easily. The class labs are hands-on, performed on the participants' own laptops, so they will be taking their actual classwork home to jumpstart their application to the real world.

### Who should attend this training?

This training is ideal for professionals who intend to utilize tools and techniques like data mining, data visualizations and predictive modeling to predict and optimize outcomes and derive value from their data. The lessons learnt will be applicable to areas such as customer analytics, targeted marketing, social media analytics, fraud detection, predictive maintenance, resource management, etc.

After the training, students will have the ability to:

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

This training is the perfect preparation and pre-requisite for the Analyst level certification exam.

## Outline: RapidMiner Basics Pt. 1

- Overview
  - Business scenario
  - Analytics
  - Data mining in the enterprise
  - CRISP-DM
- Basic usage
  - User interface
  - Creating and handling RapidMiner repositories
  - Starting a new RapidMiner project
  - Operators and processes
  - Loading data
  - Storing data, processes, and results
- EDA: Exploratory Data Analysis
  - Data Types
  - Data Hierarchy
  - Quick Summary Statistics
  - Visualizing Data
  - Charting
- Data preparation
  - Normalization and standardization
  - Basic transformations of value types
  - Handling missing values
  - Sampling
  - Filtering examples and attributes
  - Handling attribute roles
- Building better processes
  - Organizing
  - Renaming
  - Relative Path
  - Flow Control
  - Subprocesses
  - Building Blocks
  - Breakpoints
- Predictive models
  - Correlations
  - K-Nearest Neighbor
  - Naive Bayes
  - Linear Regression
  - Rules
  - Decision Trees
  - Importance of attributes
- Model evaluation
  - Applying models
  - Overfitting
  - Splitting data
  - Evaluation methods
  - Performance criteria
- Sharing and collaboration
  - Exporting images
  - RapidMiner Server

For any enquiries, please contact:

**Quandatics**

e: [contact@quandatics.com](mailto:contact@quandatics.com)

m: +6 016 236 2220

( Jaden Teo )

## Outline: RapidMiner Basics Pt. 2

- Overview
  - Business case changes
  - Intro course recap
  - Loading new data
- EDA
  - Multiple sources
  - Understanding new attributes
  - Schema relationships
- Data preparation
  - Joins
  - Aggregation
  - Multi-level aggregation
  - Pivot
  - Set Theory
  - Calculated values
  - Regular Expressions
  - Changing value types
  - Balancing data
  - Outlier detection
  - Feature selection
  - Dimensionality reduction
- Predictive models (sample varies)
  - SVM
  - Random Forest
  - K-Means Clustering
  - Neural Networks
  - Logistic Regression
  - Meta Learning
- Model evaluation
  - Advanced performance criteria
  - ROC plots
  - Comparison between models
  - Lift Chart
- Deployment
  - Sharing data, models, and processes
  - Exporting processes as web service
  - Basics of report creation
  - Managing processes and services

For any enquiries, please contact:

**Quandatics**

e: [contact@quandatics.com](mailto:contact@quandatics.com)

m: +6 016 236 2220

( Jaden Teo )