

# PREDICTIVE MODELLING WITH CLEMENTINE

■ Platform: Windows ■ Duration: 3 days ■ Advanced

## TARGET AUDIENCE

This course follows either the Introduction to Clementine and Data Mining course or the Preparing Data for Data Mining course and is essential for anyone who wishes to become familiar with the full range of modelling techniques available in SPSS Clementine to create predictive models. For people wishing to successfully build such models using Clementine, this course is an essential part of the learning process.

## PREREQUISITES

General computer literacy. Experience using Clementine, including familiarity with the Clementine environment, creating streams, reading in data files, assessing data quality and handling missing data (including the type and data audit nodes), basic data manipulation (including the derive and select nodes), and creation of models. Prior completion of the Introduction to Clementine and Data Mining course is required and completion of the Preparing Data for Data Mining course is strongly encouraged. An introductory course in statistics, or equivalent experience, would be helpful for the statistics-based modelling techniques.

## OVERVIEW

This course demonstrates how to develop models to predict categorical and continuous outcomes, using such techniques as neural networks, decision trees, logistic regression, support vector machines, and Bayesian network models. Use of the binary classifier and numeric predictor nodes to automate model selection is included. Feature selection and detection of outliers are also discussed. Expert options for each modelling node are reviewed in detail and advice is provided on when and how to use each model. You will also learn how to combine two or more models to improve prediction.

## COURSE CONTENT

Following an overview of the main features of Clementine and an introduction to essential terminology, you will proceed logically through the following topics:

- Preparing data for modelling
- Searching for data anomalies
- Selecting predictors
- Data reduction with principal components
- Decision trees
- Neural networks
- Linear regression
- Support vector machines
- Cox regression
- Time series analysis
- Logistic regression
- Discriminant analysis
- Bayesian networks
- Binary Classifier node
- Combining models to improve performance
- Getting the most from models
- Decision list