

Course : Data Analytics with R

data modeling and representation

Practical course - 4d - 28h00 - Ref. DTA

Price : 2520 € E.T.

Big Data Analytics requires mastery of fundamental data processing techniques: statistical methods, classifications, regressions, PCA... This practical course will show you, based on real data, how to use these techniques to build and evaluate models using the R language.

Teaching objectives

At the end of the training, the participant will be able to:

-  Understanding the principle of statistical modeling
-  Choosing between regression and classification according to data type
-  Evaluating the predictive performance of an algorithm
-  Create selections and rankings from large volumes of data to identify trends

Intended audience

Data center managers (Datamining, Marketing, Quality, etc.), database users and business managers.

Prerequisites

Basic knowledge of statistics and R, or have taken the courses "Statistics, mastering the fundamentals" (Ref. STA) and "R environment, data processing and analysis ..." (Ref. TDA).

Course schedule

1 A reminder of the R language

- Data types in R.
- Import-export data.
- Techniques for drawing curves and graphs.

Role-playing

Getting to grips with scripts and Notebooks.

PARTICIPANTS

Data center managers (Datamining, Marketing, Quality, etc.), database users and business managers.

PREREQUISITES

Basic knowledge of statistics and R, or have taken the courses "Statistics, mastering the fundamentals" (Ref. STA) and "R environment, data processing and analysis ..." (Ref. TDA).

TRAINER QUALIFICATIONS

The experts leading the training are specialists in the covered subjects. They have been approved by our instructional teams for both their professional knowledge and their teaching ability, for each course they teach. They have at least five to ten years of experience in their field and hold (or have held) decision-making positions in companies.

ASSESSMENT TERMS

The trainer evaluates each participant's academic progress throughout the training using multiple choice, scenarios, hands-on work and more. Participants also complete a placement test before and after the course to measure the skills they've developed.

2 Component analysis

- Principal Component Analysis.
- Correspondence Analysis.
- Multiple Correspondence Analysis.
- Factorial Analysis for Mixed Data.
- Hierarchical Principal Component Classification.

Hands-on work

Implement the reduction in the number of variables and identify the factors underlying the dimensions associated with significant variability.

3 Modeling

- Steps in building a model.
- Supervised and unsupervised algorithms.
- The choice between regression and classification.

Hands-on work

Set up dataset sampling. Carry out evaluation tests on several supplied models.

4 Model evaluation procedures

- Re-sampling techniques in training, validation and test games.
- Testing the representativeness of training data.
- Performance measurements for predictive models.
- Confusion and cost matrix, ROC and AUC curves.

Hands-on work

Set up dataset sampling. Perform evaluation tests on several supplied models.

5 Unsupervised algorithms

- Hierarchical clustering.
- Non-hierarchical clustering.
- Mixed approaches.

Hands-on work

Unsupervised clustering on multiple datasets.

6 Supervised algorithms

- The principle of univariate linear regression.
- Multivariate regression.
- Polynomial regression.
- Regularized regression.
- The Naive Bayes.
- Logistic regression.

Hands-on work

Implement regressions and classifications on several types of data.

TEACHING AIDS AND TECHNICAL RESOURCES

- The main teaching aids and instructional methods used in the training are audiovisual aids, documentation and course material, hands-on application exercises and corrected exercises for practical training courses, case studies and coverage of real cases for training seminars.
- At the end of each course or seminar, ORSYS provides participants with a course evaluation questionnaire that is analysed by our instructional teams.
- A check-in sheet for each half-day of attendance is provided at the end of the training, along with a course completion certificate if the trainee attended the entire session.

TERMS AND DEADLINES

Registration must be completed 24 hours before the start of the training.

ACCESSIBILITY FOR PEOPLE WITH DISABILITIES

Do you need special accessibility accommodations? Contact Mrs. Fosse, Disability Manager, at psh-accueil@orsys.fr to review your request and its feasibility.

7 Text data analysis

- Text data collection and pre-processing.
- Primary entity extraction, named entity extraction and referential resolution.
- Grammatical labeling, syntactic analysis, semantic analysis.
- Lemmatization. Vector representation of texts. TF-IDF weighting.

Dates and locations

REMOTE CLASS

2026 : 16 June, 1 Dec.

PARIS LA DÉFENSE

2026 : 16 June, 1 Dec.