

# Opleiding : Gegevensanalyse met R

gegevensmodellering en -representatie

*Praktijkcursus - 4d - 28u00 - Ref. DTA*

*Prijs : 2520 € V.B.*

Big Data Analytics omvat het beheersen van fundamentele gegevensverwerkingstechnieken: statistische methoden, classificaties, regressies, PCA, enz. Deze praktische cursus laat je aan de hand van echte gegevens zien hoe je deze technieken kunt gebruiken om modellen te bouwen en vervolgens te evalueren met behulp van de taal R.

## Pedagogische doelstellingen

Aan het einde van de training is de deelnemer in staat om:

- ✓ Het principe van statistische modellering begrijpen
- ✓ Kiezen tussen regressie en classificatie afhankelijk van het type gegevens
- ✓ De voorspellende prestatie van een algoritme evalueren
- ✓ Selecties en rankings maken op basis van grote hoeveelheden gegevens om trends te identificeren

## Doelgroep

Managers van datacentra (datamining, marketing, kwaliteit, enz.), databasegebruikers en bedrijfsmanagers.

## Voorafgaande vereisten

Basiskennis van statistiek en R, of de cursussen "Statistiek, de grondbeginselen beheersen" (Ref. STA) en "R-omgeving, gegevensverwerking en analyse ..." (Ref. TDA) gevolgd hebben.

## Opleidingsprogramma

### 1 Een herinnering aan de R-taal

- Gegevenstypen in R.
- Gegevens importeren-exporteren.
- Technieken voor het tekenen van krommen en grafieken.

### Rollenspel

Aan de slag met scripts en notatieblokken.

#### DEELNEMERS

Managers van datacentra (datamining, marketing, kwaliteit, enz.), databasegebruikers en bedrijfsmanagers.

#### VOORAFGAANDE VEREISTEN

Basiskennis van statistiek en R, of de cursussen "Statistiek, de grondbeginselen beheersen" (Ref. STA) en "R-omgeving, gegevensverwerking en analyse ..." (Ref. TDA) gevolgd hebben.

#### VAARDIGHEDEN VAN DE CURSUSLEIDER

De deskundigen die de cursus leiden zijn specialisten op het betreffende vakgebied. Zij werden geselecteerd door onze pedagogische teams zowel om hun vakkennis als hun pedagogische vaardigheden voor elke cursus die zij geven. Zij hebben minstens vijf tot tien jaar ervaring in hun vakgebied en oefenen of oefenden verantwoordelijke bedrijfsfuncties uit.

#### BEOORDELINGSMODALITEITEN

De cursusleider beoordeelt de pedagogische vooruitgang van de deelnemer gedurende de gehele cursus aan de hand van meerkeuzevragen, praktijksituaties, praktische opdrachten, ... De deelnemer legt ook van tevoren en naderhand een test af ter bevestiging van de verworven kennis.

## 2 Analyse van componenten

- Principale Componentenanalyse.
- Correspondentieanalyse.
- Meervoudige Correspondentieanalyse.
- Factoriële analyse voor gemengde gegevens.
- Hiërarchische classificatie op principale componenten.

### Praktisch werk

Implementatie van de vermindering van het aantal variabelen en identificatie van de factoren die ten grondslag liggen aan de dimensies die verband houden met significante variabiliteit.

## 3 Modelleren

- De stappen die nodig zijn om een model te bouwen.
- Supervised en unsupervised algoritmen.
- De keuze tussen regressie en classificatie.

### Praktisch werk

Het opzetten van steekproeven van gegevenssets. Evaluatietests uitvoeren op verschillende geleverde modellen.

## 4 Modevaluatieprocedures

- Technieken voor herbemonstering in trainings-, validatie- en testspellen.
- Testen van de representativiteit van trainingsgegevens.
- Prestatiemetingen voor voorspellende modellen.
- Verwarring en kostenmatrix, ROC- en AUC-curven.

### Praktisch werk

Het opzetten van steekproeven van gegevenssets. Evaluatietests uitvoeren op verschillende geleverde modellen.

## 5 Algoritmen zonder toezicht

- Hiërarchische clustering.
- Niet-hiërarchische clustering.
- Gemengde benaderingen.

### Praktisch werk

Ongecontroleerde verwerking van clustering op verschillende datasets.

## 6 Algoritmen onder toezicht

- Het principe van univariate lineaire regressie.
- Multivariate regressie.
- Polynomiale regressie.
- Gereguleerde regressie.
- De Naive Bayes.
- Logistische regressie.

### Praktisch werk

Implementatie van regressies en classificaties op verschillende soorten gegevens.

### PEDAGOGISCHE EN TECHNISCHE MIDDELEN

- De gebruikte pedagogische middelen en cursusmethoden zijn voornamelijk: audiovisuele hulpmiddelen, documentatie en cursusmateriaal, praktische oefeningen en correcties van de oefeningen voor praktijkstages, casestudies of reële voorbeelden voor de seminars.
- Na afloop van de stages of seminars verstrekt ORSYS de deelnemers een evaluatievragenlijst over de cursus die vervolgens door onze pedagogische teams wordt geanalyseerd.
- Na afloop van de cursus wordt een presentielijst per halve dag verstrekt, evenals een verklaring van de afronding van de cursus indien de stagiair alle sessies heeft bijgewoond.

### TOEGANGSMODALITEITEN EN TERMIJNEN

De inschrijving dient 24 uur voor aanvang van de cursus plaatsgevonden te hebben.

### TOEGANKELIJKHEID VOOR MINDERVALIDEN

Is voor u speciale toegankelijkheid vereist? Neem contact op met mevr. FOSSE, contactpersoon voor mindervaliden, via het adres psh-accueil@ORSYS.fr om uw verzoek en de haalbaarheid daarvan zo goed mogelijk te bestuderen.

## 7 Tekstgegevensanalyse

- Verzamelen en voorbereiden van tekstuele gegevens.
- Extractie van primaire entiteiten, benoemde entiteiten en referentiële resolutie.
- Grammaticale labeling, syntactische analyse, semantische analyse.
- Lemmatisering. Vectorrepresentatie van teksten. TF-IDF weging.

## Data en plaats

**KLAS OP AFSTAND**  
2026: 16 juni, 1 dec.

**PARIS LA DÉFENSE**  
2026: 16 juni, 1 dec.