Gegevensanalyse met Python
gegevensmodellering en -representatie

Praktijkcursus van 4 dagen - 28u
Ref : BDA - Prijs 2024 : € 2 860 excl. BTW

Data Analytics is een term die wordt gebruikt om het proces van gegevensanalyse uit te drukken om beslissingen te nemen. De taal Python biedt een ecosysteem voor statistische verwerking, van de constructie van analysemodellen tot de evaluatie en weergave ervan.

PEDAGOGISCHE DOELSTELLINGEN

Na afloop van de opleiding kan de cursist:

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

HANDS-ON WORK

Ontwikkelen/creëren van analyses met behulp van Python-software, met de modules pandas, NumPy, SciPy, MatPlotLib, seaborn, scikit-learn en statsmodels.

HET PROGRAMMA

laatste update: 10/2021

1) Inleiding tot modelleren

- Inleiding tot de taal Python.
- Inleiding tot Jupiter Notebook-software.
- De stappen die nodig zijn om een model te bouwen.
- Supervised en unsupervised algoritmen.
- De keuze tussen regressie en classificatie.

Installatie van Python 3, Anaconda en Jupiter Notebook.

2) Modelevaluatieprocedures

- 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.

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

3) Algoritmen onder toezicht

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

Implementatie van regressies en classificaties op verschillende soorten gegevens.

DEELNEMERS

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

VOORAFGAANDE VEREISTEN

Basiskennis van Python.
Basiskennis statistiek of afronding van de cursus "Statistics, mastering the fundamentals" (Ref. STA).

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.

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.

4) Algoritmen zonder toezicht
- Hiërarchische clustering.
- Niet-hiërarchische clustering.
- Gemengde benaderingen.

Ongecontroleerde verwerking van clustering op verschillende datasets.

5) Analyse van componenten
- Principale componentenanalyse.
- Correspondentie factoranalyse.
- Meervoudige correspondentieanalyse.
- Factoranalyse voor gemengde gegevens.
- Hiërarchische classificatie op basis van principale componenten.

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.

6) Tekstgegevensanalyse
- Verzamelen en voorbewerken van tekstuele gegevens.
- Extractie van primaire entiteiten, benoemde entiteiten en referentiële resolutie.
- Grammaticale labeling, syntactische analyse, semantische analyse.
- Lemmatisering.
- Vectorweergave van teksten.
- TF-IDF weging.
- Word2Vec.

Verken de inhoud van een tekstdatabase met behulp van latente semantische analyse.

DATA

KLAS OP AFSTAND
2025 : 18 feb, 20 mei, 19 aug, 28 okt

BRUSSEL
2025 : 18 feb, 20 mei, 19 aug, 28 okt