

Opleiding : Big Data - Python voor gegevensanalyse

Praktijkcursus - 3d - 21u00 - Ref. PBD

Prijs : 2010 € V.B.



4,6 / 5

De taal Python heeft een wetenschappelijk ecosysteem dat statistische verwerking mogelijk maakt, van de constructie van analysemodellen tot de evaluatie en representatie ervan. Deze cursus stelt je in staat om gegevens uit verschillende bronnen te analyseren met behulp van Python-bibliotheken.



Pedagogische doelstellingen

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

- ✓ Het principe van statistische modellering begrijpen
- ✓ Weten hoe je de belangrijkste tools voor gegevensverwerking en analyse voor Python gebruikt
- ✓ Weten hoe je best practice toepast bij het opschonen en voorbereiden van gegevens voorafgaand aan de analyse.
- ✓ Kiezen tussen regressie en classificatie afhankelijk van het type gegevens
- ✓ Leer hoe je een eenvoudig leermodel opzet
- ✓ Gegevens uit een bestand kunnen halen

Doelgroep

Python-ontwikkelaars, datacenterbeheerders, softwareontwikkelaars, programmeurs, gegevensanalisten, gegevenswetenschappers.

Voorafgaande vereisten

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

Praktische modaliteiten

Analyses ontwikkelen/uitvoeren met Python, met behulp van de modules Pandas, Numpy en SciPy.

Opleidingsprogramma

DEELNEMERS

Python-ontwikkelaars, datacenterbeheerders, softwareontwikkelaars, programmeurs, gegevensanalisten, gegevenswetenschappers.

VOORAFGAANDE VEREISTEN

Beheersing van Python programmeren. Basiskennis van 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.

1 Presentatie van het wetenschappelijke Python-ecosysteem

- Overzicht van het wetenschappelijke ecosysteem van Python: de essentiële bibliotheken.
- Weet waar je nieuwe boekwinkels kunt vinden en beoordeel hun duurzaamheid.
- De belangrijkste open source tools en software voor data science.

Praktisch werk

Installatie van Python 3, Anaconda en Jupiter Notebook.

2 Werken met gegevens in Python

- De wetenschappelijke basis van Python: de SciPy Stack.
- Best practices voor een goede start van uw data science project met Python.
- Wetenschappelijke bestandsformaten en bibliotheken om ze te manipuleren.
- Pandas: analyse van gegevens in tabelvorm (CSV-bestanden, Excel, enz.), statistieken, pivots, filters, zoekopdrachten, enz.
- NumPy: numerieke berekening en lineaire algebra (vectoren, matrices, afbeeldingen).
- Gegevensextractie, voorbereiding en opschoning.

Praktisch werk

Python-scripts schrijven om met gegevens uit bestanden te werken, om filters, opmaak- en opschoningsprocessen toe te passen.

3 Inleiding tot modelleren

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

Praktisch werk

Integratie van Python-scripts in de geïnstalleerde omgeving voor analyse.

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.

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.

5 Algoritmen onder toezicht

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

Praktisch werk

Implementatie van regressies en classificaties op verschillende soorten gegevens.

6 Algoritmen zonder toezicht

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

Praktisch werk

Ongecontroleerde verwerking van clustering op verschillende datasets.

Data en plaats

KLAS OP AFSTAND

2026 : 22 juni, 21 sep., 30 nov.

PARIS LA DÉFENSE

2026 : 22 juni, 21 sep., 30 nov.