

Data Science & Predictive Machine Learning

Leer beter voorspellen dankzij de meest actuele technieken

Big Data is groot, complex en gevarieerd. Data alleen voegen geen waarde toe. Waarde wordt pas toegevoegd door analyse van deze data. De opleiding Data Science & Predictive Machine Learning maakt de deelnemers bekend met de laatste ontwikkelingen op het terrein van voorspellen, modelleren en analyse en leert hoe u waarde uit uw data haalt. U wordt uitgedaagd de theorie toe te passen in het programma R.

Verschillende methoden

De opleiding behandelt breed inzetbare methoden en technieken uit de machine learning. Het accent ligt op het benutten van data voor het doen van voorspellingen. De volgende methoden worden behandeld:

- Klassieke regressie
- Penalty-methoden zoals Ridge Regression, de Lasso en het Elastic Net
- Support Vector Machines

De theorie achter deze methoden komt uitgebreid aan bod.

Zelf aan de slag in R

De opleiding besteedt veel aandacht aan het zelf kunnen werken met deze methoden. Aan de hand van datasets oefent u met de verschillende statistische analyses in het programma R. U ervaart wat de gebruiksmogelijkheden zijn in de praktijk.

Toepassingsgericht

Naast de theorie en het uitvoeren van de analyses is er ook aandacht voor de toepassing. U kunt daarbij denken aan vragen als: wat doet de methode, wanneer is het nuttig, welke vragen kan ik ermee beantwoorden en waar moet ik op letten bij het interpreteren van de resultaten?

Wanneer

Maandag 13, 20, 27 januari en 3 februari 2020

Prijs

De investering bedraagt € 2725 (vrijgesteld van BTW). Inbegrepen zijn dagarrangement en cursusmateriaal.

Waar

Novotel Brainpark Rotterdam

Direct inschrijven

www.erasmusacademie.nl/datascience

Aantal deelnemers

Het maximum aantal deelnemers bedraagt 25.

Certificaat

Op de slotdag van de opleiding ontvangt u als bewijs van deelname een certificaat.

Meer informatie

Neem contact op met Miranda Smit, Adviseur Opleidingen T: 010-4088687 E: smit@erasmusacademie.nl

Resultaten

U weet hoe u enkele belangrijke voorspellingsmethoden uit de Data Science & Predictive Machine Learning in R toepast en maakt op die manier waardevolle data-analyses. U kunt:

- Voorspellingen maken met multi-pele regressie in R en de output interpreteren;
- Enkele belangrijke diagnostische maten en methoden toepassen op data in R en daar conclusies uittrekken;
- Ridge regressie uitvoeren in R en kruisvalidatie toepassen;
- Variabele selectie toepassen m.b.v. de Lasso en het elastic net en de bijbehorende resultaten interpreteren;
- Kruisvalidatie toepassen om een goede afweging te maken tussen flexibiliteit van het model en de voorspelling in de populatie;
- Voorspellingen maken met Support Vector Machines in R en interpreteren van enkele belangrijke diagnostische maten;
- Een niet-lineaire SVM uitvoeren.

Programma

Dag 1:

Het vakgebied Data Science zetten we kort uiteen. Daarbij bespreken we in vogelvlucht enkele gangbare technieken van dit moment. Het grootste deel van deze dag besteden we aan multi-pele regressie. Aan de hand van voorbeelddatasets oefent u met de statistische taal R.

Dag 2:

Bij multi-pele regressie kunnen problemen ontstaan die te detecteren zijn met diagnostische maten. Tijdens de ochtendsessie bespreken we deze maten en oefent u ermee. In de middag maken we een begin met regressie via penalty's, in het bijzonder ridge regression. Deze methode is geschikt voor situaties met veel voorspelvariabelen, waar gewone regressie niet meer toereikend is. We introduceren het begrip kruisvalidatie om overoptimistische voorspellingen tegen te gaan.

Dag 3:

We gaan door met andere penalty-methoden zoals de Lasso en het Elastic Net. Beide zijn goed te gebruiken voor het selecteren van relevante voorspelvariabelen. We staan ook stil bij de zogenaamde bias-variance trade-off. Dat wil zeggen, de afweging tussen enerzijds een flexibele methode die goed voorspelt in de data, maar slecht in de populatie. En anderzijds een wat meer rigide methode die wat minder goed de data voorspelt, maar wel goed voorspelt in de populatie.

Dag 4:

De Support Vector Machine (SVM) is een van de basistechnieken in machine learning voor het voorspellen van een variabele met maar twee uitkomsten. Bijvoorbeeld: een klant koopt wel of niet het product, is wel of niet kredietwaardig, is wel of niet geïnteresseerd in de aanbieding van een product. De kern van deze techniek leggen we uit aan de hand van een casus. Vervolgens bespreken we de uitbreiding naar niet-lineaire voorspellingen.

Docent

Patrick Groenen

Prof.dr. Patrick Groenen is hoogleraar Statistiek aan de Erasmus School of Economics van de Erasmus Universiteit Rotterdam. Patrick focust zich op de ontwikkeling van nieuwe Data Science methoden waaronder machine learning technieken, exploratieve data analyse, visualisatie en hun toepassingen.

Voor wie?

Voor de beginnende Data Scientist met enige kennis van statistiek en statistische software, die kennis wil maken met R en verschillende voorspelmodellen uit Machine Learning maar ook voor professionals met affiniteit voor modellen en kwantitatieve data-analyses. Om de diverse statistische concepten te kunnen begrijpen heeft u kennis nodig van basale statistiek op minimaal HBO niveau. Of u heeft gelijkwaardige praktijkervaring.

Vorbereiding

Wij vragen u een laptop mee te nemen en daarop voor aanvang van de opleiding de statistische taal R en R-Studio te installeren. Beide programma's zijn vrij te downloaden op: www.r-project.org
www.rstudio.com/products/rstudio/#Desktop
We zullen voornamelijk werken in het programma R. Basiskennis daarvan is echter niet noodzakelijk.

Inschrijven is mogelijk:

www.erasmusacademie.nl/datascience

Wat vinden onze deelnemers

"Deze masterclass was zeer waardevol om mijn kennis te vergroten over het gebruik van R, de theorie achter Regressie (verschillende soorten), interpretatie van modellen en de onderlinge verschillen in toepassingsgebied."

"Goede insight in Data Science & Predictive Machine Learning. Zeer deskundige docent en prettige locatie"