

Eindtoets Data-analyse met R (5 EC)

Deze leerlijn bestaat uit de volgende cursussen:

1. Kennismaking met R (1 EC)
2. Dataframes, visualisatie en normale verdeling (1 EC)
3. Correlatie (1 EC)
4. Machine learning 1 (1 EC)
5. Machine learning 2 (1 EC)

Opdrachtschrijving

In de leerlijn Data-analyse met R heb je kennis en vaardigheden opgedaan omtrent het inladen, bewerken, analyseren en presenteren van data aan de hand van R-studio en programmeertaal R. Aan de hand van deze eindopdracht toon je aan zelfstandig machine learning en datavisualisatie op basis van een dataframe uit een extern bestand toe te kunnen passen.

Opdracht

Je selecteert een bestaand databestand of stelt deze zelf samen en leest deze in, in R studio als dataframe. Je analyseert de data en past statistische en visualisatietechnieken toe om de data op te schonen en betekenisvoller te maken. Onderstaande technieken dienen tenminste toegepast te worden:

- Dataframe inlezen
- Statistische berekeningen – tenminste drie
- Data visualisatie – tenminste een histogram
- Machine learning – tenminste één techniek

Op te leveren producten

- Script en bronbestand

Lever in een zip bestand het script en de origineel gebruikte data aan, zodat de docent en examinator de berekeningen kunnen uitvoeren.

- Rapportage

In een rapport van max. 2000 woorden, excl. voorwoord, inhoudsopgave, inleiding, geef je een beschrijving van de gebruikte data, analyses en resultaten. Voeg het rapport toe aan het zipbestand.

Beoordelingscriteria

Voor ieder criterium krijg je een cijfer tussen de 1 en 10, afhankelijk van het niveau van de uitwerking. Een 8 wordt gezien als zeer goed. Een cijfer boven de 8 wordt alleen gegeven wanneer er boven het gevraagde niveau opgeleverd wordt.

#	Data-analyse met R	Weging
1	De student toont aan geschikte datasets te kunnen selecteren of samenstellen en deze aan hand van een dataframe in te kunnen lezen.	20%
2	De student toont aan data te kunnen visualiseren waarvan tenminste 1 histogram van een deel van de data wordt gegenereerd.	20%
3	De student toont aan over een subset van de data tenminste drie statistische berekening naar keuze uit te kunnen voeren.	20%
4	De student kan tenminste één machine learning techniek naar keuze toepassen op de dataset.	20%
5	De student toont in een rapport aan de data te kunnen analyseren, de keuze van technieken te kunnen onderbouwen en resultaten te kunnen interpreteren en beschrijven.	20%