

## Statistisch modelleren en datamining (F000759)

**Cursusomvang** *(nominale waarden; effectieve waarden kunnen verschillen per opleiding)*

**Studiepunten 5.0** **Studietijd 150 u**

**Aanbodsessies en werkvormen in academiejaar 2024-2025**

A (semester 1)	Engels	Gent	hoorcollege zelfstandig werk werkcollege
----------------	--------	------	--

**Lesgevers in academiejaar 2024-2025**

Benoit, Dries	EB23	Verantwoordelijk lesgever
---------------	------	---------------------------

**Aangeboden in onderstaande opleidingen in 2024-2025**

	stptn	aanbodssessie
<a href="#">Bachelor of Science handelsingenieur</a>	5	A
<a href="#">Uitwisselingsprogramma Economie en Bedrijfskunde</a>	5	A
<a href="#">Preparatory Course Master of Science in Business Engineering</a>	5	A

**Onderwijstalen**

Engels

**Trefwoorden**

univariate en multivariate regressieanalyse, classificatie, resampling methodes, model selectie en regularisatie, boom-gebaseerde methodes, ongesuperviseerd leren, R

**Situering**

Door de digitalisering van bedrijfskundige processen van de voorbije decennia, heeft een spectaculaire toename van informatie die opgeslagen is in databases plaatsgevonden. De bedoeling van datamining is uit deze gigantische hoeveelheden data bedrijfsrelevante patronen te halen met als bedoeling de bedrijfsvoering verder te optimaliseren. De studenten worden vertrouwd gemaakt met het statistisch extraheren van informatie uit grote databases. De studenten leren hoe deze brede waaier van technieken op een correcte manier toe te passen in praktische bedrijfsapplicaties. Ze leren ook de vaardigheden om de resultaten die bekomen worden op een betekenisvolle manier te interpreteren. Alle oefeningen worden opgelost met behulp van de statistische programmeertaal R. Dit vak bouwt voort op de inhoud van het vak 'Statistiek I'

**Inhoud**

- \* Verband tussen zuiverheid, variantie en model complexiteit
- \* Regressie analyse
  - enkelvoudige regressieanalyse
  - meervoudige regressieanalyse
  - kwalitatieve predictoren
- \* Classificatie
  - logistische regressieanalyse
  - lineaire discriminant analyse
  - k-nearest neighbors
- \* Resampling methodes
  - crossvalidatie
  - bootstrap
- \* Model selectie en regularisatie
  - selectie van variabelen
  - shrinkage methodes
  - dimensie reductie
- \* Ongesuperviseerd leren

- principale componenten analyse
- clustering

### **Begincompetenties**

De studenten dienen een diepgaand begrip te hebben van de basis statistische concepten zoals behandeld in het vak 'Statistiek I'.

### **Eindcompetenties**

- 1 Datamining methodes op een correcte manier kunnen toepassen, gebruik makend van de programmeertaal R
- 2 De uitkomsten op een correcte manier kunnen interpreteren en rapporteren
- 3 Op een gefundeerde manier een techniek kiezen voor een specifiek probleem en hierbij bewust zijn van de voor- en nadelen van de gekozen aanpak.

### **Creditcontractvoorwaarde**

Toelating tot dit opleidingsonderdeel via creditcontract is mogelijk mits gunstige beoordeling van de competenties

### **Examencontractvoorwaarde**

Dit opleidingsonderdeel kan niet via examencontract gevolgd worden

### **Didactische werkvormen**

Werkcollege, Hoorcollege, Zelfstandig werk

### **Toelichtingen bij de didactische werkvormen**

De leeromgeving Ufora zal gebruikt worden om een vlotte organisatie van de verschillende didactische werkvormen voor dit vak.

Actieve deelname aan lab sessies via platform dodona is vereist.

### **Studiemateriaal**

Type: Handboek

Naam: An Introduction to Statistical Learning: with Applications in R (2nd edition)

Richtprijs: Gratis of betaald door opleiding

Optioneel: nee

Taal : Engels

ISBN : 978-1-07161-417-4

Online beschikbaar : Ja

Beschikbaar in de bibliotheek : Ja

Beschikbaar via studentenvereniging : Ja

Gebruik en levensduur binnen het opleidingsonderdeel : intensief

Gebruik en levensduur binnen de opleiding : regelmatig

Bijkomende info: Gratis (en legale) pdf downloaden via: <https://www.statlearning.com/>

### **Referenties**

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2021) An Introduction to Statistical Learning with Applications in R, Springer (Versie 2). (Opm: een gratis, legale versie van het boek kan online gevonden worden).

### **Vakinhoudelijke studiebegeleiding**

De studenten kunnen voor de vakinhoudelijke begeleiding een beroep doen op de verantwoordelijke lesgever en de assistenten. De interactieve ondersteuning verloopt via Ufora (groepsdiscussie, oefeningen). De lesgever stelt ook voorbeeldexamenvragen ter beschikking.

### **Evaluatiemomenten**

periodegebonden en niet-periodegebonden evaluatie

### **Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode**

Schriftelijke evaluatie

### **Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode**

Schriftelijke evaluatie

### **Evaluatievormen bij niet-periodegebonden evaluatie**

Participatie, Schriftelijke evaluatie open boek

### **Tweede examenkans in geval van niet-periodegebonden evaluatie**

Examen in de tweede examenperiode is niet mogelijk

### **Toelichtingen bij de evaluatievormen**

Periodegebonden evaluatie (schriftelijk examen):

- Meerkeuzevragen
  - Open vragen
- Niet-periodegebonden evaluatie (opdracht):
- Dodona oefeningen

### **Eindscoreberekening**

95% periodegebonden evaluatie (schriftelijk examen)

5% niet-periodegebonden evaluatie (opdracht)

Gevolgen van deadlines niet respecteren: Het niet respecteren van deadlines van het individueel werk en/of het groepswerk zonder gegronde en gestaafde reden (zoals ziekte) leidt tot een nul voor respectievelijk het individueel werk en/of het groepswerk.