



## Geavanceerde voorspellende analytische methoden (F000882)

Wegens Covid19 kan mogelijk afgeweken worden van de onderwijs- en evaluatievormen. Dergelijke afwijkingen zullen via Ufora worden gecommuniceerd.

Cursusomvang (nominale waarden; effectieve waarden kunnen verschillen per opleiding)  
Studiepunten 8.0 Studietijd 240 u Contacturen 80.0 u

Aanbodsessies en werkvormen in academiejaar 2020-2021

A (semester 2)	Engels	Gent		
			online werkcollege:	0.0 u
			PC-klasoefeningen	
			hoorcollege	7.5 u
			groepswork	5.0 u
			werkcollege: geleide	7.5 u
			oefeningen	
			werkcollege: PC-	55.0 u
			klasoefeningen	
			zelfstandig werk	5.0 u
			online hoorcollege	0.0 u

Lesgevers in academiejaar 2020-2021

Bogaert, Matthias

EB23 Verantwoordelijk lesgever

Aangeboden in onderstaande opleidingen in 2020-2021

stptn aanbodsessie

[Master of Science in Data Science for Business](#)

8

A

Onderwijstalen

Engels

Trefwoorden

Geavanceerde marketingmodellen, data mining, machine learning, artificiële neural netwerken, mixturemodellen voor classificatie, ensembles, computationele statistiek, R

Situering

Naast de nodige conceptuele vaardigheden, wordt van marketing managers heden ten dage immer meer een feitengebaseerde analyse en de aanwending van beslissingsmodellen gevraagd om de marketingstrategie van de onderneming te bepalen. In dit vak zullen we nagaan hoe we computerondersteunde interactieve beslissingsmodellen kunnen aanwenden. Deze cursus is een aanvulling op de cursussen analytische klantenbeheerssystemen en prijszetting en omzetbeheer. De opleiding beoogt dus een grondige wetenschappelijke studie. De algemene doelstelling is net die specialisten af te leveren die binnen marketing, op het domein van de marketinganalyse, kunnen zorgen voor de ondersteuning van de strategie en de actieplannen in de onderneming. Een bedrijfskundige vertaalslag naar de bruikbaarheid van deze analysetechnieken voor beslissingsnemers wordt hierbij beklemtoond.

Inhoud

Het toepassen van geavanceerde statistische methodes en algoritmes uit machinaal leren (zowel gecontroleerd als ongecontroleerd leren) met het oog op het oplossen van bedrijfs- en marketingproblemen staat centraal in deze cursus. De focus ligt vooral in het begrijpen en implementeren van technieken i.v.m. gecontroleerd en ongecontroleerd leren. Bij alle technieken die we bespreken, zal er uitgebreide aandacht besteed worden aan de afweging tussen interpretatie en performantie. Bijzondere aandacht zal besteed worden aan het bias-variance probleem. De volgende technieken uit gecontroleerd leren komen aan bod:

- (Niet-)Lineair regressie
- Logistische regressie
- Beslissingsbomen
- Neurale netwerken

- Bagging
- Boosting
- Random forest
- Rotation forest
- Hybride ensembles

De volgende algoritmes uit ongecontroleerd leren zullen worden onderricht:

- Hiërarchisch clusteren
- K-means clusteren
- K-medoids clusteren
- Spectraal clusteren
- Hoofdcomponentanalyse (factoranalyse)
- Zelf-organiserende mappen (Kohonen mappen)

Evaluatie van al deze technieken is eveneens cruciaal. Daarvoor zullen we de volgende methodes gebruiken: AUC, accuraatheid, F1 en lift voor classificatie en  $R^2$ , RMSEP voor regressie.

Bovendien komen data visualisatie in R aan bod, die de studenten moeten gebruiken.

### Begincompetenties

Sluiten aan op de eindtermen van de cursus analytische klantenbeheerssystemen alsook van marketinginformatiesystemen/database marketing (SQL) en “Prijszetting en Omzetbeheer”, die op hun beurt kennis veronderstellen van gevorderde statistiek en econometrie.

Basiskennis van de R programmeertaal.

### Eindcompetenties

- 1 Doorgronden van geavanceerde data mining methoden.
- 2 Een uitgebreide kennis van de R programmeertaal en deze ook effectief inzetten in reële projecten.
- 3 De technieken van modelbouw hanteren en komen tot creatieve oplossingen met de beschikbare modellen voor reële bedrijfsproblemen gebaseerd op grote datavolumes.
- 4 Een bestaand bedrijfsprobleem zo performant mogelijk oplossen in groeps- en competitieverband, inclusief het op zoek gaan naar geavanceerde methoden (R-packages).
- 5 Het nemen van gepaste bedrijfsbeslissingen op basis van de bekomen analyseresultaten en beslissingsmodellen (rekening houdend met alle betrokken partijen), alsook de gepaste communicatie verzorgen ten opzichte van technische als niet-technische professionals.
- 6 Exploreren van de kansen die nieuwe gegevensbronnen bieden.

### 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

Groepswerk, hoorcollege, zelfstandig werk, werkcollege: geleide oefeningen, werkcollege: PC-klasoefeningen, online hoorcollege, online werkcollege: PC-klasoefeningen

### Toelichtingen bij de didactische werkvormen

De nadruk ligt op de praktische toepassing van eerder verworven analytische kennis op de verschillende substantieve terreinen inzake marketing.

De nadruk ligt aldus op de oefeningen, en de analyse van diverse gegevens, evenals de formulering van marketingplannen op basis hiervan.

### Leermateriaal

- Eigen syllabus
- Wetenschappelijke artikels
- Handboeken:
  - James G., Witten D., Hastie T., Tibshirani R. (2015), An Introduction to Statistical Learning: with Applications in R, Springer.
  - Ballings M. & Van den Poel D. (2016), Analytical Customer Relationship Management Using R.

### Referenties

Wetenschappelijk artikels:

- Ballings, M., Van den Poel, D., 2015. CRM in Social Media: Predicting Increases in Facebook Usage Frequency. European Journal of Operational Research.
- Prinzie, A., Van den Poel, D., 2008. Random Forests for multiclass classification: Random MultiNomial Logit. Expert Systems with Applications 34, 1721-1732.
- Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we Need

Hundreds of Classifiers to Solve Real World Classification Problems? Journal of Machine Learning Research 15, 3133–3181.

- Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32.
- Breiman, L., 1996. Bagging predictors. Mach Learn 24, 123–140.
- Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Information Processing & Management 45, 427–437.
- Rodriguez, J.J., Kuncheva, L.I., Alonso, C.J., 2006. Rotation Forest: A New Classifier Ensemble Method. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 1619–1630.
- Dietterich, T.G., 2000. Ensemble Methods in Machine Learning, in: Multiple Classifier Systems, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 1–15.
- Ngai, E.W.T., Xiu, L., Chau, D.C.K., 2009. Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications 36, 2592–2602.

#### Vakinhoudelijke studiebegeleiding

Talrijke oefeningen worden tijdens de sessies behandeld. Bovendien worden taken opgegeven, die in groepsverband dienen te worden opgelost. Hierbij krijgen de studenten uitgebreide (studie)begeleiding en na afloop feedback (met de volledige groep, in team en individueel).

Na de talrijke periodieke tests over de programmeertaal R wordt individueel feedback gegeven.

#### Evaluatiemomenten

niet-periodegebonden evaluatie

Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode

Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode

Evaluatievormen bij niet-periodegebonden evaluatie

Schriftelijk examen, mondeling examen, werkstuk, vaardigheidstest, peer-evaluatie

Tweede examenkans in geval van niet-periodegebonden evaluatie

Examen in de tweede examenperiode is enkel mogelijk in gewijzigde vorm

#### Toelichtingen bij de evaluatievormen

Presentatie van een mini-project in voorspellende analytische methoden

Schriftelijk, waarbij gepeild wordt naar de mate waarin de student zich

1 de principes van gevorderde analytische CRM eigen heeft gemaakt,

2 de programmeertaal R beheerst,

3 de principes van geavanceerde analysetechnieken beheerst,

4 hoe de R-taal kan worden aangewend voor het modelmatig oplossen van bedrijfsproblemen,

5 het nemen van gepaste bedrijfsbeslissingen (rekening houdende met de beperkingen alsook de diverse visies van belanghebbenden) eigen heeft gemaakt.

#### Eindscoreberekening

Mondeling gedeelte (20%) Schriftelijk gedeelte (80%)

mogelijk met een correctie door middel van "peer assessment".