



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

Studiepunten 8.0 Studietijd 240 u Contacturen 80.0 u

Aanbodsessies en werkvormen in academiejaar 2017-2018

A (semester 1)	groepswerk	5.0 u
	hoorcollege	7.5 u
	zelfstandig werk	5.0 u
	werkcollege: PC-klasoefeningen	60.0 u
	demonstratie	2.5 u

Lesgevers in academiejaar 2017-2018

Van den Poel, Dirk	EB07	Verantwoordelijk lesgever
--------------------	------	---------------------------

Aangeboden in onderstaande opleidingen in 2017-2018

	stptn	aanbodsessie
Master of Science in Marketing Analysis	8	A

Onderwijstalen

Engels

Trefwoorden

Informatiesystemen, Data Warehouse, SQL, Big Data, Apache Spark/Spark SQL/Python, Machine Learning, Spark MLlib, ML pipelines

Situering

De brede doelstelling van dit vak bestaat erin studenten op een grondige manier theoretische en praktische kennis bij te brengen m.b.t. het gebruik en beheer van gegevens en informatie, dit zowel op strategisch, analyse- technisch, als organisatorisch vlak. "Nieuwe data, nieuwe methoden en nieuwe vaardigheden - hoe het allemaal samenbrengen?" is één van de onderzoeksprioriteiten voor 2016-2018 van het MSI (Marketing Science Institute) - de leidinggevende denktank in marketing (wereldwijd).

Inhoud

- 1 Belang van informatiebeheer in het algemeen: welke ontwikkelingen liggen aan de basis van het toegenomen belang van informatiegebruik?
- 2 Gegevensbronnen en -verzamelingsmethoden: over welke databronnen kan de beheerder van data vandaag/morgen beschikken, wat is big data, hoe omgaan met geautomatiseerde datacollectiemethoden als scanning en internet? Hoe omgaan met de nosql revolutie?
- 3 Opbouw van een database: welke principes liggen aan de basis van een goede opbouw van een marketing database? Hoe wordt de structuur opgebouwd (Entity Relationship diagramma's)?
- 4 Ondervraging van een database: aanleren van de SQL (structured query language) programmeertaal met oefeningen (in casu: Oracle SQL en Hive/Apache Spark/Spark SQL/Python) op grote bestaande databases.
- 5 "Feature Engineering": het opbouwen van variabelen die gebruikt worden in diverse (predictieve en prescriptieve) modellen.
- 6 Machinaal leren met behulp van Spark ML pipelines
- 7 Data visualisatie is een belangrijke functionaliteit om management te overtuigen. De PySpark omgeving ondersteunt het gebruik van Matplotlib functionaliteiten. Deze worden dan ook uitgebreid besproken.
- 8 Implementatie/integratie MIS in de organisatie: welke valkuilen kunnen zich voordoen tijdens het implementatieproces van een MIS binnen de organisatie, wat

zijn de principes van datawarehousing? Elk van deze onderwerpen wordt in de diepte behandeld op basis van een mix van interactieve discussies en real-life cases.

Begincompetenties

Sluiten aan bij de eindtermen van een basiscursus informatica. Basisnoties van het CRISP-DM (data mining) model.

Eindcompetenties

- 1 Begrijpen van de opbouw van een database.
- 2 Beheersen van de programmeertaal SQL alsook NoSQL (big data) om te komen tot een optimaal data- en informatiebeheer inclusief het bouwen van applicaties op basis van deze datastructuren met het oog op het nemen van correcte conclusies voor complexe marketingproblemen, alsook de gepaste communicatie verzorgen ten opzichte van technische als niet-technische professionals.
- 3 Het benutten van diverse externe databronnen en deze optimaal gebruiken, inclusief de nieuwe vormen (sociale media data).
- 4 Het aanwenden van Big Data Analytics algoritmen via Spark ML Pipelines op gevallenstudies om complexe marketingbeslissingsmodellen te bouwen.
- 5 Eigen onderzoeksresultaten aftoetsen met bestaande literatuurstudie van internationale toptijdschriften.
- 6 Een bedrijfsgevallenstudie uitvoeren in een internationaal en interdisciplinair team waarbinnen er verschillende niveaus van ervaring vertegenwoordigd zijn.
- 7 Een professionele, mondelinge presentatie geven omtrent een geavanceerd probleem en de bijhorende oplossing.

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

Demonstratie, groepswork, hoorcollege, zelfstandig werk, werkcollege: PC-klasoefeningen

Toelichtingen bij de didactische werkvormen

Ex cathedra uiteenzetting alsook actieve bespreking in groepsverband van de diverse technieken en modellen met interactieve oefeningen in het PC-lokaal.

Leermateriaal

Oracle(TM) SQL en PL/SQL cursusmateriaal waaronder:

- Oracle Database 10g: Introduction to SQL
- (wetenschappelijke) artikels
- Slides omtrent Apache Spark/Spark SQL/Python en Big Data/Big Data Analytics in Spark ML Pipelines
- Cases

Referenties

Van den Poel D. (2003), "Predicting Mail-Order Repeat Buying: Which Variables Matter?", Tijdschrift voor Economie en Management, 48 (3), p.371-404.

Ryza S., Laserson U. (2015), Advanced Analytics with Spark: Patterns for Learning from Data at Scale", O'Reilly.

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 studiebegeleiding en na afloop feedback (met de volledige groep, in team en individueel). Na de periodieke tests over de programmeertalen SQL en Hive/Apache Spark/Spark SQL/Python wordt collectieve alsook individuele 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 met open vragen, 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

Schriftelijke ondervraging, waarbij gepeild wordt naar de opgedane kennis m.b.t. het interpreteren van bestaande database structuren (Entity Relationship Diagramma's), de SQL en NoSQL databankbevragingstaal alsook het concept Big Data. Technische alsook bedrijfsrelevante presentatie van resultaten van een Big Data gevallenstudie. Hierbij dient gebruikt te worden gemaakt van Apache Spark SQL en Spark ML Pipelines.

Eindscoreberekening

30% op SQL oefeningen/examen en 70% op Big Data/Apache Spark/Spark SQL/Python oefening
mogelijk met een correctie door middel van "peer assessment"