



Big data (F000802)

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 6.0 Studietijd 180 u Contacturen 45.0 u

Aanbodsessies en werkvormen in academiejaar 2020-2021

A (semester 1)	Engels	Gent	werkcollege: PC- klasoefeningen	37.5 u
			hoorcollege	5.0 u
			groepswerk	5.0 u
			online hoorcollege	0.0 u
			online werkcollege: PC-klasoefeningen	0.0 u

Lesgevers in academiejaar 2020-2021

Van den Poel, Dirk EB23 Verantwoordelijk lesgever

Aangeboden in onderstaande opleidingen in 2020-2021

	stptn	aanbodsessie
Educatieve Master of Science in de economie (afstudeerrichting toegepaste economische wetenschappen)	6	A
Master of Science in Business Economics (afstudeerrichting Accountancy)	6	A
Master of Science in Business Economics (afstudeerrichting Corporate Finance)	6	A
Master of Science in Business Engineering (afstudeerrichting Data Analytics)	6	A
Master of Science in Business Engineering (afstudeerrichting Finance)	6	A
Master of Science in Business Economics (afstudeerrichting Marketing)	6	A
Master of Science in Business Engineering (afstudeerrichting Operations Management)	6	A
Uitwisselingsprogramma Economie en Bedrijfskunde	6	A

Onderwijstalen

Engels

Trefwoorden

Informatiesystemen, Data Warehouse, SQL, Big Data, Apache Spark, Spark SQL, Python, PySpark, ML pipelines (Machine Learning).

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.

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 Apache Spark/Spark SQL/Python) op grote bestaande databases.
- 5 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.

In de big data gevallenstudie komen NLP (natural language processing) en visualisatie ook uitgebreid aan bod.

Begincompetenties

Sluiten aan bij het vak Statistisch Modelleren en datamining.

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
- 3 Bouwen van applicaties op basis van deze datastructuren met het oog op het nemen van correcte conclusies voor complexe marketingproblemen
- 4 De gepaste communicatie verzorgen ten opzichte van technische als niet-technische professionals.
- 5 Optimaal gebruiken, inclusief de nieuwe vormen (sociale media data) van diverse externe databronnen
- 6 Het aanwenden van Big Data Analytics algoritmen via Spark ML Pipelines op gevallenstudies om complexe marketingbeslissingsmodellen te bouwen.
- 7 Eigen onderzoeksresultaten aftoetsen met bestaande literatuurstudie van internationale toptijdschriften
- 8 Een bedrijfsgevallenstudie uitvoeren in een internationaal en interdisciplinair team waarbinnen er verschillende niveaus van ervaring vertegenwoordigd zijn.
- 9 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

Groepswerk, hoorcollege, werkcollege: PC-klasoefeningen, online hoorcollege, online werkcollege: geleide oefeningen, online 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 cursusmateriaal waaronder
- Oracle Database 10g: Introduction to SQL
- Slides omtrent Apache Spark/Spark SQL/Python & Big Data
- Cases
- Big Data boek (gratis download van pdf): Learning Spark: Lightning-Fast Data Analytics, O'Reilly, 2nd Edition covering Apache Spark 3.0
- Tijdschriftartikel: van Dongen G. & Van den Poel D. (2020), Evaluating Stream Processing Frameworks, IEEE Transaction on Parallel and Distributed Systems

Referenties

geen

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 wordt collectieve 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

Tweede examenkans in geval van niet-periodegebonden evaluatie

Examen in de tweede examenperiode is mogelijk

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 databankbevragingstaal. Groepswerk met Apache Spark alsook het concept Big Data.

Eindscoreberekening

30% op SQL oefeningen/examen en 70% op Big Data/Apache Spark/Spark SQL/Python oefening potentieel gecorrigeerd voor "peer assessment". Van de 70% score van het groepswerk is 30% gebaseerd op individuele vragen en 40% op groepsprestatie.

Wanneer men niet geslaagd is voor beide onderdelen van de evaluatie, kan men niet meer slagen voor het geheel van het opleidingsonderdeel. Indien de eindscore toch een cijfer van tien of meer op twintig zou zijn, wordt dit teruggebracht tot 8/20.

Teams zullen op lukrake wijze worden geselecteerd om hun resultaten voor groepswerk te presenteren. Op deze wijze kunnen maximum twee bonuspunten (op 20) worden verdiend.