



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

Studiepunten 6.0      Studietijd 165 u      Contacturen 62.5 u

Aanbodssessies en werkvormen in academiejaar 2018-2019

A (semester 2)	Engels	werkcollege: PC- klasoefeningen	15.0 u
		zelfstandig werk	10.0 u
		hoorcollege	22.5 u
		begeleide zelfstudie	15.0 u

Lesgevers in academiejaar 2018-2019

Saeyns, Yvan	WE02	Verantwoordelijk lesgever
Peralta, Daniel	WE02	Medelesgever

Aangeboden in onderstaande opleidingen in 2018-2019

	stptn	aanbodssessie
<a href="#">Master of Science in de informatica</a>	6	A
<a href="#">Uitwisselingsprogramma informatica (niveau master)</a>	6	A

Onderwijstalen

Engels

Trefwoorden

Regression, classification, model building, dimension reduction, big data

Situering

Familiarize the students with the most important methods to extract information from large databases in a statistical way. The students are expected to learn how to use these techniques correctly in applications and they acquire the skills to interpret obtained results in a statistically correct manner.  
The students will also be introduced to big data and the problems this might impose. This course builds on the content of 'Analysis of continuous data' and 'Categorical data analysis' and assumes the student has acquired the skills taught in 'Statistical Computing'.

Inhoud

### **Distributed Databases**

- noSQL database systems
- Distributed DBMS
  - Distributed data storage
  - Distributed query processing
  - Distributed transaction model
  - Homogeneous and heterogeneous solutions
  - Client-server distributed databases
- Parallel DBMS
  - Parallel DBMS architectures: shared memory, shared disk, shared nothing;
  - Speedup and scale-up, e.g., use of the MapReduce processing model
  - Data replication and weak consistency models

### **Data Mining**

- Uses of data mining
- Data mining algorithms
- Associative and sequential patterns
- Data clustering
- Market basket analysis
- Data cleaning

- Data visualization

#### Begincompetenties

Required prerequisites: Have a thorough understanding of linear regression as taught in the course 'Analysis of continuous data'.

#### Eindcompetenties

- 1 Explain the techniques used for data fragmentation, replication, and allocation during the distributed database design process. [Familiarity]
- 2 Evaluate simple strategies for executing a distributed query to select the strategy that minimizes the amount of data transfer. [Assessment]
- 3 Explain how the two-phase commit protocol is used to deal with committing a transaction that accesses databases stored on multiple nodes. [Familiarity]
- 4 Describe distributed concurrency control based on the distinguished copy techniques and the voting method. [Familiarity]
- 5 Describe the three levels of software in the client-server model. [Familiarity]
- 6 Compare and contrast different uses of data mining as evidenced in both research and application. [Assessment]
- 7 Explain the value of finding associations in market basket data. [Familiarity]
- 8 Characterize the kinds of patterns that can be discovered by association rule mining. [Assessment]
- 9 Describe how to extend a relational system to find patterns using association rules. [Familiarity]
- 10 Evaluate different methodologies for effective application of data mining. [Assessment]
- 11 Identify and characterize sources of noise, redundancy, and outliers in presented data. [Assessment]
- 12 Identify mechanisms (on-line aggregation, anytime behavior, interactive visualization) to close the loop in the data mining process. [Familiarity]
- 13 Describe why the various close-the-loop processes improve the effectiveness of data mining. [Familiarity]

#### 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

Begeleide zelfstudie, hoorcollege, zelfstandig werk, werkcollege: PC-klasoefeningen

#### Toelichtingen bij de didactische werkvormen

Minerva will be used to ensure a smooth organisation and follow-up of the practical assignments.

#### Leermateriaal

A syllabus is available Geraamde totaalprijs: 10 EUR

#### Referenties

Hastie T., Tibshirani R., and Friedman J. (2001), "The Elements of Statistical Learning: Data Mining, Inference and Prediction", Springer-Verlag.

#### Vakinhoudelijke studiebegeleiding

The exercises and practical assignments are supervised by the lecturer.

#### Evaluatiemomenten

periodegebonden evaluatie

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

Mondeling examen, werkstuk

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

Mondeling examen, werkstuk

#### Evaluatievormen bij niet-periodegebonden evaluatie

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

Niet van toepassing

#### Toelichtingen bij de evaluatievormen

Project work shows the student's capability to analyse real data problems by using the methods taught in this course and interpret correctly the obtained results.

Oral examination based on a written report about the results of the student's project work.

Eindscoreberekening

100% examination