



## Datamining (C002703)

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

Studiepunten 3.0      Studietijd 80 u      Contacturen 25.0 u

Aanbodssessies en werkvormen in academiejaar 2018-2019

A (semester 1)	Engels	werkcollege: PC- klasoefeningen	10.0 u
		hoorcollege	15.0 u

Lesgevers in academiejaar 2018-2019

Saeys, Yvan      WE02      Verantwoordelijk lesgever

Aangeboden in onderstaande opleidingen in 2018-2019

	stptn	aanbodssessie
<a href="#">Master of Science in Biochemistry and Biotechnology</a>	3	A
<a href="#">Uitwisselingsprogramma biochemie en biotechnologie (niveau master)</a>	3	A

Onderwijstalen

Engels

Trefwoorden

Bioinformatica, Data mining, Data-analyse, Patroonherkenning, Classificatie, Regressie, Kenmerkselectie, Sequentie en microarray data

Situering

Dit opleidingsonderdeel biedt de student een overzicht van de technieken in het domein van data mining, en hun toepassing in de bioinformatica. Het analyseren van hoogdimensionale en complexe data is een probleem dat zich steeds meer en meer stelt in de biotechnologie/biologie, en geavanceerde computertechnieken om deze data te analyseren dringen zich dus op.

In de cursus komen zowel classificatie- als regressiemethoden aan bod, en de student wordt vertrouwd gemaakt door de methoden die gezien worden in de theorie toe te passen op bestaande problemen door middel van sessies met behulp van bestaande software en data sets.

Dit opleidingsonderdeel sluit aan bij volgende opleidingscompetenties: Ma.WE.BB.1.2, Ma.WE.BB.1.3, Ma.WE.BB.2.5, Ma.WE.BB.2.6

Inhoud

- Overzicht van data mining
  - Ontwerpcyclus van data mining algoritmen
  - Verband tussen bias, variantie, en modelcomplexiteit
- Classificatie methoden:
  - Nearest Neighbors
  - Classificatiebomen
  - Bayesiaanse methoden
  - Toepassingen in de bioinformatica
- Clustering methoden:
  - Hierarchisch clusteren
  - K-means
  - Self-organizing maps
  - Toepassingen in de bioinformatica
- Regressiemethoden:
  - Regressiebomen
  - Principal Component Regression en Partial Least Squares
- Modelselectie en evaluatie:
  - methoden om de fout te schatten

- Cross-validatie
- Bootstrap
- Ensemble methoden
- Bagging - Boosting
- Methoden om met hoogdimensionale data om te gaan:
  - Principal Component Analysis
  - Independent Component Analysis
  - Kenmerkselectie

#### Begincompetenties

Basiskennis bioinformatica, informatica en statistiek

#### Eindcompetenties

- 1 De student is in staat om voor een gegeven data mining probleem de meest geschikte methode te gebruiken om een vooropgesteld doel te verwezenlijken.
- 2 De student is in staat om recente literatuur te begrijpen, te verwerken en toe te passen.

#### 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, project, werkcollege: PC-klasoefeningen

#### Toelichtingen bij de didactische werkvormen

Theorie: software demonstraties

Oefeningen: PC-zaal, eigen implementatie en gebruik bestaande pakketten onder de vorm van practica met verplichte verslagen, projectwerk in kleine groepen

#### Leermateriaal

Powerpoint presentaties van de theorielessen, ondersteunend leermateriaal, programmacode en data zijn beschikbaar op Minerva. Geraamde totaalprijs: 10 EUR

#### Referenties

T. Mitchell (1997). Machine Learning. McGraw-Hill

R.O. Duda, P.E. Hart, and D.G. Stork. (2001) Pattern Classification, Wiley, New York

P. Baldi and S. Brunak (1998) Bioinformatics, the machine learning approach, MIT Press

#### Vakinhoudelijke studiebegeleiding

De docent voorziet een spreekuur voor vragen i.v.m. de theorie.  
Begeleide computeroefeningen.

#### Evaluatiemomenten

periodegebonden en niet-periodegebonden evaluatie

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

Mondeling examen

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

Mondeling examen

#### Evaluatievormen bij niet-periodegebonden evaluatie

Verslag

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

Examen in de tweede examenperiode is mogelijk

#### Toelichtingen bij de evaluatievormen

Periodegebonden evaluatie: presentatie van projectwerk en mondelinge ondervraging

#### Eindscoreberekening

Periodegebonden evaluatie (theorie) (50%) + niet-periodegebonden evaluatie (oefeningen) (50%) Indien de student niet is geslaagd voor de niet-periodegebonden evaluatie, wordt er een tweede kans geboden via een compenserende activiteit tussen de eerste en de tweede examenperiode.