Vak: Introductie Datamining — credits: 3

Vakcode	BFVH18IDMN
Naam	Introductie Datamining
Studiejaar	2023-2024
ECTS credits	3
Taal	Nederlands
Coördinator	D.R.M. Langers
Werkvormen	Practicum / Training
Toetsen	Introductie Datamining - Schriftelijk, organisatie ToetsCentrum

Leeruitkomsten

- Onderscheid kunnen maken tussen typen (supervised/unsupervised) machine learning problemen
- Samenstellen, opschonen en transformeren van (trainings- en test-) datasets geschikt voor datamining
- Interactief classificatie, regressie en clustering analyses kunnen uitvoeren en evalueren m.b.v. Weka
- Begrijpen van de globale werking en kenmerken van diverse "klassieke" machine learning algoritmen
  - Rule-based (ZeroR, OneR)
  - Trees (Id3, J48)
  - Naive Bayes
  - (k)-Nearest Neighbor
  - Regressie (lineair, logistisch)
  - Ensemble learning (bagging, randomisation, boosting, voting/stacking)
- Interactief diverse classificatie, regressie en clustering analyses kunnen uitvoeren en evalueren m.b.v. Weka
- Kunnen evalueren en vergelijken van de prestatie van verschillende algoritmen
- Instellingen en hyperparameters van analysemethoden kunnen optimaliseren
- Relevante attributen in datasets kunnen identificeren en selecteren
- Eenvoudige berekeningen kunnen uitvoeren aangaande informatie-inhoud, formule v. Bayes, prestatiematen

Inhoud

Veel biologische datasets zijn te onoverzichtelijk om in een oogopslag de biologische vraag erachter te kunnen oplossen. Door gebruik te maken van geautomatiseerde machine learning technieken kan in een grote, complexe dataset gezocht worden naar relevante patronen. Tijdens deze module wordt kennis gemaakt met datamining toepassingen, algoritmen en analyses. In het bijzonder wordt de werking van een aantal "klassieke" classificatie-algoritmen behandeld. Geleerd wordt hoe de prestaties van deze algoritmen kunnen worden geevalueerd, geoptimaliseerd en vergeleken om zo te komen tot een optimale analyse voor een gegeven dataset. Diverse voorbeelden van datasets worden bekeken en voorbereid t.b.v. nadere analyse.

Opgenomen in opleiding(en)

Bio-Informatica

School(s)

Instituut voor Life Science & Technology

share your talent. move the world.

De ECTS onderwijscatalogus van de Hanzehogeschool Groningen wordt met de grootst mogelijke zorg samengesteld. Het is echter mogelijk dat de inhoud van de catalogus -en de daarin vervatte informatie- verouderd, incompleet of onjuist is. Aan de inhoud van de catalogus kunnen dan ook geen rechten worden ontleend.