

Data Science bij het CIT

Er komen steeds meer databronnen beschikbaar die kunnen worden benut voor wetenschappelijk onderzoek. Door data uit verschillende bronnen te combineren, kunnen interessante inzichten ontstaan. Nieuwe technieken maken het mogelijk om in de verzamelde data waardevolle informatie te vinden, door bijvoorbeeld het ontdekken van patronen, onregelmatigheden en interessante verbanden.

Om de toenemende hoeveelheid data werkelijk te benutten, is een goede en slimme analyse van de data nodig. Het uitvoeren van zo'n analyse vraagt om de nodige expertise waar een data scientist bij kan helpen om uit de grote bulk aan informatie juist die informatie te halen die relevant is.

Nieuwe wetenschappelijke inzichten

Data science gaat over het ontdekken van verborgen patronen in grote hoeveelheden data. Waar de ICT-infrastructuur en applicaties het verzamelen van grote hoeveelheden data mogelijk maken, is het juist de data scientist die dit kan omzetten naar waardevolle informatie. Hierbij wordt een combinatie van vaardigheden toegepast op het gebied van onder andere statistiek, kunstmatige intelligentie en informatica. Het doel is om relevante informatie uit de grote hoeveelheden data te halen, met behulp van machine learning-algoritmes waarbij gebruik gemaakt kan worden van bijvoorbeeld neurale netwerken (deep learning), regressieanalyses en support vector machines.

Verskillende onderzoeksdomeinen

De activiteiten worden uitgevoerd door multidisciplinaire teams, omdat voor de diversiteit van data science-projecten verschillende kennis en vaardigheden nodig zijn. Hierbij werken de specialisten van het CIT nauw samenwerken met partners van zowel binnen als buiten de universiteit, zoals Data Science and Systems Complexity (DSSC) van de Faculteit Wiskunde en Natuurwetenschappen, KPN consulting, IT Academy Noord-Nederland en het eScience center. Zo wordt het mogelijk teams te vormen met specialisten vanuit de verschillende (onderzoeks)domeinen.

Recente projecten

De data scientists van het CIT hebben al aan verschillende projecten meegewerkt. Op dit moment lopen er tien projecten. Een paar voorbeelden:

- **Data Science for Sustainable Data Centers:** voorspellingen doen over het energieverbruik van datacenters op basis van verschillende factoren zoals processorbelasting, omgevingstemperatuur en tijdstip.

De rol van ICT in wetenschappelijk onderzoek is de afgelopen jaren sterk toegenomen. Steeds meer wetenschappers concentreren hun onderzoek rond grote datasets en hebben gespecialiseerde ICT-faciliteiten nodig om met die data om te kunnen gaan. Het CIT ondersteunt onderzoekers daarin op verschillende manieren, en sinds kort ook door middel van data science.

De modellen die uit dit onderzoek komen, kunnen worden ingezet om het energieverbruik omlaag te brengen.

- **Data science for Center for Medical Imaging:** analyses van CT-scans ter ondersteuning bij de diagnose van longkanker en ernstige kransslagaderverkalking.
- **Automated analysis of discursive practices on Twitter:** ontwikkeling van een tool om tweets te selecteren en annoteren. Wanneer er voldoende tweets zijn geannoteerd, kan met behulp van een machine learning-algoritme automatisch nieuwe tweets worden geannoteerd.
- **Identifying Dutch accents automatically:** aan de hand van geluidsfragmenten de computer leren om automatisch te herkennen waar een spreker vandaan komt.

IT Strategiefonds

Er wordt gewerkt aan een jaarlijkse call for proposals. Onderzoekers kunnen projectvoorstellen indienen die worden gefinancierd vanuit het IT-Strategiefonds van de RUG. Het streven is om de call nog voor deze zomer te openen.

Wanneer u op de hoogte gehouden wilt worden van de ontwikkelingen met betrekking tot data science en andere onderwerpen op het gebied van ICT en onderzoek, kunt u zich aanmelden voor de mailinglist van de afdeling Research and Innovation Support (RIS) van het CIT via ris@list.rug.nl. 