

The background features a series of concentric, light gray circles centered on the slide. Overlaid on these are stylized, light blue circuit-like lines with small circles at the nodes, appearing in the corners of the slide.

DATA SCIENCE – AFSPRAKEN

JENS BAETENS

WIE BEN IK?

Jens Baetens

- Gestudeerd in Gent
- Burgerlijk ingenieur computerwetenschappen : ingebedde systemen
- Doctoraat in de computerwetenschappen
- Toadi
- Technologie, Robotica, Ruimtevaart, Houtbewerking

INHOUD - KEUZETRAJECT

Data Science – 5 studiepunten

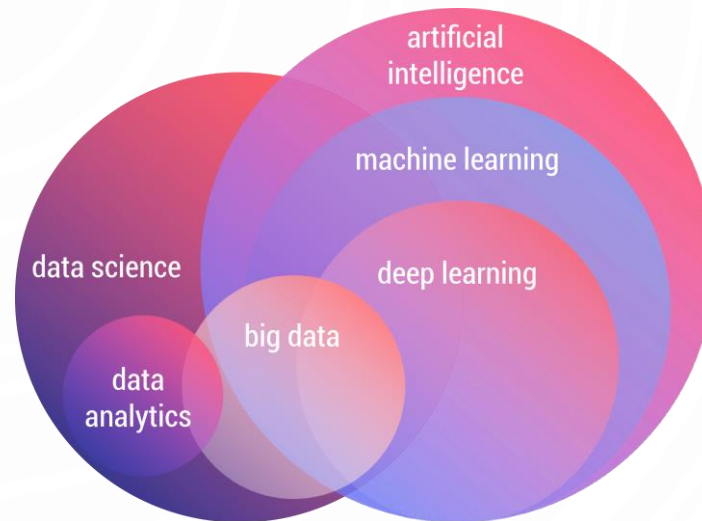
Big Data – 5 studiepunten

Machine Learning – 6 studiepunten

The diagram consists of three overlapping circles. The leftmost circle is labeled 'data science'. The bottom-right circle is labeled 'big data'. The rightmost circle is labeled 'machine learning'. The intersection of 'data science' and 'big data' is labeled 'data analytics'. The intersection of 'data science' and 'machine learning' is labeled 'machine learning'. The intersection of 'big data' and 'machine learning' is labeled 'deep learning'. The intersection of all three circles is labeled 'deep learning'.

Big Data – 5 studiepunten

Machine Learning – 6 studiepunten



ECTS – DATA SCIENCE

Data management

- Obtaining Data
- Data exploration
- Data Cleaning

Machine Learning

- Supervised vs Unsupervised
- Regressie vs Classificatie vs Clustering
- Loss functions
- Over/Under fitting
- Trainingset - Validatieset – Testset
- Hyperparameters
- Neurale Netwerken

VERLOOP

Lessen verlopen online (tot 28 oktober) en op campus (23 november-11 januari)

Twee kartalen met in totaal 12 lesblokken van 4 uur

Na de lesblokken nog tijd voorzien voor de mondelinge verdediging van het project

EVALUATIE

Project: 50%

Projectopgave volgt na de les over data-cleaning

Oefeningen: 50%

Oefeningen die tijdens en na de lessen gemaakt worden

Respecteer de deadlines hiervoor, te laat indienen is een 0 voor die oefening

Elkaar helpen mag maar kopiëren is niet toegestaan



PROJECT

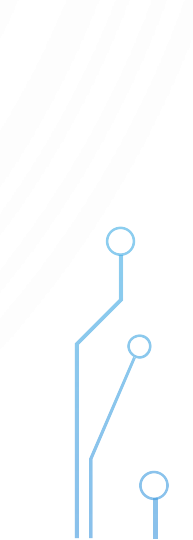
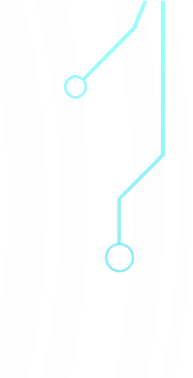

Aandeel: 50% van de totaalscore

In groepjes van 2 studenten

Bestuderen en voorspellen van olympische resultaten

Deadline: zondag **16 januari 23:59**

Verdedigingen: Dinsdag **18 en 25 januari**



STUDIEMATERIAAL

Slides en voorbeeld code

- Github repository

Slides

- Toledo

Opdrachten + Project

- Github Classroom

Tip: Maak voor jezelf een glossary aan waarin je in eigen woorden de termen uitlegt!

AFSPRAKEN

Wees op tijd

In het geval er vragen zijn buiten de lessen:

- Mag steeds via Teams en/of mail

Voor de oefeningen mag je elkaar helpen maar code kopiëren is niet toegestaan

Actief zijn in de les en oefenen zijn de beste leermethoden

The background features a series of concentric, light gray circles centered on the page. In the four corners, there are stylized circuit board traces in a light blue color, with small circles at the end of the lines, resembling electronic components or data paths.

TOOLS

GITHUB

Gebruik github noodzakelijk door classroom

Bij het indienen van oefeningen en project

- Zorg dat alles op de main branch staat
- Zorg dat alles werkt
- Voeg naar gewenst commentaar toe

Zorg ervoor dat je geen datasets commit (.csv, ... file)

- Normaal in de .gitignore file



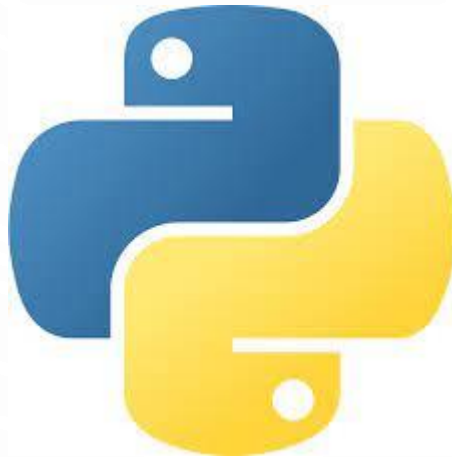
PYTHON

Scripting programming language

Installeer het via Anaconda:

- <https://www.anaconda.com/products/individual#Downloads>

Bevat een heel aantal handige packages voor data science en machine learning



JUPYTER NOTEBOOK

Web based applicatie om interactive documenten te maken

Soort van document waarin code kan geschreven en uitgevoerd worden

Standaard mee geïnstalleerd met Anaconda



KAGGLE

Website voor competities van Machine Learning / Data Science

Gaan we vaak gebruiken om datasets te downloaden

Maak hier alvast een account voor aan

The Kaggle logo, featuring the word "kaggle" in a blue, lowercase, sans-serif font. A small "TM" trademark symbol is located to the upper right of the letter "e". The logo is centered within a light gray rectangular box.

kaggle™