

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

How to participate?



Click on the projected screen to start the question

wooclap

100 %

30



www.wooclap.com/FULZUQ

Wat hoop je te leren in dit traject?



Werken met
data en
databases

AI om
beurskoers te
voorspellen

Classificatie
van bepaalde
data

AI

AI

Click on the projected screen to start the question

Machine
learning

Kunnen
omgaan met
zeer veel data

Een dating AI
maken

Leren werken
met artificiële
intelligentie

Hoe AI in
elkaar zit en
ermee leren
werken

Hoe AI in
elkaar zit, wat
we er

Toekomst
voorspellen



Data efficiënt

wooclap



100 %



24



Waarmee denk je dat ik je het meest moet helpen om te bereiken wat je ...



Python

GOTTEM

DN

Wiskunden
Python

Logica achter
de AI/grondige
Python coding
uitleg

geen idee... ik
weet nog niet
echt wat de
moeilijkheden
aan zijn

oefeningen

Python als
programmeert
a al goed
uitleggen en
nuttig en

Statistiek

oefeningen

Logica/wiskun
dig aspect

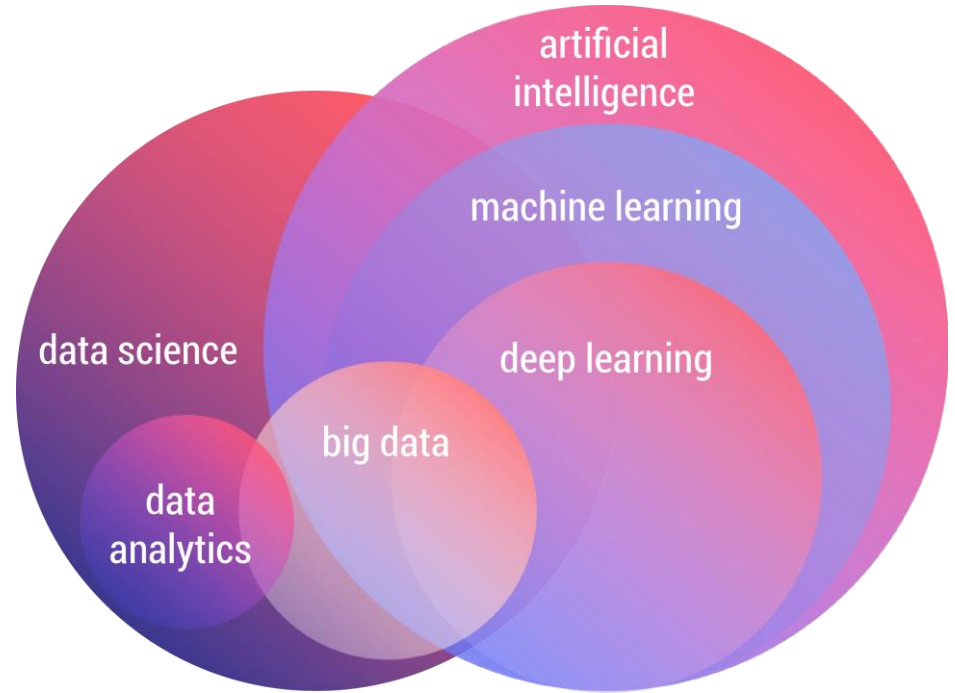
Voor al het
begrijpen en
onder de knie

Het logica



Inhoud - Keuzetraject

- Data Science – 5 studiepunten
- Big Data – 5 studiepunten
- Machine Learning – 6 studiepunten



ECTS



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

Verloop

- Lessen verlopen online (kwartaal 1) en op campus (kwartaal 2)
- 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: 40%
 - Projectopgave volgt later
 - Individuele oefeningen: 30%
 - Oefeningen die na de lessen gemaakt worden
 - Groepsoefening met bijhorende presentatie: 10%
 - Theoretische test: 20%
-
- Respecteer de deadlines hiervoor, te laat indienen is een 0 voor die oefening
 - Elkaar helpen bij oefeningen mag maar kopiëren is niet toegestaan

Project

- Aandeel: 40% van de totaalscore
- In groepjes van 2 studenten
- Opdracht over het wereldkampioenschap voetbal
 - <https://www.kaggle.com/datasets/abecklas/fifa-world-cup>
- Deadline: Zondag **08 januari 23:59**
- Verdedigingen: Dinsdag **10 en 17 januari (online)**

Studiemateriaal

- Slides en voorbeeld code
 - Github repository
- Opdrachten + Project
 - Github Classroom
- Tip: Maak voor jezelf een glossary/woordenlijst aan
 - Eigen verklaring en uitleg voor alle termen die je tegenkomt!

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

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, ... f)
 - 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
 - Quick guide: <https://www.purdue.edu/cancer-research/bigcare/Course/Pre-course/Jupyter-Notebook-Quick-Guide.pdf>
- Je kan ook werken via Visual Studio Code



Kaggle

- Website voor competities van Machine Learning / Data Science
- Gaan we vaak gebruiken om datasets te downloaden
- Maak hier alvast een account voor aan

