

## Opzet applicatie traineeship search4solutions

Thomas Mons, September 2019

Dit programma gaat de uitgebreide informatie die beschikbaar is over planeten buiten ons zonnestelsel (exoplaneten) analyseren met behulp van machine learning. Aangezien er veel informatie bekend is over de planeten, maar niet bekend is welke eigenschappen van een planeet belangrijk zijn voor bijvoorbeeld buitenaards leven gaat dit programma de bekende planeten clusteren. De hoop is dat op deze manier enkele duidelijk onderscheidbare klassen van planeten ontstaan. Vervolgens kan de aarde ook ingedeeld worden in een van de klassen, en is op deze manier een klasse planeten geïdentificeerd die het meest op de aarde lijken (en wellicht meer potentie hebben op buitenaards leven).

### Stappenplan

1. Download exoplaneten dataset van nasa  
<https://exoplanetarchive.ipac.caltech.edu/index.html>
2. Omzetten dataset naar database
3. Basisfunctionaliteit om data te visualiseren in c# (scatterplots en grafieken van eigenschappen naar keuze)
4. Selectie van nuttige eigenschappen (e.g. afstand tot ster, maar niet ontdekkende satelliet)
5. Schrijven van unsupervised cluster analysis in python met k-means clustering
  - Herschalen van parameters naar percentage tussen minimum en maximum (voor uniforme afstand)
  - Euclidische afstand gebruiken voor clusteren
  - Toevoegen clusterindeling aan database
6. Visualiseren verschillende planetenclusters in 2d of 3d weergaves (e.g. afbeelding hieronder)
7. Berekenen van eigenschappen (dichtheid cluster geeft maat voor efficiëntie algoritme)
8. Aarde (en eventueel mars) indelen bij cluster en planetenklasse vinden die het meest op aarde lijkt.

