De sleutel tot een goede algoritme-audit

Auditdienst Rijk
Fré Vink
f.t.vink@minfin.nl
16 november 2023

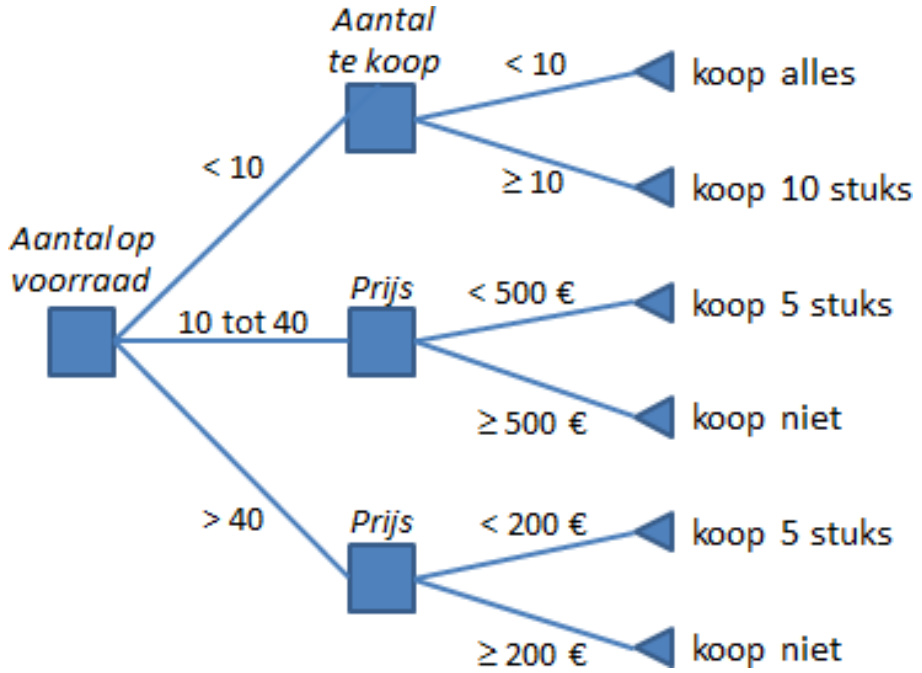
Wie heeft er alle vertrouwen in algoritmes?

Auditdienst Rijk
Ministerie van Financiën

Vertrouwen door audits

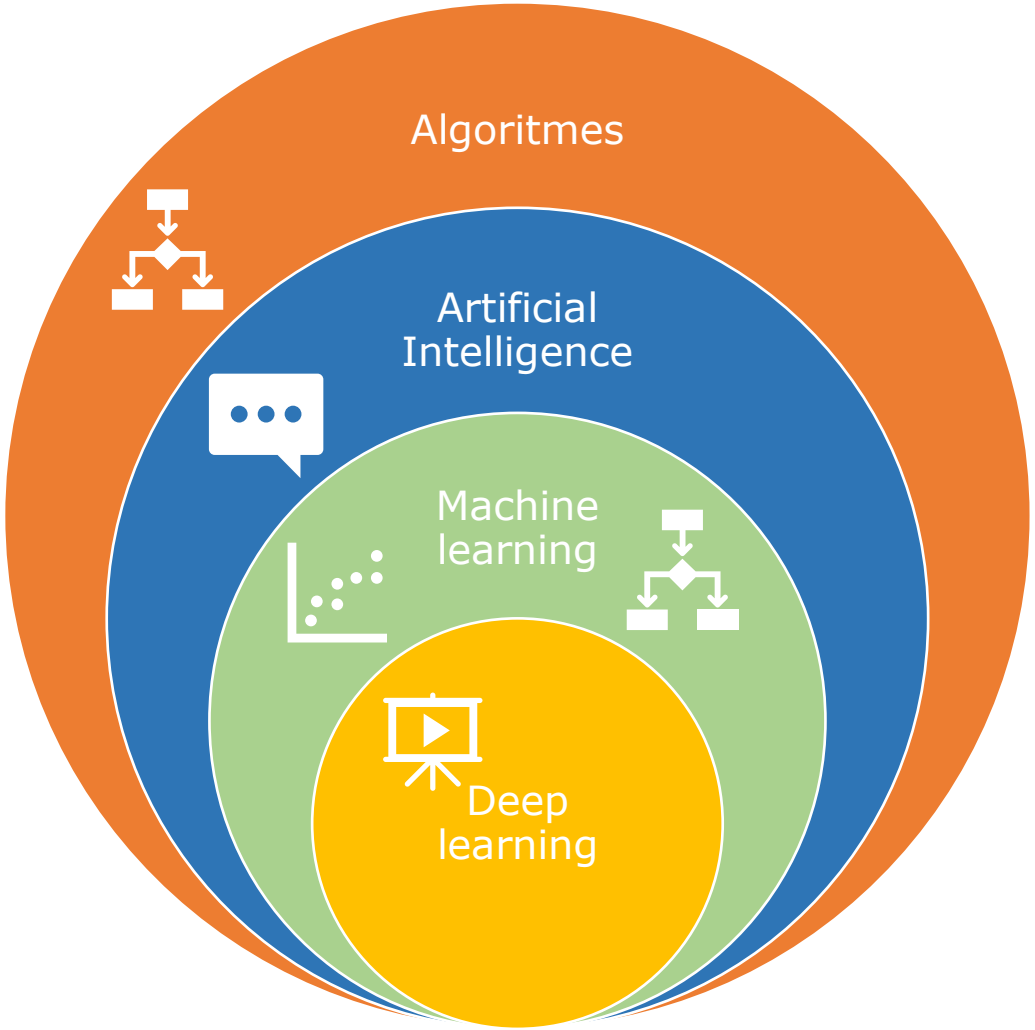
Wat is een algoritme?

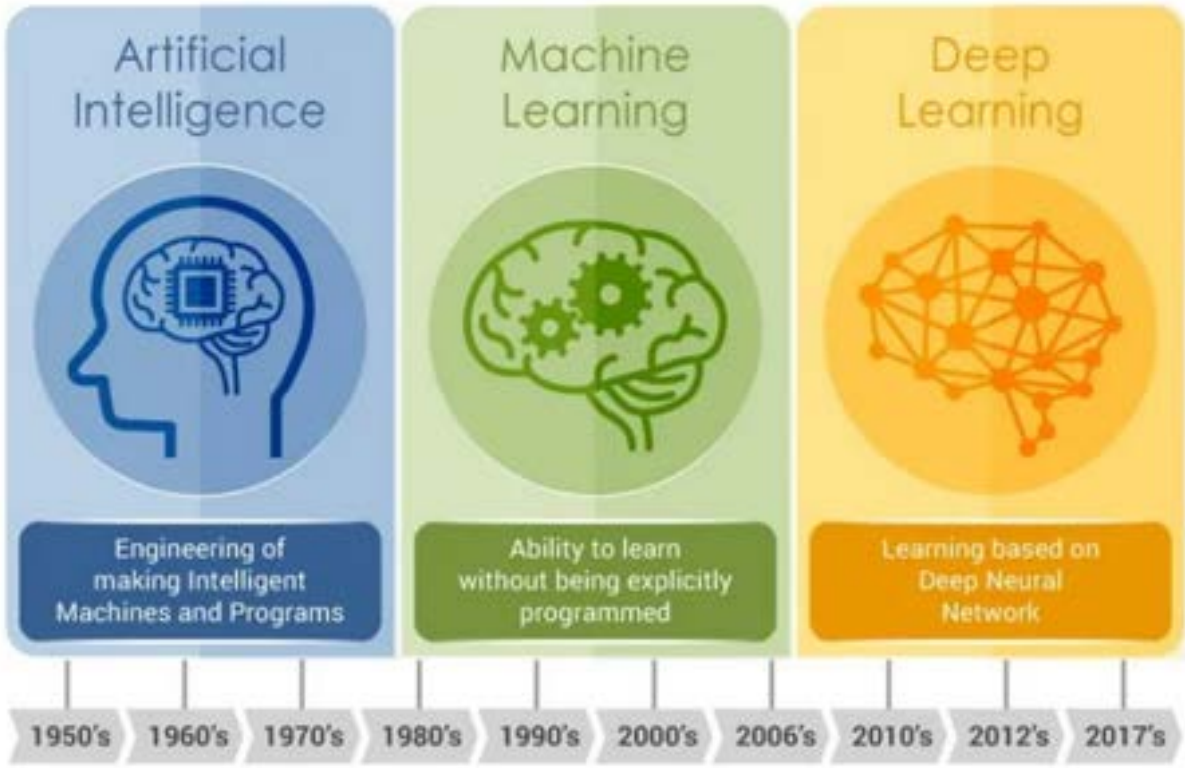
- *"een set van regels en instructies die een computer geautomatiseerd volgt bij het maken van berekeningen om een probleem op te lossen of een vraag te beantwoorden"* [1]



[1] Rapport 'Aandacht voor Algoritmes' 14 januari 2021. https://www.rekenkamer.nl/onderwerpen/algoritmes/documenten/rapporten/2021/01/26/aandacht-voor-algoritmes.

AI vs Machine Learning





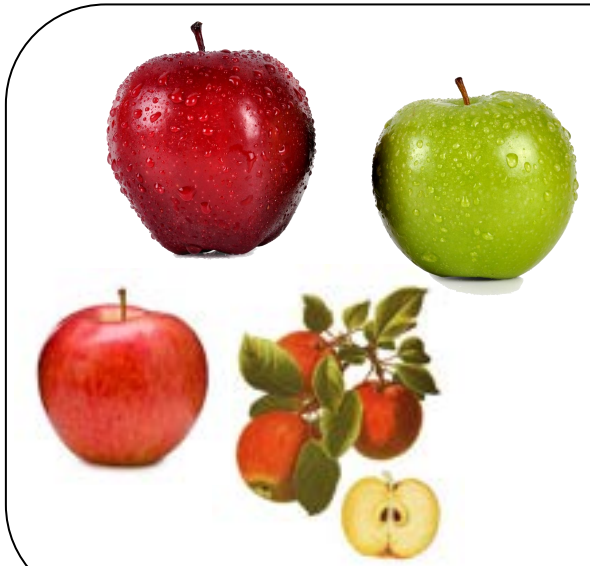
Machine Learning

https://www.cnblogs.com/dreamafar/p/6202622.html

Rule based

Object is (rond) & (rood of groen) & (heeft een steeltje) → Appel!

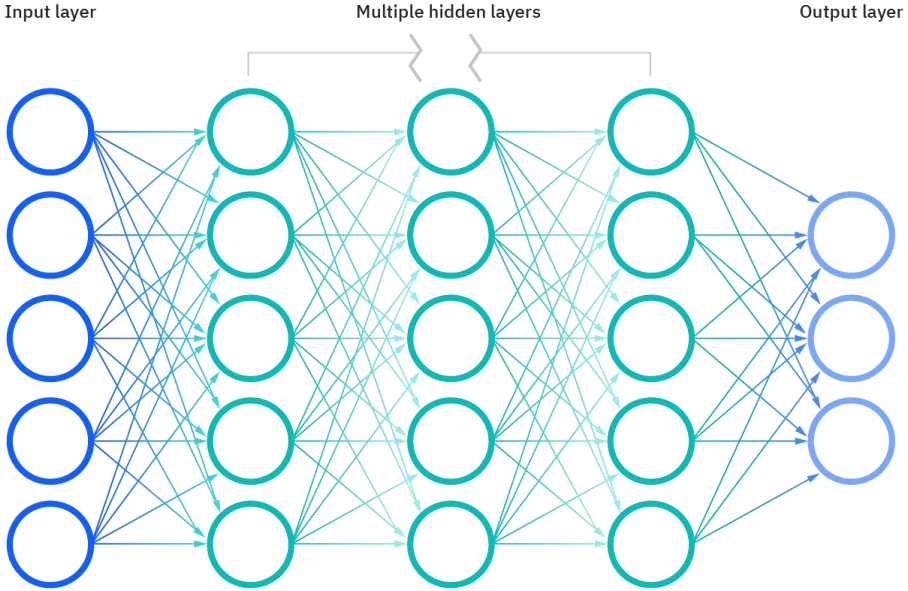
Machine learning



appel

geen appel

algoritme



Het algoritme bouwt een model van de werkelijkheid op basis van data

Algoritmespecificaties

Business rules

- Handmatige instructies voor de computer
- Bijv. wel of geen recht op zorgtoeslag
- Algoritme/AI

- Uitlegbaarheid: ★★★★★
- Eerlijkheid: ★★★★
- Duurzaamheid: ★★★★★
- Prestatie: ★★★
- Tijdsinvestering: ★

Lineaire regressie

- Lineaire relatie tussen input en output
- Bijv. huizenprijs t.o.v. oppervlakte
- Machine Learning

- Uitlegbaarheid: ★★★★
- Eerlijkheid: ★★
- Duurzaamheid: ★★★★★
- Prestatie: ★
- Tijdsinvestering: ★★★

Random forest

- Combinatie van beslisbomen
- Bijv. creditcardfraude
- Machine Learning

- Uitlegbaarheid: ★★
- Eerlijkheid: ★★
- Duurzaamheid: ★★★
- Prestatie: ★★★★
- Tijdsinvestering: ★★★★★

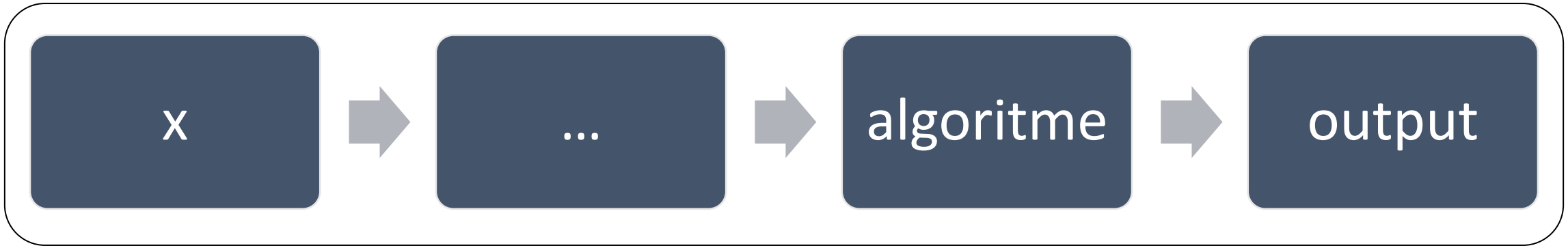
Neuraal netwerk

- Een netwerk met veel parameters als model
- Bijv. classificatie van afbeeldingen
- Deep Learning

- Uitlegbaarheid: ★
- Eerlijkheid: ★★★
- Duurzaamheid: ★
- Prestatie: ★★★★★
- Tijdsinvestering: ★★★★

Eisen aan het algoritme

Proces



Wat is nu de bedoeling?
Is het wenselijk?

Procesvereisten → eisen voor het algoritme

Uitlegbaarheid: ★★
Eerlijkheid: ★★★★
Duurzaamheid: ★★★★
Prestatie: ★★★
Tijdsinvestering: ★★

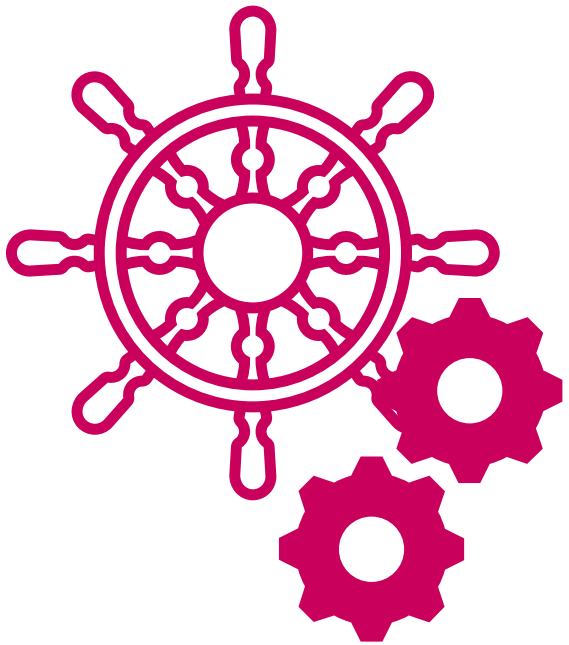
ADR onderzoekskader algoritmes

Sturing & Verantwoording
Privacy
Data & Model
Informatiebeveiliging

Sturing & Verantwoording

- SV.2: Een bewuste afweging of het algoritme het juiste middel is om het probleem op doelmatige en doeltreffende wijze op te lossen is gemaakt en vastgelegd.

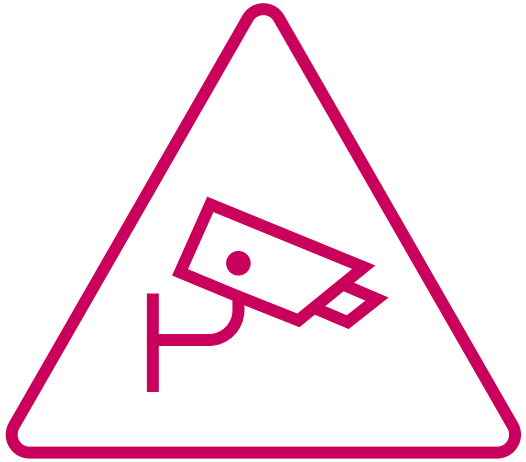
- SV.14: De inzet en werking van het algoritme is gepubliceerd en inzichtelijk voor de doelgroep. De mate van transparantie en uitlegbaarheid daarbij is afgewogen.



Privacy

- PRI.3: Passende maatregelen op basis van de gesignaleerde privacyrisico's zijn aanwezig. Deze maatregelen zijn zo veel mogelijk bij de ontwerpfase getroffen (privacy by design).

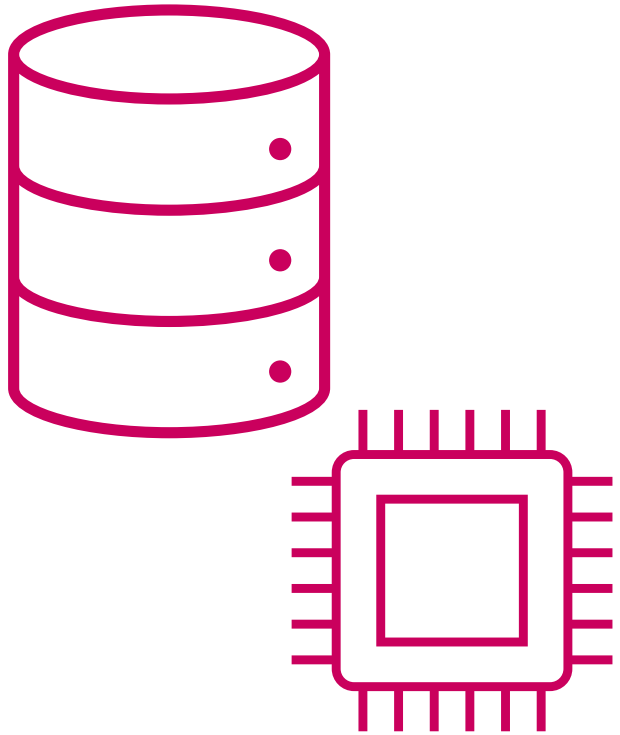
- PRI.10: Het algoritme betreft geen uitsluitend op geautomatiseerde verwerking, waaronder profilering, gebaseerd besluit waaraan voor de betrokkene juridische of anderszins aanmerkelijke gevolgen zijn verbonden.



Data & Model

- DM.1: De doelstelling van het algoritme is concreet uitgewerkt tot functionele eisen voor het algoritme. De mate waarin aan deze eisen is voldaan is bepaald.

- DM.17: De mate van geaccepteerde bias in de uitkomst is opgenomen in de functionele eisen en uitgewerkt in meetbare prestatiecriteria.



Informatiebeveiliging

- IB.7: Voor toegang vanuit een onvertrouwde omgeving dient, twee-factor authenticatie te worden gebruikt.

- IB.9: Wachtwoorden mogen niet in originele vorm (plaintext) worden opgeslagen, maar dienen versleuteld te worden.

Algoritme-auditteam

- Iemand met auditervaring
- Kennis in het team van:
  - Sturing & Verantwoording
  - Privacy
  - Data & Model
  - Informatiebeveiliging

- Aanbevolen:
  - Operational auditor
  - IT-auditor
  - Senior data scientist
  - Pas aan naar de context

Succes met jullie algoritme-audits!

Auditdienst Rijk
Fré Vink
f.t.vink@minfin.nl
16 november 2023