



Odisee
DE CO-HOGESCHOOL

Data Science – Week 1



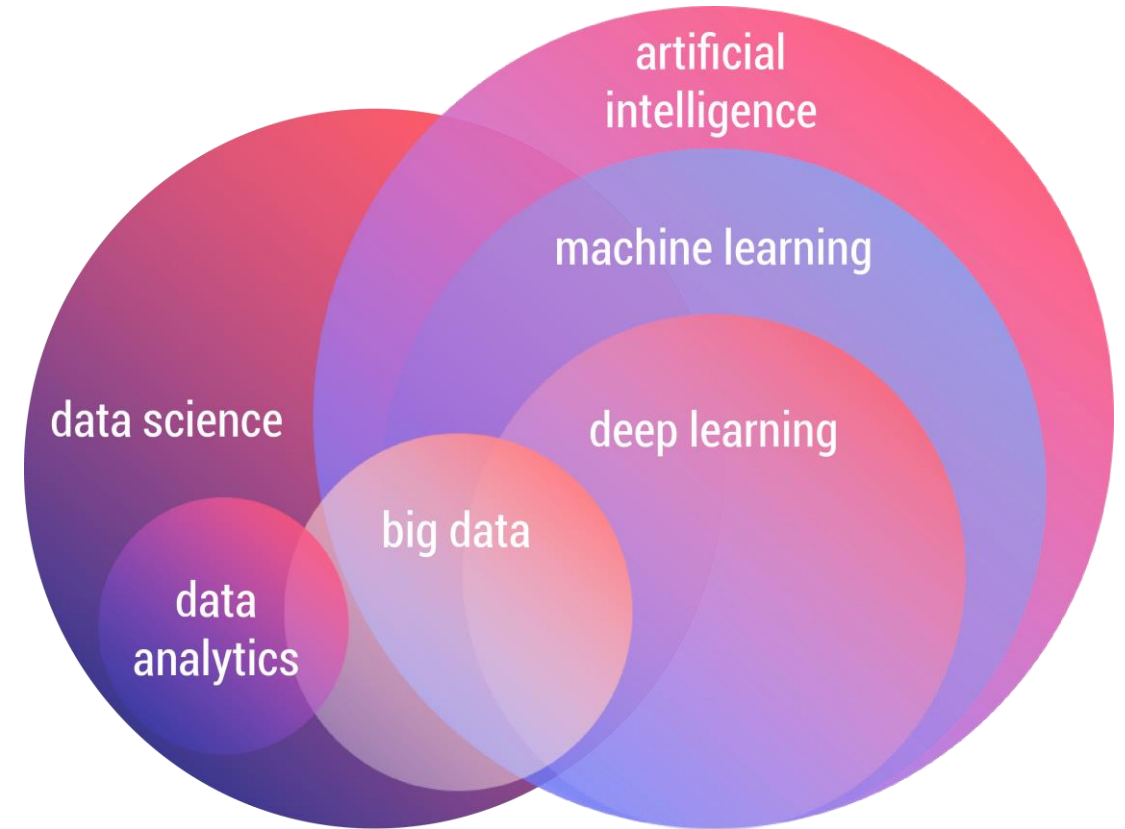
Jens Baetens

Wie ben ik?

- ▣ Burgerlijk ingenieur gestudeerd in Gent
- ▣ Gedoctoreerd in Gent
- ▣ Gewerkt bij Toadi (noemt momenteel Eeve)
- ▣ Passie voor:
 - Alles innovatief: ruimtevaart, mRNA vaccins, DNA, ...
 - Houtbewerking
 - Science-fiction genre



- 🌅 Data Science – 5 studiepunten
- Big Data – 5 studiepunten
- ▶ Machine learning – 6 studiepunten
(Niet voor microdegree)





Data Science - ECTS

▣ Data Management

- ▬ Obtaining data
- ▬ Data Exploration
- ▬ Data Cleaning

▣ Machine Learning

- ▬ Supervised vs unsupervised
- ▬ Regressie vs classificatie vs Clustering
- ▬ Loss functions
- ▬ Over/underfitting
- ▬ Training- vs validatie- vs testdataset
- ▬ Hyperparameters

How to participate?



Click on the projected screen to start the question

 [Copy participation link](#)

wooclap

 100 % 

61 





Go to **wooclap.com** and use the code **FULZUQ**



Wat verwacht je dat het moeilijkste zal zijn tijdens dit keuzetraject?



MEDESTUDENTEN WAKKER WORDEN VOOR 8U30 🤖 WISKUNDIG ASPECT
HET TRAJECT ZELF MAKEN 2 SEM OP CAMPUS CODEREN WAKKER WORDEN SCHRIJVEN
DATABANKEN WISKUNDE
SQL TIME MANAGEMENT MACHINE LEARNING INNOVATIEF ZIJN
GROUPSWERKEN PYTHON PYTHON DLYAA WAAROM OM 8 30 LES VOOR
PYTHON LIBRAIRIES GEBRUIKEN ENORME DATA
GROEPWERK WAT EEN IDEE 😊 JUISTE CODE SCHRIJVEN HET WISKUNDE

Click on the projected screen to start the question



wooclap



100 %



29





Afspraken en planning





Verloop van de lessen

- ▣ Twee kwartalen met in totaal 12 lesblokken van 4 uur
 - ▬ Na de lessen is er tijd voorzien voor mondelinge verdediging van het project (online)
- ▣ Lessen verlopen:
 - ▬ Kwartaal 1: online
 - ▬ Kwartaal 2: op campus
- ▣ Zie lessenrooster voor exacte datums



Studiemateriaal

- ▣ Slides en voorbeeld code
 - ▬ Zie Toledo -> Studiemateriaal -> Inhoud voor de link naar de github repository
- ▣ Opdrachten + Project
 - ▬ Via Github classroom
 - ▬ Link via Toledo
- ▣ Tip: maak voor jezelf een woordenlijst met korte beschrijving aan
 - ▬ Zal handig zijn om te leren voor de schriftelijke test



Evaluatie

- ▣ Theoretische test: 30%
 - ▣ Individuele oefeningen: 20%
 - ▣ Project: 50%
-
- ▣ Respecteer de deadlines: te laat ingediend resulteert in een 0
 - ▣ Elkaar helpen mag maar overnemen is niet toegestaan
-
- ▣ Herkansing bestaat uit het uitwerken van een opdracht:
 - 50% op de opdracht
 - 50% op de mondelinge bespreking



Project

- ▣ Opdracht: Voorspel welke partij welke gemeente zal winnen bij de verkiezingen volgend jaar
 - ▬ <https://classroom.github.com/a/IJTuvGFv>
- ▣ Aandeel: 50% van de score
- ▣ In groepjes van max 2 studenten
 - ▬ Zorg er wel voor dat iedereen alle code kent tijdens de mondelinge bespreking
- ▣ Deadline:
 - ▬ Code: Zondag na de laatste les
 - ▬ Mondelinge besprekingen: In de twee weken erna



Vragen?

- ▣ Indien er vragen zijn mag je me steeds contacteren via teams of mail
 - Meestal antwoord ik binnen de 2-3 dagen. Indien het langer duurt, stuur gerust een reminder (eventueel via het andere kanaal).



Tools



Github

▣ Gebruikt voor:

- ▬ Leerstof
- ▬ Oefeningen
- ▬ Project

▣ Verplicht door gebruik classroom

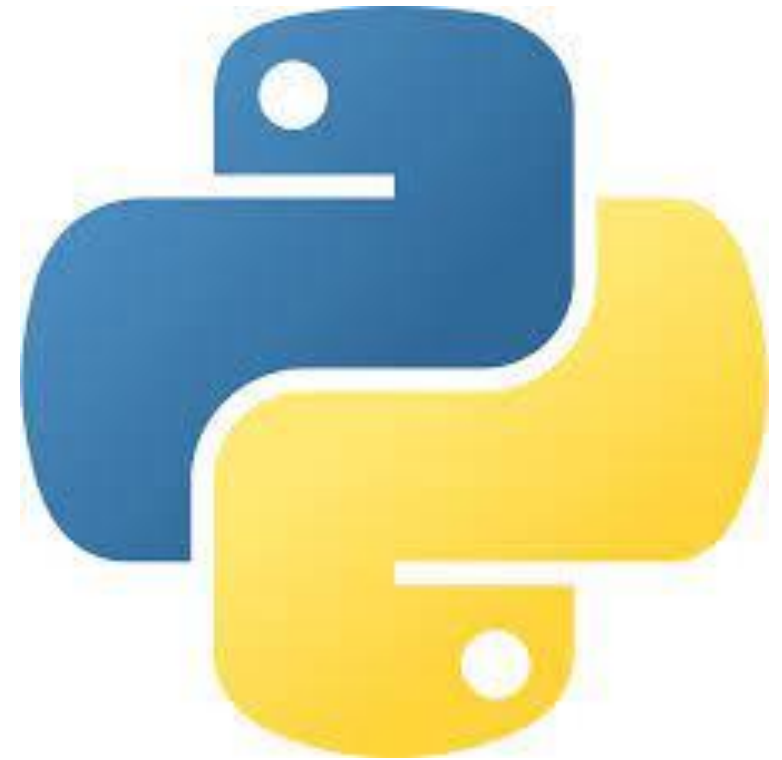
▣ Let op:

- ▬ Zorg dat je laatste versie zeker gecommit is
 - Add – commit – push
- ▬ Je commit best geen datasets (.csv, .xlsx, ...)



Python

- ▣ Scripting programming language
- ▣ Heel populair in Data Science/AI
- ▣ Installeer het via Anaconda
 - <https://www.anaconda.com/products/individual#Downloads>
- ▣ Bevat veel handige packages om het programmeerwerk te vereenvoudigen



Jupyter notebook

- ▣ Web-based applicatie om interactieve documenten te maken
 - ▬ Code die tekst met markdown bevat
- ▣ Geïnstalleerd samen met Anaconda
 - ▬ Quick guide: <https://www.purdue.edu/cancer-research/bigcare/Course/Pre-course/Jupyter-Notebook-Quick-Guide.pdf>
- ▣ Of via het commando
 - ▬ `pip install jupyter`
- ▣ Visual studio code kan dit openen





Kaggle

- ▣ Website voor competities van Machine Learning
- ▣ Gebruikt tijdens dit vak voor
 - Datasets
 - Extra informatie / tutorials / oefeningen
- ▣ Maak hier alvast een account voor aan tijdens een pauze



kaggle™



Data science – Wat is het?

2 MILLION+
VIEWS

WHAT IS DATA SCIENCE?



Data Science project - voorbeeld

Go to wooclap.com and use the code **FULZUQ**



Ik wil weten wie dit weekend de voetbalwedstrijd gaat winne...



ALLE UITSLAGEN VAN JAREN TERUG
VOORGAANDE SCORES ALLE SCORE FABRIZIO ROMANO
RONALDO DE GOAT STADIUM VORM VAN HET TEAM BLESSURES ELKE CLUB SUPPORTERS
VERSCHILLENDE WEDSTRIJDEN VAN SPELERS 16 HISTORIEK PLOEGEN
BETFIRST METEO GERAKEN WIN % TEAMS TEAM RATIO ANDERE COACH MESSI > RONALDO
AANWEZIGE SPELERS VORIGE BEIDE SCORES VERLIEZ VAAK PERFORMANTIE
VORIGE WEDSTRIJDEN INFO PYTHON DE VOETBALTEAMS HISTORIE
WEDSTRIJDSTATISTIEKEN
DE VORIGE WEDSTRIJDEN BINNEN OF BUITEN WEDSTRIJD



wooclap



100 %



24





Data Science voorbeeld – verwerken van de verzamelde data

- ▣ Spoor fouten op: Verkeerde ingave, labelling, ruis, ...
- ▣ Zorg dat de privacy in de data gegarandeerd is
- ▣ Reduceer het volume: verwijder onnodige data/duplicaten/ ...
- ▣ Zoek naar reeds bestaande patronen
 - ▣ ML is een zeer krachtige tool om patronen te zoeken



Data Science voorbeeld – model bouwen op basis van de beschikbare data

- ▣ Kies 1 of meerdere ML-technieken
- ▣ Train met de verzamelde data
- ▣ Valideer de resultaten en vergelijk verscheidene technieken

Data Science voorbeeld – Communiceer over je resultaten

- ▣ Hoe goed werkt het model?
- ▣ Wat kunnen we eruit leren?
- ▣ Doe dit op maat van je doelpubliek!

Landen	Odds*
Engeland	7.50
Belgie	8.00
Frankrijk	4.50
Nederland	15.00
Spanje	8.50
Duitsland	11.00
Portugal	8.50
Italie	7.50



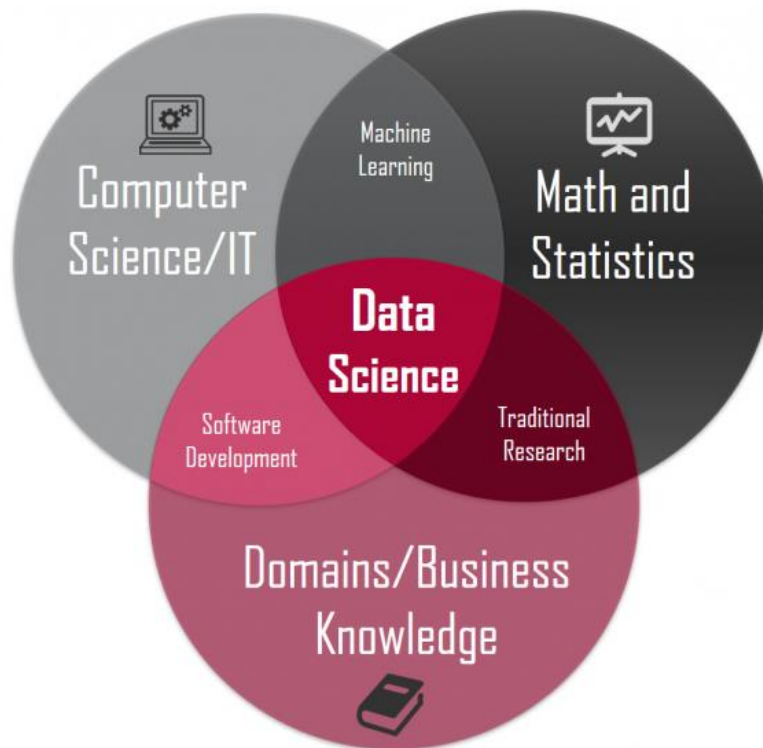
Wat is Data Science?

- ▣ Data Science is het vermogen om de juiste data
 - te selecteren
 - te begrijpen
 - te verwerken
 - te visualiseren

met als doel een bepaalde waarde uit de data te halen

Wat zijn de benodigde vaardigheden?

- ▣ Combinatie van IT, statistiek en domein kennis





Wat is het nut van Data Science?

- ▣ Hebben jullie in de zomer een voorbeeld gehoord?
 - Over Data Science, AI, ML?



colruyt laagste prijzen

HEALTH
IT ANALYTICS
xtelligent HEALTHCARE MEDIA

[Home](#) [News](#) [Features](#) [Interviews](#)

[Population Health](#) [Precision Medicine](#) [Quality & Governance](#) [Tools & Strategies](#) [Focus on AI](#) [Care](#)

AI Pain Care Tool May Increase Access, Reduce Costs for Chronic Pain Patients

A cognitive behavioral therapy intervention for chronic pain, personalized for patients using artificial intelligence, achieved similar results as standard therapies.

Bewusteloze chauffeur rijdt 23 kilometer op linkerrijstrook op E314 tussen Houthalen en Lummen

Een man van 41 is gisteren bewusteloos geraakt terwijl hij op de autosnelweg E314 reed tussen Houthalen en Lummen. Door de adaptieve cruisecontrol en de lane-assist die de auto op hetzelfde rijvak houdt, is de man 23 kilometer verder kunnen rijden vooraleer de politie hem heeft kunnen tegenhouden.

BBC

[Sign in](#)

[Home](#)

[News](#)

[Sport](#)

[Reel](#)

[Worklife](#)


NEWS

[Home](#) [War in Ukraine](#) [Coronavirus](#) [Climate](#) [Video](#) [World](#) [UK](#) [Business](#) [Tech](#) [Science](#) [Stories](#)

[Tech](#)

Meta's chatbot says the company 'exploits people'

By James Clayton
North America technology reporter



AI vermindert fors de onderhoudskosten en -inspanningen voor de 341 windmolens in de Noordzee

Use case

Maandag, 9 november 2020

In samenwerking met: [VUB](#), [De Blauwe Cluster](#)

341 windmolens zijn intussen operationeel in acht offshore parken voor de Belgische kust. Eind 2020 zullen het er 399 zijn en zullen ze energie leveren voor maar liefst 2,2 miljoen gezinnen. Wat we daarbij vergeten is dat deze energiecentrales niet zomaar blijven draaien. De opvolging en het onderhoud zijn intensief en uitdagend waarbij de kosten snel oplopen. Maar door nieuwe AI-modellen komt er een revolutie aan.

Rode Duivels geblesseerd? Ekipa voorspelt en voorkomt sportblessures met AI

Use case

Dinsdag, 29 juni 2021

In samenwerking met: [VLAIO](#)

De sportblessures van Kevin De Bruyne en Eden Hazard hebben de Rode Duivels zeker niet geholpen tijdens het afgelopen EK. Misschien brengt de AI-oplossing van Ekipa volgend tornooi raad. Hun software schat blessures van voetballers op voorhand in.



Staat het verkeerslicht op groen of rood? Deze Vlaamse AI-app vertelt het aan blinden en slechtzienden

Use case

Maandag, 30 mei 2022

Rateltickers aan verkeerslichten begeleiden vandaag mensen met een visuele beperking door het verkeer. Maar vaak zijn deze tikkende bakjes er niet of werken ze niet. Drie Antwerpse vrienden lanceren de oplossing. Hun app loodst slechtzienden veilig door het verkeer.

Financiële sector

- Financiële sector
 - ─ Fraude detecteren
 - ─ Risico's inschatten
 - ─ Beleggen op de beurs



Beeld ter illustratie © Getty Images

Fransen belastingdienst heeft al 20.000 niet-aangegeven zwembaden opgespoord dankzij artificiële intelligentie

FROST LAW

[HOME](#) [ATTORNEYS](#) [PRACTICE AREAS](#) [ARTICLES](#) [BLOG](#) [REQUEST A CONSULTATION](#)

| 202-505-6076

[Home](#) » [Blogs](#) » The IRS's Use of Data Analytics to Combat Tax Evasion

The IRS's Use of Data Analytics to Combat Tax Evasion

The IRS understands that the tax code is long and confusing. Therefore, the agency is willing to cut people slack if they make an honest mistake on their taxes. This reasonable attempt to comply with tax code is not a crime but referred to as negligence. However, tax evasion – willfully not paying a tax liability – is illegal. People try to avoid paying taxes by underreporting or omitting income, claiming false deductions and credits, concealing assets or improperly claiming tax credits or exemptions.

The IRS conducts investigations into alleged tax code violations through its [IRS Criminal Investigation Division](#) (IRS-CI). The IRS is committed to making the most effective use of their resources to ensure compliance with the tax code.

Healthcare

- ▣ Beter diagnoses
- ▣ Opvolgen van patiënten via chatbots/sensors
- ▣ Ontlasten personeel zodat ze kunnen focussen op hun kerntaken

AI versnelt opkomst van zichzelf vermeerderende mRNA-vaccins

De nieuwe generatie mRNA-vaccins heeft minder bijwerkingen zoals koorts, hoofdpijn en vermoeidheid door toepassing van AI bij de ontwikkeling ervan, aldus Wageningen University & Research.

#bijwerking, #corona, #mRNA, #vaccin



NIEUWS
**Benelux-primeur: AI helpt
Belgische ziekenhuizen bij
diagnose kanker**

Persbericht

Autisme voorspellen, kan dat?

Zou het mogelijk zijn? Zouden we met de *state of the art in machine learning*, de beschikbaarheid van grote hoeveelheden data en onze huidige kennis van het menselijke brein op vroege leeftijd kunnen voorspellen welke kinderen een risico lopen op het ontwikkelen van autisme? Onderzoek aan de Universiteit Gent suggereert dat die aanpak, nl. het gebruiken van steeds grotere datasets om mentale aandoeningen te voorspellen en te diagnosticeren, het paard achter de wagen spant. Waarom is dit het geval?



6 Data Science Use Cases in Healthcare



Data Science for
Medical Imaging



Data Science for
Genomics



Data Science for
Drug Discovery



Predictive
Analytics



Tracking and
Preventing Diseases



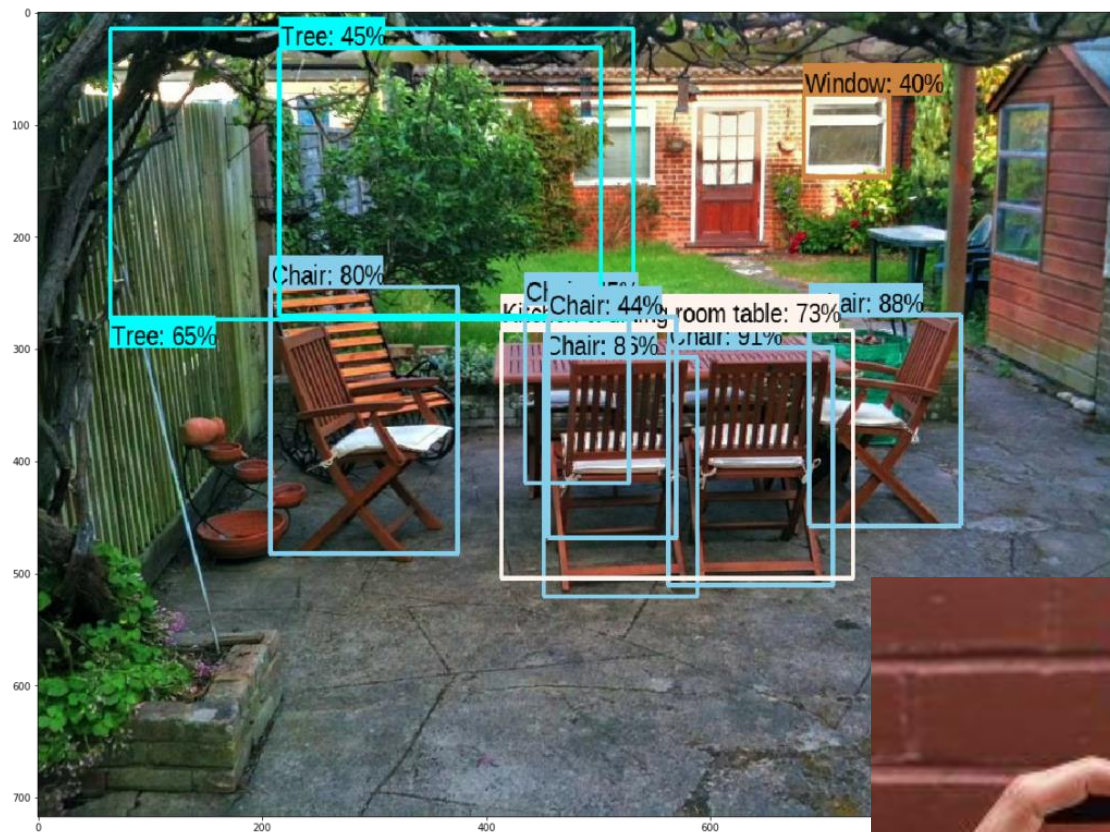
Data Science for
Wearables

Marketing

- ▣ Gepersonaliseerde reclame
- ▣ Klantenbinding
- ▣ Individuele behandeling van type klant
 - ▣ Soorten reclame en opvolging



Computer vision



A person riding a motorcycle on a dirt road.



A group of young people playing a game of frisbee.



A herd of elephants walking across a dry grass field.

Machine learning use cases in retail



Demand Prediction



Price Formation



Logistics



Merchandizing



Personalized Offers



Fraud Detection



Churn Prediction



Location Optimization

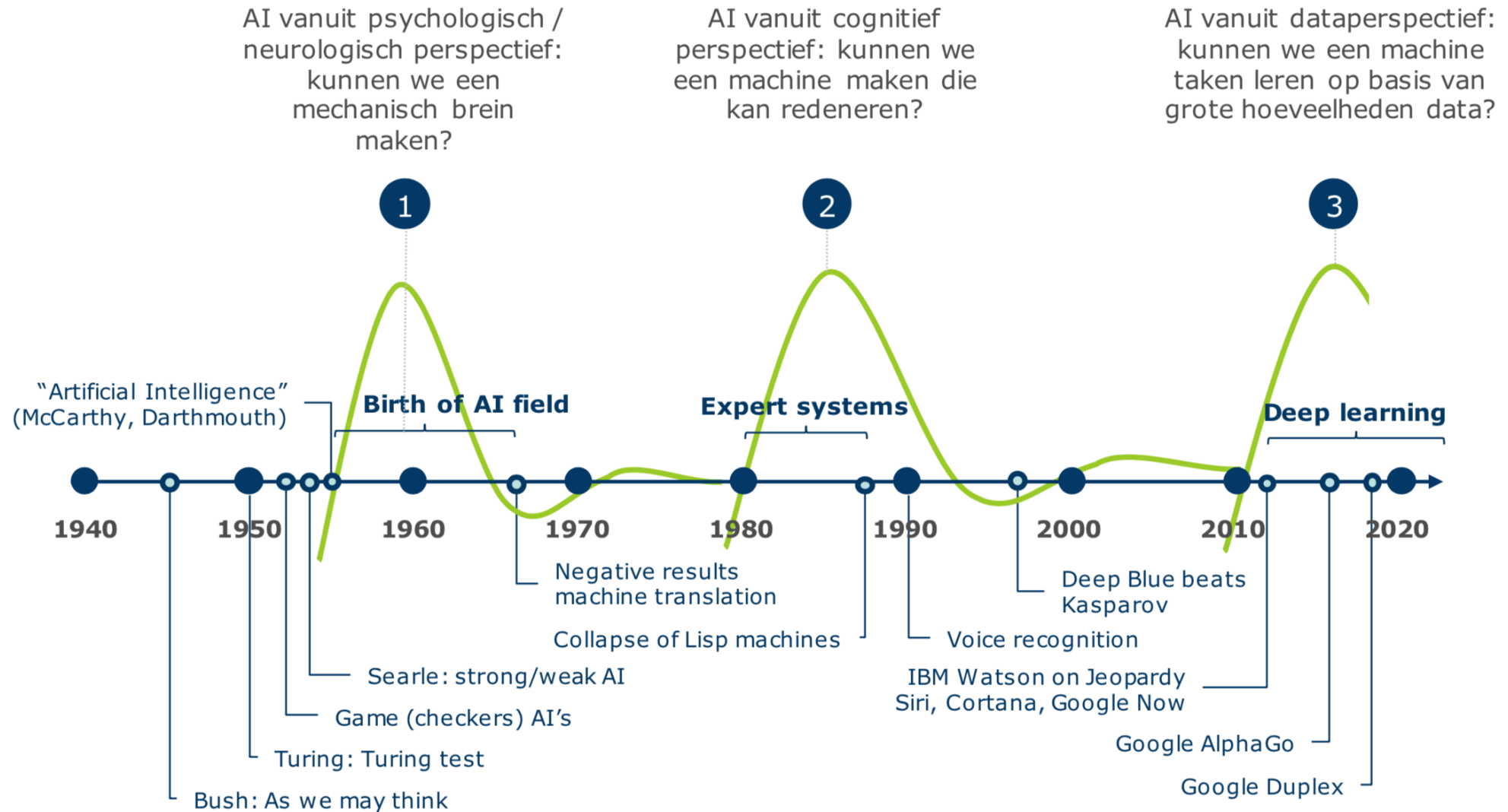


Sentiment Analysis



Document Work Automation

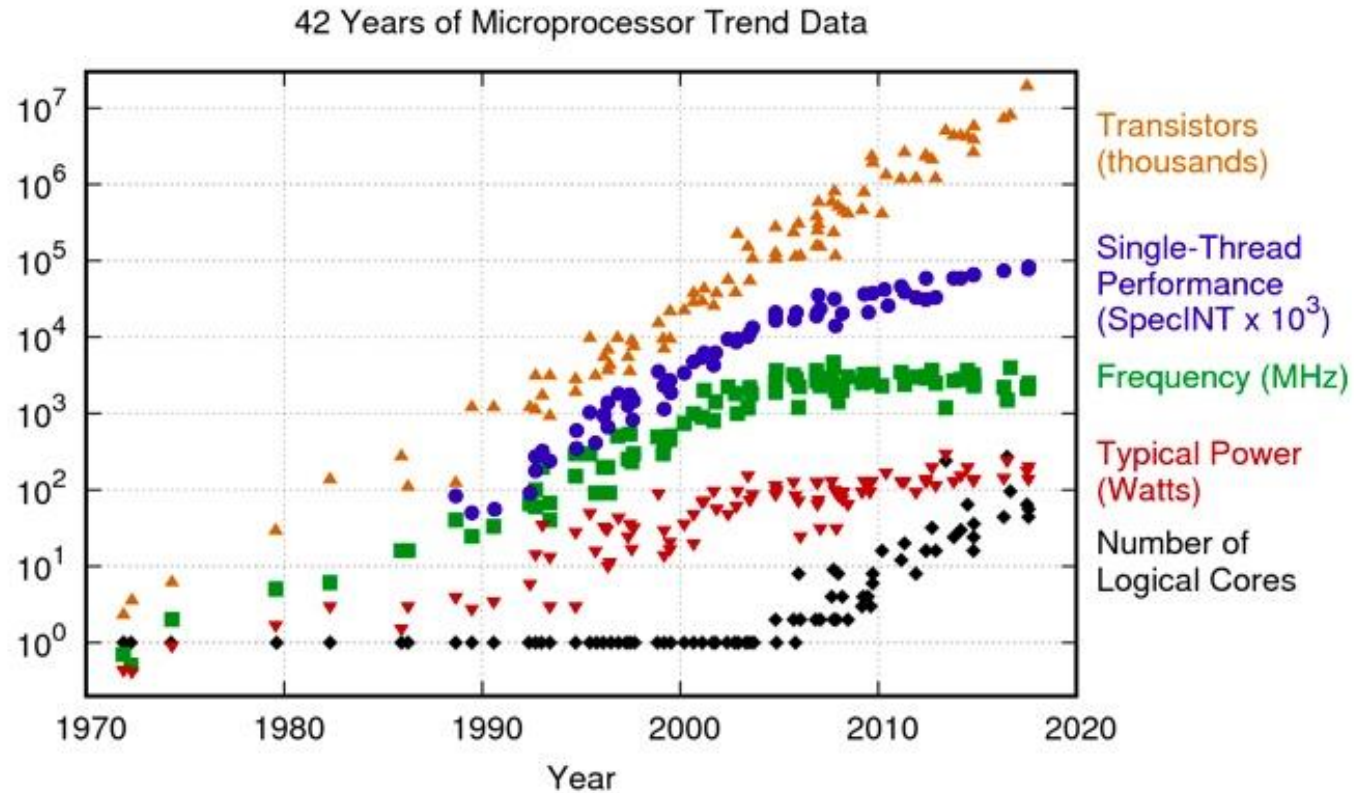
Geschiedenis van AI/Data Science



Waarom zoveel interesse in AI nu?

■ Snellere hardware

■ GPU / TPU



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

<https://www.karlrupp.net/2018/02/42-years-of-microprocessor-trend-data/>

Waarom zoveel interesse in AI nu?

- Meer data beschikbaar
- Internet of Things
- Social Media
- Prijs van data-opslag is laag

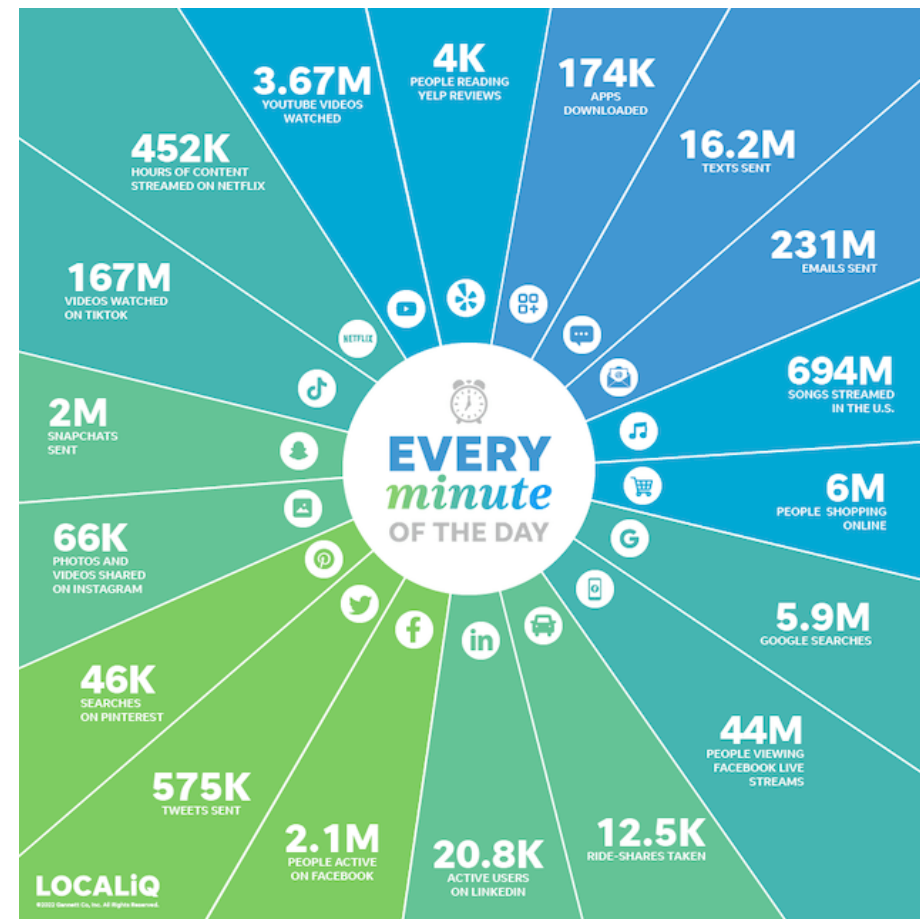
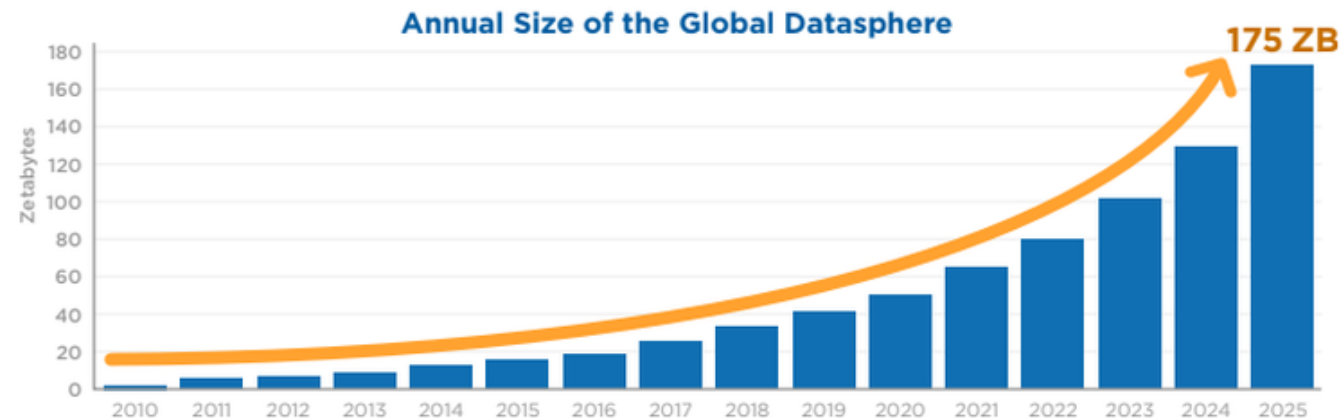


Figure 1 - Annual Size of the Global Datasphere



Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018



Waarom zoveel interesse in AI nu?

- ▣ Beter en toegankelijker algoritmes
- ▣ Open source frameworks
 - ▬ Scikit-Learn / Pandas / Tensorflow / ...

ENGAGEMENT PROCESS

Step 1: Build the Data Model



Step 2: Define The Report



Step 3: Generate SQL commands



Step 4: Create Report



The data warehouse is a "schema-on-load" approach because the data schema must be defined and built prior to loading data into the data warehouse. Without an underlying data model, the BI tools will not work.

Business Intelligence Questions

What happened?

Descriptive Analysis

Standard Reporting

Business Intelligence Analyst



Data Science Questions

Why? What will happen?
What should I do?

Data Scientist



Predictive Analysis

Prescriptive Analysis

VS

DIFFERENCE BETWEEN BUSINESS INTELLIGENCE AND DATA SCIENCE

CHARACTERISTICS

Focus	Reports, KPIs, trends	Patterns, correlations, models
Process	Static, comparative	Exploratory, experimentation, visual
Data Sources	Pre-planned, added slowly	On the fly, as-needed
Transform	Up front, carefully planned	In-database, on-demand, enrichment
Data quality	Single version of truth	"Good enough," probabilities
Data model	Schema on load	Schema on query
Analytics	Retrospective, Descriptive	Predictive, Prescriptive, Preventative

ENGAGEMENT PROCESS

Step 1: Define Hypothesis to Test



Step 2: Gather Data



Step 3: Build Data Model



Step 4: Explore the Data



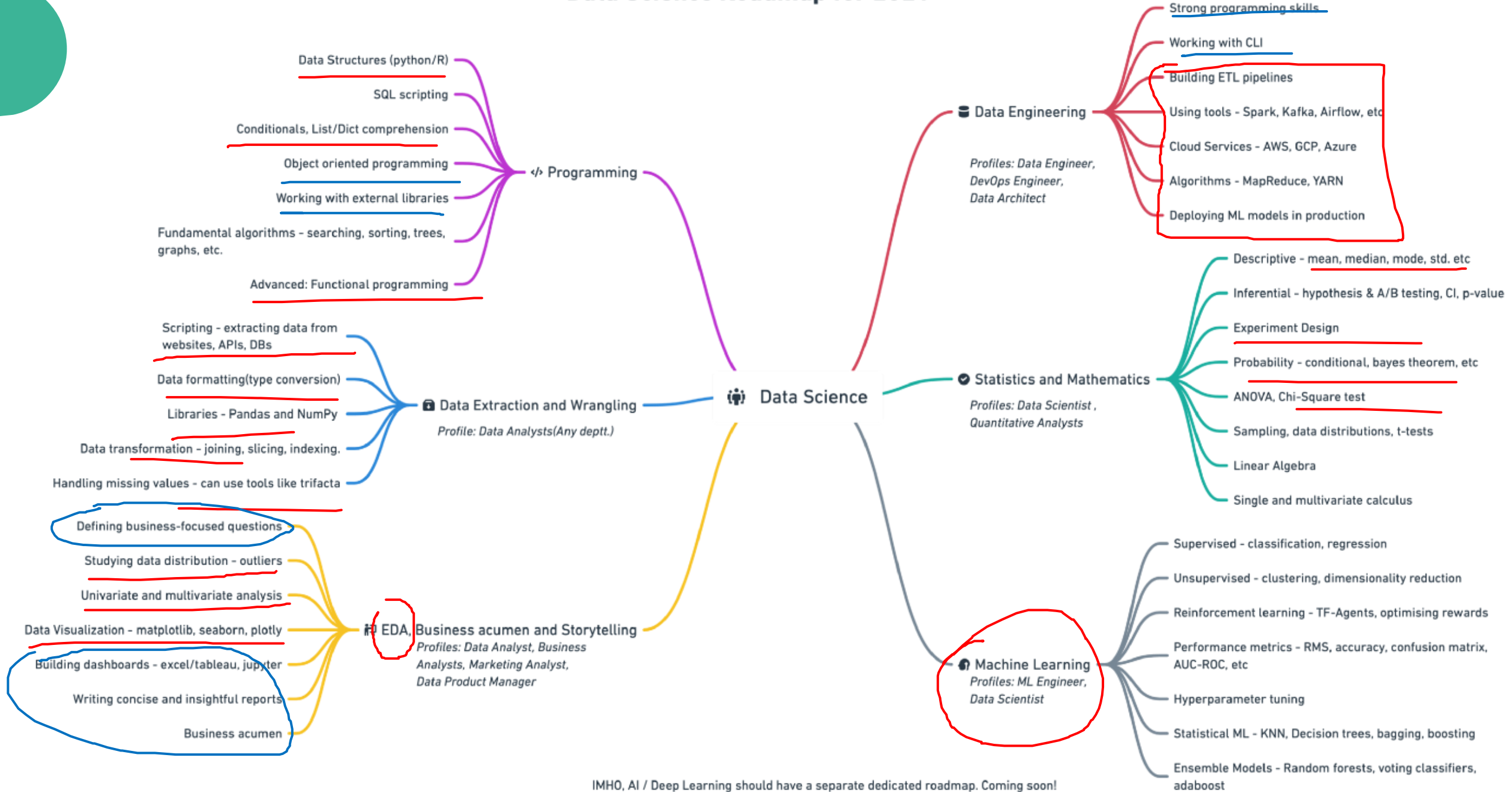
Step 5: Build and Refine Analytic Models



Step 6: Ascertain Goodness of Fit



Data Science Roadmap for 2021



IMHO, AI / Deep Learning should have a separate dedicated roadmap. Coming soon!



Python introduction

▣ Ga naar:

- ▬ <https://www.kaggle.com/learn/python>
- ▬ Volg de volledige tutorial voor een python introductie
- ▬ Deze tutorials zijn deel van de leerstof van dit vak en helpen om vlot te kunnen beginnen
 - ▣ Volg ze dus volledig tegen volgende les



Python oefening

- ▣ Opgave: <https://classroom.github.com/a/SNaNyKWl>
- ▣ Maak de oefening individueel tegen volgende week
- ▣ Dit is een opwarmer en wordt niet geëvalueerd
 - ▣ Controle of python goed geïnstalleerd is, hoe oefeningen te beginnen en in te dienen