



Data Science ***introduction***

Jens Baetens



Wat is Data Science?



Welke voetbalploeg gaat winnen?

- Stap 1: Welke data is er nodig?
 - + Welke periode bekijken we?
 - + Welke bronnen gebruiken we?

Verwerken 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

Model opbouwen voor de data

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

Communiqueer en visualiseer

- 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



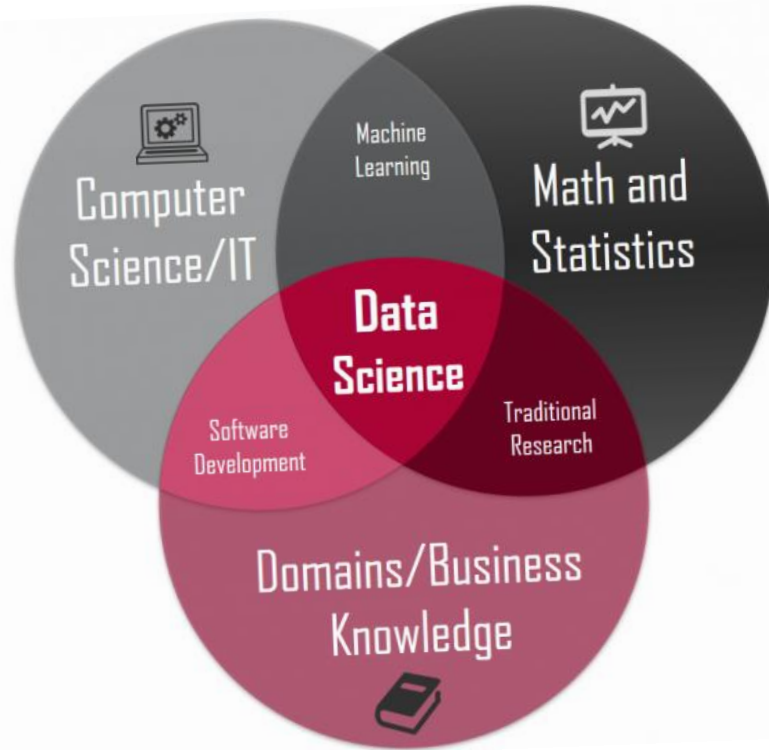
Wat is Data Science?

Definitie

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



- Combinatie van IT, statistiek en domein kennis



Waarom is data science belangrijk?

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



**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 Story

Tech

Meta's chatbot says the company 'exploits people'

By James Clayton
North America technology reporter

AI vermindert fors de onderhoudskost 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.

<https://www.digitaletoeekomst.be/nl/artificiele-intelligentie/verhalen/ai-vermindert-fors-onderhoudskost-en-inspanningen-windmolens-noordzee>

Rode Duivels geblesseerd? Eqipa 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 Eqipa 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

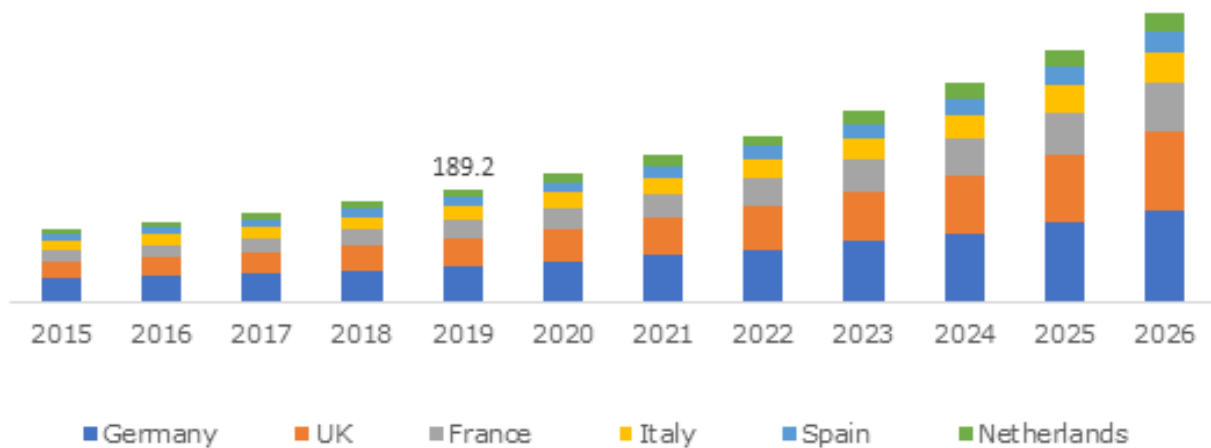
Use case

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.

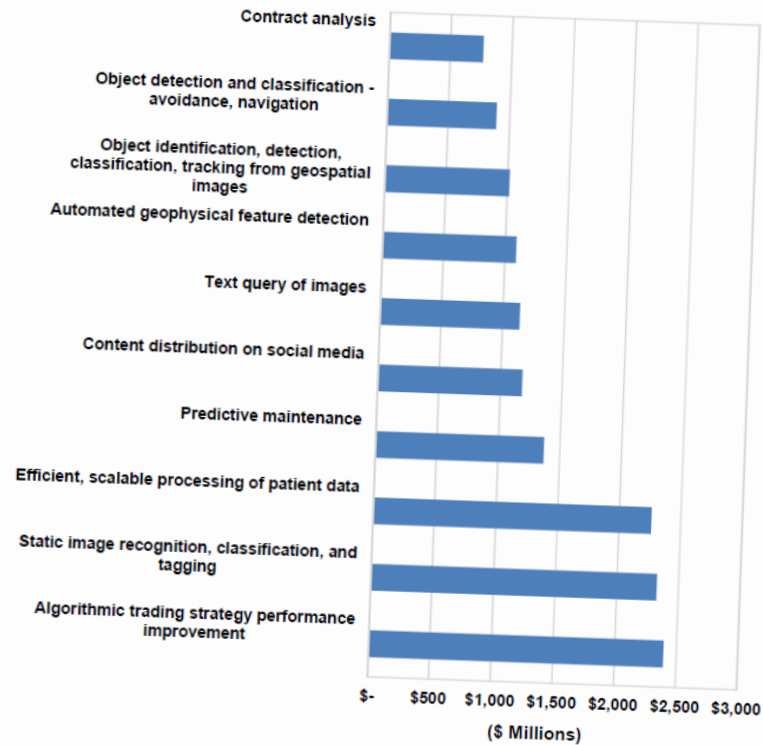
<https://www.digitaletekomst.be/nl/artificiele-intelligentie/verhalen/OKO-app-voor-slechtzienden-verkeerslicht-groen-of-rood>

Europe Location Analytics Market Share, By Region, 2015 - 2026 (USD Million)



Welke

markten?



(Source: Tractica)



More examples of use cases

Financiële sector

- Fraude detecteren
- Risico's inschatten
- Beleggen op de beurs



[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 deduct 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 commit making the most effective use of their resources to ensure compliance with the tax code.

<https://www.districtofcolumbiataxattorney.com/blog/2019/12/the-irss-use-of-data-analytics-to-combat-tax-evasion/>

More examples of use cases

Gezondheidszorg

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



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

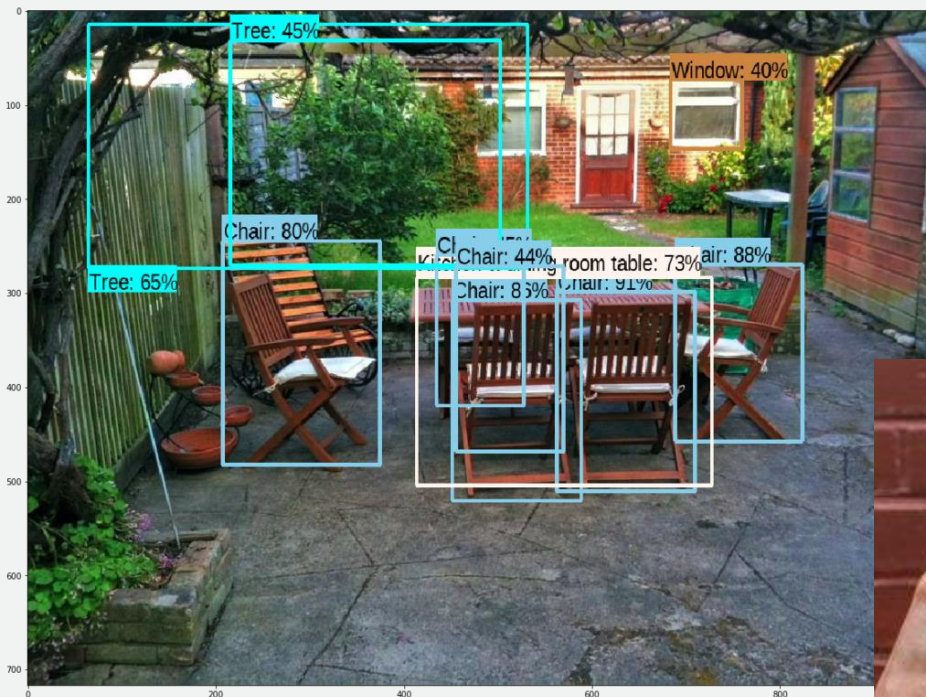
More examples of use cases

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



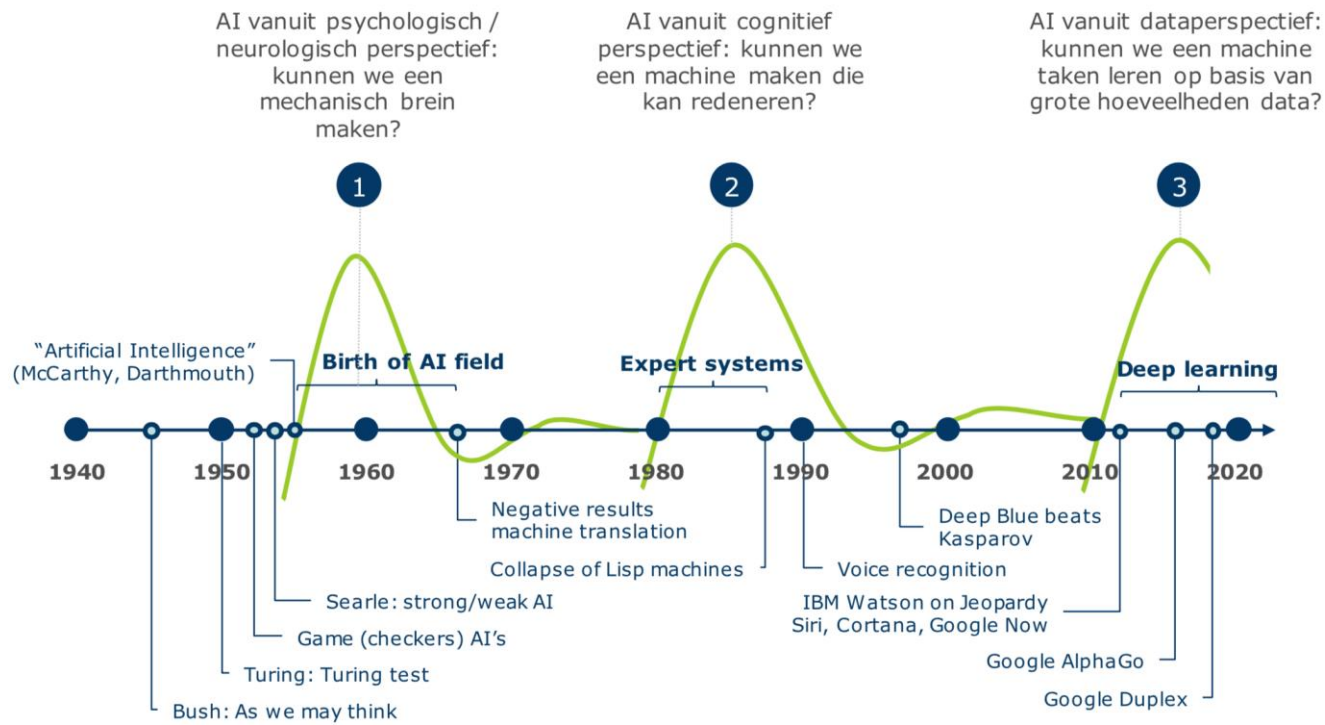
Sentiment Analysis



Document Work

Retail

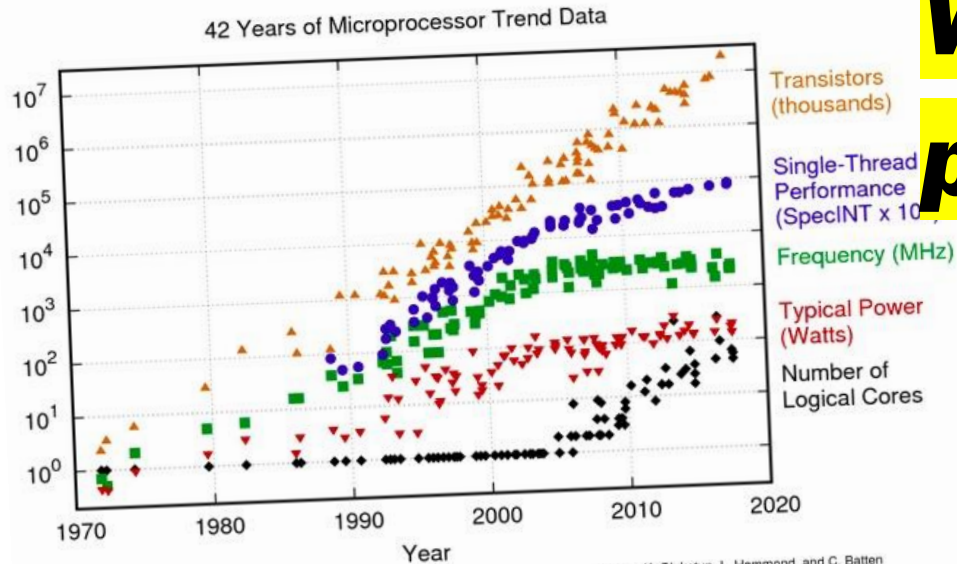




History



Waarom nu zo popular?



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/>

- Snellere hardware
+ GPU / TPU



Waarom nu?

Meer data beschikbaar

Internet of Things

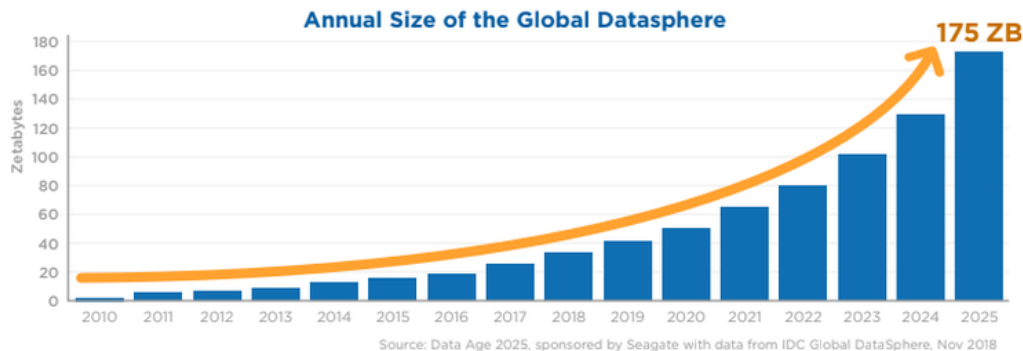
Social Media

Prijs van dataopslag is laag

2021 This Is What Happens In An Internet Minute



Figure 1 - Annual Size of the Global Datasphere



Waarom nu?

- Betere en toegankelijke 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, BI tools will not work.

Business Intelligence Questions

What happened?

Descriptive Analysis

Standard Reporting

Business Intelligence Analyst



VS

Data Science Questions

Why? What will happen?
What should I do?

Predictive Analysis

Prescriptive Analysis

Data Scientist



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



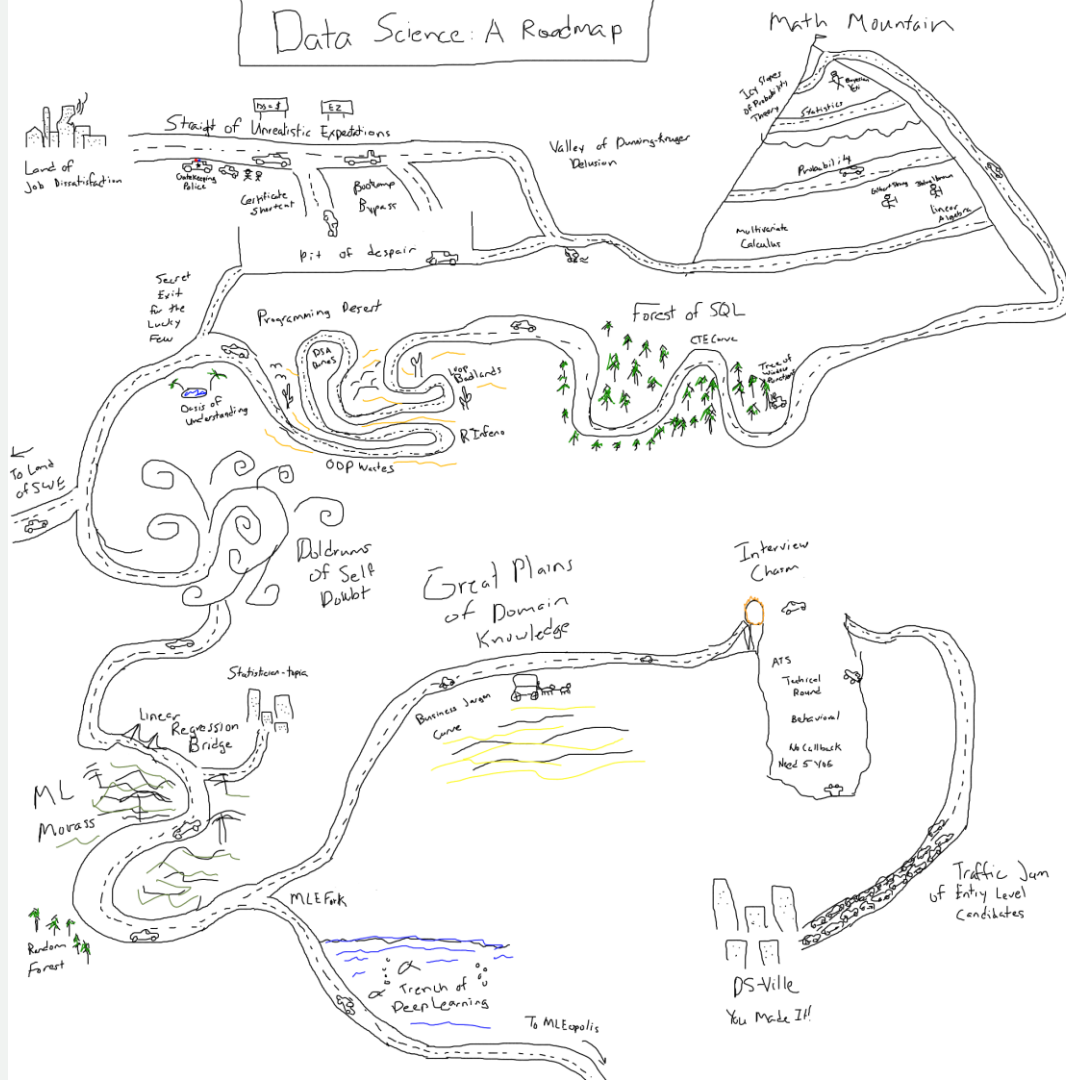
Step 5: Build and Refine Analytic Models



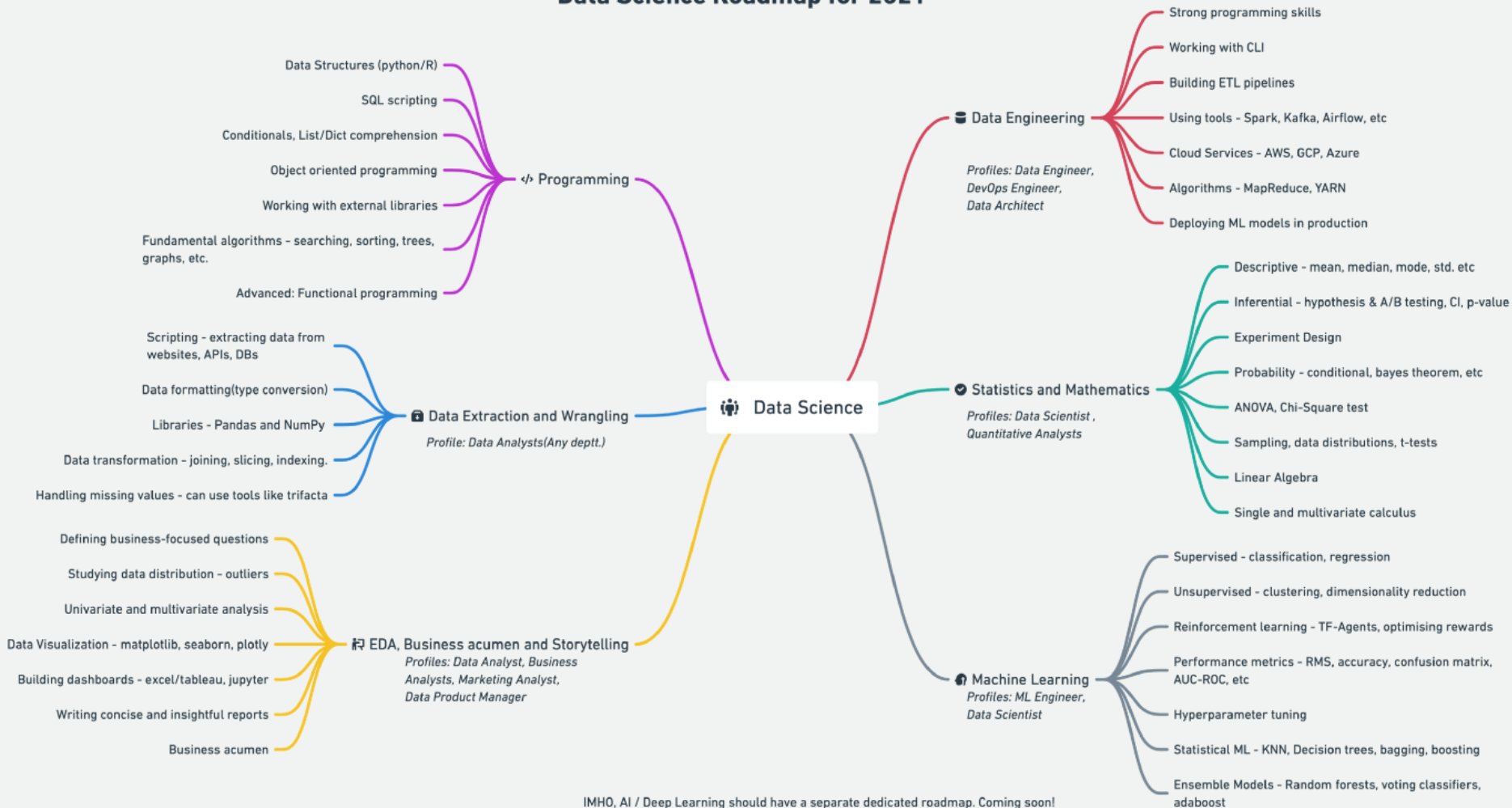
Step 6: Ascertain Goodness of Fit



Data Science: A Roadmap



Data Science Roadmap for 2021



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