

Data Modelling

JENS BAETENS

Definitie

?
set van parameters die een output
bepalen van een ~~if~~ input

Opbouwen van een model om waarde en inzichten uit de data te halen en een antwoord te bieden op de gewenste vraag.

Stappen:

① - Kies een of meerdere ML technieken om te gebruiken

② - Trainen van het model op basis van de data → zoeken naar juiste frequentie
radiozender

③ - Evalueren van het model op nieuwe data

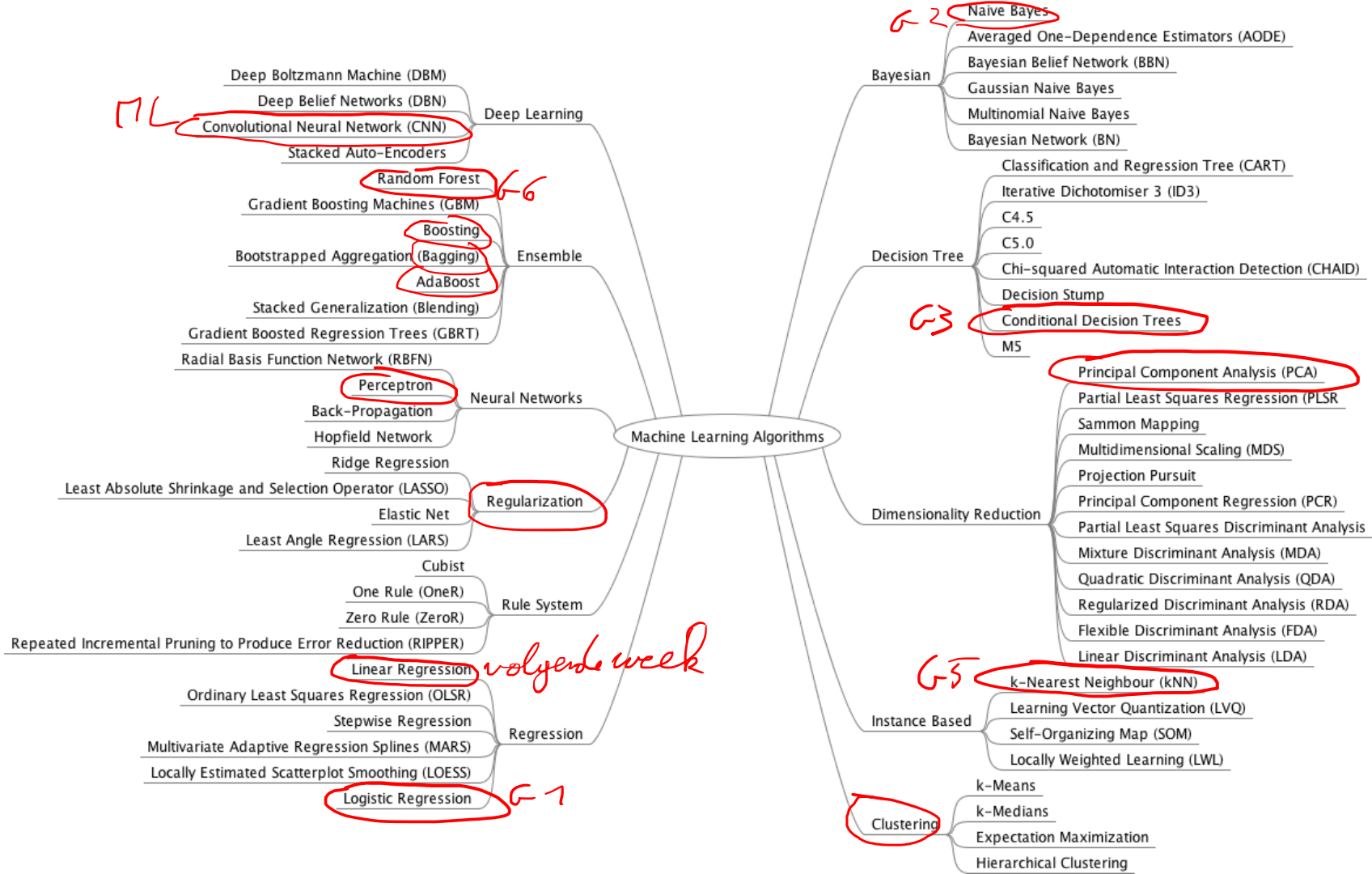
↳ data niet gebruikt voor training

parameters → kan je niet kiezen

→ optimale waarden gezocht door het model

hyperparameters → kan je kiezen

→ bepalen hoe het model werkt



Soorten technieken

Gebruik van data en labels

Supervised ML

"ground truth"

Gebruik van data

planten clusteren

Unsupervised ML

Gebruik van regels

geen data?

Reinforcement ML

zoals een kind of hond opvoeden

Supervised vs unsupervised

Supervised Learning

labelled data = feature + labels
data = features ground truth
Labels
-> probe te
voorspellen

features

Inputs

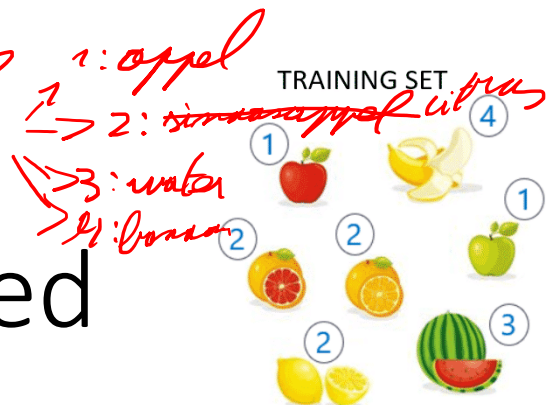
Labels

# Avg. Area I...	# Avg. Area ...	# Avg. Area ...	# Avg. Area ...	# Area Popul...	# Price	▲ Address
79545.45857431678	5.682861321615587	7.009188142792237	4.09	23086.800502686456	1059033.5578701235	208 Michael Ferry Apt. 674 Laurabury, NE 37010-5101
79248.64245482568	6.0028998082752425	6.730821019094919	3.09	40173.07217364482	1505890.91484695	188 Johnson Views Suite 079 Lake Kathleen, CA 48958
61287.067178656784	5.865889840310001	8.512727430375099	5.13	36882.15939970458	1058987.9878760849	9127 Elizabeth Stravenue Danielstown, WI 06482-3489
63345.24004622798	7.1882360945186425	5.586728664827653	3.26	34310.24283090706	1260616.8066294468	USS Barnett FPO AP 44820
59982.197225708034	5.040554523106283	7.839387785120487	4.23	26354.109472103148	630943.4893385402	USNS Raymond FPO AE 09386

rij
observatie
(training)
example

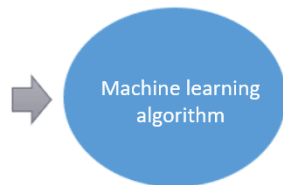
foto's fruit = data

labels = cijfers
= klassen



cijfertjes = pixels

Feature vectors



Supervised Learning

TEST SET



zonder labels



RESULTS

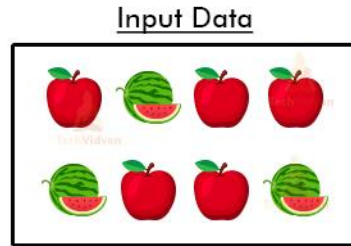


voorspeld

→ of fruit rot is of niet / tot bepaalde categorie

Unsupervised learning

↳ clustering / groepering
probleem



Model

cluster 1: apples

Output

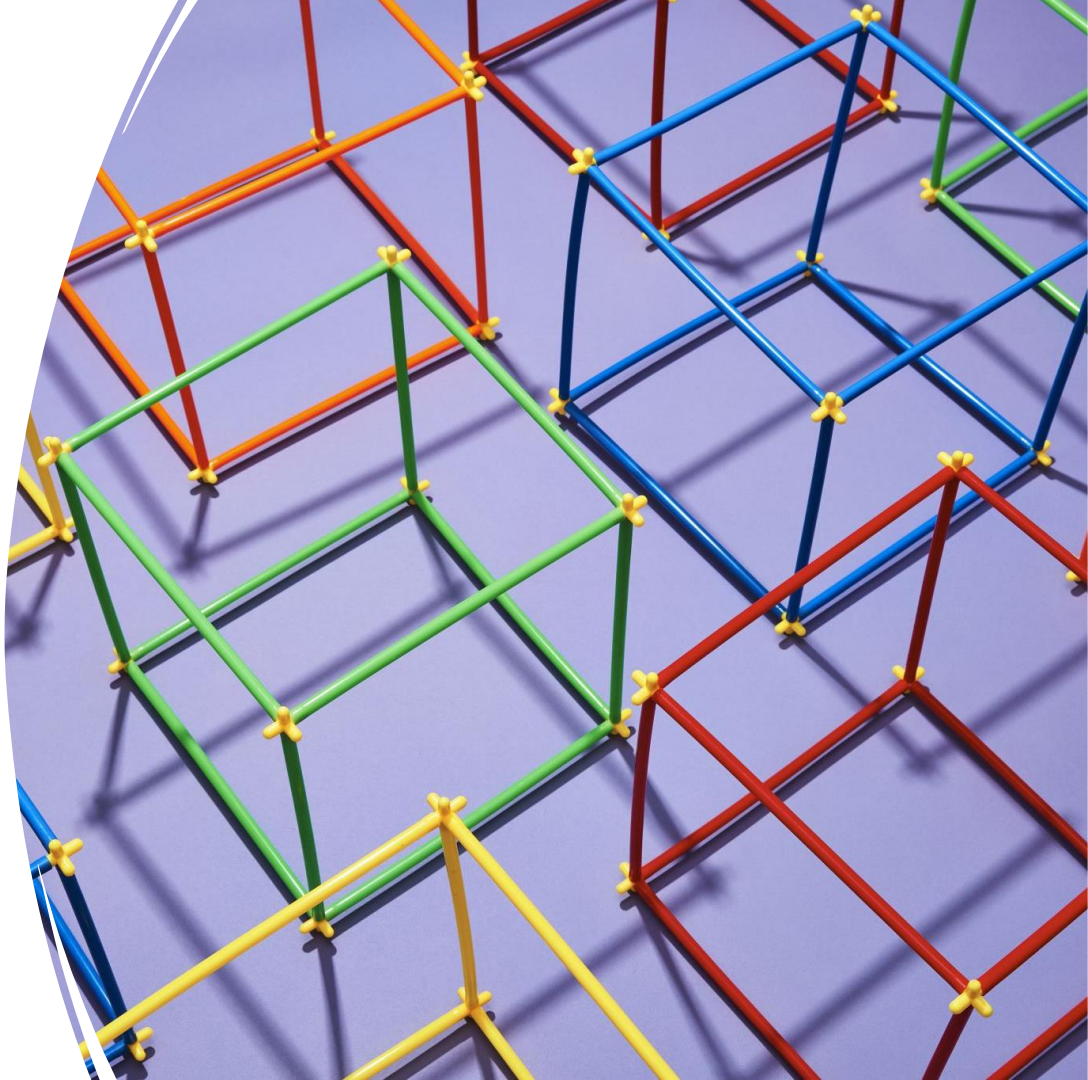


cluster 2
watermeloenen

Reinforcement learning

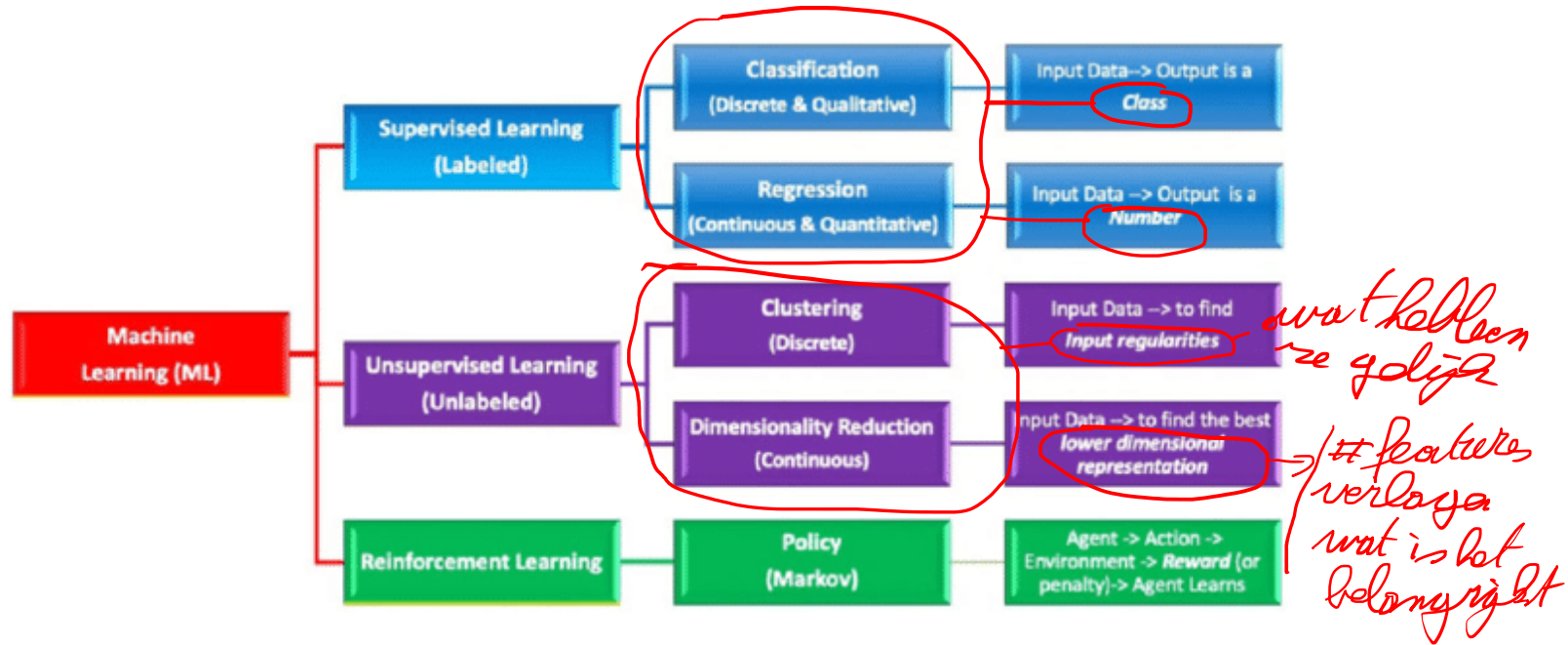
Omvat technieken om modellen te bouwen die leren wat de beste actie om een zo hoog mogelijke score te bekomen.

Geen data gebruikt maar wel de regels van de omgeving waarin het moet werken



Reinforcement learning

TECHNIEKEN



Soorten technieken - Doel

Hoeveel gaat het zijn?

→ Regression techniques

Is het type A of B?

→ Classification techniques

Zijn deze gelijk?

→ Clustering techniques

Is het vreemd?

→ Anomaly detection techniques

Wat moet ik doen?

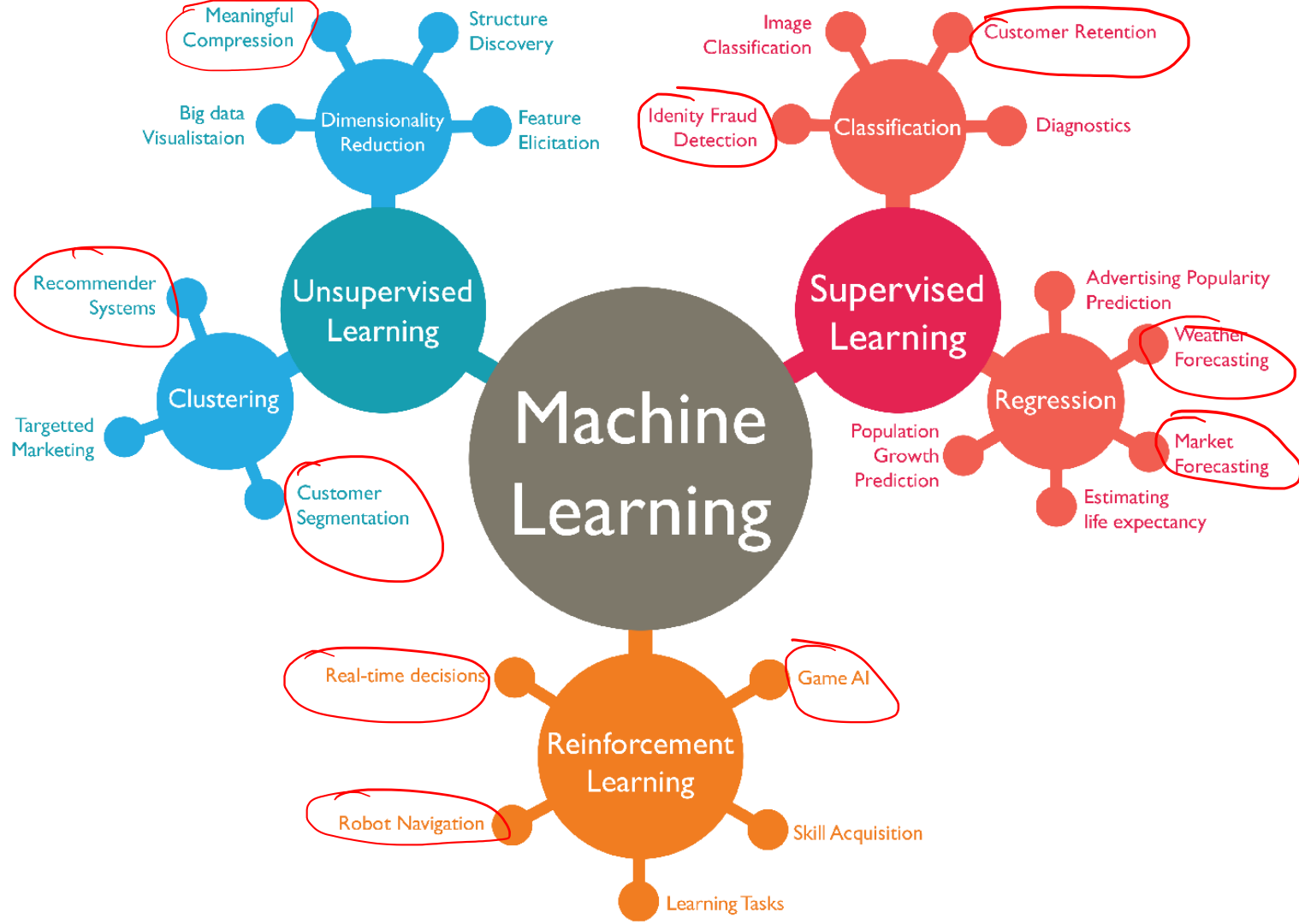
→ Recommendation techniques

Welke zijn het belangrijkste?

→ Dimensionality reduction

supervised

unsupervised



How to participate?



Click on the projected screen to start the question





Een supervised learning model maakt gebruik van

①

Enkel van bestaande data

0%

0



②

De bestaande data worden gebruikt om

Click on the projected screen to start the question

0%

0



③

Enkel van het resultaat van een actie

0%

0





Welke vraag wordt opgelost door welke soort technieken

Hoeveel?

1

A

Anomaly detection

Welk type?

2

B

Classification

Is dit gelijkaardig aan iets anders?

3

C

Clustering

Is dit resultaat vreemd?

4

D

Recommendation

Wat zijn de opties?

5

E

Regression

