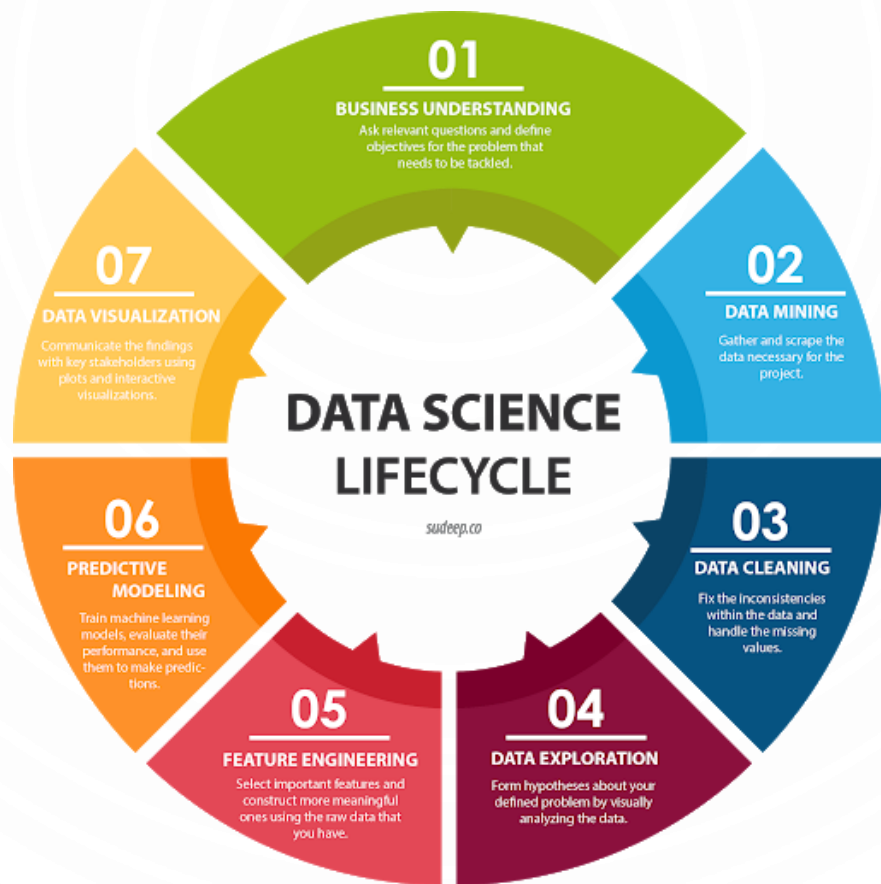


The background features a series of concentric, light gray circles centered in the middle. Overlaid on these are stylized, light blue circuit-like lines with small circles at the ends, resembling a network or data flow. These lines are positioned in the corners and along the edges of the frame.

HERHALING





Wat is het verschil tussen Data Science en Machine learning?

Welke stappen moeten er vaak uitgevoerd worden bij data cleaning?

Wat betekent Exploratory Data Analysis?





Supervised ML technieken

Unsupervised ML technieken

Regressie

Classificatie

Trainen

Voorspellen

Metrieken

Regularisatie





Wat is trainen van een model?

Wat is het nut van een splitsing te maken tussen trainingsdata en testdata?

Wat is overfitting?

Wat is underfitting?

Wat is hyperparameter tuning? Hoe kan je dit automatiseren?

Waarvoor worden pipelines gebruikt?

Hoe evalueer je een model?



CLASSIFICATION – NAIVE BAYES

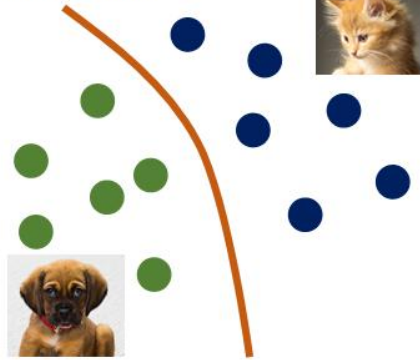
JENS BAETENS

The image features a light gray background with a subtle pattern of concentric circles. In the four corners, there are decorative elements resembling circuit board traces or neural network connections, consisting of thin blue lines and small circles.

WAAR DENK JE AAN BIJ BAYES?

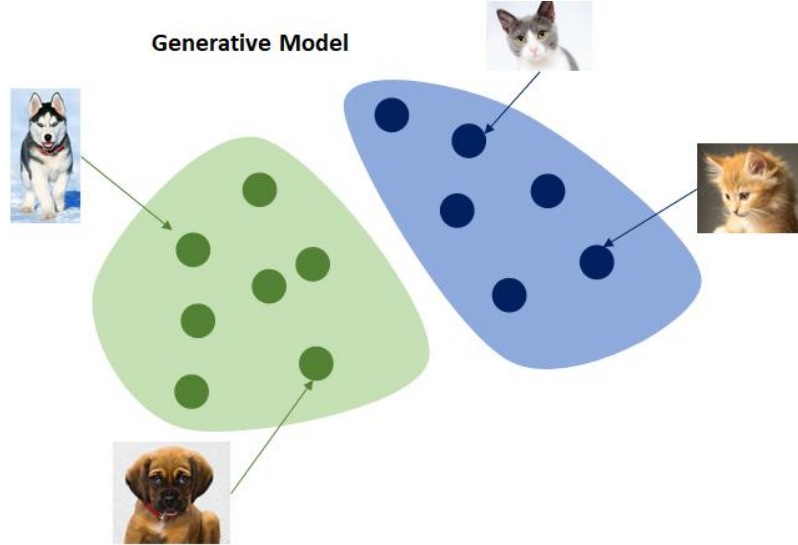
GENERATIEVE VS DISCRIMANTIEVE CLASSIFIER

Discriminant Model



$$P(y | x)$$

Generative Model



$$P(x | y) \text{ en } p(y)$$

VOORBEELD

Stel: Er is een test om kanker op te sporen die

- Sensitiviteit van 90%: Test is positief als je kanker hebt
- Specificiteit van 90%: Test is negatief als je geen kanker hebt

De kans dat iemand willekeurig kanker heeft is echter 1%

Wat is de kans dat je kanker hebt als de test positief is?

VOORBEELD

$$P(\text{Kanker}|\text{Positief}) = \frac{P(\text{Positief}|\text{Kanker})P(\text{Kanker})}{P(\text{Positief})}$$

$$P(\text{Kanker}|\text{Positief}) = \frac{0.9*0.01}{0.01*0.9+0.1*0.99} = 0.08333$$

Prior: $P(\text{Kanker})$

Likelihood: $P(\text{Positief} | \text{Kanker})$

Marginal: $P(\text{Positief})$

Posterior: $P(\text{Kanker} | \text{Positief})$

GAUSSIAN NAIVE BAYES CLASSIFIER

"Gaussian" because this is a normal distribution

This is our prior belief

$$P(\text{class} | \text{data}) = \frac{P(\text{data} | \text{class}) \times P(\text{class})}{P(\text{data})}$$

We don't calculate this in naive bayes classifiers

ChrisAlbon