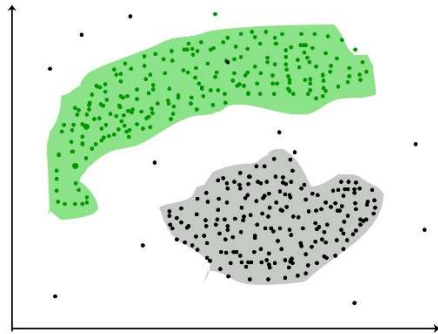


## Clustering

Clustering is het verdelen van data punten in zo klein mogelijke groepen, waarbij de punten in de data groepen op elkaar lijken en er tegelijk een duidelijk verschil zit tussen groepen. Als er een groep gevormd wordt moet er afstand zijn tot een andere groep, anders zullen de twee groepen samengevoegd worden tot één groep. Als er losse datapunten zijn die niet in de buurt van andere datapunten staan, worden ze ook genegeerd aangezien ze niet gegroepeerd kunnen worden.



De groepen die uit de clustering komen betekenen op zich niks, aangezien er uiteindelijk door de gebruiker ervan gekozen moet worden wat die groepen daadwerkelijk betekenen. Hierom is er ook geen duidelijk criterium waarop gebaseerd wordt of een clustering goed werkt, aangezien er geen duidelijke goed of fout is bij het groeperen van data.

Bron:

(Feb. 2020), <https://www.geeksforgeeks.org/clustering-in-machine-learning/>