

# NUTTIGE BEGRIPPEN

In dit themanummer over Big Data zult u geregeld onderstaande termen tegenkomen. Ter introductie alvast een korte uitleg.

## BIG DATA

Iedereen heeft het over Big Data. Toch bestaat er vooralsnog geen definitie waarover consensus bestaat. Big Data bestaan in ieder geval uit grote hoeveelheden digitale data, zoals cijfers, tekst, foto's, en video's. Volgens het internationale onderzoeksbureau Gartner bestaan Big Data uit grote hoeveelheden data, die met een snel tempo verzameld worden én die zo divers en gevarieerd zijn (de data bestaan uit heel veel variabelen) dat ze niet op traditionele wijze verwerkt en geanalyseerd kunnen worden.

## DATAMINING

Datamining is het herkennen van patronen in grote dataverzamelingen. Net als in een mijn wordt er tussen de 'kale' data gezocht naar waardevolle verbanden. Met behulp van datamining kunnen bedrijven bijvoorbeeld op basis van het koopgedrag van hun klanten de marketingstrategie aanpassen.

## ALGORITMES

Een algoritme kun je vergelijken met allemaal stappen die een gebruiker, bijvoorbeeld een computersysteem of jijzelf, in een bepaalde volgorde moet uitvoeren om iets te bereiken. Een computerprogramma bevat algoritmes, maar een kookboek ook. Recepten zijn algoritmes. Bij een kookboek is de kok de gebruiker van het algoritme en gebruikt hij of zij de stappen uit het recept om een maaltijd te maken.

## ARTIFICIAL INTELLIGENCE (AI)

Artificial Intelligence, ofwel kunstmatige intelligentie, heeft als doel om machines

te maken die de menselijke intelligentie minstens evenaren. De wetenschap van de AI houdt zich dan ook bezig met het creëren van een artefact (niet-natuurlijk verschijnsel) dat een bepaalde vorm van intelligentie vertoont. Wat daarin precies de definitie van 'intelligentie' is, blijft vooralsnog onduidelijk.

Vaak denken we bij AI aan de toekomst, bijvoorbeeld aan robots die op mensen lijken en volledig zelf denken. Maar AI is al jaren in ons midden. Denk aan de spraakassistenten Siri, Alexa en Google Home. Deze assistenten communiceren namelijk zelf met jou en beschikken dus over een bepaalde vorm van intelligentie.

## MACHINE LEARNING

Machine learning is een discipline binnen de kunstmatige intelligentie die zich bezighoudt met het ontwikkelen van algoritmes en andere technieken waarmee computers kunnen leren van data. Er zijn grofweg twee manieren waarop een computer iets kan leren. Ten eerste kan een computer leren gegevens op een correcte manier te classificeren. Hij krijgt daartoe eerst een aantal voorbeelden te zien. Bijvoorbeeld duizend foto's van honden en katten, ofwel de input, met de daarbij behorende output: het label 'hond' of 'kat'. Vervolgens krijgt de computer enkel nog input en moet hij op basis van wat hij geleerd heeft het juist label generen. De tweede manier waarop een computer iets kan leren is minder gecontroleerde. De computer krijgt zelf grote hoeveelheden data te zien, met verschillende variabelen, en gaat zelf op zoek naar patronen. Bij-

voorbeeld of het tijdstip waarop treinreizigers inchecken, samenhangt met hun keuze om koffie of thee op het perron te kopen.

## KUNSTMATIGE NEURALE NETWERKEN

Deze netwerken zijn een vorm van machine learning en zijn geïnspireerd op de werking van neuronen. Op hoe hersencellen met elkaar verbonden zijn en informatie met elkaar uitwisselen. Wij kunnen leren omdat ons brein op basis van goede en slechte ervaringen de verbindingen tussen hersencellen kan versterken of verzwakken. Met kunstmatige neurale netwerken kun je aan de hand van datzelfde principe van alles maken, bijvoorbeeld een computerprogramma dat ieder mens zal verslaan met schaken.

## DEEP LEARNING

Voor deep learning gebruik je kunstmatige neurale netwerken die uit heel veel lagen 'neuronen' bestaan. Het is daardoor een heel 'diep' netwerk. Hoe meer lagen, hoe meer verschillende kenmerken het netwerk kan destilleren uit de input die je hem geeft. Stel dat hij een foto van een kat te zien krijgt, dan zullen de verschillende lagen herkennen dat er een vacht te zien is, welke kleur die vacht heeft, wat voor neus er te zien is, welke kleur ogen... Op basis van die gegevens kan het netwerk bepalen dat hij een kat ziet, zonder dat je hem vooraf een voorbeeld hebt moeten geven.

Veel leesplezier!