

Machine Learning

Jeroen Arts | Fontys Hogescholen Techniek en Logistiek, Mechatronica | 31-10-2022

# Voorwoord

Artificial Intelligence (AI) speelt steeds een grotere rol in de wereld. Je ziet steeds vaker dat bedrijven gebruik maken van een AI om informatie te verwerken. De meeste mensen zullen bij AI denken aan Marvels Ultron of Terminators Skynet. Dit geeft geen goed beeld van de werkelijkheid.

Dit literatuuronderzoek wordt geschreven naar aanleiding van het vak PRP3. Dit vindt plaats in het derde semester van onze opleiding. Het doel van dit literatuuronderzoek is om studenten van de opleiding Mechatronica of een vergelijkbare opleiding, een informatieve blik te geven op wat Machine Learning, het leerproces van de AI, is en een basisbegrip van hoe het werkt.

Venlo, november 2022

Jeroen Arts

# Samenvatting

Machine Learning en AI zijn bekende begrippen waarvan veel mensen niet weten hoe het daadwerkelijk werkt. AI komt steeds vaker voor en wordt steeds meer gebruikt. Om een beter beeld te geven wordt dit literatuuronderzoek geschreven.

De term AI vindt zijn oorsprong in 1940, bedacht door Alan Turning, Intelligentie gemaakt door een persoon. Dit groeide snel door naar Machine Learning een geheel nieuw tak van de wetenschappen. AI zijn in principe slimme algoritme.

Machine Learning kan worden onderverdeeld in drie groepen. Reinforcement Learning, Supervised Learning en Unsupervised Learning.

Reinforcement Learning gaat in op het leren door te doen. Het is de methode die het meest lijkt op wat mensen zelf doen. Het bekendste algoritme voor Reinforcement Learning is Q-learning gemaakt door Watkins in 1989. Bekende voorbeelden van Reinforcement Learning zijn de DeepBlue en AlphaGo supercomputers die grootmeesters in hun spellen hebben verslagen.

Supervised Learning is leren zoals een kind zal doen in een klaslokaal. De leraar legt iets uit en maakt de verbanden die de leerlingen moeten leren. Dit is dus het leren, door gebruikt te maken van menselijke ervaring. Supervised learning kan verder verdeeld worden in Classification en Regression. Beide methodes gebruiken bekende datapunten om achteraf nieuwe data te voorspellen of om verbanden te leggen in de data.

Unsupervised learning is wanneer een leraar informatie geeft maar het verband tussen de informatie niet uitlegt en dus de leerling zelf de verbanden laat vinden. Unsupervised Learning kan verder worden verdeeld in Clustering en Dimensional Reduction. De eerste methode maakt verbanden in onbekende data en de tweede methode zet de complexe data om in minder complexe data.

Alle methodes hebben hun voor- en nadelen. Zo is Reinforcement Learning goed wanneer je geen dataset hebt maar een omgeving om in te werken. Het is slecht wanneer de beloningen niet goed gedefinieerd zijn en de leerling alleen zoveel mogelijk beloningen wil krijgen. Supervised en Unsupervised learning hebben ook hun eigen voor- en nadelen en de één is meestal goed in wat de ander slecht in is.

Inhoud

[Voorwoord 2](#_Toc124161168)

[Samenvatting 3](#_Toc124161169)

[1. Inleiding 6](#_Toc124161170)

[2. Wat is AI? 7](#_Toc124161171)

[3. Wat is Machine Learning? 8](#_Toc124161172)

[4. Reinforcement Learning (RL) 9](#_Toc124161173)

[4.1 Wat is Reinforcement Learning? 9](#_Toc124161174)

[4.2 Hoe werkt Reinforcement Learning? 9](#_Toc124161175)

[4.3 Toepassingen van Reinforcement Learning. 10](#_Toc124161176)

[5. Supervised Learning 11](#_Toc124161177)

[5.1 Wat is Supervised Learning? 11](#_Toc124161178)

[5.2 Classification 11](#_Toc124161179)

[5.2.1 Wat is Classification? 11](#_Toc124161180)

[5.2.2 Hoe werkt Classification? 11](#_Toc124161181)

[5.2.3 De toepassingen van Classification 12](#_Toc124161182)

[5.3 Regression 13](#_Toc124161183)

[5.3.1 Wat is Regression? 13](#_Toc124161184)

[5.3.2 Hoe werkt Regression? 13](#_Toc124161185)

[5.3.3 Toepassingen van Regression 14](#_Toc124161186)

[6. Unsupervised Learning (UL) 15](#_Toc124161187)

[6.1 Wat is Unsupervised Learning? 15](#_Toc124161188)

[6.2 Clustering 15](#_Toc124161189)

[6.2.1 Wat is Clustering? 15](#_Toc124161190)

[6.2.2 Hoe werkt Clustering? 16](#_Toc124161191)

[6.2.3 Toepassingen van Clustering 17](#_Toc124161192)

[6.3 Dimensionality Reduction (DR) 17](#_Toc124161193)

[6.3.1 Wat is Dimensionality Reduction? 17](#_Toc124161194)

[6.3.2 Hoe werkt Dimensionality Reduction? 18](#_Toc124161195)

[6.3.3 Toepassingen van Dimensionality Reduction 18](#_Toc124161196)

[7 Voor- en nadelen 19](#_Toc124161197)

[7.1 Reinforcement Learning 19](#_Toc124161198)

[7.2 Supervised Learning 19](#_Toc124161199)

[7.3 Unsupervised Learning 19](#_Toc124161200)

[7.4 Conclusie 19](#_Toc124161201)

[8. Slotwoord 21](#_Toc124161202)

[Literatuurlijst 22](#_Toc124161203)

[Bijlage 1: Reflectie 24](#_Toc124161204)

# Inleiding

Iedereen heeft er wel eens over gehoord: AI, Kunstmatige Intelligentie (Artifical Intelligence). AI komt voor in films, series, boeken en spellen. Ook in de werkelijkheid wordt AI toegepast voor een veelvoud van toepassingen. Maar hoeveel mensen weten hoe het werkt? Is wat we zien op televisie ook echt hoe het er in de werkelijkheid aan toe gaat? Een AI wordt gemaakt en op het internet losgelaten. Het wordt steeds slimmer en slimmer, totdat het zo slim is dat mensen het niet meer bij kunnen houden. Maar hoe wordt een AI nou zo slim, hoe leert de AI alles wat het weet? Wat is nou Machine Learning, het leerproces van een AI?

Om te begrijpen hoe een AI leert, moet het duidelijk zijn wat een AI is. In hoofdstuk 2 wordt uitgelegd wat een AI is. Daarna wordt in hoofdstuk 3 ingegaan op wat Machine Learning is. Vervolgens worden in hoofdstukken 4, 5 en 6 de drie grote leermethodieken uitgelegd.

Het doel van dit verslag is om een informatieve blik te geven op wat Machine Learning is en hoe het werkt. Na het lezen van dit verslag zal een persoon zonder kennis van AI weten wat Machine Learning is en enigszins begrijpen hoe het werkt. Met deze informatie zouden mensen die geïnteresseerd zijn in AI een keuze kunnen maken in hoe ze een AI willen trainen.

# Wat is AI?

AI is ontstaan in Engeland tijdens de 2e Wereld Oorlog in 1940, toen Alan Turning een machine ontwikkelde om de Enigma Code van Duitsland te breken. Tien jaar later bracht hij een artikel uit over “Computer Machinery and Intelligence” (Kaplan & Haenlein, 2019). Dit is tevens ook het ontstaan van de Turning Test die gebruikt wordt om de intelligentie van een AI te meten.

AI staat voor Artificial Intelligence, intelligentie gemaakt door een ander. De term Artificial Intelligence werd voor het eerst officieel gebruikt door een onderzoeksteam van het Dartmouth College in de Verenigde Staten in het jaar 1957 (Dick, 2019).

Maar de vraag is nog steeds, wat is AI? Er zijn verschillende definities voor AI. Een van die definities luidt dat AI van methodes gebruik maakt gebaseerd op het intelligente gedrag van mensen en dieren om complexe problemen op te lossen (Coppin, 2004, p. 4). Een AI is dus een slimme computer of algoritme dat in staat is te leren en zichzelf te verbeteren.

# Wat is Machine Learning?

Machine Learning is een onderdeel van AI dat gericht is op het maken van systemen die van data kunnen leren of data kunnen gebruiken om beter te worden. AI en Machine Learning worden vaak samen gebruikt, maar het zijn twee verschillende dingen. Terwijl Machine Learning altijd onder AI valt, valt AI niet altijd onder Machine Learning. Het doel van Machine Learning is om computers nieuwe dingen te laten leren die er niet hard in geprogrammeerd worden zoals we bijvoorbeeld zien bij een montagelijn.

Zoals eerder vernoemd zijn er drie specifieke leermethodes die onder Machine Learning vallen: Reinforcement, Supervised en Unsupervised.

In hoofdstuk 4 wordt Reinforcement learning behandeld. Bij deze methode wordt gebruik gemaakt van positieve of negatieve beloningen. De AI krijgt de taak om een zo hoog mogelijk score te halen. De AI zal dus langzaam leren wat goed is en wat fout is.

In hoofdstuk 5 wordt Supervised learning behandeld. Bij deze methode wordt de AI getraind met specifieke informatie. Er zal dus een trainer zijn die tegen de AI zegt; deze afbeelding laat een kat zien en deze afbeelding laat een hond zien. De AI leert langzaam aan het verschil tussen katten en honden te zien totdat de AI vanuit zichzelf kan zeggen dat dit een afbeelding van een kat en dat een afbeelding van een hond is.

In hoofdstuk 6 wordt Unsupervised learning behandeld. Bij Supervised learning wordt er gezegd tegen de AI wat de afbeelding laat zien. Bij Unsupervised learning wordt dit niet gedaan. De AI moet zelf gaan kijken, wat bij elkaar hoort. De AI zal alle afbeeldingen gaan vergelijken en alles wat op elkaar lijkt, bij elkaar zetten. Achteraf zal een persoon zeggen; ‘Deze afbeeldingen die je bij elkaar hebt gezet zijn van katten’. De AI weet nu dus, dat die groep op katten lijkt. Op deze manier zal de AI leren om het verband te vinden tussen dingen die gelijk zijn.

# Reinforcement Learning (RL)

In hoofdstuk 3 is een korte beschrijving gegeven over RL. In dit hoofdstuk zal er dieper ingegaan worden over wat RL is, hoe het werkt en waarvoor het gebruikt wordt. De voor- en nadelen van deze methodiek zullen behandeld worden in hoofdstuk 7.

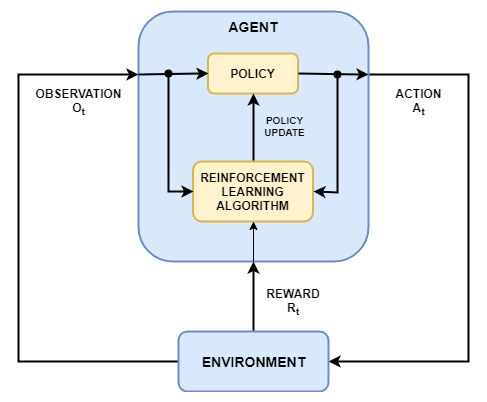
## 4.1 Wat is Reinforcement Learning?

Reinforcement Learning (RL) is leren hoe je bepaalde situaties aan acties kunt koppelen. Bij RL wordt niet gezegd wat je moet doen. Alles hangt af van de acties die de leerling uitvoert. Elke actie heeft een reactie en dus de uitgevoerde acties van de leerling kunnen een positieve of negatieve reactie hebben. In de meeste gevallen wil je dat de leerling alleen positieve reacties krijgt op zijn acties. Bij RL zal de leerling zelf moeten achterhalen welke actie tot die positieve reactie leidt. De leerling zal niet meteen, bij de eerste positieve reactie, weten wat gewenst is en wat niet. Alleen door trial-and-error zal de leerling leren welke acties de gewenste reactie opleveren. (Sutton & Barto, 2018, pp. 1-4).

Om het anders te beschrijven kan er gekeken worden naar jonge kinderen. Kinderen zijn heel nieuwsgierig, ze zullen overal op klimmen en alles aanraken. Als kind zelf, zat ik ooit bij mijn grootouders te kijken naar het vuur in de kachel in de woonkamer. Het vuur zag er heel mooi uit en dus wilde ik het aanraken. Ik drukte mijn vinger op de kachel waarin al een tijdje een vuur brandde. Het is dus niet ver gezocht dat ik mijn vinger flink verbrandde. Dit is Reinforcement Learning. Een actie wordt uitgevoerd en de leerling leert gebaseerd op het resultaat van de actie, zowel positief als negatief. Ik leerde om niet meer kachels aan te raken als er vuur in zit door middel van een negatieve beloning.

## 4.2 Hoe werkt Reinforcement Learning?

Zoals eerder vernoemd, werkt RL op het principe van doen en ervaren. Het is een methode waarbij de leerling zelf moet leren zonder dat hij informatie gegeven krijgt. We zullen nu verder ingaan op hoe je dat voor elkaar krijgt. In figuur 4.1 is een diagram te zien dat het principe van RL weergeeft.



Figuur 4.1 RL Diagram (Matworks, 2022)

Het is al eerder vernoemd, dat bij RL de leerling de omgeving observeert en daarop gebaseerd een actie maakt. Die actie creëert een reactie in de omgeving en de leerling krijgt een positieve of negatieve beloning. Deze cirkel herhaalt zichzelf constant. Nadat de leerling talloze keren door de cirkel heengegaan is, zal deze leren om de reactie van een bepaalde actie in een bepaalde omgeving te leren voorspellen. Op dat moment is het doel van de leerling bereikt. Om te zorgen dat de leerling sneller leert, kun je extra feedback geven aan het RL Algoritme. Door observaties van vorige omgevingen en de acties die de leerling neemt te vergelijken, zal de leerling sneller op het punt komen waarbij die reacties kan gaan voorspellen. (Ribeiro, 2002)

Op het gebied van RL-algoritmes is Q-Learning het meest gebruikt en begrepen. Q-Learning is voor het eerst geïntroduceerd door Christopher Watkins in 1989. Het doel van Q-learning is het vinden van de optimale beloning Q. Dat wil zeggen dat de gehele algoritme op zoek is naar de Q die het meest oplevert en dus het beste presteert. Het algoritme werkt op dezelfde manier als is afgebeeld in afbeelding 4.1. De huidige situatie wordt bekeken, een actie wordt gemaakt, het resultaat wordt bestudeerd. Een beloning wordt gegeven gebaseerd op het resultaat en Q wordt aangepast als de nieuwe beloning beter is dan de oude beloning. Als dat niet het geval is dan zal het algoritme de situatie opnieuw bekijken en een nieuwe actie kiezen totdat er een betere Q tevoorschijn komt. (Zie ook Watkins, 1989 en Watkins & Dayan, 1992)

## 4.3 Toepassingen van Reinforcement Learning.

Voor diegenen die zich bezighouden met Machine Learning zal de naam AlphaGo niet onbekend zijn. AlphaGo, die gebruik maakt van de AlphaZero algoritme, is een van de meest bekende toepassingen van Reinforcement learning. AlphaGo is de supercomputer die voor het eerst in de geschiedenis een professionele GO-speler versloeg in een wedstrijd. Het is ook de eerste computer die de wereldkampioen GO-spelen versloeg. De supercomputer leerde GO spelen door constant tegen zichzelf wedstrijden te spelen. Voordat AlphaGo begon met leren, werd hem de regels uitgelegd. Daarna zijn alle vooruitgangen vanuit AlphaGo zelf gekomen totdat het goed genoeg was om de beste GO-speler ter wereld te worden. (AlphaGo, 2017)

Aangezien het doel van RL is om robots te voorzien van de mogelijkheid om te leren en te groeien, zal het niet verrassend zijn dat RL wordt toegepast in de robotica. Zo worden er in productielijnen robotarmen gebruikt die zichzelf hun taak hebben aangeleerd.

# Supervised Learning

In hoofdstuk 3 is een korte beschrijving gegeven over Supervised Learning. In dit hoofdstuk zal er dieper ingegaan worden over wat Supervised Learning is, hoe het werkt en waarvoor het gebruikt wordt. De voor- en nadelen van deze methodiek zullen behandeld worden in hoofdstuk 7.

## 5.1 Wat is Supervised Learning?

Supervised Learning (SL) is het leren door gebruik te maken van menselijke ervaringen. Bij SL is er duidelijk een leraar en een leerling. Het doel van SL is een leerling, zoals bij een mens, te laten leren. Een leraar vertelt of laat iets zien en de leerling zal hierdoor nieuwe kennis en vaardigheden krijgen. SL kan worden onderverdeeld in Classification en Regression.

Als voorbeeld nemen we kinderen samen met hun ouders die naar vuur staan te kijken, bijvoorbeeld een open haard of een kampvuur. Een kind ziet een mooi vuur en wil het aanraken. De ouders zien dat en zeggen ‘Doe dat niet, je zult je vingers verbranden’. Ze zullen waarschijnlijk ook het kind tegenhouden. Dit is waar SL om draait. Iemand die de kennis en ervaringen heeft zal tegen iemand, die de kennis en ervaring niet heeft, zeggen wat er gebeurt als bepaalde acties worden uitgevoerd.

## 5.2 Classification

Van de twee subcategorieën is Classification de eerste waarvan de werking en toepassing behandeld zal worden.

### 5.2.1 Wat is Classification?

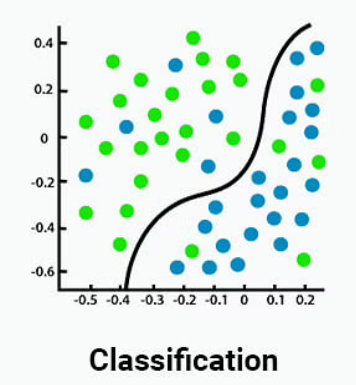
Classification is het werken met gelabelde informatie. Voordat de leerling gaat leren, wordt letterlijk aan de leerling verteld wat hij ziet. Een foto van een kat zal dus altijd een kat zijn en de foto van een hond zal altijd een hond zijn.

### 5.2.2 Hoe werkt Classification?

Wanneer een Classification-leerling wordt getraind zal die werken met bekende groepen data. Er zal tegen hem gezegd worden, deze groep hoort bij elkaar. De leerling weet alleen dat die groep bij elkaar hoort en niet waarom dat zo is. De reden voor de groepering zal de leerling zelf moeten bedenken. Er kan bijvoorbeeld gekeken worden naar de kleur of de vorm bij afbeeldingen. De leerling zal conclusies trekken waarom de groep een groep is. Vanuit deze conclusie zal de leerling later zelf groeperingen kunnen maken met niet gelabelde data. Maar hoe gaat dat te werk?

Bij Classification zijn data van groot belang. Anders dan met RL, waarbij je geen tot weinig data nodig hebt, hangt bij Classification alles af van de kwaliteit van de data die gebruikt wordt om de leerling te trainen. Er is ook een bepaald percentage onvoorspelbaarheid bij Classification. De documentaire ‘Coded Bias’ uit 2020 laat zien hoe een gezichtsherkenning AI alleen gezichten van blanke mensen waarneemt. Omdat deze AI alleen getraind was op afbeeldingen en gezichten van blanke mensen, begon deze AI de kleur van het gezicht mee te nemen. Doordat er ook naar de kleur van het gezicht werd gekeken, dacht de AI dat donkere mensen geen mensen waren.

In Afbeelding 5.1 is een voorbeeld van Classification te zien.



Afbeelding 5.1 Een voorbeeld van Classification (Terra, 2022)

Alle punten in deze afbeelding stellen data voor. De waarde van deze data is onbelangrijk. Nadat alle datapunten in een grafiek verwerkt zijn, is het mogelijk om een lijn te trekken door de grafiek. Hierdoor heb je, je twee verschillende datawaardes gescheiden. De leerling kan nu makkelijk zeggen dat alles links van de lijn X is en alles rechts van de lijn Y. Dit gaat natuurlijk niet met 100% nauwkeurigheid, er zullen altijd een paar datapunten zijn die buiten de lijnen vallen.

### 5.2.3 De toepassingen van Classification

Classification wordt vaak gebruikt voor het detecteren van fraude bij bijvoorbeeld creditcards. Alle betalingen die met een creditcard worden uitgevoerd worden bijgehouden. Met deze data kan een simpele Classification AI getraind worden. Wanneer deze AI opeens afschriften ziet die buiten de gewone marges vallen, kan er meteen aan de bel getrokken worden. Dit zorgt ervoor dat een menselijk element niet meer nodig is om alle afschriften te controleren. Fraude wordt hierdoor dus eerder ontdekt.

Een ander voorbeeld van de toepassing van Classification is het herkennen van ziektes zoals huidkanker of Alzheimer. AI kan getraind worden op afbeeldingen van huidkanker en gezonde huid. De AI zal leren om het verschil te vinden tussen huidkanker en gezonde huid, ook als het verschil met het menselijk oog niet te zien is. Hierdoor zou je dus met een app kunnen kijken ‘Is deze pigmentvlek huidkanker of is het dat niet?’ Hierdoor zou je een bezoek aan de huidspecialist kunnen besparen of juist met spoed een afspraak maken met de huidspecialist om te bevestigen of het inderdaad huidkanker is. We hebben het ook over de ziekte van Alzheimer gehad. Er zijn AI die getraind zijn op MRI-scans van mensen met Alzheimer en mensen zonder Alzheimer. De AI leerde op een gegeven moment, om met 89,22% zekerheid Alzheimer te vinden in de MRI-scans. (Mahmood & Ghimire, 2013)

## 5.3 Regression

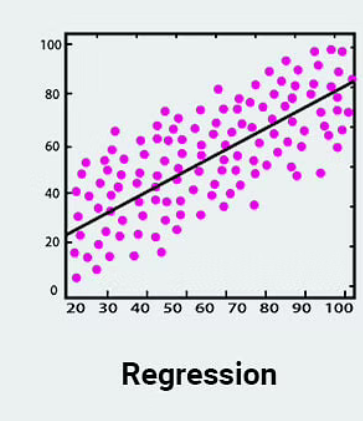
Van de twee subcategorieën van SL is Regression de tweede waarvan de werking en toepassing in deze paragraaf behandeld wordt.

### 5.3.1 Wat is Regression?

Regression learning is een techniek die voor twee theorieën wordt gebruikt. Regression kan gebruikt worden voor prognoses en voorspellingen maar ook voor het leggen van verbanden tussen afhankelijke en onafhankelijke variabelen. (Maulad & Abdulazeez, 2020)

### 5.3.2 Hoe werkt Regression?

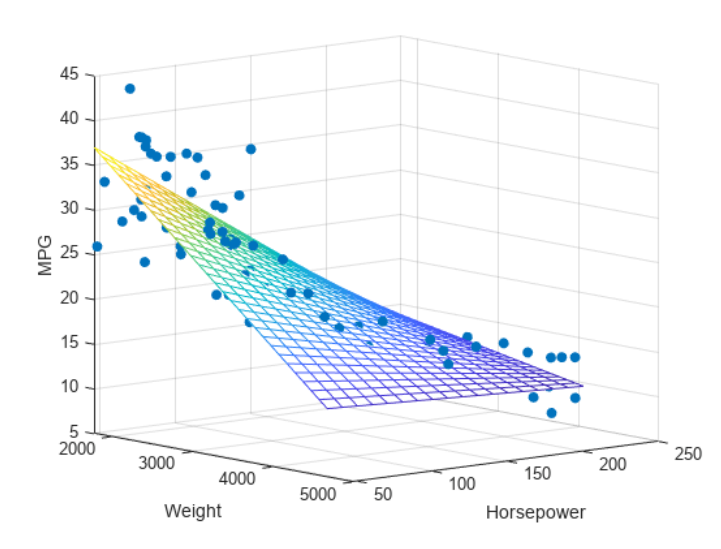
Bij Regression kunnen verbanden worden gelegd tussen vergelijkbare datapunten. In afbeelding 5.2 is een voorbeeld van Simple Linear Regression te zien.



Afbeelding 5.2 Een voorbeeld van Simpel Linear Regression (Terra, 2022)

Bij Simple Linear Regression wordt er één rechte lijn door de datapunten getrokken om de fout tussen de lijn en de data punten zo klein mogelijk te houden. Dit is anders dan bij Classification, waar de lijn buigt om zoveel mogelijk van dezelfde datapunten aan een kant te hebben. Simple Linear Regression wordt gebruikt wanneer er in de dataset maar één onafhankelijke variabele is.

Wanneer er meerdere onafhankelijke variabelen zijn gaan we van Simple Linear Regression naar Multiple Linear Regression. Afbeelding 5.3 is een voorbeeld van Multiple Linear Regression.



Afbeelding 5.3 Een voorbeeld van Multiple Linear Regression (Mathworks, Multiple linear regression, 2006)

Anders dan bij Simple Linear, kijkt Multiple Linear naar het vlak waar zich de meeste datapunten bevinden. Het voorbeeld in afbeelding 5.3 laat het verband zien tussen de mijlen per gallon en het gewicht en paardenkracht van een auto. Als er meer dan twee onafhankelijke variabelen aanwezig zijn, dan kan de Multiple Linear niet meer in een grafiek worden weergegeven. De variabelen worden dan weergegeven in een tabel. Simple Linear en Multiple Linear volgen hetzelfde principe binnen de Regression. (Tranmer, Murphy, Elliot, & Pampaka, 2020)

### 5.3.3 Toepassingen van Regression

Regression wordt vooral gebruikt voor het voorspellen van data. Denk hierbij aan het voorspellen van het weer of levensverwachtingen. Het kan ook gebruikt worden voor het voorspellen van de bevolkingsgroei. Regression wordt dus voornamelijk gebruikt als je de leerling wil gaan trainen om data of omstandigheden te voorspellen.

# Unsupervised Learning (UL)

In hoofdstuk 3 is een korte beschrijving gegeven van Unsupervised Learning. In dit hoofdstuk zal er dieper worden ingegaan op wat UL is, hoe het werkt en waarvoor het gebruikt wordt. De voor- en nadelen van deze methodiek zullen behandeld worden in hoofdstuk 7.

## 6.1 Wat is Unsupervised Learning?

Het principe van UL is het tegenovergestelde van SL. In hoofdstuk 5 is er gezegd dat SL het leren is door gebruik te maken van menselijke ervaringen en kennis. Bij UL wordt de menselijke ervaring en kennis niet gebruikt. De leerling zal zelf alles moeten achterhalen. Pas helemaal op het einde van het leertraject van de leerling, zal een leraar komen kijken of dat wat de leerling geleerd heeft wel het juiste is. Data kan gezien worden als een grafiek met rij X en kolom Y. Bij UL zal de data die wordt gegeven alleen X *of* Y bevatten. De ontbrekende component wordt door de leerling bedacht. (Qin, 2020)

UL kan onderverdeeld worden in drie methodieken; Clustering, Association en Dimensionality Reduction.

## 6.2 Clustering

Clustering is een van de drie methodieken die behandeld worden op het gebied van UL. Ten eerste zal er kort beschreven worden wat Clustering is. Daarna zullen er twee algoritmes worden besproken.

### Wat is Clustering?

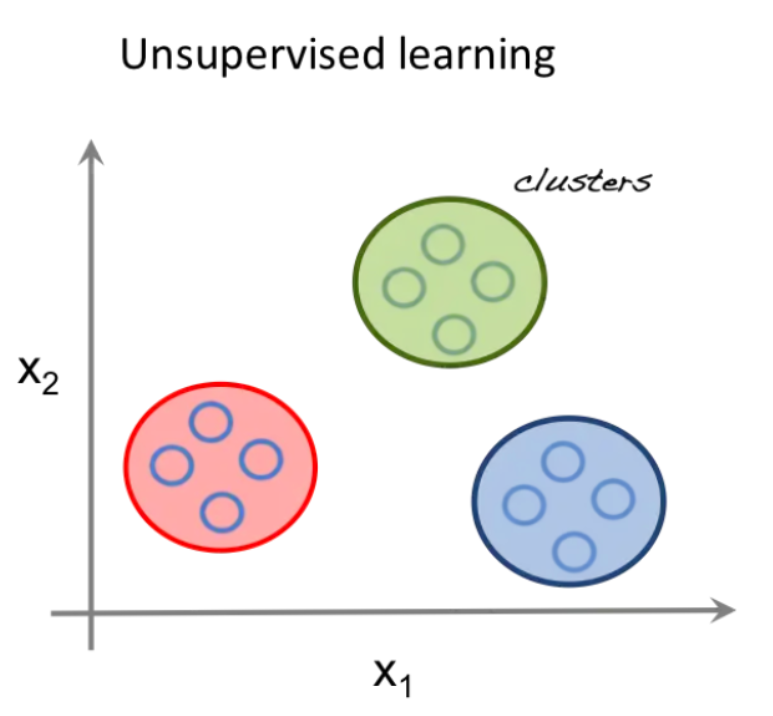
Clustering is een methode om verbanden te vinden in een groep data. Bij Clustering zal gekeken worden naar natuurlijke groepen in de data. Denk hierbij bijvoorbeeld aan de kleur van een appel of de vorm van een stuk fruit zoals te zien is in afbeelding 6.1. Daarin is te zien hoe een berg aardappels wordt gegroepeerd (Clustered) gebaseerd of hun kleur.

Afbeelding met fruit, noot, verschillend, groente

Automatisch gegenereerde beschrijving

Afbeelding 6.1 Voorbeeld clustering met aardappels (Suresha, 2021)

Clustering is dus het organiseren van niet-gelabelde data in groepen die iets gemeen met elkaar hebben. In afbeelding 6.2 is een ander voorbeeld te zien van Clustering.



Afbeelding 6.2 clustering in UL (Youngson, 2020)

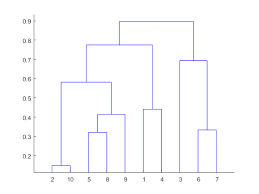
De gemeenschap van de verschillende data punten is dat deze dicht bij elkaar liggen op de onbekende assen X1 en X2. Alle data punten zullen wellicht iets met elkaar te maken hebben, maar de UL-leerling heeft deze gebaseerd op positie aan elkaar gekoppeld. (Serra & Tagliaferri, 2019)

### Hoe werkt Clustering?

Zoals met alle ML methodes wordt er gebruik gemaakt van algoritmen. Er zijn veel verschillende algoritmes gemaakt voor Clustering. Een paar van deze algoritmes zijn K-means en Hierachical.

Bij K-means clustering is het doel om te achterhalen hoe vaak het nummer K voorkomt in de data die is gegeven. Hierbij is K een willekeurig verband die een cluster vormt. Bij K-means clustering zijn er een paar parameters van belang. Dit zijn: het centerpunt van de cluster en de startlocatie van de cluster. Het startpunt van de cluster kan op een willekeurige positie worden geplaatst. Als een leerling gebruik maakt van K-means om verbanden te vinden in een dataset, dan zal deze van tevoren gezegd worden dat er n-aantal K’s gevonden moeten worden. Bij deze K’s moet dan het centerpunt worden weergegeven, zodat de leraar achteraf kan zien wat de leerling heeft geleerd. (Károly, Fullér, & Galambos, 2018) Afbeelding 6.1 van hoofdstuk 6.2.1 is een basiscluster zoals verkregen wordt met een K-means-algoritme.

Bij Hierarchical clustering worden er geen willekeurige clusters gemaakt maar volgen deze een specifieke hiërarchie. Waar bij K-means het nodig was om te zeggen hoeveel cluster er nodig zijn is dit bij Hierarchical niet het geval. Hierarchical clustering maakt als eindresultaat een dendrogram. Een voorbeeld van een dendrogram is te zien in afbeelding 6.2. (Károly, Fullér, & Galambos, 2018)



Afbeelding 6.2 Standaard dendrogram gemaakt in Mathlab ( (Mathworks, Dendrogram, 2022)

Een dendrogram wordt van boven naar beneden gelezen. Helemaal boven aan is de oorspronkelijke data. Verder naar beneden wordt de data opgesplitst in groepen die op elkaar lijken. Hoe verder je naar beneden gaat des te meer de data wordt opgesplitst in verschillende groepen. Dit is het einddoel van de Hierarchical Clustering methode.

Deze twee algoritmes zijn niet de enige algoritmes die gebruikt kunnen worden voor Clustering. Andere algoritmes zijn bijvoorbeeld Fuzzy Clustering, Decision Tree’s, Expectation maximization, Adaptive Resonance en meer. (Zie ook Károly, Fullér, & Galambos, 2018) Deze algoritmes worden in dit document niet verder besproken.

### Toepassingen van Clustering

Clustering zie je vaak terug in de datatechnologie. Het wordt hierin voornamelijk gebruikt om grote databases te doorgronden en nuttige data te vinden. Ook kun je het terugvinden in bijvoorbeeld spamfilters of zoekmachines. Het kan ook gebruikt worden voor gerichte marketing waarbij de Clustering AI kijkt naar wat jij aan het zoeken bent en dan alle dingen die daar op lijken aan jou voorstelt. Clustering wordt dus gebruikt om data te vinden in grote databases.

## 6.3 Dimensionality Reduction (DR)

Dimensionality Reduction is de tweede UL-methode die we behandelen. We beginnen met een beschrijving van wat DR is. Daarna gaan we verder met de werking van en als laatste de toepassingen van DR.

### 6.3.1 Wat is Dimensionality Reduction?

Dimensionality Reduction is een methode om met hoog-dimensionale data te werken. Hoog-dimensionale data is data met veel variabelen. Hoog-dimensionale data komt vaak voor in de medische zorg waar veel verschillende biometrische gegevens van patiënten nodig zijn. Vanwege de grote hoeveelheid variabelen in hoog-dimensionale data, zijn er maar weinig soorten algoritmes die met deze datasets kunnen werken. Om die data bruikbaar te maken is het nodig om de dimensionaliteit van de data omlaag te brengen. Wat van groot belang is, is dat de data hierdoor zo min mogelijk verandert.

### 6.3.2 Hoe werkt Dimensionality Reduction?

De precieze werking van DR is zwaar afhankelijk van de algoritme die gebruikt wordt. Het principe blijft wel hetzelfde. Wanneer je hoog-dimensionale data hebt, dat betekent data met honderden, duizenden of miljoenen variabelen. Laat je een DR AI de dimensionaliteit omlaag brengen, dan krijg je data met minder variabelen maar nog steeds met dezelfde structuur. DR moet een goede balans vinden van het aantal variabelen, zodat de AI zelf begrijpt wat de data betekent. Is het aantal variabelen te laag, dan kan de betekenis van de data veranderen en niet relevant zijn. Het probleem is dat er niet één DR AI is die alle verschillende soorten datasets kan opschonen. Voor elke soort dataset moet er een DR getraind worden. Wanneer een bedrijf maar met één soort hoge dimensionale dataset werkt dan is dit goed te doen, maar wanneer er meerdere soorten hoge dimensionale datasets gebruikt worden, dan zijn er ook meerdere DR AI’s nodig.

### Toepassingen van Dimensionality Reduction

DR wordt voornamelijk toegepast in de datatechnieken. Het is een methode om databases op te schonen en data bruikbaar te maken. In de praktijk kan een bedrijf of persoon eerst een DR AI loslaten op een database die hij gekregen of gemaakt heeft en vervolgens de nieuwe schonere data gebruiken om een AI te trainen. Schone data is data die dusdanig opgeschoond is dat deze gebruikt kan worden voor jouw doeleinden.

# Voor- en nadelen

In de vorige drie hoofstukken zijn drie Machine Learning methodes besproken. In dit hoofdstuk worden de voor- en nadelen van de verschillende methodes besproken.

## 7.1 Reinforcement Learning

Een van de grootste voordelen van RL is dat er geen tot weinig datasets nodig zijn om een RL-leerling te trainen. Waar de andere methodes allemaal databases en veel informatie nodig hebben, heeft een RL-leerling alleen een omgeving nodig om in te werken. Alle informatie die hij nodig heeft krijgt hij uit de omgeving. Een goed getrainde RL-leerling is ook beter in zijn vakgebied dan een mens ooit kan zijn. Hiervoor kan er gekeken worden naar Deepblue en AlphaGo, die beter zijn geworden in het spelen van schaken en Go dan de zittende grootmeesters.

Een van de grootste nadelen van RL is dat er geen onderscheid gemaakt wordt tussen exploratie en exploitatie. De leerling wil zoveel mogelijk positieve reacties krijgen. Terwijl dit voor de leerling gewenst is, is dit niet altijd het geval voor de leraar. Als voorbeeld van exploitatie kan er gekeken worden naar een AI die, door middel van RL, Tetrix leerde te spelen. De leerling krijgt positieve reacties door zo lang mogelijk in het spel te blijven. Door exploratie kwam de leerling erachter dat hij het spel op pauze kon zetten. Terwijl het spel op pauze stond kreeg de leerling nog steeds positieve reacties.

## 7.2 Supervised Learning

Een voordeel van SL is dat we altijd precies weten waar de AI naar kijkt. Aangezien wij de data en de labels geven, weten we altijd dat de AI gaat kijken naar die labels binnen de data.

Kijkend naar de nadelen van SL wordt het snel duidelijk dat SL niet goed is met complexe taken. Wanneer de leraar niet weet wat het verband is kan hij de leerling geen verband laten leren. Met UL zou dit wel mogelijk zijn. Een ander groot nadeel is dat je afhankelijk bent van je datasets. Als je dataset niet goed of compleet is, dan gaat de precisie van de SL-leerling ook achteruit.

## 7.3 Unsupervised Learning

Een van de grootste voordelen van UL is dat het verbanden kan zien die mensen zelf niet kunnen visualiseren. Een goed voorbeeld hiervan is de AI die Alzheimer kan herkennen gebaseerd op MRI-scans. We weten niet waar de AI naar kijkt maar het is wel 90% accuraat. Een ander voordeel is dat het makkelijker is om ongelabelde data te vinden dan de gelabelde data die nodig is voor SL.

Een nadeel van UL is dat, anders dan met SL, we niet altijd weten waar de AI naar kijkt. Hierdoor kan het leerproces waardeloos worden, als de AI de verkeerde verbanden gaat maken. Dit is het grootste nadeel van UL en het zorgt er ook voor dat de leraar geen invloed heeft op het eindresultaat van de UL-leerling.

## 7.4 Conclusie

Elke methode heeft zijn eigen voor- en nadelen. De voor- en nadelen die behandeld zijn, zijn globaal voor de drie methodes. Het kan zijn dat er een specifieke algoritme is voor één van de methodes die een groot nadeel van die methode kan oplossen. Maar dat brengt meestal andere nadelen met zich mee. Er kan ook niet geconcludeerd worden dat een bepaalde methode beter is dan de andere, aangezien het afhankelijk is van het doel waarvoor de AI wordt toegepast en de aanwezigheid of afwezigheid van data.

# Slotwoord

In het begin van dit verslag werd gezegd dat het doel is om lezers te informeren over Machine learning en te zorgen dat ze een basisbegrip hebben van Machine Learning en de methodes die gebruikt worden. Met dit literatuuronderzoek is er een beeld gegeven van wat Machine Learning is. Maar hoe het werkt is een ander verhaal. Het is niet mogelijk om de diepere werking en alle algoritmes die gebruikt kunnen worden uitteleggen in dit onderzoek. Als deze informatie toch gewenst is, dan raad ik aan om in de literatuurlijst te kijken en de bronnen op te zoeken. Vele van deze bronnen gaan een heel stuk dieper in op de wiskunde achter de algoritmes en de mogelijkheden van de AI’s.

Machine Learning en AI zullen steeds een grotere rol gaan spelen in de samenleving. Persoonlijk vind ik het van belang dat mensen weten wat er mogelijk is met AI en hoe het gebruikt wordt. Vanuit Fontys zijn er minors en studie waarin AI een rol speelt en deze raad ik aan als AI interessant is voor de lezer.

# Literatuurlijst

*AlphaGo*. (2017). Opgeroepen op November 1, 2022, van Deepmind: https://www.deepmind.com/research/highlighted-research/alphago

Coppin, B. (2004). *Artificial Intelligence Illuminated.* Sudbury, Canada: Jones and Bartlett Publishers. Opgeroepen op October 24, 2022

Dick, S. (2019, Juni 23). *Artificial Intelligence.* Harvard Data Science Review. doi:10.1162/99608f92.92fe150c

Kaplan, A., & Haenlein, M. (2019, Augustus). A brief history of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*(4), pp. 5-14. doi:10.1177/0008125619864925

Károly, A. I., Fullér, R., & Galambos, P. (2018). Unsupervised Clustering for Deep Learning: A tutorial survey. *Acta Polytechnica Hungarica*(8). doi:10.12700/APH.15.8.2018.8.2

Mahmood, R., & Ghimire, B. (2013, July 9). Automatic Detection and Classification of Alzheimer's Disease. *International Conference on Systems, Signals and Image Processing*(20). doi:10.1109/IWSSIP.2013.6623471

Mathworks. (2022). Dendrogram. Opgeroepen op Oktober 7, 2022, van https://nl.mathworks.com/help/stats/dendrogram.html

Mathworks. (sd). *Multiple linear regression.* Mathlab. Opgeroepen op November 21, 2022, van https://nl.mathworks.com/help/stats/regress.html

Matworks. (2022). What Is Reinforcement Learning? Opgeroepen op October 23, 2022, van https://nl.mathworks.com/help/reinforcement-learning/ug/what-is-reinforcement-learning.html

Maulad, D. H., & Abdulazeez, A. M. (2020). A review on Linear Regression Comprehensive in Machine Learning. *Journal of Applies Science and Technology Trends*(4), 104-147. Opgeroepen op November 2, 2022, van https://jastt.org/index.php/jasttpath/article/download/57/20

Qin, T. (2020). Machine Learning Basics. In *Dual Learning* (pp. 11-23). Singapore: Springer. doi:https://doi.org/10.1007/978-981-15-8884-6\_2

Ribeiro, C. (2002, Mei). Reinforcement Learning Agents. *Artificial Intelligence Review*(17), 223-250. doi:10.1023/A:1015008417172

Serra, A., & Tagliaferri, R. (2019). *Unsupervised Learning: Clustering.* Opgeroepen op Oktober 7, 2022, van https://www.researchgate.net/profile/Angela-Serra-3/publication/322543367\_Unsupervised\_Learning\_Clustering/links/5a6f7680458515015e616155/Unsupervised-Learning-Clustering.pdf

Suresha, H. (2021, Januari 12). Cluster analysys in Machine Learning. Opgeroepen op Oktober 7, 2022, van https://medium.com/mlearning-ai/cluster-analysis-6757d6c6acc9

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning* (2e ed.). MIT Press,. Opgeroepen op October 26, 2022

Terra, J. (sd). *Classification.* https://www.simplilearn.com/regression-vs-classification-in-machine-learning-article.

Tranmer, M., Murphy, J., Elliot, M., & Pampaka, M. (2020). Multiple Linear Regression. (2). Opgeroepen op November 21, 2022, van http://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/working-papers/2020/multiple-linear-regression.pdf

Watkins, C. J. (1989). *Learning from Delayed Rewads.* Universiteit van Londen. londen: King's College. Opgeroepen op October 22, 2022, van https://d1wqtxts1xzle7.cloudfront.net/50360235/Learning\_from\_delayed\_rewards\_20161116-28282-v2pwvq-with-cover-page-v2.pdf?Expires=1667393292&Signature=cezNhPSViN3aGLjPJp0SeOBX4KJpWLDJ4cSgXcszCLWuUrwDHN37hD74Z7TiIsndSSHRhTji4M6-wIUcKWKt5W8sjTQO7JM76JeXcShX

Watkins, C., & Dayan, P. (1992, Mei). Q-Learning. *Machine Learning*(8), pp. 279–292. doi:https://doi.org/10.1007/BF00992698

Youngson, M. (2020, Oktober 5). Unsupervised Learning. Opgeroepen op November 9, 2022, van https://mark-youngson5.medium.com/artificial-intelligence-the-beginning-of-a-new-era-3e3838807887

# Bijlage 1: Reflectie

Deze reflectie geschreven met behulp van de STARR-methode.

**S**ituatie

In het 3de semester van de opleiding Mechatronica moeten voor het vak PRP3 een literatuuronderzoek doen. Het schrijven en bepalen van deze opdracht is een eis om aan de stage in het 3de jaar te mogen beginnen. Dit wordt voornamelijk zelfstandig gedaan,

**T**aak

Mijn taak was het schrijven van een literatuuronderzoek met een onderwerp dat wij zelf interessant vonden en aansluit op de opleiding. Dit is een zelfstandige opdracht. Voor vragen kunnen we terecht bij de docent maar er wordt van ons verwacht dat wij zelf aan de slag gaan met informatie vinden en het schrijven van het verslag. Het doel was dus zelfstandig werken, informatie werven en rapporteren.

**A**ctie

Aangezien dit mijn 4de poging is om het literatuuronderzoek te schrijven, heb ik kritisch gekeken naar een onderwerp dat mij interesseert en mijn interesse kan houden. Tegelijkertijd met dit verslag deed ik ook de Minor ‘AI For Society’. Ik heb het onderwerp Machine Learning gekozen omdat ik toch al bezig ben met Machine Learning. Door die associatie met de minor kan in makkelijker mijn interesse in het verslag behouden.

Ik heb wanneer ik tijd had aan het verslag gewerkt. Er was niet een bepaalde vaste dag die ik gebruikte, aangezien ik lessen had vanuit de minor en ik ook aan persoonlijke en groepsprojecten moest werken, had ik niet een standaard dag om aan het literatuuronderzoek te werken.

Ik had geen vaste structuur tijdens het schrijven van het verslag. Ik schreef hoofdstukken die mij op dat moment interesseerden. Tijdens het schrijven zocht ik meer informatie en bronnen om te gebruiken. Het ene moment was ik bezig met het schrijven van een bepaald hoofdstuk en dan vond ik goede informatie voor een ander hoofdstuk en ga ik verder met het schrijven van dat hoofdstuk.

**R**esultaat

Mijn wisselvallige manier van schrijven zorgt ervoor dat ik veel tijd verlies. Elke keer als ik stop met schrijven van een hoofdstuk om opeens naar een ander hoofdstuk te gaan, verlies ik tijd omdat ik opnieuw in het schrijven moet komen. Wanneer dit gecombineerd wordt met mijn moeite om te beginnen aan iets nieuws, zorgt het ervoor dat ik veel energie en tijd verlies. Mijn latere start met het daadwerkelijk schrijven hielp ook niet mee.

**R**eflectie

De eerste vraag bij een reflectie is meestal ‘Hoe vond je dat je het hebt gedaan?’. Als dat vergeleken wordt met mijn andere pogingen tot het schrijven van een literatuuronderzoek, dan heb ik het 100% beter gedaan. Ik ben alleen zelf niet 100% tevreden. Ik heb een verslag geschreven dat aan de minimumeisen voldoet Ik heb iets om in te leveren. Maar heb ik het goed gedaan? Nee. Geen vaste planning, geen structuur tijdens het schrijven en veel verspilde tijd en energie. Ik heb hier wel van geleerd. Structuur is belangrijk, vooral voor mij. Ik heb die structuur nodig om te kunnen beginnen en om bezig te blijven. Dit is iets wat ik terugzie in andere projecten en opdrachten en waar ik al lang mee stoei. Dit verslag heeft dus bevestigd wat ik al van mijzelf weet en laat weer zien dat ik daaraan moet werken.