



Project AI (INFPRJ01)

Versie 1.0
Final Version
1 Maart 2023

Onderzoeksverslag

Artificial Intelligence

Hakan Dalama (0963609)
(Gizem Sazak, Rahul Udhayua, Remco de Zeeuw, Hoes Fortuin)

Rotterdam University of Applied Sciences

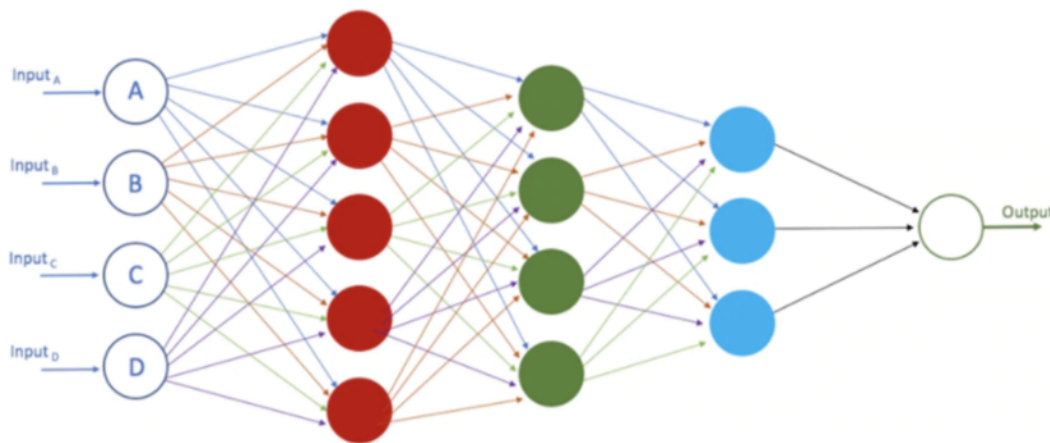
Wat is Deep Learning?

Laten we de volgende definitie erbij pakken:

“Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example.” (MathWorks, sd)

Deep Learning is dus min of meer het aanleren van een computer om te leren zoals mensen dat doen. Een goed voorbeeld hiervan is een zelfrijdende auto die verkeersborden kan herkennen en hier op de juiste manier op reageren (bv. Stoppen voor een stopbord, snelheid aanpassen aan een maximum snelheidsbord, etc.). (MathWorks, sd)

Deep Learning verschilt van andere soorten AI door het gebruik van zogenaamde lagen. Als er meerdere opdrachten tegelijk worden uitgevoerd (elk in een eigen laag) spreken we over Deep Learning. Dit geheel van lagen wordt een neuraal netwerk genoemd. (Meer, 2019)



Figuur 1: Neuraal network (Tierney, sd)

Wat is Machine Learning?

Net als bij Deep Learning pakken we er eerst een definitie bij:

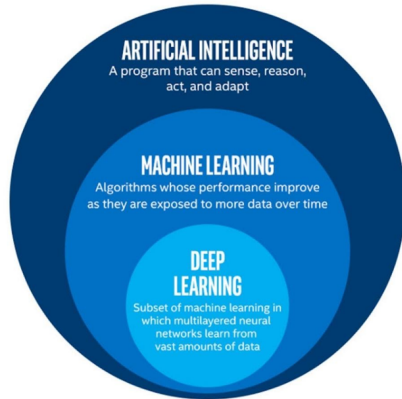
“Machine learning is een vorm van kunstmatige intelligentie (AI) die is gericht op het bouwen van systemen die van de verwerkte data kunnen leren of data gebruiken om beter te presteren.” (Oracle, sd)

Daarnaast valt Machine Learning op te delen in twee varianten, namelijk supervised learning en unsupervised learning. (Oracle, sd)

Aangezien dit onderzoek vooral gericht is op Deep Learning, gaan we hier verder niet op in.

Verschil tussen Machine Learning en Deep Learning

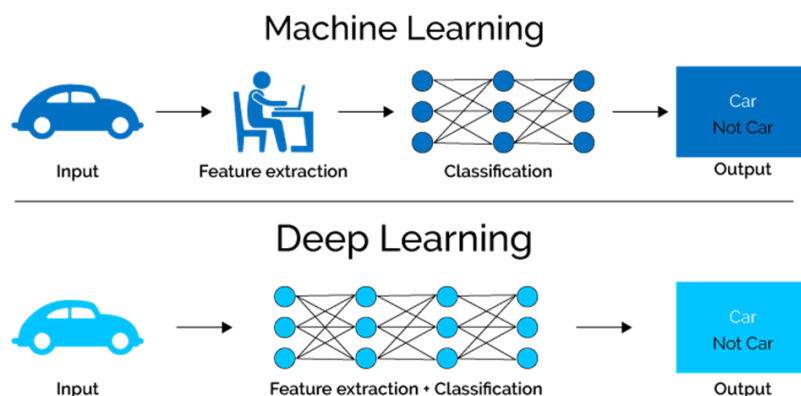
Machine Learning en Deep Learning zijn erg verwant aan elkaar. Sterker nog, Deep Learning is een onderdeel van Machine Learning, zoals te zien in onderstaand figuur.



Figuur 2: Relatie tussen Machine Learning en Deep Learning (Oracle, sd)

Machine Learning en Deep Learning delen dus dezelfde werking. Het voordeel van Deep Learning is dat het gebruik maakt van zogenaamde layers (lagen), waardoor er meer data tegelijkertijd verwerkt kan worden. Dit heeft echter als nadeel dat het veel meer rekenkracht kost, vooral van de GPU (videokaart). (Pilli, sd)

Ook maakt Deep Learning gebruik van neurale netwerken, waardoor het systeem uiteindelijk zelf patronen leert herkennen. Machine Learning is in dat opzicht een stuk minder complex, en kan geen patronen leren herkennen (zie ook figuur 4). (Pilli, sd)



Figuur 3: Verschillen tussen Machine Learning en Deep Learning (Pilli, sd)

	Deep Learning	Machine Learning
Wordt helaas en ten onrechte verward met AI/Kunstmatige Intelligentie	V	V
Maakt gebruik van neurale netwerken	V	O
Draait om ontwikkelen algoritmes	O	V
Leert dankzij het herkennen van patronen	V	O

Figuur 4: Deep Learning en Machine Learning met elkaar vergeleken (Meer, 2019)

Automatic Text Classification

Text Classification is een methode binnen Machine Learning (en dus ook Deep Learning) die tekst kan sorteren op categorieën op basis van verschillende criteria. Een goed voorbeeld hiervan is het sorteren van nieuwsberichten op onderwerp. Deze methode lijkt in eerste instantie ideaal voor de doeleinden van dit project. Deze methode kan zowel handmatig als automatisch uitgevoerd worden (supervised en unsupervised). Ook biedt Python genoeg keuze wat betreft libraries die Text Classification ondersteunen. Enkele voorbeelden hiervan zijn Scikit-Learn en NLTK. Met TensorFlow is het ook mogelijk om Text Classification toe te passen in Deep Learning. (Pascual, 2019)

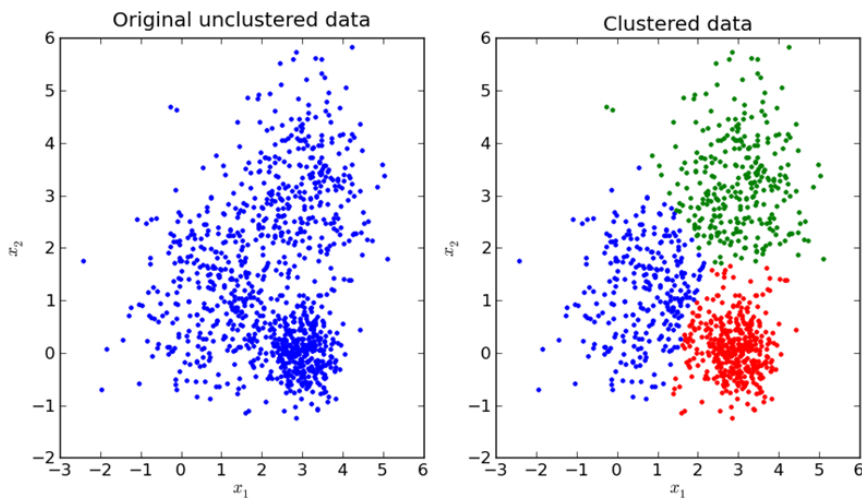
Text Classification werkt als volgt: Je “voert” het systeem een aantal goede voorbeelden. Na een tijdje gaat het systeem leren wat wat is, en heeft het geen voorbeelden meer nodig. Het systeem is dan goed genoeg getraind om zelfstandig te kunnen functioneren. (Pascual, 2019)

Supervised Text Classification

In deze vorm van Text Classification geeft men gelabelde data aan het systeem. Het systeem wordt hiermee getraind, en zal steeds beter worden in het zelfstandig sorteren van de data. Een goed voorbeeld hiervan is een spamfilter in een emailontvanger. Na een tijdje volgt de testperiode, waarin het systeem data krijgt zonder labels, waarna het systeem zelf de data moet indelen op categorie. Als dit goed gaat is het systeem voldoende getraind. Zo niet, dan zal er nog meer gelabelde data nodig zijn. (KDnuggets, sd)

Unsupervised Text Classification

Unsupervised Text Classification wordt uitgevoerd zonder externe informatie te geven aan het systeem. De data heeft dus geen labels, zoals dit wel het geval is bij de supervised method. Bij deze method probeert het algoritme structuur in de data te vinden. Het algoritme maakt op basis van de resultaten clusters. Alle punten die dicht bij elkaar liggen worden gezien als categorie. Op die manier bouwt het algoritme alle mogelijke categorieën op. (KDnuggets, sd)



Figuur 5: Geclusterde data na unsupervised learning

Kunnen wij Deep Learning in ons project gebruiken?

Om die vraag te beantwoorden zullen we moeten kijken naar de precieze werking van Deep Learning. Zoals hierboven gezegd is het een manier om het menselijk leren te simuleren. Het is voor ons natuurlijk heel makkelijk om een negatieve comment te herkennen. Een computer kan dit echter niet uit zichzelf.

Deep Learning zou een goede manier zijn om onze AI te trainen. Een van de manieren waarop Deep Learning werkt is het “voeren” van gelabelde data, om zo de AI te trainen. Op basis hiervan wordt een model opgesteld waarmee de AI in de toekomst geen label meer nodig heeft. Hoe meer training (gelabelde data, of comments in dit geval), des te beter het model zal worden. (MathWorks, sd)

Daarnaast heeft Deep Learning, in tegenstelling tot Machine Learning, geen menselijke input nodig. Ook dit maakt deze methode meer geschikt voor ons project.

Echter, omdat Deep Learning krachtigere hardware nodig heeft, is het onwaarschijnlijk dat we dit kunnen laten uitvoeren door onze eigen PC's. We zullen dus gebruik moeten maken van cloud computing als we Deep Learning willen inzetten. Hogeschool Rotterdam biedt ons deze mogelijkheid aan, als we dit zouden willen. (Pilli, sd)

Welke programmeertaal is het meest geschikt?

Voor Machine Learning/Deep Learning zijn er redelijk wat programmeertalen geschikt, maar de drie die er met kop en schouders bovenuit steken zijn C#, R en Python. In dit onderzoek zullen wij alleen deze drie talen onder de loep nemen. (SSA DATA, 2020)

C#

In principe is dit een goede keuze voor Machine Learning. Echter, het aanbod van libraries met betrekking tot AI is beperkt. Ook is C# voornamelijk gericht op het Windows-platform, en zal het dus vrij lastig zijn om een cross-platform applicatie te bouwen met deze taal. (SSA DATA, 2020)

R

Deze taal is vooral gericht op het verwerken van data. Ideaal voor AI dus. Echter zijn de mogelijkheden buiten dataverwerking om vrij minimaal, en is de taal ook berucht om zijn traagheid in het maken van berekeningen. (SSA DATA, 2020)

Python

Deze taal wordt het meest gebruikt bij AI-gerelateerde applicaties. Python heeft een ruim aanbod aan libraries voor AI, is makkelijk te leren, en werkt op vrijwel alle platforms. (SSA DATA, 2020)

Aan de hand van bovenstaande informatie is Python voor ons project het meest geschikt. Sommige van onze projectleden hebben al ervaring met Python, ook dit weegt mee in onze beslissing.

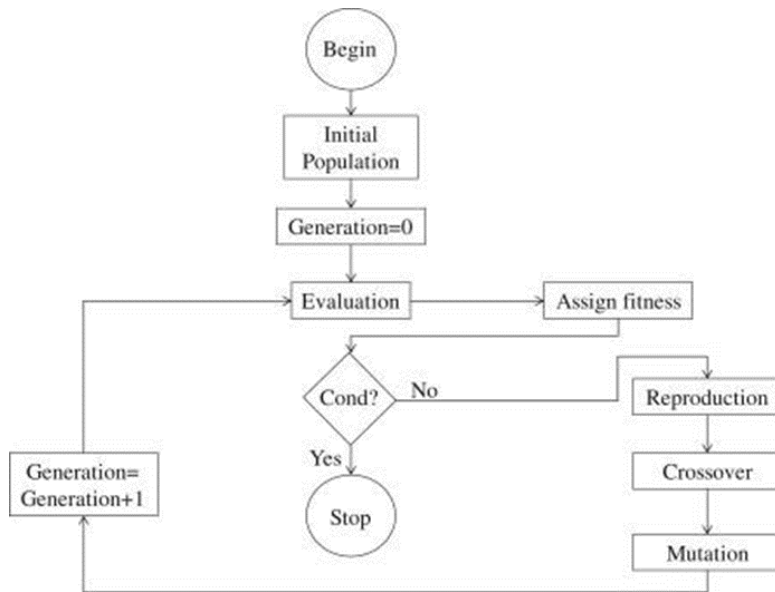
Wat is Genetic Algorithm?

Genetic Algorithm is gebaseerd op de Darwin theorie. Het is een langzaam proces waarbij steeds de twee beste resultaten worden gekozen. Dit zijn de parents. Vervolgens wordt er gebaseerd op die parents weer een nieuwe generatie gemaakt. Van die nieuwe generatie worden de twee besten weer gekozen, enzovoort. Deze methode is langzaam, maar zorgt er wel voor dat de AI zich steeds verbetert. (Gad, 2018)

Dit proces kent een aantal stappen:

1. Initial Population – Dit is een willekeurige populatie gebaseerd op de data.
2. Fitness function – Zoeken naar het fitness niveau van alle uitkomsten.
3. Selection – Selecteer de twee uitkomsten met de hoogste fitness.
4. Cross-over – Creëer een nieuwe populatie gebaseerd op de zojuist gekozen parents.
5. Mutation – Mutatie uitvoeren voor een of meerdere eigenschappen van de nieuwe generatie. Zo ontstaat er meer diversiteit. (Choudhary, sd)

Voer vervolgens stap 2-5 uit voor elke opeenvolgende generatie, net zolang tot de nieuwere generaties niet meer verbeteren of totdat er verzadiging is opgetreden. (Choudhary, sd)



Figuur 6: Genetic Algorithm uitgebeeld (Science Direct, sd)

Wanneer gebruik je Genetic Algorithms

Je gebruikt Genetic Algorithms als je iets wilt optimaliseren. In veel gevallen in AI wordt iets niet verbeterd, maar met GA kun je elke generatie verbeteringen doorvoeren. Op die manier creëer je steeds een efficiënter product, met een hogere accuracy. (Sommerville, 2021)

Kunnen wij Genetic Algorithms gebruiken in ons project?

Wij kunnen dit zeker in ons project gebruiken. Met de methode die we tot nu toe hebben gebruikt krijgen we slechts één resultaat gebaseerd op een bepaalde dataset. Dit betekent dat het systeem zich daarna niet meer verbeterd, en dus niet meer bijleert. Op die manier zal de efficiëntie nooit toenemen. Zelfs met een dataset van duizenden records was de efficiëntie niet hoger dan 75%. Als we Genetic Algorithms gebruiken zal deze efficiëntie veel hoger kunnen worden naarmate de generaties toenemen.

Wat is Reinforcement Learning?

Reinforcement Learning is een vorm van kunstmatige intelligentie (AI). Het is een manier van zelflerende vermogen onder de categorie Machine Learning. Het doel is om het ideale gedrag te ontdekken voor de best haalbare prestaties.

Een aantal voorbeelden uit de praktijk zijn bijvoorbeeld:

- Robots producten laten uitpakken
- Prijzen automatisch veranderen
- Automatisch beleggen

Het werkt met observatie, ontdekking en een trial and error beloningssysteem. Een voorbeeld uit de praktijk kan zijn een hond, als een hond iets goeds doet wordt hij beloond met een snoepje bijvoorbeeld. Als de hond niet doet wat hij hoort te doen krijgt de hond geen beloning en zal dit dan ook hoogstwaarschijnlijk niet meer doen.

Basic reinforcement is gemodelleerd als een "Markov decision process" (**MDP**):

Types of Reinforcement

Positief

Positieve Reinforcement wordt gedefinieerd als wanneer een gebeurtenis optreedt als gevolg van een bepaald gedrag, de kracht en de frequentie van het gedrag toeneemt. Met andere woorden, het heeft een positief effect op gedrag .GeeksforGeeks. (2020, 18 mei).

Voordelen van Reinforcement learning zijn:

- Maximaliseert prestaties
- Ondersteunt verandering gedurende een lange periode

Nadelen van Reinforcement learning:

- Te veel versterking kan leiden tot een overbelasting van toestanden, waardoor de resultaten kunnen verminderen Sutton, R. S., & Barto, A. G. (2018).

Negatief

Negatieve Reinforcement wordt gedefinieerd als het versterken van een gedrag doordat een negatieve toestand wordt gestopt of vermeden.

Voordelen van Reinforcement learning:

- Verhoogt gedrag
- Verzet tegen de minimale prestatie norm

Nadelen van Reinforcement learning:

- Het biedt alleen genoeg om aan het minimum gedrag te voldoen

Kan het gebruikt worden voor het product?

Waarvoor wordt Reinforcement Learning normaliter gebruikt?

Reinforcement Learning is een vorm van technologie dat gebruikt wordt om producten beter te kunnen beheren (optimaliseren) , implementeren (personaliseren), realiseren (efficiënter maken) en ontwerpen (verbeteren van applicatie) (2019, 11 juli). APPelit. Q-learning, double Q learning en policy gradient methods zijn algorithms onder Reinforcement Learning

Kan het specifiek gebruik worden voor ons product?

Er is een mogelijkheid om Automated Text Classification te gebruiken voor ons project.

Automated Text Classification kan taken toewijzen tot 1 of meer member Gupta, S. (2018, 21 juni)

Bronnen

Choudhary, A. (sd). Generic Algorithm in Machine Learning using Python. DataScience+. <https://datascienceplus.com/genetic-algorithm-in-machine-learning-using-python/>

Gad, A. (2018, juli 3). Introduction to Optimization with Genetic Algorithm. Towards Data Science. <https://towardsdatascience.com/introduction-to-optimization-with-genetic-algorithm-2f5001d9964b>

KDNuggets. (sd). Automated Text Classification Using Machine Learning. KDNuggets. <https://www.kdnuggets.com/2018/01/automated-text-classification-machine-learning.html>

MathWorks. (sd). What Is Deep Learning? MathWorks. <https://www.mathworks.com/discovery/deep-learning.html>

Meer, B. v. (2019, mei 31). Wat is deep learning? Een korte introductie tot deep learning. Moqod. <https://moqod.com/nl/wat-is-deep-learning-een-korte-introductie-tot-deep-learning/>

Oracle. (sd). Wat is machine learning. Oracle. <https://www.oracle.com/nl/data-science/machine-learning/what-is-machine-learning/>

Pascual, F. (2019, oktober 14). Text Classification with Python. MonkeyLearn. <https://monkeylearn.com/blog/text-classification-with-python/#:~:text=Text%20classification%20is%20the%20automatic,categories%20to%20unstructured%20text%20data.&text=That%27s%20a%20lot%20of%20social,companies%20make%20data%2Dbased%20decisions.>

Pilli, N. (sd). Difference between Machine Learning and Deep Learning. Morioh. <https://morioh.com/p/ed56b4fdbf1c>

Science Direct. (sd). Genetic Algorithm. <https://www.sciencedirect.com/topics/engineering/genetic-algorithm>

Sommerville, G. (2021, februari 4). Using genetic algorithms on AWS for optimization problems. AWS. <https://aws.amazon.com/blogs/machine-learning/using-genetic-algorithms-on-aws-for-optimization-problems/>

SSA DATA. (2020, januari 20). TOP Programming Languages For Machine Learning (Python, C#, R). Geraadpleegd van <https://www.ssa-data.com/blog/archive/best-programming-languages-for-machine-learning/>

Tierney, B. (sd). Understanding, Building and Using Neural Network Machine Learning Models using Oracle 18c. Geraadpleegd van <https://developer.oracle.com/databases/neural-network-machine-learning.html>

GeeksforGeeks. (2020, 18 mei). Reinforcement learning. Geraadpleegd van <https://www.geeksforgeeks.org/what-is-reinforcement-learning/>

Gupta, S. (2018, 21 juni). Automated Text Classification Using Machine Learning. Medium. Geraadpleegd van <https://towardsdatascience.com/automated-text-classification-using-machine-learning-3df4f4f9570b>

Reinforcement learning: slimmere & succesvoller bedrijf. (2019, 11 juli). APPelit. Geraadpleegd van <https://www.appelit.com/reinforcement-learning/>

Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning. Amsterdam University Press.

Wikipedia contributors. (2021, 1 maart). Reinforcement learning. Wikipedia. Geraadpleegd van https://en.wikipedia.org/wiki/Reinforcement_learning

Glen, S. (1 januari 2017). "Probabilistic: Definition, Models and Theory Explained". Geraadpleegd van <https://www.statisticshowto.com/probabilistic/>

Ghahramani, Z. (2017). "Probabilistic machine learning and AI". Geraadpleegd van <https://www.microsoft.com/en-us/research/wp-content/uploads/2017/03/Ghahramani.pdf>

Ratnayaka, G. (29 april 2020). "Probability and Machine Learning? — Part 1- Probabilistic vs Non-Probabilistic Machine Learning Models". Geraadpleegd van <https://gathikaapoorwa.medium.com/probability-and-machine-learning-570815bad29d#:~:text=Some%20examples%20for%20probabilistic%20models,input%20data%20instance%20belong%20to>

Ministry of Education (24 september 2013). "Deterministic and probabilistic models". Geraadpleegd van <https://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/AOs-by-level/AO-S8-4/Deterministic-and-probabilistic-models>

McGarigal, K. (2008). "Analysis of Environmental Data Conceptual Foundations: deterministic models". Geraadpleegd van <https://www.umass.edu/landeco/teaching/ecodata/schedule/deterministic.pdf>