

TWEE POPULAIRE DIENSTEN DIE MACHINE LEARNING ALGORITMES GEBRUIKEN

Zonder het zelf te beseffen worden we dagelijks geconfronteerd met toepassingen die heel uiteenlopende algoritmes bevatten. Maar wat is een algoritme juist? Een **algoritme** bestaat uit een reeks instructies, die als ze correct uitgevoerd worden een probleem kunnen oplossen of een taak kunnen voltooien. Een algoritme wordt daarom vaak omschreven als een stappenplan. Wanneer een algoritme wordt uitgevoerd door een computer, spreken we van een **computeralgoritme**. **AI-algoritmes** verwijzen naar de groep van algoritmes die binnen het vakgebied van artificiële intelligentie (AI) worden gebruikt. Wanneer algoritmes een model leren op basis van data, spreken we over **algoritmes voor machinaal leren**.

Wanneer je informatie leest of hoort over algoritmes, hoor je vaak ook de woorden model en techniek opduiken. Een **model** beschrijft de opgedane kennis en informatie door het algoritme die wordt gebruikt om beslissingen te nemen en/of voorspellingen te maken. **Techniek voor machinaal leren** verwijst naar de vele strategieën die worden gebruikt binnen het vakgebied van machinaal leren. Een voorbeeld van dergelijke techniek is het koppelen van de voorkeuren van luisteraars aan de voorkeuren van andere luisteraars die naar vergelijkbare muziek luisteren. Aan de hand van deze techniek, kan een algoritme voor machinaal leren aanbevelingen voor de luisteraar presenteren.

Nu alle begrippen werden geduid om verwarring te vermijden, gaan we in deze brAlnfood dieper in op algoritmes voor machinaal leren aan de hand van twee concrete voorbeelden: Netflix en navigatie-applicaties. Ben je benieuwd hoe deze twee diensten gebruik maken van algoritmes voor machinaal leren én waarom? Lees dan zeker verder.

Kenniscentrum Data & Maatschappij (2021). Twee populaire algoritmes voor machinaal leren. brAlnfood van het Kenniscentrum Data & Maatschappij. Brussel: Kenniscentrum Data & Maatschappij.

Dit document is beschikbaar onder een CC BY 4.0 licentie.



RECOMMENDER SYSTEEM

Met Netflix kun je op elk moment uit een grote database van films en series kijken wat je maar wil. Hoe langer het duurt om een keuze te maken, hoe groter de kans dat je Netflix weer afzet. Voor Netflix is het dus van cruciaal belang dat jij snel de juiste keuze kan maken. Om je te helpen kiezen tussen de vele opties gebruikt Netflix een **'hybrid recommender system'**, een combinatie van twee soorten algoritmes die jou aanbevelingen geven over welke film of serie te kijken:

- **'Collaborative filtering'**: bij dit algoritme worden gebruikers gegroepeerd op basis van vergelijkbaar gedrag en maakt het aanbevelingen gebaseerd op de karakteristieken van de groep.
- **'Content-based'**: dit algoritme gebruikt de informatie van een item om vergelijkbare items aan te bevelen aan een specifieke gebruiker.

DATA

Hiervoor gebruikt Netflix de volgende **data**:

- **Alle interacties die je met het platform hebt** (bv. je kijkgeschiedenis, je waardering van video's, op welke apparaten je kijkt, welke aanbevelingen je negeert);
- **De data van andere gebruikers met vergelijkbare interesses en smaak binnen Netflix**;
- **Informatie over de videos** (bv. titels, acteurs, genre, jaar van uitkomst);
- **Geen demografische informatie** (bv. leeftijd en geslacht).

PERSONALISATIE

De startpagina van het platform bestaat uit een grote matrix met meerdere rijen die elk verschillende films/series bevatten. Elke rij wordt op drie manieren **gepersonaliseerd**:

- **De thema's** van de rijen (bv. 'romantische komedies');
- **De films/series** binnen dat thema **die je mogelijk interesseren**;
- **De volgorde** waarin je de films/series ziet.

Hieronder beschrijven we **een aantal veelvoorkomende rijen met hun algoritmes**.

NETFLIX

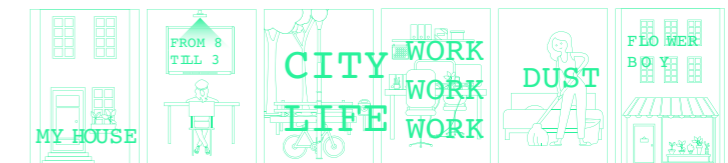
POPULAIR OP NETFLIX



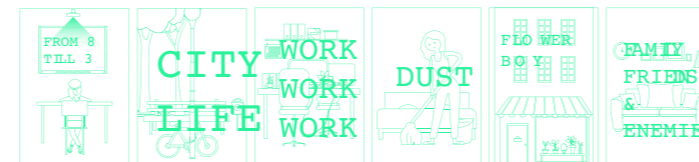
De **'Personalized Video Ranker'** (PVR) zorgt ervoor dat binnen een subset (bv. **een genre**) de volgorde van de aanbevelingen voor jou persoonlijk relevant zijn. Zo komt het dat jij en ik andere titels zien in de rij **'Populair op Netflix'**. De PVR schikt de titels namelijk gebaseerd op wat bij elke kijker past.

De **'Trending ranker'** identificeert twee soorten trends: terugkerende (bv. meer interesse in romantische films rond Valentijnsdag) en eenmalige trends (bv. meer interesse in natuurdocumentaires na een natuurramp). Samen met personalisatie op basis van jouw kijkgeschiedenis worden deze trends aan jou in de **'Trending'-rij** gepresenteerd.

TRENDING

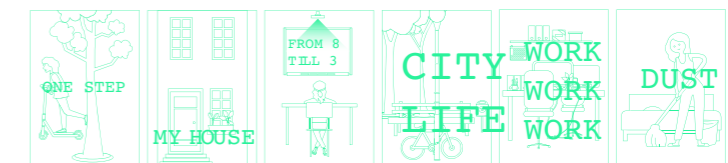


VERDERKIJKEN VOOR ...



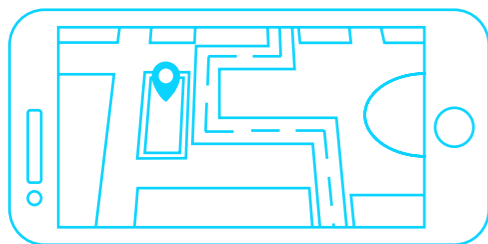
De **'Continue Watching ranker'** sorteert de films/series die je recent hebt bekeken op basis van de verwachting of je deze film/serie verder wil kijken of niet. Dit gebeurt op basis van wanneer je de film of serie uitzet, en of je sindsdien andere titels hebt bekeken en op welke apparaten.

THEMARIJEN



Het **'Page Generation algoritme'** bepaalt welke **thema-rijen** op jouw startpagina verschijnen zodat die relevant en divers zijn. Kijk je bijvoorbeeld andere soorten films/series op verschillende momenten van de dag of gebruiken meerdere mensen hetzelfde account, dan houdt het algoritme hier rekening mee.

NAVIGATIE-APPS



Navigatie-applicaties (bv. Google Maps, Apple Maps, Waze) helpen miljoenen mensen van punt A tot bij punt B te geraken. Navigatie-apps moeten niet alleen hun gebruikers tot hun bestemming leiden, ze moeten ook de **snelste route kunnen berekenen** of een **route die afgestemd is op het vervoermiddel of andere voorkeuren van de gebruiker**.

DATA

Om verschillende routes te kunnen aanbieden, gebruiken navigatie-apps **data** uit verschillende bronnen, zoals:

- **Live verkeersinformatie**, afkomstig van de toestellen van gebruikers;
- **Historische verkeersdata**;
- **Informatie over snelheidslimieten en wegenwerken**;
- **Informatie over het type, de vorm en de kwaliteit van de weg**.

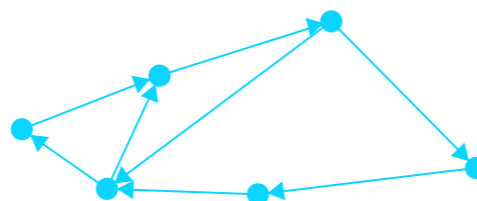
Voor die laatste groep van data, haalt Google Maps ook informatie uit Google Street View of satellietdata. Die data worden via verschillende machine learning-modellen verwerkt tot voorspellingen van verkeersstromen.

DIJKSTRA'S ALGORITME

De applicatie gaat op zoek naar het kortste tijdstraject. Om deze berekening te maken, worden zeer geavanceerde algoritmes gehanteerd, maar veel van deze algoritmes vallen terug op het **concept van Dijkstra**.

Het algoritme van Dijkstra, dat in 1959 door de Nederlandse informaticus Edsger Dijkstra werd bedacht, bepaalt op een slimme manier **de kortste weg van een gegeven beginpunt naar elke knoop van een netwerk**.

EEN GRAAF



Het algoritme kan niet meteen op een kaart toegepast worden, omdat deze te veel informatie bevat. Het zal daarom werken met **een representatie die bestaat uit punten en verbindingen tussen deze punten**. Deze representatie wordt in wiskunde een graaf genoemd.

Er zijn allerlei specifieke soorten van grafen mogelijk, vaak afhankelijk van de eigenschappen van de verbindingen. De verbindingen kunnen gericht zijn (bv. bij eenrichtingsverkeer), in dat geval spreken we over een **gerichte graaf**. De verbindingen kunnen ook ongericht zijn, we spreken dan over een **ongerichte graaf**.

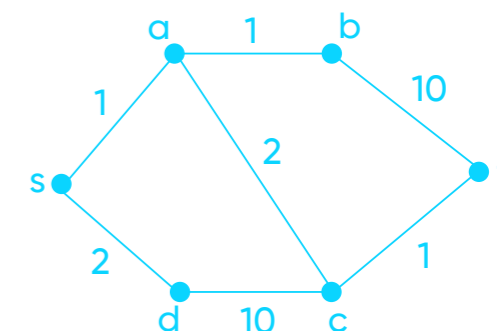
WAARDENETWERK

Beeld je in dat je huidige positie een knooppunt is, net zoals je eindbestemming. Tussen die twee verschillende posities liggen verschillende andere knooppunten, die met elkaar verbonden zijn. In de graaf heeft **iedere verbinding een waarde**. Die waarde representeert een 'afstand' of een 'kost', die staat voor de inspanning die nodig is om naar het andere knooppunt te geraken. Bij navigatie-apps representeren die waarden **de reële afstand of de hoeveelheid verkeer**.

GULZIGE METHODE

Het Dijkstra-algoritme werkt met een 'gulzige methode'. Deze methode **gaat stap voor stap alle knooppunten in het netwerk af**. Om steeds de volgende stap te bepalen wordt gekozen voor een aansluiting met een ander punt dat onmiddellijk het grootste voordeel geeft. Dit leidt vaak niet direct tot een optimale eindoplossing, maar uiteindelijk zal het algoritme toch de **globale optimale oplossing van vertrekpunt tot eindbestemming** voorstellen.

VOORBEELD



In deze eenvoudige voorstelling zal het algoritme eerst van 's' naar 'a' gaan, en vervolgens van 'a' naar 'b'. Als kortste pad zal echter de reeks 's-a-c-t' voorgesteld worden.

Stel dat een navigatie-app een route moet berekenen tussen Brussel (vertrekpunt) en Antwerpen (bestemming). Als de toepassing zuiver met de Dijkstra-methode zou werken, dan zou het algoritme in alle richtingen rond Brussel beginnen te zoeken. Dat zou een inefficiënte inspanning zijn en te veel tijd en rekenkracht kosten. Navigatie-apps werken met meer **geavanceerde algoritmes**, die vanuit het concept van de Dijkstra-methode vertrekken, waardoor er specifiek **in de richting van de bestemming** gezocht kan worden.

Bronnen Netflix

Hunt, N., & Gomez-Urbe, C. A. (2016). The Netflix Recommender System: Algorithms Business Value and Innovation. *Journal, ACM Transactions on Management Information Systems (TMIS)*, 6(4). [\[paper\]](#)

Netflix. Hoe het aanbevelingssysteem van Netflix werkt. [\[website\]](#)

Goltsman, K. (2017). Under the Hood of Netflix Recommender System. [\[blogpost van DataScience Foundation\]](#)

Bronnen navigatie-applicaties

Fitro, A., P Sulistio Ilham, A., B Saeful, O., & Frenadiana, I. (2018). Shortest Path Finding in Geographical Information Systems using Node Combination and Dijkstra Algorithm. *International Journal of Mechanical Engineering and Technology* 9(2), 2018, pp. 755-760. [\[paper\]](#)

Lanning, D. R., Harrell, G. K., & Wang, J. (2014). Dijkstra's algorithm and Google maps. *Proceedings of the 2014 ACM Southeast Regional Conference on - ACM SE '14*, 1. [\[paper\]](#)

Lau, J. (2020). Google Maps 101: How AI helps predict traffic and determine routes. [\[blogpost van Google\]](#)

Byrne, M. (2015). The Simple, Elegant Algorithm that Makes Google Maps Possible. [\[blogpost van vice\]](#)

Velden, L. (Department of Mathematics, Technical University of Munich) (2014). Der Dijkstra-Algorithmus. The classic among shortest path algorithms. [\[website\]](#)

Bitesize Data Science (2019). Navigation Apps: The Algorithms that Get You from Here to There. [\[blogpost van Bitesize Data Science\]](#)