

Machine learning als extra versnelling voor de wetenschap

Ronald Kok

Data is booming, vrijwel alle sectoren lijken steeds meer te focussen op het vergaren van data. De financiële sector doet hier uiteraard ook volop aan mee. We spreken professor Philip Hans Franses van de Erasmus School of Economics over de belangrijkste ontwikkelingen op het gebied van data-analyse. We spreken hem over de transformatie van het in de wetenschap zoeken naar 1 exacte oplossing naar het via een gigantisch spectrum aan data zoeken naar juist meerdere oplossingen.

De afdeling econometrie in Rotterdam is in haar vakgebied de grootste ter wereld. Jan Tinbergen was grondlegger van de econometrie en kreeg daar in 1969 samen met Ragnar Frisch de Nobelprijs voor. Samen met de andere 5 universiteiten waar econometrie wordt aangeboden zit Nederland ver voorop in de internationale kopgroep.

Data-analyse niets nieuws

Franses heeft vanuit zijn vak econometrie data-analyse (ofwel machine learning) steeds meer zien ontwikkelen. Van de zogenaamde neurale netwerken in de jaren '90 tot de meest geavanceerde kunstmatige intelligentie anno 2019. Waar we in de media veel tromgeroffel horen over zaken zoals 'big data' en 'algoritmes' tempert Franses deze euforie. De bomen groeien volgens hem zeker niet tot in de hemel en robots gaan absoluut niet op korte termijn ons leven domineren. *"Een algoritme is niets anders dan een soort beslissingsregel. Bijvoorbeeld: als het regent dan neem ik een paraplu mee is in weze ook een algoritme"* aldus de professor. *"Data-analyses, keuzes in data, het beheren van data is iets dat we al decennia doen"*. Wat dat betreft is er dus volgens hem absoluut niets nieuws onder de 'data-zon'. Denkt u bijvoorbeeld aan de verzekeraars en supermarkten, die dolgraag zoveel mogelijk informatie over hun klanten willen weten. Als voorbeeld verwijst Franses naar de streepjescode op producten. Dit was niet alleen een uitkomst voor de logistiek want er was dan duidelijk dat een product uit de schappen was maar er kon ook gelijk inzicht worden gecreëerd in de verkopen.

Als voorbeeld haalt Franses de bonuskaart aan. Met een bonuskaart kan deze informatie weer gekoppeld worden met heel veel persoonlijke variabelen en dit biedt volgens de professor zeeën van mogelijkheden. Voorheen was deze data simpelweg veel te groot voor zelfs de beste computers. Met de technologische ontwikkelingen staan de poorten nu dus wijd open om voor alle sectoren heel veel data te vergaren en deze ook optimaal te analyseren. Bedrijven, wetenschappers, gebruikers et cetera zouden dan op basis van data-analyse enorm veel progressie kunnen maken. Machine learning kunnen we dus zien als een extra versnelling voor alle

denkbare wetenschappen. Hiermee benadrukt Franses dat machine learning absoluut niet gezien moet worden als een vervanging maar is het juist een enorme mogelijkheid om bestaande wetenschappen verder uit te bouwen.

Van exact naar optimalisatie voorspellingen

Bij de opleiding econometrie ziet Franses een verdere uitdieping van de materie. Het gaat volgens hem steeds minder om het berekenen van één exact antwoord en steeds meer om de beste voorspellende variabelen te vinden. *"Natuurlijk blijven klassieke financiële modellen zoals het CAPM als een paal boven water staan."*



Philip Hans Franses Franses, geboren in 1963 in universiteitsstad Wageningen studeerde econometrie in Groningen en behaalde onder begeleiding van professor Teun Kloek in 1991 zijn PhD met zijn onderzoek *"Model selection and seasonality in time series"*. Sinds 1996 is Franses professor bij Erasmus Universiteit Rotterdam. Franses is ook verbonden aan de Chiang Mai University en de University of Western Australia. Tot februari dit jaar was Franses decaan bij de Erasmus School of Economics.

Dergelijke regressiemodellen zijn nog steeds het fundament van de beleggingswetenschappen. Data-analyse biedt mogelijkheden om nog accuratere voorspellingen te doen" aldus Franses. Volgens hem is machine learning niets anders dan een extra toevoeging op de bestaande modellen. Verder was het voor een student vroeger vaak de opdracht om een model met een handjevol variabelen in kaart te brengen. Nu kunnen dat wel 3.000 variabelen zijn. De student moet dan op zoek gaan naar de beste voorspellende variabelen. Dit brengt de materie volgens hem dan ook zo steeds dichterbij de praktijk.

Moderne computerprogramma's voor permanente educatie

Vergeleken met nog geen 10 jaar terug is door de entree van machine learning het academische onderwijs behoorlijk op de schop gegaan. Bij diverse opleidingen worden nu ook programmeertalen zoals Python aangeboden. Waar vroeger Excel nog het belangrijkste dataprogramma was voor de economiestudent is dit programma al totaal ingehaald. *"Excel, eigenlijk een Lotus boekhoudprogramma is leuk om kleine hoeveelheden op te slaan maar meer ook niet"* aldus Franses. Het gros van de VBA en CFA leden zal echter nog heel veel met Excel bezig zijn maar in weze is het niet meer van deze tijd en al zeker heel onbruikbaar voor echte data-analyse aldus Franses. Als voorbeeld haalde Franses een auto in de garage aan: nu gaat er een stekker in die eerst de data uitleest en is er dus ook veel software- en analysekennis nodig die eerst überhaupt nooit nodig was in een garage. Om up to date te zijn met de laatste academische technieken ziet Franses zeker een behoefte om machine learning toe te voegen aan de permanente educatie.

Data-analyse blijft toch mensenwerk

Waar we in veel media vaak de vrees horen dat op den duur alles door robots (algoritmes) bepaald gaat worden valt dit volgens Franses zeker wel mee. Machine learning is al in behoorlijke mate kunstmatig intelligent maar het zijn nog steeds mensen die uiteindelijk aan de knoppen zitten. Het zal volgens hem nog heel erg lang

duren voordat 'machines' echt de overhand krijgen. Hij onderschrijft dit dat veel ontwikkelingen nooit lineair gaan en ook vaak weer verstoord worden. Als voorbeeld haalt Franses het 'Internet of Things' (IoT) aan. *"Privacy-issues kunnen veel van deze ontwikkelingen stevig op de rem trappen"* aldus Franses. De 'echte' wereld is volgens hem nog verre van programmeerbaar want er zijn zo veel variabelen waar we amper grip op hebben. Franses vergelijkt dit met schaken en dammen waarbij de keuzemogelijkheden aftelbaar zijn en dus programmeerbaar. In de 'echte' wereld zijn de keuzes echter niet aftelbaar en blijft er dus altijd ruimte open.

Stimulans voor beleggers

Het grote voordeel van machine learning is dat het emotioneel is. *"Beleggers reageren vaak emotioneel en handelen vaak erg irrationeel. Handelsbeslissingen op basis van machine learning kunnen hier zeker uitkomst bieden"* stelt Franses. Als voorbeeld haalde Franses de vrij recente euforie over de BRIC landen aan. Voor veel analisten en fondshuizen waren de landen Brazilië, Rusland, India en China absoluut de toekomst voor beleggers maar ondertussen hebben we gezien dat dit heel anders kan uitpakken. Een machine luistert niet naar al die mooie verhalen maar blijft rationeel zo stelt Franses. Een machine kan wel optimisme over bepaalde beleggingen meewegen maar hij kan daar ook gelijk een bepaalde onzekerheidsmarge bij houden.

Op jacht naar de beste methoden

Dagelijks komen er nieuwe data-technieken bij en hierin schuilt natuurlijk het risico van door de bomen het bos niet meer zien. Het is voor de wetenschap zaak om de beste methoden in kaart te brengen. Enkele sterke data-methoden zijn onder andere die van Lasso en Elastic Net. *"Je kunt deze wedloop zien als het bereiken van de Champions League voor analysemethoden"*. Als data-analist is het dan de uitdaging om per vakgebied de Champions League spelers te gebruiken. Dit zal voor de komende jaren in dataland zeker één van de belangrijkste taken zijn zo besluit Franses.